Quality & Quantity
https://doi.org/10.1007/s11135-025-02461-8

Using formative assessment approaches to enhance rigor and quality in research

J. Patrick Biddix[1] · Brian Bourke[2]

Received: 5 February 2025 / Accepted: 19 October 2025

Abstract
This article suggests a formative assessment approach to enhance rigor and quality in research across paradigms. Acknowledging that no strategy can fully address the compromises inherent in research design, data collection, or analysis, the article recommends formative practices to address these challenges early in the research process. It explores the development of validity as a central concept in cross-paradigm research evaluation and critiques the limited adoption of standardized methods for assessing rigor and quality. The proposed approach enables researchers to address potential issues as they arise, fostering transparency and enhancing the credibility of findings. It identifies specific methods to improve rigor through alignment (visual model) and accuracy (pilot study), and quality through credibility (triangulation) and dependability (audit trail). By incorporating these strategies, researchers can strengthen methodology, ensure reliable results, and promote transparency. The article concludes with recommendations for implementing these approaches across various research traditions to improve rigor and quality.

Keywords Visual model · Pilot · Triangulation · Audit trail · Transparency · Validity

1 Introduction

No strategy or measure fully accounts for the necessary compromises to design, data collection, or analysis in research (Cook and Campbell 1979; Howe 1988). In well-conducted studies, researchers acknowledge challenges to rigor and quality but often rely on

✉ J. Patrick Biddix
pbiddix@utk.edu

Brian Bourke
bbourke@murraystate.edu

[1] Department of Educational Leadership and Policy Studies (ELPS), College of Education, Health and Human Services (CEHHS), University of Tennessee, 1122 Volunteer Blvd., 316 Bailey Education Complex, Knoxville, TN 37996-3430, USA

[2] Department of Educational Studies, Leadership and Counseling, College of Education and Human Services, Murray State University, 3201 Alexander Hall, Murray, KY 42071, USA

Published online: 20 November 2025

post hoc or reactive strategies to address them. Efforts to identify common language and approaches for evaluating rigor and quality in research have seen limited adoption (Morse 2015), despite proposals for alternative terminology (Lincoln and Guba 1985; Shenton 2004), broad approaches (Barker and Pistrang 2005), collections of recommended strategies (Onwuegbuzie and Leech 2007), and conceptual frameworks (Dellinger and Leech 2007; Poortman and Schildkamp 2012). Most recommendations default to validity language (Maxwell 1992) or contrast based on paradigmatic traditions, epistemological stances, and methodologies (Howe 1988; Maxwell, 2004; Shadish et al. 2002). As a result, there is little consensus on standardized methods for documenting and strengthening rigor and quality in research (Biddix 2018).

Morse et al. (2002) observed, “standards and criteria applied at the end of the study cannot direct the research as it is conducted, and thus cannot be used proactively to manage threats to reliability and validity” (p. 16). A proactive approach involves incorporating formative assessment practices throughout the research process, enabling researchers to address challenges early to enhance rigor and quality as a “continuous endeavor” (Chang 2023). Documenting these practices improves transparency and strengthens credibility for the research audience (Akkerman et al. 2008; Poortman and Schildkamp 2012). Through the following subsections, we provide clear and direct definitions of formative assessment, accuracy, and quality. We also clearly articulate the unique contribution that the article makes to assessing quality in research.

1.1 Defining formative assessment

Formative assessment reflects approaches intended to evaluate efforts in-process rather than after their conclusion (Bennett 2011). Based on application of assessing student learning within classroom settings, Black and Wiliam (2009) offered a theory of formative assessment that situates it as on-going reiterative processes of learning and applying lessons to subsequent learning. In applying this concept to conducting quality research, we situate formative assessment as a continuous component of the research process.

1.2 Defining rigor

Rigor in research suggests that procedures are carried out in a manner that reflects methodological accuracy. The concept of rigor is central to the argument of this article, as the use of formative assessment is intended to confirm that rigorous methodological practices are followed in conducting research. The visual model and audit trail elements offered later in this article provide clear approaches to enacting and asserting methodological rigor, enhancing the quality of the study.

1.3 Defining quality

Throughout the article we use the word quality in connection to perceptions of credibility and transparency of research products. Quality is also intertwined with accuracy, as both serve as surrogates for asserting that rigor has been part of the research process (Morse et al. 2002). The researcher is responsible for persuading their audience that their findings and conclusions are worthy of their attention. Additionally, by ensuring that research is conducted in methodologically sound ways, and by clearly documenting those procedures, researchers can be better situated to demonstrate the soundness of their work.

1.4 Original contributions

This article describes the need for commonly accepted approaches to proactively evaluate and document research practices. It begins with a definition and historical overview of validity, a central concept in debates surrounding cross-paradigm research evaluation. Drawing on an extensive review of literature, the article advocates for a formative assessment approach to monitor and refine research practices, identifying four specific strategies to enhance rigor and quality across paradigms. It concludes with essential recommendations for implementation.

2 Review of literature

This section offers an examination of pertinent literature in key areas of validity, the role of formative assessment as an evaluation framework, rigor, and quality. We highlight that while research has often used validity as the primary indicator of a study's soundness, and therefore quality, we argue that a more thorough and overarching approach is warranted given the complexities of contemporary research designs and practices. From there we explore the concept of formative assessment which typically shows up in educational contexts, but we propose its application as a framework to approach ongoing evaluation during the research process. In addition, we offer a deeper explanation of the intertwined concepts of rigor and quality. Through the examination of literature in these key areas, we provide a foundation for understanding how a formative assessment approach can enhance research practices.

2.1 Validity as an evaluative framework

Validity is defined as ensuring accuracy to enhance relevance, meaning, and appropriateness. It is important to acknowledge that the term is not universally accepted across all research paradigms. Campbell and Stanley (1963) described validity as a "basic minimum" practice with origins in the testing movement related to educational research of the 1950s (Cook and Campbell 1979). Although researchers have used different terminology to describe the need, process, and outcomes of evaluating research, there is general conceptual agreement on some key points about the importance of ensuring and demonstrating a study is sound. Cypress (2017), echoing Maxwell (1992), noted, "Validity is not a single, fixed, or universal concept but rather a contingent construct, inescapably grounded in the processes and intentions of particular research methodologies." (p. 257). Sale et al. (2002) and Poortman and Schildkamp (2012) argued that, despite epistemological differences, a common underlying logic of inquiry applies across studies, allowing for shared practices. For Morse et al. (2002), validity is better understood as a verification strategy that focuses on the inferences or interpretations drawn from the data, rather than the data itself. Onwuegbuzie and Leech (2007) preferred the term legitimation over validity in qualitative inquiry, describing its purpose similarly to Maxwell's (1992) contingency phrasing. They noted, "assessing legitimation does not lead to a dichotomous outcome (i.e., valid vs. invalid), but represents an issue of level or degree…Such strategies either help to evaluate legitimation or to increase legitimation, or both" (pp. 238–239).

Guba (1981) highlighted the challenges of applying traditional criteria for evaluating research to constructivist paradigms, noting: "The more one insists on rigor (internal validity), and assures it by control of the sort possible in a laboratory, the less relevance (external validity) one can expect, for the results will apply only in another laboratory" (p. 78). As an alternative, Lincoln and Guba (1985) advanced the concept of "trustworthiness" and identified four aspects comparable to quantitative practices. They defined the term as the extent to which data and analysis are credible (internal validity), transferable (external validity), dependable (reliability), and confirmable (objectivity). Sandelowski (1993) situated validity (and trustworthiness) as a theoretical and persuasive process "whereby the scientist is viewed as having made those practices visible and, therefore, auditable" (p. 2). Morse et al. (2002) critiqued the timing of validity practices, arguing that methods for evaluating trustworthiness often are implemented after a study's completion, which hinders the ability to verify accuracy throughout the process.

2.2 Formative assessment in evaluative practice

Sovacool et al. (2018) referred to the search for common evaluative guidelines as a "fool's errand" due to the diversity and complexity of methods and traditions. However, they noted, "there are practices and guidelines that can *improve* the quality of research and increase the probability of positive impact" (p. 12). Others have suggested that common evaluative criteria are, at best, provisional and evolving. Barker and Pistrang (2005) acknowledged this endeavor is "almost certainly contentious" but believed it may assist in achieving "a new integration of existing ideas" (p. 203). Tashakkori and Teddlie (1998) noted how attention to establishing common concepts can promote quality in inferences, noting "such confidence is a direct result of your ability to 'defend' your conclusion (to other experts, the reader of your research, and so on) by ruling out alternative explanations for your obtained relationships" (p. 68). Morse et al. (2002) argued that identifying common evaluative concepts for research is hindered by the overuse of terms for variations and specific situation, stating, "presently, this situation is confusing and has resulted in a deteriorating ability to actually discern rigor" (p. 15).

Rather than defining separate procedures for qualitative and quantitative traditions (e.g., Babbie 2020; Poorman and Schildkamp 2012; Shenton 2004), proposing summative criteria applied after data collection (Morse et al. 2002), or leaning on validity concepts or language (Maxwell 1992, 2013; Sovacool et al. 2018), this article advocates for a formative assessment approach utilizing a set of common practices. This approach emphasizes continuous evaluation to monitor progress, identify challenges, and refine the research process throughout the study. Greene et al. (1989) distinguished formative from summative evaluation as a "process versus product" intent. In an assessment context, formative evaluation can be identified as "assessment for learning" and is linked with continuous monitoring (Black and Wiliam 2009). This aligns with the focus of program evaluation on comprehensive internal analyses. Bennett (2011) characterized formative assessment as the integration of instrumentation and process, describing formative approaches as "moments of contingency" implemented for the purpose of regulation. They further distinguished decisions based on formative assessment as either intentions (collected but not used) or actions (used for adjustments). A key principle in outlining effective assessment tasks is the ability to support meaningful inferences about learning (Quinlan and Pitt 2021). When applied to the formative assessment of research, this focus enables researchers to draw inferences about rigor and quality. Building on this foundation and grounded in the core concept of validity, a framework using formative assessment is proposed to monitor and evaluate research rigor and quality effectively.

2.3 Rigor as a formative evaluation practice

Formative rigor is defined as a deliberate approach to planning, data collection, and analysis to ensure accuracy. It emphasizes appropriateness, soundness, validity, and consistency during the conceptual, planning, and testing phases of a study. Focusing on rigor involves establishing clear goals, aligning goals with research questions, identifying appropriate data, and applying sound analytic procedures. A rigorous study incorporates multiple principles of good research design, including aligning the research problem, purpose, and questions with the methodology and ensuring the data collection tools and procedures (e.g., survey, form, questionnaire) are accurate. Rigor poses two formative questions:

1. Are the research problem, theory, questions, and method aligned? (Alignment)
2. Do the research procedures accurately identify the problem? (Accuracy)

2.4 Quality as a formative evaluation practice

Formative quality is defined as the implementation of principles and practices to enhance credibility and transparency. It emphasizes trustworthiness, transferability, dependability, and confirmability by systematically documenting decisions from conceptualization to interpretation. It also involves exploring alternative approaches to confirm, challenge, or expand conclusions (Hesse-Biber 2012). A quality study is precisely documented and described (Giraldo and Hernández 2022) to allow for reproducibility, at least conceptually, under similar conditions. Quality can also be demonstrated through comparison by presenting supplemental or additional data to provide confirmation and strengthen trustworthiness. Lincoln and Guba (1985) described quality as an obligation of the researcher "to persuade his or her audience (including self) that the findings of the inquiry are worth paying attention to, worth taking account of" (p. 290). Quality poses two formative questions:

1. Do multiple inquiry approaches support the credibility of the findings? (Credibility)
2. Are the research procedures and decisions precisely documented? (Transparency)

3 Framework development—formative rigor and quality approaches

This section overviews four approaches linked to formative rigor and quality. Each approach is related to evaluative questions, associated concepts from the literature, implementation timing, and specific practices. Table 1 provides a summary.

Table 1 Formative rigor and quality methods with implementation timing and essential recommendations

Evaluative practice	Formative approach	Evaluative question	Associated terminology	Implementation timing	Essential recommendations
Rigor	Visual model	Are the research problem, theory, questions, and method aligned?	Alignment Appropriateness Soundness	Conceptual and planning — a priori (before data collection)	Develop early and maintain across stages to document alignment. Reference and revise as the study progresses; retain earlier drafts as artifacts supporting an audit trail
Rigor	Pilot study	Do the research procedures accurately identify the phenomenon?	Accuracy validity consistency	Feasibility and testing — a priori (before study launch)	Conduct pilots to test sampling, instruments, and analyses. Use findings to refine feasibility, reliability, and equity considerations before launching the main study
Quality	Triangulation	Do multiple inquiry approaches support the credibility of the findings?	Credibility confirmability applicability	Data collection and confirmation — a posteriori (after study launch)	Apply after study launch by comparing sources and methods. Use structured documentation to evaluate convergence and guide interpretation
Quality	Audit trail	Are the research procedures and decisions precisely documented?	Transparency dependability reliability	Data collection and confirmation — *in continuo* (maintain continuously)	Maintain across all stages by recording key decisions, timestamping documents, and retaining artifacts to enhance dependability and transparency

The pilot study is presented here as a single category (including instrument, sampling, and analysis). In Fig. 1, these pilots are disaggregated to illustrate their distinct sequencing within the study timeline

3.1 Visual model

A visual model is a diagram that lists and monitors the procedures of a study to ensure alignment of the research aims and study design with the methodology. Barker and Pistrang (2005) defined alignment as the clear explanation of context and purpose, accompanied by the selection of appropriate methods, which can be enhanced by visual mapping. Visual models, also called research or design diagrams, frequently are used in mixed methods studies to represent the conceptual and methodological nuances of a design (Teddlie and Tashakkori 2003). Utilizing a visual model helps identify and address potential gaps in a study by mapping the research process, highlighting areas that may raise concerns about the validity of the conclusions (Creamer 2021). They also can be used to show associative, causal, or temporal relations with constructs (Moher et al. 2010; Stamenkov 2023).

Shenton (2004) referred to visual models as diagrammatic audit trails, which use a traditional (data-oriented) or a holistic (theoretical) approach to trace research decisions and procedures to ensure alignment and transparency. In quantitative research, this approach often is reflected in the documentation for randomized trials. Moher et al. (2010), emphasizing the importance of clear, complete, and transparent reporting, recommended using a diagram to illustrate adherence to CONSORT (Consolidated Standards of Reporting Trials). Diagrams are intended to promote internal and external validity, which should be established during planning and maintained throughout the study. For quantitative research, Gilbert et al. (2013) as well as Creamer (2021, 2024) recommended diagramming procedures when using complex methods and sources. Leech and Onwuegbuzie (2010) suggested "outlining" for mixed methods, using diagrammed process models to illustrate connections between research planning and implementation to enhance coherence.

A robust visual model demonstrates rigor by providing a formative account of the study, ensuring appropriate choices and sound implementation. It documents the conceptual, design, and revision procedures, allowing for ongoing adjustments throughout the process (Creamer 2021). At the conclusion of the study, a completed visual model can serve as a visual audit trail, guiding the reader as an "auditor" in following the logic of inquiry (Morse et al. 2002; Poortman and Schildkamp 2012). This also can promote reporting transparency by demonstrating how theoretical framing and empirical evidence develop in a structured and deliberate manner (Akkerman et al. 2008), reinforcing rigor and supporting the "consistent, conscientious, and reasoned conceptual development of the issue under investigation" (Øby 2024, p. 2). Pragmatically, Ivankova et al. (2006) proposed ten rules for diagramming visual models while Creamer (2024) more recently differentiated types of visual displays such as those incorporating temporal or physical mapping elements to enhance validity. As an internal product, the visual model ensures consistency and accuracy, while as an external product, it can be published as an appendix with the study.

3.2 Pilot study

A pilot study is the process of implementing research procedures on a smaller scale prior to collecting final data (Malmqvist et al. 2019). As a formative rigor approach, a pilot study can be beneficial for ensuring accuracy and readability of questions, evaluating measures to ensure they represent the concepts they are intended to evidence, and verifying the analytic procedures yield results aligned with the research problem, purpose, and design. A pilot study can address internal efficacy and external credibility or trustworthiness of a study (Fowler 2014), which are contingent considerations of validity (Cypress 2017). Attending to validity means demonstrating that data collection process and tool/s accurately measure, assess, or uncover what they are intended to reveal. A pilot study process also enables researchers to demonstrate alignment by making connections from the literature review and theoretical framework into the methodology and through findings (Giraldo and Hernández 2022; Marshall and Rossman 2014). A robust pilot study is a three-part process that includes sampling, instrument testing, and preliminary analyses.

3.2.1 Sampling pilot

A sampling pilot is used to apply data collection procedures to ensure the sources accurately and adequately represent the concepts studied. This process is important for confirming the data collection plan is feasible and that intended data sources (e.g., surveys, database extraction, interviews, observations) can be accessed and obtained. The sampling approach varies based on the research methodology and data availability, but it should always be "intended to maximize efficiency and validity" (Palinkas et al. 2015, p. 2). Probabilistic or random sampling methods (e.g., simple or stratified) are generally more suitable for quantitative research (Fowler 2014). Qualitative studies often benefit from purposive approaches, such as critical case sampling, which identifies extreme or highly informative cases (Patton 2014). For mixed methods studies, Palinkas et al. (2015) recommended using a narrowly focused, purposive qualitative sample to complement a broader focused, probability quantitative sample.

When data are collected from individuals, pilot participants complete the instrument and provide feedback on question wording, accuracy in measuring the concept(s) studied, and other aspects of the data collection process (Tabachnick, Fidell and Ullman 2013; Turner 2010). For example, Chang (2023) implemented interviews to validate scale items by contextualizing participant experiences. They noted that this approach supports deeper and broader interpretations across time, occasion, and context. Alternatively, the instrument might be distributed with a different representative group (Giraldo and Hernández 2022), at a different time or location (Seidman 2006). In his ethnographic study of seafarer communities, Sampson (2004) found simply walking on the cargo ship and familiarizing himself with the locations and items helped him address unanticipated challenges before full data collection. He reflected, "I learned some things that I had not anticipated and, in the way of all things, found that it was harder than I expected to learn about the things that initially seemed the most obvious and least problematic" (p. 393). Sampson later concluded, "The experience of this pilot study and my earlier experiences of entering the 'field' of the ship relatively 'blind' have convinced me that pilot work is invaluable in conducting ethnographic, as well as other forms of qualitative research (p. 398).

3.2.2 Instrument pilot

An instrument pilot involves evaluating the data collection tools to ensure they accurately and consistently identify the concepts studied. This may involve testing surveys, data logs, observation forms, questionnaires, or other data collection procedures (Fowler 2014). Several evaluative principles can be applied, even when the researcher is the instrument (Emerson et al. 2011; Patton 2014). This holistic view aligns with both Fowler's (2014) and Field's (2013) differentiation of internal efficacy and external extrapolation, as well as Lincoln and Guba's (1985) and later Shenton's (2004) parallel forms of ensuring validity, and Seidman's (2006) concept of evaluating authenticity through "questioning" the instrument. Desimone (2009) described the importance of pilot testing in both qualitative and quantitative methods, stating, "Analogous to establishing interrater reliability, interviews and surveys should undergo proper development and piloting, which may include focus groups, critical review, cognitive laboratory interviews, and pretesting" (p. 192). However, Field cautioned it is nearly impossible for an instrument to be 100% valid. As a result, validity often is described as the "degree" to which a concept is accurately measured or represented. An instrument pilot provides several forms of validity.

1. Content validity considers the entire instrument, asking whether the measures fully and accurately represent the concepts (i.e., theoretical-based empirical indicators) it is designed to capture (Babbie 2020; Denzin 2017). This process is sometimes also referred to as expert review (Denzin and Lincoln 2011). Face validity addresses a similar question but relies on non-content experts for validation.
2. Theoretical validity assesses both descriptive validity (concept accuracy) and interpretive validity (relationships between concepts) to ensure the explanation derived from the data aligns with the research problem. In quantitative research, this process can evidence general causation, (i.e., seeking broad, replicable patterns), while in qualitative research this can evidence local causation (i.e., exploring unique, context-specific processes) (Johnson, et al. 2019; Maxwell 1992). Procedurally, this may involve testing different interpretations using alternative theories or comparing individual items or questions with other tests measuring similar concepts (Campbell and Stanley 1963; Malmqvist et al. 2019).
3. Criterion validity considers the individual items, questions, or factors of an instrument, questioning the extent to which a measure aligns with a "gold standard" of measures for this concept, if one exists. It asks, how well does this instrument perform against others that measure the same variables? This technique can be challenging with qualitative studies, which are more context- and content-specific and as a result, seldom standardized (Patton 2014). An alternate way to evaluate criterion validity for a survey or questionnaire is to ask several of the questions in the same way in a similar context and examine the results.
4. Equity-centered validity considers the research context, including the history of the location, the program, and people, and how power is formally and informally distributed (Waterman et al. 2023). Culturally responsive evaluation practices emphasize conducting research *with* communities, rather than *on* them (Kim 2010; Hood et al. 2015). When the researcher is acting as the instrument, piloting the instrument helps assess field readiness and ethical considerations, including cultural competency (Sousa and Rojjanasrirat 2011). An equity-centered validity approach may involve inviting the target population or cultural experts to review instrumentation, helping prevent harm and identifying outliers or overlooked perspectives that may further perpetuate marginalization (Waterman et al. 2023).

3.2.3 Analysis pilot

An analysis pilot involves analyzing pilot data to determine if the data accurately and reliably capture the intended concepts. This helps establish measurement validity as data are collected and verified "during the course of research… as an attribute of the process (means) of reaching those end results" (Mertens 2019, p. 77). Quantitative researchers may use several forms of reliability, including test–retest (consistency over time), parallel forms (comparison of similar instruments) and internal consistency (consistency across items) (Fowler 2014). Qualitative researchers may implement inter-rater or inter-coder reliability, where multiple observers collect or review data and compare observations or themes (O'Conner and Joffe 2020). An analysis pilot also enhances efficiency by allowing researchers to confirm that the identified constructs or phenomena are accurately captured before proceeding with a full study. For example, pilot survey responses that are too similar may indicate an inadequate measure or the need to broaden the scale (Artino et al. 2014). Similarly, pilot interview responses that reveal additional insights may suggest the need for an updated protocol (Palinkas et al. 2015).

One analysis pilot method that can be used across multiple types of direct interaction research is member checking or asking participants to confirm the authenticity or accuracy of data and/or conclusions (Lincoln and Guba 1985). Procedurally this involves seeking participant input on the data, categories and themes, interpretations and conclusions to "assessing intentionality" and offer an opportunity to correct errors or provide additional information. Johnson (1997) referred to member checking as "participant feedback" and noted a key purpose is to reduce miscommunication and ensure accuracy. Morse (2015) suggested that member checking should be integrated throughout a study as a formative process, cautioning that summative checks could put the researcher in a difficult position if the participant disagrees with the interpretation.

3.3 Triangulation

Triangulation is defined as the use of an additional inquiry approach to enhance the credibility, confirmability, and applicability of findings. As Mathison (1988) noted, "Good research practice obligates the researcher to triangulate, that is, to use multiple methods, data sources, and researchers to enhance the validity of research findings" (p. 13). Onwuegbuzie and Leech (2007) regarded triangulation as "[reducing] the possibility of chance associations, as well as of systematic biases prevailing (p. 239). In this evaluative context, triangulation aligns with the concept of cross-validation (Maxwell 2013; Patton 2014). By integrating additional inquiry approaches, researchers can complement or verify findings, thereby improving research quality and strengthening credibility across different settings. Conceptually, this definition resonates with Flick's (2017) perspective on triangulated approaches as a framework for enhancing research quality and depth. However, it does not require that the approaches be independent or integrated, as in a mixed methods design[1] (Creswell and Creswell 2017; Tashakkori et al. 2020). This narrower, evaluative view of triangulation (Silverman 1985; Bryman 2002), reflects Stamenkov's (2022, 2023) view of additional inquiry through triangulation as "value-added tool" for enhancing research quality.

Several implementation considerations apply to this definition. In addition to methods triangulation, Bryman's (2016) caution about evaluating the compatibility and limitations of different approaches extends to other forms of inquiry, such as involving multiple investigators, applying different theories, or using various data sources. Careful interpretation of different approaches, informed by factors like equity-centered validity and culturally responsive evaluation (Waterman et al. 2023), should take into account the contexts and

[1] Contemporary mixed methods scholars argue that the term triangulation contradicts broader goals of exploration, discovery, and understanding subjective meanings (Fetters and Molina-Azorin 2017). While Greene et al. (1989) viewed triangulation as a strategy for promoting confirmation and corroboration, they considered its use in research to be a "constrained-narrow" design. Bryman (2001; 2002) argued the primary goals of triangulation are aligned with achieving convergent validity, critiquing it as a naively realist position. This view aligns with Flick's (2017) perspective that a strong triangulation program should generate additional knowledge, rather than merely confirming what is already known through the convergence of findings.

mechanisms through which data is produced. Additionally, the rationale and timing for incorporating triangulation must be clearly explained. It is most effective during the design phase, allowing researchers to document and adjust the procedures, while maintaining flexibility to explore alternative explanations. However, drawing on classic mixed methods designs (Creswell 2004; Teddlie and Tashakkori 2003), inquiry approaches can be implemented sequentially, simultaneously, or with an embedded or nested design.

Four inquiry approaches are proposed as formative practices to enhance quality and credibility across methodologies, building on Denzin's (1970) foundational recommendations. These align with Lincoln and Guba's (1985) view of triangulation as a strategy to strengthen credibility (or "believability"). While several works cited focus on comprehensive design, they offer valuable insights for implementing formative evaluation.

1. Method triangulation refers to employing various methods to generate data. Examples include combining responses from surveys and existing datasets (within-methods), utilizing interviews and observations (within-methods), or integrating survey responses with interviews (between-methods). Shenton (2004) noted the importance of considering the limitations and opportunities of both within- and/or between-method designs, as "distinct characteristics also result in individual strengths" (p. 65). Arias Valencia (2022) provided pragmatic advice for implementation, emphasizing consistency in simultaneous or sequential approaches. Regardless of the strategy, they noted as a quality evaluation "data should converge, *i.e.*, they should not contradict, although conserving their multiple variations." (p. 7)
2. Investigator triangulation involves two or more researchers who independently collect data or conduct analyses for comparison and validation (Denzin 1989). Flick (2017) recommended using investigators from differing theoretical, epistemological, and/or methodological backgrounds. Archibald (2016) cited potential "multilevel" benefits of involving additional researchers – especially in mixed methods designs, where it can be used to compensate for lone investigators deficits, enhance inference quality, and explore tensions influencing study design, analysis, and interpretation.
3. Theory triangulation incorporates multiple theories to support or refute data interpretations. Addressing plausible rival explanations to the research question and finding evidence to reject them creates more credible findings (Yin 2009). Flick (2017) described this as often co-occurring with investigator triangulation, where different perspectives can be brought to approach a problem on several levels. Thurmond (2001) regarded the purpose of theoretical triangulation as increasing confidence by testing data findings against an opposing hypothesis or theoretical explanation. However, the concepts must be clearly defined and justified as relevant to the phenomenon under study (Greene and McClintock 1985; Flick 2017).
4. Data triangulation involves collecting information from different individuals or sources and is a verification practice. Shenton (2004), drawing on Van Maanen's (1979) recommendations for cross-checking data among informants, emphasized that this approach allows individual data points to be verified against others, leading to a richer understanding of the data and findings. Chang (2023) demonstrated this approach by verifying survey items with interview data. Working with diverse informants or a range of documents also can help corroborate findings, enhancing the credibility and depth of interpretations (Patton 2014). Noble and Heale (2019) further noted that incorporating

multiple data sources is effective for refuting, confirming, or explaining findings, thereby enhancing confidence in the results.This consideration aligns with Campbell and Fiske's (1959) assertion that validity should be demonstrated through both convergent and discriminant evidence.

4 Audit trail

An audit trail is a narrative account of key decisions related to research design and data analysis. An audit trail evidences quality by demonstrating transparency, providing dependability, and promoting reliability through the systematic documentation of decisions (Morse et al. 2002; Morse 2015). Sandelowski (1993) emphasized making research practices visible by "[leaving] a clear decision trail" that allows another researcher to "follow the progression of events in the study and understand their logic" (p. 34). Such documentation enables researchers to verify a study's accuracy and/or to replicate or potentially reproduce results with a similar data sample (Kirk and Miller 1986; Nowell et al. 2017). Replication in research is sometimes criticized as a fallacy (LeBeau et al. 2021) or considered unsuitable for interpretive studies (Morse 2015). However, LeBeau, Ellison, and Aloe emphasized the distinction between reproducibility and replicability, stating, "a reproducible analysis does not ensure that the same result will be obtained if another sample is obtained from the sample population. An analysis framework that is reproducible will allow for replication analyses to be done more efficiently" (pp. 201–202).

Lincoln and Guba (1982) described an "educational audit" journal as a reliability technique to verify that inquiry products are substantiated by data, emphasizing the importance of both process and product for ensuring trustworthy results. Seale (1999), Tracy (2010), and Twining et al. (2017), regarded an audit trail as essential for transparency. Twining et al. noted it should provide an "explicit description" of the operations as related to the empirical data, allowing the reader to "follow each step of the analysis" (p. A6). For quantitative analysis, documentation should include key methodological decisions about sampling, data cleaning, variable manipulation, assumptions related to statistical tests, and model specification (Field 2013). For qualitative analysis, it typically focuses on the coding scheme to provide a detailed account of how the researchers generated or produced and then processed (i.e., made meaning of) findings (Bowen 2009). Glaser and Strauss (1967) suggested "criteria of judgment be based instead on the detailed elements of the actual strategies used for collecting, coding, analyzing, and presenting data" (p. 224). Creswell and Miller (2000) recommended keeping a separate document to record the stages of a study from conceptualization to publication. Morrow (2005) described this as "carefully tracking the emerging research design" and maintaining a "detailed chronology of research activities" (p. 252). This aligns with Onwuegbuzie and Leech's (2007) view that documenting and clearly delineating processes is essential to promote legitimacy.

5 Discussion—essential recommendations

This section provides four essential recommendations that should be minimally implemented to enhance rigor and quality across research and assessment studies. Additional guidance on implementation and timing is also included.

5.1 Visual model

A visual model should be developed early and maintained to ensure rigor from conceptualization through dissemination. Initial drafting occurs before data collection, but the model should be updated across stages to document alignment among research questions, design, and findings. As an active document, the visual model is continually referenced, annotated, and revised as the study progresses. Earlier drafts can be retained as study artifacts and serve as the foundation for an audit trail. This is particularly useful in complex studies where the researcher may need to revisit the study's goals for alignment as it progresses (Poortman and Schildkamp 2012). A visual model can also be referenced during external review and included in the revisions process as a figure or appendix. Complete examples of visual models are uncommon in published research aside from diagrams of conceptual frameworks, path models, or mixed methods design figures. Plano-Clark and Creswell (2008) provided related examples for creating research diagrams, using Greene et al.'s (1989) guidance for methodological notations. Carcary's (2009) descriptions of audit trail components are also useful for identifying which aspects of a study to document. Recently, Creamer and Edwards (2022) edited a special edition of *Methods in Psychology*, which focused on approaches to using visual models in mixed methods research.

5.2 Pilot study

A pilot study, consisting minimally of instrument and analysis pilots, should be implemented to promote rigor during the feasibility and testing phases of a study. A pilot study can be used to demonstrate content, face, theoretical, criterion, and equity-based validity. Further, it can promote efficiency by helping to ensure that the design, planning, and analytic procedures are efficacious once implemented for a fuller scale project. Testing the instrument and analyzing the data it generates helps to ensure yield usable, research-supportive, and/or theoretically aligned results. By piloting with a sample approximating the population, feedback can be sought from majority and non-majority pilot participants to promote equity-centered and ethically considered research (Waterman et al. 2023). Desimone (2009) recommended several methods for producing sound survey, interview, or observation instruments. In the notes of the article, they stressed the importance of wording, order, format, and response choices as influencing self-reported behaviors and attitudes in interviews and surveys.

5.3 Triangulation

A triangulation inquiry approach should be included as a value-added tool (Stamenkov 2022, 2023) to verify findings and strengthen credibility. Researchers can choose from various triangulation approaches based on study goals and efficiency, including using multiple methods, involving additional investigators, exploring alternative theoretical explanations, or gathering data from different sources. Researchers should carefully consider and report convergence, inconsistencies, and contradictions when interpreting findings (Mathison 1988). Poteat et al. (2017) demonstrated how to incorporate multiple data sources in a review of research on Gay-Straight Alliances (GSA). The researchers supplemented youth-reported data with advisor-reported data, records, and input from additional stakeholders to provide a more comprehensive view of how GSAs operate within in a larger school and district context. They found using multiple data sources helped account for variability in activities and experiences. Morgan (2019) recommended using a cross tabulated table to document the results of triangulation strategies, which can be adapted to various forms of inquiry. Campbell et al. (2020) proposed a color-coding system to show data convergence and divergence within and across methods, which could also be used with the table and incorporated as an audit trail to demonstrate quality. For Stamenkov (2023), visual modeling is most effective when paired with triangulation, where the researcher maps the theory and logic of the study and then seeks to verify or validate the findings using additional inquiry approaches.

5.4 Implications, limitations, and future directions

We now offer a critical examination of how formative assessment frameworks intersect with emerging issues, such as equity-centered validity, open-science transparency, and methodological pluralism. This examination is offered through three novel insights of equity in formative practices, transparency and reproducibility, and adaptive frameworks for mixed methods.

5.4.1 Equity in formative practices

Equity in formative practices suggests proactive and continuous efforts being made to ensure all individuals have access to resources they need. One element to address is equity-centered validity, which includes considerations about the research context, including the history of a place and its people, along with power structures. An audit trail can serve to bring about awareness of power dynamics and prompt culturally responsive adjustments, which rely on the researcher being keenly aware of their positionality (Bourke 2014). When utilizing both sampling and instrument pilots, researchers can receive crucial feedback from members of dominant and marginalized populations.

5.4.2 Transparency and reproducibility

Visual modeling can provide actionable documentation that fully supports open-science principles. The use of the proposed formative assessment process enhances transparency provided the researcher addresses their positionality within the research project and process (Bourke 2014). While Bourke (2014) notes that a researcher's positionality does have an influence on both the research process and the product, the influence of their subjectivities does not preclude the reproducibility of a study; much in the same way a change of context where a subsequent study is conducted does not preclude reproducibility.

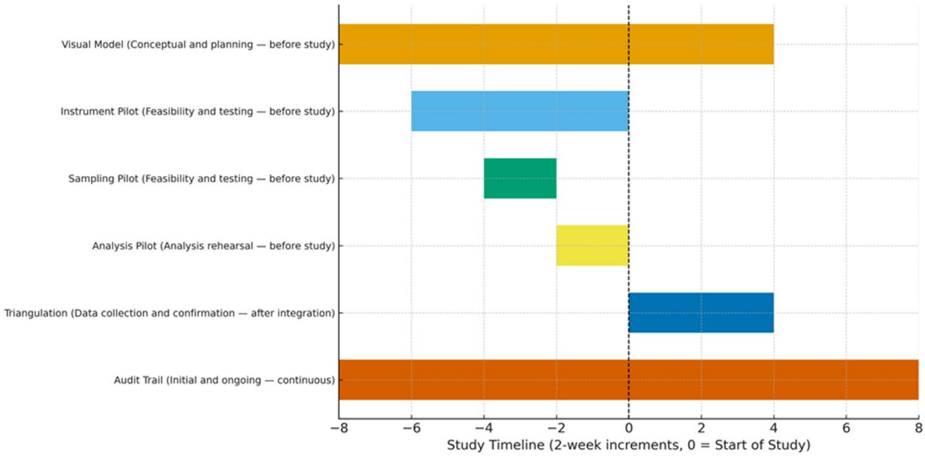

Fig. 1 Procedural workflows for formative practices. *Note* This figure depicts a sample study as a Gantt-style chart in 2-week increments. Visual Model (–8 to + 4), Instrument Pilot (–6 to 0), Sampling Pilot (–4 to –2), Analysis Pilot (–2 to 0), Triangulation (0 to 4), and Audit Trail (–8 to + 8, continuous). Negative numbers indicate weeks prior to the start of study

5.4.3 Adaptive frameworks for mixed methods

Rooted in the proposed formative assessment approach is a perspective that no research method in inherently superior to any other, and that having an understanding of a breadth of methodological approaches and practices enables a researcher to better adapt in their research practice (Biddix 2018). By embracing the applicability of a variety of approaches as an adaptive strategy, researchers are able to respond to complex research settings, along with emergent findings.

5.4.4 Timing of formative assessment activities

Each of the aforementioned activities play crucial roles in the formative assessment framework we have proposed. To aid in its adoption and use in research, we suggest offer timing of each activity relative to the start of data collection. Figure 1 provides a visual representation of this timeline, presented in 2-week increments as a sample workflow to illustrate how formative practices may be sequenced. Actual timing may vary by study. The researcher should conduction the instrument pilot two to four weeks prior to data collection, with sampling pilot activities occurring one to two weeks before study launch. As the results of those a priori efforts are integrated into the research design, the researcher should utilize an analysis pilot, which occurs concurrently, where triangulation should be a posterior implementation, taking place immediately after each integration of the analysis pilot. Finally, the audit trail should be maintained throughout the entire timeline.

The proposed timeline is intended to offer the researcher a quantitatively driven plan that can aid the utilization of the formative assessment processes. Our hope is that this helps the researcher move from abstract, qualitative descriptions to more direct, actionable timings of key activities.

6 Final considerations

While Wolcott (1990) contended, “I do not accept validity as a valid criterion for guiding or judging my work” (p. 148), he acknowledged its value in raising important questions. Proposing a common set of formative practices for ensuring rigor and quality offers standardized methods for evaluating rigor throughout the research process, independent of method-specific procedures. Additionally, this approach emphasizes the importance of implementation timing, advocating for continuous monitoring at all stages of the research process.

In reviewing the research on validity, rigor, and quality across research traditions, transparency emerged as a unifying principle. Transparency involves openly sharing methodological details, decisions, and criteria to ensure clarity and reproducibility. Clear reporting strengthens credibility, while a lack of transparency can undermine trust in the findings (Aguinis et al. 2018; Øby 2024). Øby (2024) noted, “The capacity for other researchers to draw similar conclusions to the original authors about a study’s results is hindered when transparency is lacking. Conversely, methodological rigor that is transparently demonstrated through clear reporting supports claims of validity” (p. 2).

The four formative approaches proposed in this article align with the principles of transparency by revealing challenges, problems, and necessary corrections through ongoing monitoring during the research process. Uncovering these “moments of contingency” (Bennett 2011) enables timely adjustments and fosters meaningful insights during the research process (Quinlan and Pitt 2021), which enrich our understanding of the research phenomenon. This continuous assessment not only identifies and addresses issues promptly but also demonstrates transparency for reviewers and readers through clear documentation, assuring the integrity of the research process while promoting rigor and quality.

Author contributions The lead author conceptualized the study and drafted the initial manuscript. Both authors contributed to revisions and approved the final version. Generative AI (ChatGPT) was used to support language refinement and organization. The authors carefully reviewed the output to ensure accuracy and integrity.

Funding The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Conflict of interest The author has no relevant financial or non-financial interests to disclose.

References

Akkerman, S., Admiraal, W., Brekelmans, M., Oost, H.: Auditing quality of research in social sciences. Qual. Quant. **42**, 257–274 (2008)

Aguinis, H., Ramani, R. S., & Alabduljader, N. (2018). What you see is what you get?Enhancing methodological transparency in management research. Academy of ManagementAnnals, *12*(1), 83–110.

Archibald, M.M.: Investigator triangulation: a collaborative strategy with potential for mixed methods research. J. Mix. Methods Res. **10**(3), 228–250 (2016)

Arias Valencia, M.M.: Principles, scope, and limitations of the methodological triangulation. Investig. Educ. Enferm. **40**(2), e03 (2022). https://doi.org/10.17533/udea.iee.v40n2e03

Artino, A.R., La Rochelle, J.S., Dezee, K.J., Gehlbach, H.: Developing questionnaires for educational research: AMEE guide no. 87. Med. Teach. **36**(6), 463–474 (2014)

Babbie, E.R.: The Practice of Social Research, 15th edn. Cengage Learning, Boston (2020)

Barker, C., Pistrang, N.: Quality criteria under methodological pluralism: implications for conducting and evaluating research. Am. J. Community Psychol. **35**, 201–212 (2005)

Bennett, R.E.: Formative assessment: a critical review. Assess. Educ. Princ. Policy Pract. **18**(1), 5–25 (2011)

Biddix, J.P.: Research Methods and Applications for Student Affairs. Jossey-Bass, Hoboken (2018)

Black, P., Wiliam, D.: Developing the theory of formative assessment. Edu Assess Evaluat Account **21**, 5–31 (2009)

Bourke, B.: Positionality: reflecting on the research process. Qual. Rep. **19**(33), 1–9 (2014)

Bowen, G.A.: Supporting a grounded theory with an audit trail: an illustration. Int. J. Soc. Res. Methodol. **12**(4), 305–316 (2009)

Bryman, A.: Social research methods. Oxford University Press, Oxford (2001)

Bryman, A.: Triangulation. In: Lewis-Beck, M., Bryman, A.E., Liao, T.F. (eds.) The SAGE Encyclopedia of Social Science Research Methods, pp. 1123–1125. SAGE Publications, Los Angeles (2002)

Bryman, A.: Social Research Methods, 5th edn. Oxford University Press, Oxford (2016)

Campbell, D.T., Fiske, D.W.: Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol. Bull. **56**(2), 81–105 (1959)

Campbell, D.T., Stanley, J.C., Gage, N.L.: Experimental and Quasi-Experimental Designs for Research. Houghton, Mifflin and Company, Boston (1963)

Campbell, R., Goodman-Williams, R., Feeney, H., Fehler-Cabral, G.: Assessing triangulation across methodologies, methods, and stakeholder groups: the joys, woes, and politics of interpreting convergent and divergent data. Am. J. Eval. **41**(1), 125–144 (2020)

Carcary, M.: The research audit trial—enhancing trustworthiness in qualitative inquiry. Electron. J. Bus. Res. Methods **7**(1), 11–24 (2009)

Chang, W.C.: Validation of the teaching equity enactment scenario scale in Singapore: a mixed-methods convergent study. Qual. Quant. **57**(6), 5257–5282 (2023)

Cook, T.D., Campbell, D.T.: Quasi-Experimentation: Design and Analysis Issues for Field Settings. Houghton Mifflin, Boston (1979)

Creamer, E.G.: Leveraging an integrated visual display for case-based analysis in mixed method research. Int J Qual Methods (2021). https://doi.org/10.1177/16094069211059000

Creamer, E.G.: Visual Displays in Qualitative and Mixed Method Research: A Comprehensive Guide. Taylor & Francis, Milton Park (2024)

Creamer, E.G., Edwards, C.D.: Introduction to the special issue, innovative approaches to visual methods in mixed method research in psychological fields. Methods Psychol. **6**, 100090 (2022)

Creswell, J.W., Creswell, J.D.: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. SAGE Publications, Los Angeles (2017)

Creswell, J. W., Fetters, M. D., & Ivankova, N. V. (2004). Designing a mixed methods study inprimary care. The Annals of Family Medicine, *2*(1), 7–12.

Creswell, J.W., Miller, D.L.: Determining validity in qualitative inquiry. Theory Pract. **39**(3), 124–130 (2000)

Cypress, B.S.: Rigor or reliability and validity in qualitative research: perspectives, strategies, reconceptualization, and recommendations. Dimens. Crit. Care Nurs. **36**(4), 253–263 (2017)

Dellinger, A.B., Leech, N.L.: Toward a unified validation framework in mixed methods research. J. Mixed Methods Res. **1**(4), 309–332 (2007)

Denzin, N.K.: The Research Act: A Theoretical Introduction to Sociological Methods. Aldine Publishing Company, New York City (1970)

Denzin, N.K.: The Research Act: A Theoretical Introduction to Sociological Methods, 3rd edn. Prentice Hall, Upper Saddle River (1989)

Denzin, N.K.: The Research Act: A Theoretical Introduction to Sociological Methods. Routledge, Milton Park (2017)

Denzin, N.K., Lincoln, Y.S.: The SAGE Handbook of Qualitative Research. SAGE Publications, Los Angeles (2011)
Desimone, L.M.: Improving impact studies of teachers' professional development: toward better conceptualizations and measures. Educ. Res. **38**(3), 181–199 (2009)
Emerson, R.M., Fretz, R.I., Shaw, L.L.: Writing Ethnographic Fieldnotes. University of Chicago Press, Chicago (2011)
Fetters, M.D., Molina-Azorin, J.F.: The journal of mixed methods research starts a new decade: perspectives of past editors on the current state of the field and future directions. J. Mixed Methods Res. **11**(4), 423–432 (2017)
Field, A.: Discovering Statistics Using IBM SPSS Statistics, 4th edn. SAGE Publications, Los Angeles (2013)
Flick, U.: Mantras and myths: the disenchantment of mixed-methods research and revisiting triangulation as a perspective. Qual. Inq. **23**(1), 46–57 (2017). https://doi.org/10.1177/1077800416655827
Fowler, F.J.: Survey Research Methods, 5th edn. SAGE Publications, Los Angeles (2014)
Gilbert, L.S., Jackson, K., di Gregorio, S.: Tools for analyzing qualitative data: the history and relevance of qualitative data analysis software. In: Spector, J., Merrill, M., Elen, J., Bishop, M. (eds.) Handbook of Research on Educational Communications and Technology, pp. 221–236. Springer, New York (2013)
Giraldo, C.G., Hernández, G.C.: A methodological proposal for a strategic design investigation. Qual. Quant. **56**(5), 3325–3336 (2022)
Glaser, B., Strauss, A.: The Discovery of Grounded Theory Strategies for Qualitative Research. Sociology Press, Mill Valley (1967)
Greene, J.C., McClintock, C.: Triangulation in evaluation: design and analysis issues. Eval. Rev. **9**(5), 523–545 (1985)
Greene, J.C., Caracelli, V.J., Graham, W.F.: Toward a conceptual framework for mixed-method evaluation designs. Educ. Eval. Policy Anal. **11**(3), 255–274 (1989)
Guba, E.: Criteria for assessing the trustworthiness of naturalistic inquiries. Educ. Technol. Res. Dev. **29**(2), 75–91 (1981)
Hesse-Biber, S. (2012). Feminist approaches to triangulation: Uncovering subjugatedknowledge and fostering social change in mixed methods research. Journal of Mixed MethodsResearch, *6*(2), 137–146.
Hood, S., Hopson, R.K., Kirkhart, K.E.: Utilization-Focused Evaluation: A Comprehensive Guide to Theory and Practice. SAGE Publications, Los Angeles (2015)
Howe, K.R.: Against the quantitative-qualitative incompatibility thesis or dogmas die hard. Educ. Res. **17**(8), 10–16 (1988)
Ivankova, N.V., Creswell, J.W., Stick, S.L.: Using mixed-methods sequential explanatory design: from theory to practice. Field Methods **18**(1), 3–20 (2006)
Johnson, R.B.: Examining the validity structure of qualitative research. Education **118**(2), 282–292 (1997)
Johnson, R.B., Russo, F., Schoonenboom, J.: Causation in mixed methods research: the meeting of philosophy, science, and practice. J. Mixed Methods Res. **13**(2), 143–162 (2019)
Kim, J.: Quantitative Research Methods in the Social Sciences. Oxford University Press, Oxford (2010)
Kirk, J., Miller, M.L.: Reliability and Validity in Qualitative Research. SAGE Publications, Los Angeles (1986)
LeBeau, B., Ellison, S., Aloe, A.M.: Reproducible analyses in education research. Rev. Res. Educ. **45**(1), 195–222 (2021)
Leech, N.L., Onwuegbuzie, A.J.: Guidelines for conducting and reporting mixed research in the field of counseling and beyond. J. Couns. Dev. **88**(1), 61–69 (2010)
Lincoln, Y.S., Guba, E.G.: Naturalistic Inquiry. SAGE Publications, Los Angeles (1982)
Lincoln, Y.S., Guba, E.G.: Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches. Jossey-Bass, San Francisco (1985)
Malmqvist, J., Hellberg, K., Möllås, G., Rose, R., Shevlin, M.: Conducting the pilot study: a neglected part of the research process? Methodological findings supporting the importance of piloting in qualitative research studies. Int J Qual Methods **18**, 1–11 (2019)
Marshall, C., Rossman, G.B.: Designing Qualitative Research. SAGE Publications, Los Angeles (2014)
Mathison, S.: Why triangulate? Educ. Res. **17**(2), 13–17 (1988)
Maxwell, J.: Understanding and validity in qualitative research. Harv. Educ. Rev. **62**(3), 279–301 (1992)
Maxwell, J.A.: Causal explanation, qualitative research, and scientific inquiry in education. Educ. Res. **33**(2), 3–11 (2004)
Maxwell, J.A.: Qualitative Research Design: An Interactive Approach, 3rd edn. SAGE Publications, Los Angeles (2013)
Mertens, D.M.: Research Methods in Education and Psychology: Integrating Diversity with Quantitative & Qualitative Approaches. SAGE Publications, Los Angeles (2019)

Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gøtzsche, P.C., Devereaux, P.J., Altman, D.G.: CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BJM **340**, 1–28 (2010)
Morgan, D.L.: Commentary—after triangulation, what next? J. Mixed Methods Res. **13**(1), 6–11 (2019)
Morrow, S.L.: Quality and trustworthiness in qualitative research in counseling psychology. J. Couns. Psychol. **52**(2), 250–260 (2005)
Morse, J.M.: Critical analysis of strategies for determining rigor in qualitative inquiry. Qual. Health Res. **25**(9), 1212–1222 (2015)
Morse, J.M., Barrett, M., Mayan, M., Olson, K., Spiers, J.: Verification strategies for establishing reliability and validity in qualitative research. Int J Qual Methods **1**(2), 13–22 (2002)
Noble, H., Heale, R.: Triangulation in research, with examples. Evid Based Nurs **22**(3), 67–68 (2019)
Nowell, L.S., Norris, J.M., White, D.E., Moules, N.J.: Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods (2017). https://doi.org/10.1177/1609406917733847
O'Connor, C., Joffe, H.: Intercoder reliability in qualitative research: debates and practical guidelines. Int J Qual Methods **19**, 1609406919899220 (2020)
Øby, E.: Assessing transparency and methodological precision in variable measurement within organizational research: implications for validity. Qual. Quant. **59**(Suppl 1), S489–S506 (2024)
Onwuegbuzie, A.J., Leech, N.L.: Validity and qualitative research: an oxymoron? Qual. Quant. **41**, 233–249 (2007)
Palinkas, L.A., Horwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N., Hoagwood, K.: Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm. Policy Ment. Health Ment. Health Serv. Res. **42**, 533–544 (2015)
Patton, M.Q.: Qualitative Research & Evaluation Methods: Integrating Theory and Practice. SAGE Publications, Los Angeles (2014)
Plano-Clark, V.L., Creswell, J.W.: The Mixed Methods Reader. SAGE Publications, Los Angeles (2008)
Poortman, C.L., Schildkamp, K.: Alternative quality standards in qualitative research? Qual. Quant. **46**, 1727–1751 (2012)
Poteat, V.P., Yoshikawa, H., Calzo, J.P., Russell, S.T., Horn, S.: Gay-straight alliances as settings for youth inclusion and development: Future conceptual and methodological directions for research on these and other student groups in schools. Educ. Res. **46**(9), 508–516 (2017)
Quinlan, K.M., Pitt, E.: Towards signature assessment and feedback practices: a taxonomy of discipline-specific elements of assessment for learning. Assess. Educ. **28**(2), 191–207 (2021)
Sale, J.E., Lohfeld, L.H., Brazil, K.: Revisiting the quantitative-qualitative debate: implications for mixed-methods research. Qual. Quant. **36**, 43–53 (2002)
Sampson, H.: The role of triangulation in qualitative research. Nurse Res. **11**(1), 20–29 (2004)
Sandelowski, M.: Rigor or rigor mortis: the problem of rigor in qualitative research revisited. Adv. Nurs. Sci. **16**(2), 1–8 (1993)
Seale, C.: Quality in qualitative research. Qual. Inq. **5**(4), 465–478 (1999)
Seidman, I.: Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences. Teachers College Press, New York City (2006)
Shadish, W.R., Cook, T.D., Campbell, D.T. Experimental and quasi-experimental designs for generalized causal inference. Houghton, Mifflin and Company (2002)
Shenton, A.K.: Strategies for ensuring trustworthiness in qualitative research projects. Educ. Inf. **22**(2), 63–75 (2004)
Silverman, D.: Qualitative Methodology and Sociology: Describing the Social World. SAGE Publications, Los Angeles (1985)
Sousa, V.D., Rojjanasrirat, W.: Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J. Eval. Clin. Pract. **17**(2), 268–274 (2011)
Sovacool, B.K., Axsen, J., Sorrell, S.: Promoting novelty, rigor, and style in energy social science: towards codes of practice for appropriate methods and research design. Energy Res. Soc. Sci. **45**, 12–42 (2018)
Stamenkov, L.: The role of triangulation in qualitative research. Nurse Res. **11**(1), 20–29 (2022)
Stamenkov, G.: Recommendations for improving research quality: relationships among constructs, verbs in hypotheses, theoretical perspectives, and triangulation. Qual. Quant. **57**(3), 2923–2946 (2023)
Tabachnick, B.G., Fidell, L.S., Ullman, J.B.: Using Multivariate Statistics, pp. 497–516. Pearson, Boston (2013)
Tashakkori, A., Teddlie, C.: Mixed Methodology: Combining Qualitative and Quantitative Approaches. SAGE Publications, Los Angeles (1998)
Tashakkori, A., Johnson, R.B., Teddlie, C.: Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences. SAGE publications, Los Angeles (2020)

Teddlie, C., Tashakkori, A.: Handbook of Mixed Methods in Social & Behavioral Research. SAGE Publications, Los Angeles (2003)
Thurmond, V.A.: The point of triangulation. J. Nurs. Scholarsh. **33**(3), 253–258 (2001)
Tracy, S.J.: Qualitative quality: Eight “big-tent” criteria for excellent qualitative research. Qual. Inq. **16**(10), 837–851 (2010)
Turner, D.W.: Qualitative interview design: a practical guide for novice investigators. Qual. Rep. **15**(3), 754–760 (2010)
Twining, P., Heller, R.S., Nussbaum, M., Tsai, C.C.: Some guidance on conducting and reporting qualitative studies. Comput. Educ. **106**, A1–A9 (2017)
Van Maanen, J.: The fact of fiction in organizational ethnography. Admin. Sci. q. **24**(4), 539–550 (1979)
Waterman, S.J., Baker, G.R., Henning, G.W., Lundquist, A.E.: Equity-centered assessment: varying approaches and lenses. In: Henning, G., et al. (eds.) Reframing Assessment to Center Equity, pp. 112–128. Routledge, Milton Park (2023)
Wolcott, H.: On seeking–and rejecting–validity in qualitative research. In: Eisner, E.W., Peshkin, A. (eds.) Qualitative Inquiry in Education: The Continuing Debate, pp. 121–152. Teachers College Press, New York (1990)
Yin, R.K.: Case Study Research: Design and Methods, 4th edn. SAGE Publications, Los Angeles (2009)