



Article Info/Makale Bilgisi

Received/Geliş: 03.06.2025 Accepted/Kabul: 18.07.2025

DOI: 10.30794/pausbed.1712967

Translated Article/Çeviri Makalesi

This article is the English translation of the original Turkish article:

Arslan, E. (2022). "Nitel Araştırmalarda Geçerlilik ve Güvenilirlik," *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 51 (Özel Sayı 1), Ö395-Ö407. <https://doi.org/10.30794/pausbed.1116878>.

Published with the author's permission.

VALIDITY AND RELIABILITY IN QUALITATIVE RESEARCH

Erdal ARSLAN**

Abstract

The ability to demonstrate the validity and reliability of research findings is one of the important factors that determine the value of scientific research. Validity and reliability or trustworthiness are fundamental issues in scientific research whether qualitative, quantitative, or mixed research. It is a necessity for researchers to describe the criteria they have taken into consideration for the findings obtained during the validity and reliability of the research process. Even though currently, the validity and reliability of qualitative research findings are still controversial in some academic environments, particularly among positivists, qualitative research has lately gained momentum among researchers. Conceptual complexity in literature and unclear statements are misleading to researchers who are interested in qualitative research, such as newbies and students. Based on this viewpoint, the current study is aimed to present the specific information on validity and reliability in qualitative research, from an analytical perspective. In this context, the concepts of internal validity and external validity, internal consistency reliability, and external reliability used in quantitative research that are equivalents to credibility, transferability, dependability, and confirmability in qualitative research are scrutinized. In addition, techniques to provide each criterion that researchers can use in their research are also detailed.

Keywords: *Validity and Reliability, Trustworthiness, Credibility, Transferability, Dependability, Confirmability.*

NİTEL ARAŞTIRMALARDA GEÇERLİLİK VE GÜVENİLİRLİK

Öz

Bilimsel araştırmaların değerini belirleyen unsurlardan biri araştırma bulgularının geçerlilik ve güvenilirliğinin ortaya koyulabilmesi yeteneğine bağlıdır. Nitel, nicel veya karma araştırmalarda olması fark etmeksizin geçerlilik ve güvenilirliğin veya güven duyulabilirliğin sağlanması temel bir sorun olmaktadır. Araştırma sonuçlarının hangi bilimsel temellerle elde edildiğinin kanıtlandığı bu süreçte, araştırmacıların elde ettikleri sonuçları hangi ölçütleri dikkate alarak sağladığını açıklaması bir zorunluluktur. Günümüzde bazı akademik çevrelerce bilimselliği hâlâ tartışıldığı hâlde, son yıllarda araştırmacıların artan bir biçimde ilgi göstermeye de devam ettiği nitel araştırmalarda geçerlilik ve güvenilirlik konusu ile ilgili soru işaretleri hâlâ devam etmektedir. Literatürde yer alan kavram karışıklıkları ve yeterince net olmayan ifadeler ise başta nitel araştırmaları henüz öğrenmeye başlayanlar olmak üzere nitel araştırmalar ile ilgilenen araştırmacılar için yanıltıcı olmaktadır. Bu düşünceden hareketle mevcut araştırmada, nitel araştırmalarda araştırmacıların geçerlilik ve güvenilirlik konusuyla ilgili faydalanabileceği bilgilerin analitik bir bakış açısıyla sunulması amaçlanmıştır. Bu bağlamda araştırmada, nicel araştırmalarda kullanılan iç geçerlilik ve dış geçerlilik ile iç tutarlılık güvenilirliği ve dış güvenilirlik kavramlarının nitel araştırmalardaki karşılığı olan inandırıcılık, aktarılabilirlik, tutarlılık ve doğrulanabilirlik kavramları açıklanmıştır. Buna ek olarak araştırmacıların kendi araştırmalarında kullanabilecekleri her bir kriteri sağlayacak teknikler de detaylandırılmıştır.

Anahtar kelimeler: *Geçerlilik ve Güvenilirlik, Güven Duyulabilirlik, İnandırıcılık, Aktarılabilirlik, Tutarlılık, Doğrulanabilirlik.*

*Dr., erdalarslan09@gmail.com. (<https://orcid.org/0000-0002-5791-2073>)

INTRODUCTION

Qualitative research is positioned as a paradigm that rejects the notion of a single objective reality and instead posits that reality is constructed through multiple interpretations shaped by context and the researcher's perspective (Yin, 2011; Ravitch & Carl, 2019). The aim of the qualitative researcher, therefore, should be to design the research process through systematic and analytical procedures in order to produce trustworthy findings that can effectively respond to biased criticisms often directed at interpretative approaches. Although this perspective is not unfamiliar to social scientists, the social world has largely been approached from a positivist lens. From this standpoint, reality is believed to be derived from observable, tangible, and measurable sensory data, and unobservable mechanisms are typically excluded from consideration within the system (Schwandt, 2015; Blaikie, 1993).

There has been a growing shift toward qualitative research in the social sciences in recent years, supported by increasing interest from a variety of academic disciplines. This increased attention has more prominently revived the longstanding debate concerning the scientific credibility of such research. The growing interest in qualitative research is partly grounded in the understanding that qualitative research, much like quantitative research, fulfills methodological criteria referred to as trustworthiness in the qualitative paradigm and as validity and reliability in the quantitative tradition. While quantitative researchers typically deal with numerical datasets, qualitative researchers engage heavily with words, images, and contextual settings. Moreover, whereas researchers in quantitative studies tend to remain detached from the process, qualitative researchers are directly involved and often play a central role in shaping the study (Braun & Clarke, 2013, p. 4). Such characteristics, among many others, do not constitute an obstacle to the epistemological contributions of qualitative research. As in quantitative studies, qualitative inquiry employs clearly defined criteria and methodological strategies to ensure and enhance trustworthiness.

The value of a scientific study partly depends on the researcher's ability to demonstrate the validity and reliability of its findings. Irrespective of the methods or disciplinary approaches used for data collection and analysis, scientific inquiry aims to produce original and meaningful outcomes. Therefore, ensuring the reliability and validity of results is essential across all fields in which research is conducted (LeCompte & Goetz, 1982, p. 31). Regardless of the research type (qualitative, quantitative, or mixed) or the type of data collected (quantitative, qualitative, or both) researchers are expected to implement appropriate procedures to ensure validity and reliability. While such techniques have long been established for quantitative research, it was only in recent decades that they were developed for qualitative research, primarily through the works of Guba (1981) and Guba and Lincoln (1985). Given the assumption that the meaning and relevance of a concept become more clear through its examination alongside contrasting counterpart, this study explains the notion of trustworthiness in qualitative research, as introduced by Guba (1981, 1990) and Lincoln and Guba (1985), alongside the corresponding concepts of validity and reliability in quantitative research, occasionally drawing comparisons between the two techniques.

According to the Higher Education Information Management System for the 2020–2021 academic year, Turkey had a total of 449,717 postgraduate students, including 343,569 in master's and 106,148 in doctoral programs. This study aims to provide a clear and focused resource, free from peripheral discussions, by presenting key insights on validity, reliability, and trustworthiness, primarily for postgraduate students and researchers interested in qualitative inquiry. To this end, the author has drawn upon the foundational work of key scholars in the field, including Guba and Lincoln, seminal authors who introduced the conceptual framework of trustworthiness, as well as Denzin, Miles, Huberman, Saldaña, Braun, Clarke, Patton, Creswell, Maxwell, and Yin.

Qualitative research has gained significant attention among social science researchers due to its ability to provide in-depth insights into specific topics. However, such studies are often criticized for various reasons, including the way the research is framed, the lack of transparency in the process, concerns about the reliability of findings, or insufficient theoretical grounding. These methodological concerns have led many top-tier journals in the social sciences, particularly those ranked as Q1, to reject qualitative research submissions (Czernek-Marszałek & McCabe, 2022). One way for researchers to address this issue is by systematically implementing the procedures referred to as validity and reliability in quantitative research, and as trustworthiness in qualitative

research, to produce credible results. At this point, it becomes essential, particularly for postgraduate students and early-career researchers conducting qualitative studies, to be familiar with the core questions related to the criteria of trustworthiness (Guba, 1981, 1985, 1990). Ravitch and Carl (2019) offer a clear explanation of the trustworthiness criteria and the types of questions researchers are expected to consider in addressing each criterion, as outlined in Table 1.

Table 1: Reflective questions related to criteria of validity, reliability, OR trustworthiness

Criterion	Key questions for researchers to consider
Credibility	How can I design a study that seeks complexity and engages with the real-world intricacies of groups, contexts, or communities?
	How do my research field selection criteria and sampling strategy contribute to the development of a meaningful and grounded context?
	How well do my research questions align with my methodology?
	Have I developed a rich and complementary dataset by drawing from multiple sources?
	How do I interpret and make use of the patterns I observe in the data?
	How will I interpret and make sense of my data while addressing my own assumptions and potential bias?
	How will I bring together the pieces that emerge through data analysis into a meaningful whole?
	What role(s) do participants play in shaping the study and supporting or challenging my interpretations?
Transferability	How do I identify the contextual factors that shape and mediate my study?
	Do I provide readers with sufficient contextual information and framing to help them make sense of the findings?
	How do I interpret and present my data in ways that are contextually grounded and meaningful?
	How do I describe the research setting, the participants, and the specific features of the environment? Is the description detailed enough?
	Have I adequately explained the contextual relevance and embeddedness of my analysis and interpretations?
Dependability	What informed my choice of research methods, and how did I select them?
	What makes these methods well suited to address my research questions?
	How do my methods align with my research questions?
	Have I structured the study with intentionality and rigor? What specific decisions reflect this? In what ways could the design be strengthened?
	What potential critiques may emerge concerning my research design, data collection, or analytic process? How have I anticipated and responded to them?
	Have I engaged in critical dialogue with advisors or peers about my design, including possible methodological or conceptual limitations?
Confirmability	Do I have a predetermined roadmap, and does it influence the findings when applied to the data? If so, in what ways? How can I mitigate this influence?
	Could someone else have reached similar results or interpretations? In what ways might their perspective differ?
	Who can I involve in the study to critically assess my interpretations?
	At what stages should I involve peers to reflect on my subjectivity and position in the study?

Source: Ravitch & Carl, 2019.

Since validity and consistency cannot be achieved without dependability, and credibility and transferability cannot be ensured without confirmability, this study first elaborates on the core criteria for validity, namely internal validity (credibility) and external validity (transferability), and for reliability, including internal consistency (dependability) and external reliability (confirmability). In addition, techniques used to ensure these criteria in qualitative research are discussed in detail.

VALIDITY: CREDIBILITY AND TRANSFERABILITY

Although validity plays a pivotal role in scientific research and is frequently emphasized (Maxwell, 2013, p. 162), it is often overlooked in qualitative studies. Validity, more commonly associated with quantitative research, refers to the extent to which a study accurately measures what it intends to measure. While the concept has long been linked to quantitative or positivist traditions, it has also gained prominence in interpretivist and qualitative research over the past few decades. In a given study, the question “How does the researcher know that their subjective interpretations are accurate or valid?” points directly to the issue of validity. Creswell (2016, p. 409) defines validity in qualitative research as the extent to which findings are accurate or credible. Similarly, Silverman (2013, p. 534) describes validity as a concept that refers to the believability of interpretations in qualitative inquiry, while Goodman (2008) defines it as the extent to which a study demonstrates what it claims to reveal.

Definitions of validity for qualitative inquiry often focus on the concepts of “truth” and “accuracy.” It is therefore evident that both the suitability of the data collection instrument in relation to the research topic and purpose, as well as the researcher’s objectivity play a significant role in ensuring validity. In line with this perspective, Yin (2011, p. 78) states that validity relies on the appropriate collection and interpretation of data so that findings can accurately reflect and represent the real world being studied.

In order to contribute to the theoretical or practical development of a field, researchers require to carry out their studies rigorously. Research should provide findings and interpretations that convey a sense of credibility to readers, practitioners, and other scholars. The applied nature of most social research makes it essential for the outcomes to be perceived as trustworthy (Merriam, 2009, p. 210). At this point, Denzin and Lincoln (2018, p. 241) draw attention to the authenticity of social research and the relevance of its focus by raising a set of questions, such as how the accuracy of research findings is understood in relation to socially constructed realities, how individuals respond to these findings with trust, and how such results inform action toward members of the society in which the research was conducted. Considering the dynamic nature of human experience and the existence of multiple realities, it is difficult to claim definite answers; yet, as in other methodological traditions, qualitative research includes a range of strategies intended to enhance validity.

The concept of validity, which is widely used in scientific research, is subject to debate among qualitative researchers due to its association with the positivist paradigm or quantitative research (Denzin and Lincoln, 1998, pp. 21–22; Erlandson, Harris, Skipper, and Allen, 1993, p. 131; Merriam, 2009, p. 211). Instead, researchers emphasize the importance of using terms that refer to validity but are more consistent with the nature of qualitative research. From this perspective, Lincoln and Guba (1985) introduced the widely accepted concepts of credibility and transferability as conceptual equivalents to validity in quantitative research. Accordingly, credibility corresponds to internal validity, while transferability reflects the external validity.

Credibility (internal validity)

Lincoln and Guba (1985, p. 290) explain the concept of internal validity by drawing on the definition provided by Cook and Campbell (2002 [1979], p. 34). According to this view, credibility is described as the best available representation of the accuracy or plausibility of a given inference. It concerns the extent to which research findings correspond with reality, whether they reflect what actually exists, and whether researchers are observing or measuring what they believe they are investigating (Merriam, 2009, p. 213; Lincoln and Guba, 1985, p. 290). Guba (1981, pp. 84–86) and Lincoln and Guba (1985, p. 301) argue that credibility is one of the key components in establishing trustworthiness in qualitative research and propose several techniques to achieve it. These include prolonged engagement, persistent observation, and triangulation, which contribute to generating findings and interpretations that are considered credible. Peer debriefing offers an external check on the research process by engaging colleagues in reviewing the inquiry. Negative case analysis supports the refinement of research questions and interpretations as data are collected. Referential adequacy enables the verification of initial findings and interpretations by comparing them with raw data. Finally, member checking enables the validation of both findings and interpretations through direct feedback from the participants themselves, the co-constructors of multiple realities.

Prolonged engagement refers to the investment of time in the field to gain an in-depth understanding of the culture, identify potential misinformation arising from the researcher or participants, and establish trust (Guba and Lincoln, 1985, p. 301; Patton, 2015, p. 989). It involves the researcher spending sufficient time with participants in their natural settings to understand their surroundings and cultural context. This process helps correct possible misunderstandings and enables more accurate interpretations. Through spending time with participants, the researcher builds trust, which in turn encourages participants to be more open and expressive during interviews.

Persistent observation focuses on identifying the aspects of the phenomenon that are most relevant to the research problem and exploring them in depth. While prolonged engagement contributes to the scope of the inquiry, persistent observation enhances depth (Lincoln and Guba, 1985, p. 304).

Triangulation is a technique used to enhance the validity and richness of research findings by incorporating different types of qualitative and/or quantitative data along with various data collection methods such as interviews, observations, tests, or surveys (Lincoln and Guba, 1985, p. 306; Silverman, 2013, p. 449). By diversifying both the types of data and the methods of data collection, the researcher is able to compare sources and assess the extent to which the data support each other. When necessary, the researcher navigates across data sources, allowing for deeper exploration of the inquiry.

Peer debriefing involves the review of the research by a peer who is not involved in the study, in order to ensure that no issues or biases are overlooked in the researcher's analysis and interpretations. It helps maintain a neutral and clear perspective and supports reflection on subjective concerns that may influence participation in or interpretation of the fieldwork (Miles and Jozefowicz-Simbeni, 2010, p. 737).

Negative case analysis consists of considering alternative interpretations of data and reevaluating them by referring to data segments that tend to contradict the researcher's reconstruction of reality. These contrasting pieces of evidence are critically examined and integrated into the analysis to consider alternative interpretations. This process continues until the contrasts between such cases and the broader dataset are resolved or are no longer conceptually meaningful. In simpler terms, it emphasizes the systematic consideration of contradictory findings that emerge during analysis, serving as a form of reconciliation that contributes to refine the results and strengthen the conceptual coherence of the study (Phillips, 2014, p. 552).

Referential adequacy addresses a step that enables preliminary findings and interpretations to be checked against archived raw data (Lincoln and Guba, 1985, p. 301). Since all data are to be evaluated in relation to their context, the collection of materials that provide a holistic view of the setting is essential. Videotapes, documents, photographs, and other materials that offer a snapshot of the studied context serve as a supportive background, helping readers to gain a richer contextual understanding of the researcher's analyses and interpretations (Erlandson et al., 1993, p. 31).

Member checking, also referred to as participant validation or respondent verification, involves of having participants review and confirm the researcher's interpretations. This technique typically entails presenting a draft version of the research report, either in written or oral form, or selected parts of the analysis to the participants, and asking them to comment on the accuracy and validity of the findings (Braun & Clarke, 2013, p. 282). It is used particularly when verifying data that are not directly observable or measurable, such as personal memories or subjective narratives. For example, in a historical context, older participants sometimes have difficulty recalling precise dates, places, or details required by the researcher. In such cases, organizing a focus group discussion helps participants trigger one another's memories and improve the overall accuracy of the data. When all factual information has already been verified, there is no need to seek additional confirmation, and the findings remain as reported (Morse, 2018, p. 1397). However, in some cases, carrying out member checking is not always feasible, particularly when participants have limited formal education, face socioeconomic disadvantages, or when the research topic is highly technical or complex.

Transferability (External Validity)

In a qualitative study, transferability refers to the applicability of findings to other contexts, participant groups, or cases that share fundamental similarities with the original research setting (Guba, 1981; Lincoln & Guba, 1985; Merriam, 2009; Miles, Huberman, and Saldaña, 2014). In this regard, the concern lies in the concept of generalization, which qualitative research does not adopt as a primary focus. The type of generalization addressed within the context of external validity in qualitative research differs from that in quantitative research, which relies on probability-based sampling and rigorous data collection to produce findings representative of a larger population. Braun and Clarke (2013, p. 280) emphasize the relevance of applying findings from a specific study to broader or different individuals or groups. In traditional research, the researcher is responsible for ensuring generalization to the broader population, whereas in qualitative research, establishing and demonstrating transferability falls to the researcher who applies the findings to different settings (Erlandson et al., 1993, p. 33). In a qualitative study, ensuring transferability depends on providing detailed descriptions of the specific context, participants, research setting, and conditions of the study. This enables the reader to evaluate the potential applicability of the findings to other contexts or participant groups (Braun and Clarke, 2013, p. 282; Arslan, 2021, pp. 69–70).

Thick description is a technique that enables readers to make informed judgments about the transferability of research findings across different contexts. Researchers address this criterion by thoroughly explaining the contextual factors that influence their studies. Transferability depends on the degree of similarity between the context in which the research was conducted and the contexts in which findings might be applied in the future. Therefore, the researcher is required to present detailed accounts of the data within its context to support judgments about transferability. A well-crafted description immerses the reader in the research setting. Descriptions of specific images, sounds, and relationships within the setting create a scene in the reader's mind that closely resembles the experience itself, potentially producing a *déjà vu* effect (Erlandson et al., 1993:33). Guba (1981) emphasizes the fundamental role of having sufficient information about both the source and target contexts to assess whether transferability is achievable. He argues that when a significant similarity exists between two contexts, generalizing research from one to the other is rational; however, this assumption requires empirical inquiry.

Purposive sampling is considered a limitation in quantitative research, as it restricts generalization. Although this technique falls under the category of non-probability sampling in quantitative research, it functions in qualitative research much like probability sampling does in quantitative studies, as it allows for a purposeful and context-specific selection process. Purposive sampling involves the inclusion of individuals or groups directly related to the research topic, aiming to ensure a rich and accurate dataset. While in traditional quantitative research, the unequal chance of participation is regarded as manipulation, in qualitative studies, this appears as an advantage. For instance, in a study that investigates narratives of a specific experience, it is appropriate for the researcher to interview those who have undergone that particular experience. The inclusion of participants or settings that are directly related to the research subject is closely associated with the validity of the study. Lincoln and Guba (1985) state that, just like in thick description, the researcher must present the broadest possible range of information in order to achieve transferability. Therefore, using purposive sampling is a reasonable strategy to boost the transferability as well as the quality of the research.

Reflexive journal, fieldnotes or memos refer to techniques used by researchers to document various forms of personal information either daily or as needed (Janesick, 2014:307; Lincoln and Guba, 1985:327). Depending on the type of data, these entries may concern participants, the research setting, the distinctiveness of a participant's statement, or the emotions experienced by the researcher during the interview. Reflexive journals affect not only the credibility of the study but its transferability, dependability, and confirmability (Erlandson et al., 1993:143). These journals documenting a range of research-related processes, support the category development (Matteson, 2021:121). Guba and Lincoln (1985:327) highlight three essentials that a reflexive journal is expected to contain: the daily schedule and organization of the study; personal notes that reflect the researcher's emotions and thoughts during the process; and a methodological log that records decisions and the rationale behind them.

Reflexive journals serve as records of observations made by the researcher during the data collection phase. These entries may be produced either concurrently with the process or retrospectively, depending on the data collection design and the use of tools such as audio or video recorders (Gibson and Brown, 2009). Field-based observational and interview notes assist the researcher in retrieving context-specific insights and recalling emotionally salient moments from the fieldwork. This facilitates a more authentic and compelling representation of the findings. While a qualitative researcher might overlook certain procedural details, emotional responses tend to persist. To ensure access to these affective dimensions, they must be systematically recorded at the time of occurrence.

RELIABILITY: DEPENDABILITY AND CONFIRMABILITY

Credibility in qualitative research is contingent upon dependability, just as validity presupposes reliability (Lincoln and Guba, 1985:316). It is possible to achieve a high level of reliability in a study without establishing any form of validity (Kirk and Miller, 1986:20); however, the significance of a measurement or assessment in terms of reliability depends on its validity. In this respect, reliability and validity are tightly interconnected constructs that collectively underpin methodological rigor. Traditionally, reliability refers to the probability that identical procedures, when applied by different researchers to distinct participant groups, will produce consistent outcomes. It indicates the extent to which findings can be replicated or reproduced (Denzin and Lincoln, 1998:186). Merriam (2009:220) notes that reliability concerns the replicability of research outcomes, yet highlights the inherent challenges of replication in the social sciences, where human behavior is dynamic and unpredictable. However, given the complex and unpredictable nature of the social reality, the approaches to assessing replicability in qualitative research diverge from conventional approaches (Arslan, 2021:70).

Reliability is the capacity of a measurement instrument to produce consistent outcomes at each application instance, as well as the degree to which it reflects the subject matter and maintains the established standards. However, such conceptualizations align closely with quantitative traditions. In quantitative research, reliability is typically addressed through techniques that demonstrate whether the same study, conducted in identical contexts with the same approaches and participants, yields comparable results (Shenton, 2004:71). In qualitative research, however, ensuring reliability requires the researcher to maintain a rigorous approach and a consistent system throughout the data collection process, depending on the nature of the data source. To address this requirement Lincoln and Guba (1985; Guba, 1981) proposed the concepts, dependability and confirmability as qualitative counterparts to conventional reliability. Accordingly, dependability corresponds to internal reliability in quantitative research, while confirmability serves as its external equivalent.

Dependability (Internal reliability)

Guba (1981) introduced the concept of dependability as an alternative to reliability used in the positivist tradition and, following Kellinger (1973), considers it equivalent to the notions of credibility, stability, consistency, predictability, and accuracy. Although Guba (1981) and Lincoln and Guba (1985) discussed reliability using terminology similar to that in the positivist tradition, the understanding of reliability in qualitative research, expressed as dependability, differs in essence. While quantitative research involves measuring or testing a specific phenomenon, statistically calculating findings, and making sense of numerical data, qualitative research, on the other hand, focuses on inquiry, interpretation of expressions, and evaluation of meanings. In this regard, dependability refers to the degree of coherence among meanings derived from data such as field notes or transcripts, based on similar observations conducted by different researchers in the same context (LeCompte and Goetz, 1982; Franklin, Cody, and Ballan, 2010). In this sense, reliability highlights the extent to which data collection and analysis procedures generate the same responses across different participants during the research process (Kirk and Miller, 1986:19).

To claim dependability, the results are expected to align with data obtained in various forms, and the findings are understood to reflect the meanings conveyed by the data source (Arslan, 2021:70). This requirement is consistent with Silverman's (2013) recommendation for researchers to organize their codes in a table and to transcribe interview records in full. In this context, researchers aiming to demonstrate reliability are expected to implement and report all procedures rigorously, beginning with the interview and continuing through the analysis.

Additionally, Lincoln and Guba (1985) propose alternatives to validity and reliability in qualitative research and emphasize the use of stepwise replication and inquiry audit, both of which align with the techniques mentioned above.

In qualitative research, *stepwise replication* serves as an equivalent to the split-half reliability test used in quantitative studies. It involves splitting the research team into two equivalent groups, each analyzing a separate half of the same data set. Given the interpretive nature of qualitative research, both teams compare the emerging categories during the process as they emerge, prior to reaching final conclusions. This comparison allows researchers to exchange ideas at key stages and decide on subsequent steps during the development of findings. Documenting and reporting these communication sessions is essential (Guba, 1981:87). The advisor may serve as a suitable support for a master's student; for a doctoral student, this role may be fulfilled by the advisor and committee members, while in co-authored studies, colleagues can contribute to forming such a team.

Inter-rater reliability refers to the comparison of coding results to determine the extent of agreement between two or more coders who analyze a qualitative dataset and assign codes (Creswell, 2016:576). This practice is frequently criticized by qualitative researchers, aligning with positivist traditions, as well as introducing quantification into qualitative research. The comparison and evaluation of different coding outputs generated by independent coders are assessed using Cohen's Kappa statistic. A value of 0.80 or above is generally considered to indicate strong agreement and acceptable reliability (Guest, MacQueen, and Namey, 2012:14; Yardley, 2008:250). Kappa calculations are performed through statistical software commonly used in quantitative research, such as SPSS, as well as qualitative data analysis programs like MAXQDA and NVivo, which have become increasingly popular in recent years.

Finally, the *audit trail* is also considered one of the dependability criteria (Lincoln and Guba, 1985:301; Denzin and Lincoln, 2018:1380). Accordingly, the explanations regarding inquiry audit provided in the following section (see confirmability) are also applicable here. In this context, audit trail is used not only to address the dependability criterion but confirmability, thus improving the overall trustworthiness of the research.

Confirmability (external reliability)

Confirmability, one of the criteria developed by Lincoln and Guba (1985:324), refers to the degree to which research findings derive from the focus of the study itself rather than from the biases or judgments of the researcher. In classical terminology, this corresponds to objectivity. As a significant indicator of trustworthiness in qualitative research, confirmability is considered the equivalent of external reliability in quantitative traditions (Miles, Huberman, and Saldaña, 2014:272).

Ensuring objectivity in data collection and analysis processes is a critical concern in scientific research. It is often argued that objectivity, which is widely practiced in research, is not easily applicable to qualitative studies due to the potential for researchers to develop a connection with participants or the data (Corbin and Strauss, 2015:95). However, this perspective is often regarded as biased, given the nature of qualitative research involves prolonged engagement with the social world, developing an in-depth understanding, and observing the phenomenon within natural settings. This extended interaction, recognized as a key indicator of trustworthiness, further reinforces the interpretive orientation of qualitative inquiry. While complete objectivity is challenging to achieve in social science research, it remains possible to uphold to a certain extent. Qualitative researchers approach their studies with a conscious awareness of this underlying assumption. Such reflexivity has contributed to the development and recognition of qualitative research as a rigorous and credible tradition within these fields.

Lincoln and Guba (1985:292) identify intersubjective agreement as the primary criterion for establishing confirmability, highlighting when multiple researchers reach a shared understanding of a phenomenon, their collective judgments are considered objective. Furthermore, the transparent and systematic presentation of the methodological procedures followed by the researcher is recognized as another significant criterion. These considerations underscore the conceptual alignment between objectivity and confirmability in qualitative inquiry.

Scientific research generally pursues objectivity through methodological frameworks that emphasize transparency, detailed procedures, replicability, and the minimization of researcher bias. However, qualitative researchers approach this assumption with a different understanding. Qualitative researchers often regard such objectivity as an illusion, recognizing that methodology is inherently tied to those who construct and apply it. Instead of pursuing this type of objectivity, they prioritize the confirmability of the data. To this end, what matters is the ability to trace all constructs, arguments, and empirical claims back to their sources, supported by a clear elaboration of the explicit and implicit reasoning that underpins a coherent and well-substantiated interpretation of the findings (Guba, 1981; Lincoln and Guba, 1985).

Guba (1981:87) and Lincoln and Guba (1985:318) positioned confirmability as an alternative to the widely accepted criterion of objectivity in scientific research, offering a framework supported by various scholars (Erlandson et al., 1993; Miles, Huberman, and Saldaña, 2014) that clarifies how this criterion can be addressed within qualitative inquiry. In addition to the discussion above, confirmability in qualitative research can be addressed through techniques such as audit trail, inquiry audit, or confirmatory audit, which essentially refer to the same approach and are often used interchangeably.

Audit trail or *inquiry audit* refers to a technique grounded in the external review of a research process. Drawing on a business metaphor, Guba (1981) illustrates this method by drawing an analogy between the role of the researcher and that of an auditor, who is responsible for accomplishing two essential tasks. The first involves examining the financial documentation procedures to assess compliance with ethical standards in accounting. The second concerns verifying the accuracy of the recorded data. This verification process begins with ensuring each financial entry is properly justified. Subsequently, the auditor reviews the financial statements to determine whether profitability is accurately represented. In a similar vein, the research auditor examines the findings, conclusions, and implications presented in the study to evaluate the consistency with data and the overall coherence of the analysis (Lincoln and Guba, 1985:318). Based on this understanding, qualitative research requires systematic, transparent, and well-justified documentation addressing any potential ambiguities. When the process reviewed by an external expert, the procedures are expected to demonstrate clarity, consistency, and clear grounding of the conclusions in empirical evidence. As inquiry audits are conducted by a reader or an auditor, the first step of this process in scientific research, particularly in graduate and doctoral theses, is fulfilled through a detailed schematic representation of the methodology section. The second step involves providing a thorough explanation of how the study was completed, including the analytical procedures, findings, conclusions, and appendices. Often described as transparency by some researchers (Roller and Lavrakas, 2015), this technique is also recognized as an effective approach to strengthening dependability in qualitative research (Guba, 1981; Lincoln and Guba, 1985; Denzin and Lincoln, 2018). The key distinction lies in the focus: confirmability involves the verification of outcomes, whereas dependability concerns the coherence and consistency of the research process.

The steps followed to ensure trustworthiness in a study conducted by Arslan (2021:71–72) set a concrete example for the aforementioned principles:

1. Initially, the researcher's competence in the field of application was assessed.
2. Data collection, analysis, and interpretation stages were clearly documented, ensuring procedural transparency.
3. The interviews and analysis were conducted with a strong awareness of the researcher's role, while carefully bracketing prior assumptions and potential biases.
4. The categories and corresponding interpretations developed during the research were continuously reviewed by an expert throughout the process.
5. To enhance credibility, frequent direct quotations from participants were included in the findings.
6. The emerging codes, categories, and concepts were periodically shared with experts for critical feedback and necessary adjustments.

7. All interview recordings and transcripts were stored in digital environments to ensure data preservation and auditability.

8. The researcher's interpretations and inferences regarding the final themes were presented to co-researchers for review and approval.

9. Particular attention was paid to selecting the secondary coder, who was an academic with substantial professional experience in engaging with tourists from diverse backgrounds.

CONCLUSION

Criticisms regarding the validity and reliability or more precisely, the trustworthiness of qualitative research in the social sciences are not a recent phenomenon. However, the growing interest in qualitative research in recent years has raised these debates once again. In scientific inquiry, questions concerning how knowledge is generated and which methodological procedures are employed remain central concerns. This has once again highlighted the significance of validity and reliability, or trustworthiness in qualitative research.

Just as in quantitative research, qualitative studies can also produce valid and reliable results. Given the frequent criticism that qualitative research lacks rigor or relies heavily on subjective interpretations, it becomes essential for researchers to ensure the trustworthiness of the study (Erlandson, Harris, Skipper, and Allen, 1993:131).

Throughout this review, aiming to elaborate validity, reliability, or more precisely trustworthiness in qualitative research within the social sciences, four key criteria including credibility, transferability, dependability and confirmability were introduced as conceptual counterparts to conventional measures, alongside thirteen relevant techniques (see Table 2). Accordingly, a researcher aiming to enhance the trustworthiness of a qualitative study may utilize the criteria of credibility, transferability, dependability, and confirmability. In this regard, credibility, as the qualitative equivalent of internal validity, refers to the extent to which the findings of a qualitative study are perceived as credible by individuals other than the researcher.

The techniques for ensuring credibility covered in this study include prolonged engagement, enabling the researcher to evaluate diverse social structures and identify salient aspects of the research context; persistent observation, facilitating in-depth examination of distinctive characteristics related to the phenomenon under investigation; peer debriefing, offering external evaluation of the objectivity and accuracy of the researcher's interpretations; triangulation, supporting the authenticity of the findings by integrating multiple data sources, methods, and theoretical perspectives; negative case analysis, addressing conflicting data points to enhance interpretative depth; member checking, involving participants in the validation of the findings; and referential adequacy, allowing preliminary assessment of early interpretations once sufficient data have been gathered. The study also discusses transferability as a further validity criterion. Developed as an alternative to external validity, this criterion refers to the applicability of the study's design and/or findings to different participant groups, locations, and time frames by other researchers or individuals interested in the study. Researchers aiming to meet the transferability criterion in their studies require thick description that involves detailed explanations of the phenomena examined within the study, as well as thorough documentation of data collection, analysis, and interpretation of findings. They also require the incorporation of purposeful sampling techniques that enable the selection of data sources that are aligned with the study.

Within the scope of the present study, a further focal point concerns strategies for enhancing trustworthiness in qualitative social science research. Accordingly, enhancing trustworthiness researchers require addressing the dependability criterion, equivalent to internal reliability, and confirmability, equivalent to external reliability. In order to achieve the dependability criterion, researchers are expected to apply techniques such as stepwise replication, a repeated evaluation procedure widely recognized in the literature; audit trail, a strategy for verifying the internal consistency of the research process; and inter-coder agreement, the extent of convergence achieved by multiple researchers during data analysis. As a final point, the study discusses confirmability, commonly defined in the literature as the qualitative equivalent of external reliability. Accordingly, in order to ensure

the confirmability criterion in qualitative research, previously introduced outcome-focused audit trails and triangulation techniques are recognized as appropriate strategies. Further, reflective journals, where researchers record detailed notes on data collection and analysis procedures, offer an additional means of supporting the implementation of all four criteria.

Table 2: Criteria and techniques for validity and reliability OR trustworthiness

Validity		Reliability	
Internal Validity	External Validity	Internal Reliability	External Reliability
<ul style="list-style-type: none"> ✓ Prolonged Engagement ✓ Persistent Observation ✓ Triangulation ✓ Peer Debriefing ✓ Negative Case Analysis ✓ Referential Adequacy ✓ Member Checking ✓ Reflective Journal 	<ul style="list-style-type: none"> ✓ Thick Description ✓ Purposeful Sampling ✓ Reflective Journal 	<ul style="list-style-type: none"> ✓ Audit Trail / Inquiry Audit ✓ Stepwise Replication ✓ Inter-coder Agreement ✓ Reflective Journal 	<ul style="list-style-type: none"> ✓ Audit Trail / Confirmatory Audit ✓ Triangulation ✓ Reflective Journal

Source: Adapted from Guba, 1981; Guba and Lincoln, 1985.

Building on Guba's (1981, p. 88) recommendations, researchers aiming to ensure the trustworthiness of a qualitative study are expected to address essential criteria, such as triangulation or member checking for credibility, audit trail for dependability, and triangulation and reflective journaling for confirmability. In light of this perspective, it can be asserted that the application of at least one dedicated technique for each of the four trustworthiness criteria across the research process contributes significantly to establishing the validity and reliability or more precisely, the trustworthiness of a qualitative study. Furthermore, it is imperative for researchers to acknowledge the interrelated nature of these criteria, as achieving trustworthy outcomes ultimately depends on maintaining methodological rigor throughout the inquiry.

Ethical Statement

This article is the English translation of the original article entitled "Nitel Araştırmalarda Geçerlilik ve Güvenilirlik", authored by Erdal Arslan and published in the Pamukkale University Journal of Social Sciences Institute, Special Issue 1, Issue 51, 2022, pages Ö395–Ö407 (DOI: 10.30794/pausbed.1116878).

This translated version has been prepared with the author's permission and does not include any changes to the content, scope, or findings of the original work. The translation aims to make the study accessible to a broader international audience.

The author declares that the translation adheres to ethical standards and publication principles and does not constitute duplicate publication.

REFERENCES

- Arslan, E. (2021). *Turistik Tüketimin Kimlik İnşasındaki Rolü*. (Yayımlanmamış Doktora Tezi). Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, Eskişehir, Türkiye.
- Arslan, E. (2022). "Nitel Araştırmalarda Geçerlilik ve Güvenilirlik," *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 51 (Özel Sayı 1), Ö395–Ö407. <https://doi.org/10.30794/pausbed.1116878>.
- Blaikie, N. (1993). *Approaches to social enquiry: Advancing knowledge*. USA, Polit Press.
- Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Sage.
- Cook, T.D. & Campbell, D.T. (1979). *Quasi-experimentation: design and analysis issues for field settings*. Rand McNally.
- Creswell, J. W. (2016). *30 essential skills for the qualitative researcher*. Sage.
- Czernek-Marszałek, K., & McCabe, S. (2022). Why qualitative papers get rejected by Annals of Tourism Research?. *Annals of Tourism Research*, 92(103438), 1-3.

- Denzin, N. K. & Lincoln, Y. S. (2018). *The SAGE handbook of qualitative research*. Sage.
- Denzin, N. K. & Lincoln, Y.S. (2008). *Introduction: Entering the Field of Qualitative Research*, (Editörler: Denzin, N. K.; Lincoln, Y. S), İçinde *The landscape of qualitative research: Theories and Issues*, ss.1-34. Sage.
- Erlandson, D. A., Harris, E. L., Skipper, B. L., & Allen, S. D. (1993). *Doing naturalistic inquiry: A guide to methods*. SAGE Publications.
- Franklin, C., Cody, P. A. & Ballan, M. (2010). Reliability and Validity in Qualitative Research (Editör: Thyer, B. A.), İçinde, *The handbook of social work research methods*, ss.273-292. SAGE Publications.
- Gibson, W., & Brown, A. (2009). *Working with qualitative data*. Sage.
- Goodman, S. (2008). The generalizability of discursive research. *Qualitative research in psychology*, 5(4), 265-275.
- Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal*, 29(2), 75-91.
- Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. Sage.
- Janesick, V. J. (2014). Oral history interviewing: Issues and possibilities. (Editör: Leavy, P.), İçinde, *The Oxford handbook of qualitative research*, ss. 300-314. Oxford University Press.
- Kirk, J., & Miller, M. L. (1986). *Reliability and validity in qualitative research*. Sage.
- LeCompte, M. D., & Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. *Review of educational research*, 52(1), 31-60.
- Lincoln, Y. & Guba, E. (1985). *Naturalistic inquiry: Establishing Trustworthiness*, Beverly Hills.
- Matteson, S. M. (2021). Chex Mix™ Data Analysis Activity. *College Teaching*, 69(3), 121-125.
- Maxwell, J. A. (2012). *Qualitative research design: An interactive approach*. Sage
- Merriam, S. B., (2009). *Qualitative research: A guide to design and implementation: revised and expanded from qualitative research and case study applications in education*. The Jossey-Bass Higher and Adult Education Series.
- Miles, B. W., & Jozefowicz-Simbeni, D. M. H. (2010). Naturalistic inquiry, (Editör: Thyer, B. A.), İçinde, *The handbook of social work research methods*, ss.722-745. Sage.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. (3. Baskı). Sage.
- Morse, J. M. (2018). Reframing rigor in qualitative inquiry. In Denzin, N., Lincoln, Y. (Eds.), *SAGE handbook of qualitative inquiry* (5th ed.). Sage.
- Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice*. Sage.
- Phillips, B. D. (2014). Qualitative research disaster, (Editör: Leavy, P.), İçinde, *The Oxford handbook of qualitative research*, ss. 533-556. Oxford University Press.
- Ravitch, S. M., & Carl, N. M. (2019). *Qualitative research: Bridging the conceptual, theoretical, and methodological*. Sage Publications.
- Roller, M. R., & Lavrakas, P. J. (2015). *Applied qualitative research design: A total quality framework approach*. Guilford Publications.
- Schwandt, T. A. (2015). *The SAGE dictionary of qualitative inquiry*. (4. Baskı). SAGE publications.
- Sharan, B. (2009). *Qualitative research: A guide to design and implementation*. Wiley & Sons.
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for information*, 22(2), 63-75.
- Silverman, D. (2013). *Doing qualitative research: A practical handbook*. (4. Baskı). Sage.
- Straus, A., & Corbin, J. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. (4. Baskı). SAGE Publications.
- Yardley, L. (2008). Demonstrating validity in qualitative psychology. (Editör: Smith, J. A.), *Qualitative psychology: A practical guide to research methods*. ss.235–251. Sage.
- Yin, R. K. (2011). *Qualitative research from start to finish*. The Guilford Press.