Macrothink Institute™

International Research in Education
ISSN 2327-5499
2023, Vol. 11, No. 1

An Integrative Qualitative Framework: Improving Research Through Strategic Mapping

David C. Coker

Department of Advanced Education Programs,

College of Education, Fort Hays State University, Hays, Kansas, USA

Received: January 8, 2023 Accepted: April 17, 2023 Published: April 20, 2023

doi: 10.5296/ire.v11i1.20921 URL: https://doi.org/10.5296/ire.v11i1.20921

Abstract

Outside of education, many academic disciplines developed and promulgated checklists to evaluate the quality of qualitative studies. This article used an embedded mixed methods approach in three stages: development of a conceptual framework reviewing 62 guidelines; a content analysis of key areas of the guidelines; and review of dissertations to see how practices compare to model guidelines. Using 15 educational administration dissertations in the United States, a thematic analysis examined components of the three frameworks: planning the study, in the field, and reporting results. Many researchers failed to adequately describe methods in all phases, with lack of intensity in fieldwork being a ubiquitous problem. Developing validity and reliability should move from a post hoc procedure to a central component in all stages of design. The implications of the frameworks move beyond a rigid checklist and provide direction to develop strategic mapping for designing research in an iterative fashion to optimize representativeness of findings.

Keywords: qualitative framework, dissertation design, validity and reliability, research problems, mixed methods

1. Introduction

1.1 Introduce the Problem

Qualitative research has become popular, but the standards to design and evaluate quality remain debatable in educational administration and other fields. For example, Greenhalgh et al. (2016) in medicine recognized difficulties in qualitative research, as many admitted research was often of poor quality and lacked standards (Ullrich et al., 2020). Four points were important to developing rigorous dissertations: theory, paradigm, reflexivity, and power (Gringeri et al., 2013), but there was often little erudition beyond the superficial (Knapp, 2017). The following article offers hopes and promises to create a systematic improvement of a large order of magnitude by creating strategic mapping to develop optimal processes through all stages of the research.

The purpose of the following paper was to provide a holistic, iterative design to develop a flexible model to produce and report valid and reliable qualitative research. An assumption underlies all premises: Criteria need negotiated, owned, and qualified to produce research which displays a realist performance of the phenomenon researched and cannot claim a checklist used was the reason for validity and reliability (Anfara, et al., 2002; Loh, 2013; Morse, 2015; Norris, 1997). The research was neither prescriptive nor proscriptive; every situation will be different, determinative, and ultimately judgmental. Other researchers have used checklists to build allegiance (Munthe-Kaas et al., 2019; Stenfors et al., 2020; Yadav, 2021), as if qualitative research was really quantitative research; the current research recognized qualitative research proclaimed flexibility in theory but often lacked in practice, and the recommendations were to produce a living theoretical model which adapts and adopts to each study's requirements (cf, Coker, 2021).

A startling claim predominated most educational administration dissertations, and one can easily verify: Most everyone proved their theories, positionality, and beliefs (Grbich, 2012, called the problem recolonization; Lester et al., 2020, citing Braun and Clarke, called the problem turning the questions into answers). Replication or divergent findings in studies were startling anomalies, and one found everything and everyone produced a coherence which aligned with the researcher and to previous studies (Frias-Navarro et al., 2020; Kerasovitis. 2020; Perry et al., 2022). Writing for an audience and not to an audience created an echo chamber which neither honored the mutlivocality nor divergences of what were often tragically short data runs which moved with directionality as opposed to orientation (Gehman et al., 2018; Maxwell, 1992; van Manen, 2007). Merriam (2009) called for an iterative, creative practice which sought to reject and refine findings, but her suggestions were often not heeded.

1.2 Importance of the Problem

Liao and Hitchcock (2018) reported the hows and whats of reliability and validity of qualitative research were often vague or missing. This article purposefully used the terms reliable and valid, but one could substitute trustworthy/credible; the point was the research considered true and representative (Golafshani, 2003; Yonge & Stewin, 1988). There were

three aims to accomplish the purpose of the following research. In a multidisciplinary framework, what were design frameworks to develop model qualitative research in educational administration dissertations? What were model recommendations from the extant literature? The third question was how did educational administration dissertations apply the principles of qualitative research in relationship to the frameworks and suggestions from a literature review to their own studies? The three sections were integrated: The development of a design framework to guide researchers in planning, carrying out, and evaluating qualitative research elucidated by how researchers completing educational administration dissertations carried out a research project, compared to the extant literature. A discussion follows, with recommendations.

1.3 Background

Qualitative research was allegedly post-positivist, yet Gehman et al. (2018) pointed out theory often defied practices: Most qualitativist presented the methodologies as more rigid and linear than quantitative research. Research generally can be divided between exploration and exploitation (Gringeri et al. 2013), though one often finds everything as pure frequentist reports devoid of context and overly smooth. Quality and styles varied widely, with the problem educational research often lacking value and rigorous implementation (Coker, 2022b).

Experts and leading pundits debated validity and reliability, but most all agreed research should have a realist aspect—describing a phenomenon in rich, thorough details as opposed to pure fiction (Engel & Kuzel, 1992; Onwuegbuzie & Leech, 2007). Within the aspects of validity and reliability, research based in theory was the expectation, though sometimes researchers led with theoretical arrogance or the misuse of micro and macroprocesses beyond the findings and outside the level of analysis (Gioia, 2021; Sartori, 1970). From detailed methodology books to many flavors of each sect of qualitative research, little was clear except researchers should be systematic and rigorous.

“A methodology is not a cookbook; rather, it provides scholars with orienting principles and tools that always need to be modified and customized” (Gehman et al., 2018, pp. 298). Research should be guided by a review and analysis of the literature, with a question which was messy and ill defined (Gehman et al., 2018; Oplatka, 2021). Questions generally were framed in one of two ways: within cases or cross cases (Onwuegbuzie & Leech, 2007). Theoretical sensitizing moved the researcher beyond one’s own boundaries and examined broad factors, causes, and drivers of the phenomenon under study (Jamie & Rathbone, 2021). Qualitative research should be a process that opens dialogues, offers insights, and suggests future research to confirm, reject, and transform society. One way to seek improvement is through strategic mapping; instead of a checklist, a framework provides an iterative, useful way to map and challenge oneself through the entire research process (Hannes et al., 2015).

2. Method

The study used an embedded mixed methods design in three stages (Creswell & Clark, 2017),

as shown in Figure 1. Three qualitative frameworks were developed using thematic analysis. The frameworks were divided into three stages: planning the study, in the field, and reporting results (following recommendations by Brooks & Normore, 2015). Frameworks from many disciplines informed the development of the three frameworks. There was a horizontalism applied to analyzing the checklists and recommendations; every idea was used to construct a framework. Unlike other checklists, such as COREQ (Tong et al., 2007), the intent was to use every idea. All frameworks were downloaded into Microsoft Word and reviewed, annotated, and highlighted. Following the three stages, all ideas were sorted. Tables were constructed, and the framework was catalogued to organize the data.

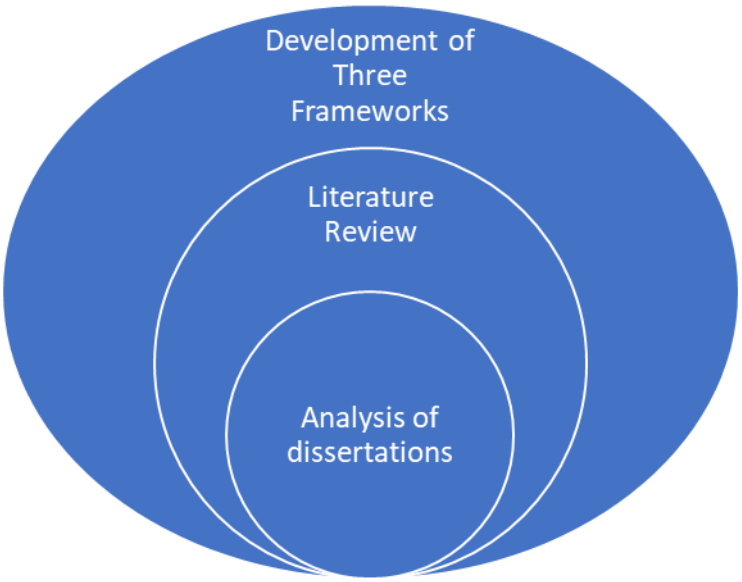

Figure 1. The embedded research design to analyze qualitative frameworks.

Next, there was a literature review around the following topics: methodology, ethics, and sample; data collection, the coding cycle, and themes; validity and reliability, audit and member checks, saturation, triangulation, reflexivity, transferability and generalizability, and limitations. A thematic analysis was conducted by reviewing, annotating, and highlighting. There was constant questioning and comparisons of different researchers' findings. All findings were sorted by topic into Microsoft Word, and categories were formed by similarities, differences, and degrees. Then all the topics of the literature review were examined by the thematic analysis of dissertations in educational administration.

During the formation of the frameworks, 15 educational administration dissertations were analyzed to explain the processes and operationalization of qualitative research. Thematic analysis was conducted (Braun & Clarke, 2006; Fereday & Muir-Cochrane, 2006), with a coding schema of reading/annotating, in vivo, descriptive, memoing, aha moments, questioning, categorization/relationship development, themes, and metathemes (Coker, 2021).

All data received a geocode to track location, and intermittent thematic formation was used to track the development of themes while formally questioning, debating, and challenging findings (Coker, 2022a). The frameworks and an extensive literature review guided the dissertation phase and compared the differences between theory and practice. My position as a researcher, frequent peer reviewer of qualitative studies, and someone who completed a dissertation, informed the findings.

2.1 Sample

The development of the three frameworks involved searching for checklists and quality methods in qualitative frameworks across all disciplines. The frameworks were coded by recommendation, and a choice was made to develop three frameworks of before, during, and after. The use of 62 different sources produced a representative list for the recommendations from diverse perspectives and fields.

There were 15 dissertations from 15 different universities in the United States, all in educational administration. The original goal was 15-20 dissertations, as previous research and practices suggested key themes that would be representative would need at least 6-8 in the sample (Guest et al., 2006). All dissertations were from 2010-2020, and the criteria were the following: all dissertations were strictly qualitative in nature, traditional, in English, completely online, and full text available. To achieve variation, there were 7 phenomenological, 6 grounded theory, 1 narrative inquiry, and 1 autoethnography in the study. Of the 15 dissertations, 13 were Ed.D. and 2 were Ph.D. All data from the dissertations were downloaded in Microsoft Word, and data analysis took place in Microsoft Excel.

The results have three parts. The first part presents the three frameworks: planning the study, in the field, and reporting the findings. Then the findings were used with an extensive literature review to analyze the different components of the qualitative dissertations that were interwoven within the three frameworks. Finally, a metatheme was presented.

3. Results

3.1 Planning the Research

Table 1 presents steps to develop and plan a qualitative research study. Though numbered and ordered in a linear fashion, the processes should be iterative, chaotic, and continuously revised (Knapp, 2017). Decisions must be made and revised throughout a study, with coherence being the step that binds all decisions together. There was significant overlap among all stages, but the conceptual meanings evolve as a study gets underway. For example, ethics and reflexivity start from the beginning and run through until publication, but how and what one does was markedly different at each stage. A horizontalism permeated all steps of the research: If an author saw a position as valuable, the idea was included as a piece of the puzzle. A key concept for all researchers was that not every point would be applicable to all situations; what a phenomenologist does and believes will be different from a ground theorist, etc. Methodology, ethics, and sample were examined in depth by exploring dissertations and

a literature review.

Table 1. Comprehensive framework to plan qualitative research studies.

Planning the Study		
Coherence	**Research Design cont'd . . .**	**Methodology and Methods**
1.1 Alignment of Study: a. meaningful, b. scholarly purpose, c. grounded in literature, d. consistency, e. connected to paradigm/theoretical, f. fit with RQ-data collection-analysis.[1, 4, 17, 22, 36, 39, 40-42, 47-48, 50, 54] **Ethics** **2.1 Ethical Standards**: a. procedural, b. situational & culturally specific [b1. harm, b2. end results], c. relational ethics [c1. researcher's own characteristics, c2. attitudes, c3. biases, etc.], d. exiting. **2.2 Maintenance**. [3-4, 16, 18-21, 44, 48, 54, 60-61] **2.3 Human Subjects:** a. informed consent, b. anonymity c. confidentiality, d. respect for participants.[16, 18, 25, 40, 55] **Reflexivity** **3.1 Conflicts of Interests.**[16] **3.2 Reflexivity:** a. neutral/disinterested, b. sincerity [b1. genuine, b2., authentic], c. positionality/orientation, d. bidirectionality, e. cultural/theoretical,	**4.3 Purpose**: a. aims/desired results by interests/priorities, b. goals/objectives, c. audience, d. analysis needs, e. utility, f. terminology defined. **4.4 Background/Literature Review:** a. grounded, b. relevant, c. gaps [new information: c1. knowledge, c2. insights, c3. perspectives]. **4.5 Rigor Strategies.** **4.6 Feedforward**: a. new problems, b. new questions, c. new practices. **4.7 Synthesis:** all components.[4, 6-7, 14, 16-18, 20-21, 23-25, 35, 39-42, 44-45, 47, 49-51, 54, 56-57, 59, 61] **4.8 Framework Approach**: a. assumptions, b. congruence with design, c. theoretical or conceptual framework/perspective, d. sensitizing concepts, e. application [e1. logical, e2. parsimonious, e3. complete, e4. useful, e5. clear]. **4.9 Paradigmatic Lens**: a. research tradition, b. philosophical underpinnings [b1. ontology, b2. epistemology, b3. deontology].[3, 12, 16-19, 22, 32-33, 54-55,	**5.1 Qualitative inquiry:** a. exploratory, b. confirmatory, c. descriptive, d. hybrids. **5.2 Grand Tour Question(s).** **5.3 Methodology:** a. analytic orientation approach defined & described, b. congruity w/ RQ-collection **5.4 Methods:** a. identified, b. appropriate, c. best practices, d. follows exemplars, e. uses journalism questions to develop, f. outlined/described, g. fidelity. **5.5 Data Collection**: a. field work approach, b. interview strategy [b1. kinds, b2. guide, b3. pilots], c. prolonged immersion / engagement, d. planning tools (e.g., PICO, computers, etc.), e. multivocality/verstehen. **5.6 Methodological Limitations.** **5.7 Reproducibility.** [1, 6-7, 15, 19-21, 23, 25, 32, 39-42, 44-45, 49, 51, 53-54, 57-59, 61] **Sample** **6.1 Sample:** a. rationale/relevance, b. setting, c. stakeholders,

f. self-reflexivity [f1. subjective values, f2. biases, f3. inclinations of researcher].[4, 6, 19, 30, 32, 40, 42, 44, 48, 50, 55]

3.3 Analytic Lens:

a. researcher's characteristics,
b. epistemology,
c. theoretical.[6, 16, 50]

3.4 Biases:

a. self as bias [a1. assumptions, a2. interests, a3. reasons],
b. methods to redress [b1. described, b2. developed b3. possibilities: journaling, bracketing, etc.],
c. openness,
d. reactivity,
e. circumvention,
f. restraint to preserve validity & reliability.[22, 30, 39, 40, 54, 58, 61]

Research Design

4.1 Systematic Design.

4.2 Topic:

a. focus,
b. interests,
c. guidance,
d. relevant,
e. timely,
f. significant,
g. interesting,
h. evocative.

57-61

4.10 Context:

a. rationale,
b. importance/significance,
c. data source,
d. end-users consulted.[14, 16, 51, 60-61]

4.11 Results:

a. primarily linguistic,
b. data collection plan/integrity,
c. emic/etic/hybrid.[21, 37, 43, 52]

4.12 Validity Plan:

a. verification,
b. credibility procedures / quality control,
c. optimization,
d. sincerity,
e. transparency,
f. appropriateness,
g. alternatives considered,
h. adaptable,
i. triangulation,
j. saturation,
k. defensible/justified.[4, 11, 17, 22, 25, 41, 47, 57, 61]

4.13 Timeline.[17, 24]

d. units of analysis [d1. characteristics, d2. groups, d3. location, d4. time frame],
e. strategy/how approached [e1. purposive, e2. convenience, e3. snowballing, e4. random, e5. theoretical, e6. adequacy, e7. bias],
f. recruitment/systematic selection using frame/frame error,
g. availability,
h. relationship with researcher,
i. size,
j. transparency conveyed,
k. data/theoretical saturation,
l. generalizability,
m. description,
n. defensible.[2-3, 6-7, 13, 16-17, 20, 23-25, 37, 39-41, 47-49, 55, 57-58, 61]

Note. Appendix A lists references cited and can be used for further reading. RQ = Research question.

3.1.1 Findings about Methodology

Exploration and explication dominated research questions and purposes, with celebrations, impact, and theory development much less common. Descriptions used the following verbiage: perceive, believe, lived experiences, and identify, etc. There was, on average, 3-4 questions per dissertation, with how or what most frequently used. Most questions were about specific situations, though how generalizable was not mentioned in most introductions. There was a positive expectation in most questions.

Describing the methods followed a standard routine: Drop a name, state some steps, and mention a coding cycle, generally with open coding, some single-word categories promenaded as methods, and voila, one had a method. Creswell and Glaser were mentioned most often. Most dissertations stated five cycles in coding and analysis, though the presentations were often unclear and incomplete. Many of the methods were either incoherent or presented as routine steps; for example, phenomenology seemed more akin to thematic analysis in many dissertations. Others used grounded theory with no theory developed, and one was even more bizarre: deductively testing categories from a theory. There was a complete absence of hybridity.

3.1.2 Ethics

Ethics were poorly considered and described, with most dissertations lacking a formal section, a finding that seemed all too common in educational administration research (Berkovich & Grinshtain, 2021). The most salient description about the ethics sections were pro forma and de minimus. Run-of-the-mill steps mentioned IRB, archival records keeping, consent, voluntary, and anonymous/confidential. Minor inclusions were checking biases, transcript accuracy, and member checks. There were two exemplary practices: a bill of rights and including discussion of ethics/informed consent within the interview guide. There was evidence from other research that informed consent was lacking (von Unger, 2016).

A major problem in most dissertations was backyard research, or researching topics or themes where the researcher was part of the community being researched (Glesne & Peshkin, 1992). Some researchers called for the democratization of research (Karnieli-Miller et al., 2009), but the ideas were problematic: What if the participants wanted the researcher to write in a way totally different from the holistic perspective or at odds with the researcher? Two major issues, cycling back to the Glesne and Peshkin (1992) article, challenged the ideas: Researchers who were a part of the community of participants under research never found oneself as anything but infallible, ennobled, and enlightened; the second major conundrum was the power differential: With whom were participants speaking? The researcher, a fellow colleague, or a personal friend? Not once did researchers consider such a major issue.

Ethics should be a central, messy topic needing questioned and considered through all stages of research (von Unger, 2016), yet most dissertations treated ethics as a trivial concern. Within the context of educational research, some problems arose from this analysis: researcher relationships were often unethical and biased (e.g., a principal asking subordinates to evaluate her practices, an instructional coach asking teachers to rate his practices, etc.) and masking was poorly considered (e.g., sometimes names were not redacted, one intimately involved could figure out the participant, and the appendices showed names of schools and locations, etc.). Not one dissertation considered traditional issues, such as anxiety/distress, exploitation, misrepresentation, possible harm to researcher/participants, and identification of participants (Richards & Schwartz, 2002). Tooley and Darby (1998) found problems with partisan researchers, while Harrison et al. (2001) and Brinkman and Kvale (2005) stated reciprocity and too much rapport needed considered and questioned. Often one can see relational aspects meant participants might tell researchers what they wanted to hear through

a variety of processes (Orb et al., 2001; Roulston & Shelton, 2015; Sikes, 2000).

3.1.3 Findings about Sampling

Sampling in most dissertations claimed purposive, yet the many uses were often contradictory and lacked detail, a finding echoed in the literature (Carlsen & Glenton, 2011; Gentles et al., 2015; Guetterman, 2015). While none claimed convenience, most researchers' samples' key criterion was the following: readily available (see Luborsky & Rubinstein, 1995). Ironically, samples were designed to be small, but then there were claims this delimitation was a limitation. A stark minority considered population, sample frame, relevance, and variance (Flick, 2019; Kennedy, 1979). Most samples claimed two to three criteria, with an average sample size of 13.9 (SD = 5.7).

There were no random samples, and infrequent methods used were snowball sampling and theoretical sampling (though theoretical sampling did not match the guidelines or spirit of grounded theory studies, as many researchers decided from the onset; see Gentles et al., 2015 and Onwuegbuzie & Leech, 2007). Subgroups and split samples were infrequent. There were no mentions of attrition and changes needed after beginning the study. Variance was not considered, and few had exemplars of criteria tables for inclusionary and exclusionary factors.

There was little erudition in the rationale and development of samples beyond picking a number (Mason, 2010). The researchers embraced the role of the ultimate insider, and with little surprise, the researchers found their friends and colleagues agreed with the researcher. The ethical implications were not considered, and developing a robust model to select the sample should be promoted to be one of the most important, consequential decisions of a study. Rules of thumbs dominated number selection as firm guidelines which displaced adaptation (Guetterman, 2015; Vasileiou et al., 2018), with relational sampling the major purpose (either directly or by proxy, i.e., an assistant principal interviewing other assistant principals to ask if they see value in their jobs, etc.). Relational samples, a part of "backyard research," were predicated on power imbalances which used friends and colleagues within one's school or district (Amalia et al., 2015; Kim, 2015); the conflicts of interest were obvious and of concern, such as asking subordinates or colleagues questions if they perceived their or the researcher's work as valuable.

"As the field of qualitative research has developed over time too much of it has, in my view, been based on unjustified, accidental or opportunistic sampling strategies" (V. Anderson, 2017, pp. 2). Developing a theoretically and researched-based sampling strategy with criteria should be the linchpin of all studies to design a study which creates the best interpretive analysis (Coker, 2022a; Malterud et al., 2016; Onwuegbuzie & Leech, 2007; Oplatka, 2021). The purpose and research questions should guide the development of a sample, but there should be flexibility as data collection takes place. Researchers need to reject relational as the driving criterion.

3.2 In the Field

Table 2 presents a framework to plan fieldwork. All the different sections describe a diverse

set of procedures to collect and analyze data in an iterative fashion that will be responsive to the needs and direction of the study. While the three tables were presented discretely, all stages should be considered and negotiated concurrently; one decision creates a ripple effect throughout the entire study. Systematicity in all steps must be continually negotiated. The following sections were examined in the dissertations: data collection, the coding cycle, and themes.

Table 2. Comprehensive framework to conduct fieldwork, data collection, and analysis.

In the Field		
Data Collection **7.1 Action Plan** a. plan, conduct, manage, b. connect proposal-study-findings, c. alignment research question, d. data formatting/processing, e. execution.[16, 17, 20, 34, 47] **7.2 Methods** a. field notes, b. instruments, c. technologies/recording, d. interview scheme, fitting, e. transcripts, f. visual data, g. analytic lens, h. ongoing conversation with literature, i. empathy as observation strategy.[2, 5, 10, 16-17, 23-24, 43, 57-58, 61] **7.3 Initial Analysis** a. inductive/deductive/hybrid, b. sample [b1. appropriate, b2. adequate, b3. theoretical sampling, b4. confidentiality/anonymity, b5. demographic reporting criteria], c. data saturation, d. theoretically linked, e. transparent/pathway, f. systematic, g. complete/sufficiency [rigor: g1. rich, g2. participants free to speak, g3. context sufficient/clearly described, g4. context bias, g5. units of analysis/levels of	**7.3 Initial Analysis cont'd . . .** j. constant comparison, k. delimitations, l. accurate descriptions/fidelity, m. nonparticipation/presence, n. attrition.[4, 6, 10, 16, 18, 24, 33, 37, 40, 53, 57-58, 60-61] **7.4 Integrity** a. appropriate/reliable, b. fidelity/perspective.[3, 18, 31 53] **7.5 Prolonged engagement & observation** a. persistent, b. repeat interviews, c. adequacy/sufficiency d. immersion[1, 22, 48, 55, 58, 61] **Analysis** **8.1 Systematicity** a. inductive-deductive-abductive, b. alertness/coherence background with bracketing, c. rigorous/valid themes or theories, d. coding [d1. prereading (familiarization & annotations), d2. coding schema/process, d3. evolutionary-devolutionary, d4. descriptions, d5. coding tree, d6. saturation/triangulation, d7. memo/memo bank, d7. questions, d8. constant comparison, d9. immersion, d10. data-driven/grounded, d11. narratives/abstracts],	**8.1 Systematicity cont'd . . .** f. effective application & interaction collection—analysis g. coding alternatives [g1. key word/linguistic, g2. summarization, g3. crystallization, g4. computer analysis, etc.].[2, 5, 10-12, 14, 16-17, 23, 25, 33, 36-37, 39-40, 43-44, 58, 61] **8.2 Theming/Theorizing** a. concepts/categories defined, b. sorting, c. finding what's missing, d. completeness-sufficiency-detailed, e. ongoing attention, f. emerging themes/theories, g. staying grounded/open, h. theoretical sensitivity, i. multiple iterations, j. adopting/abandoning preconceived notions, k. constant comparison, l. researcher relationship & responsiveness to data, m. data interpretation level & units of analysis, n. incorporation of literature.[5, 10-11, 13, 28, 32, 37, 40, 49, 55, 57] **8.3 Software**.[5, 14, 25, 34, 58] **8.4 Validity** a. common method variance, b. context bias, c. trustworthiness, d. biases.[16, 18, 29]

participation], h. triangulation [h1. internal/external, h2. data, h3. researcher, h4. theory, h5. method], i. external/internal validity,	e. context [e1. ongoing attention, e2. sample described, e3. polydimensional/multivocality, e4. nonlinear/chaotic],	

Note. Appendix A lists references cited and can be used for further reading.

3.2.1 Findings about Data Collection

Data collection in the dissertations was where all the planning finally became action. How and what processes were done determined the value of the entire dissertation. Three major themes enveloped the data collection method: how to collect data, length of time in the field, and the iterative nature.

How to collect data and length of time in the field were problematic. All dissertations claimed semistructured interviews (range of 4-25 questions), though some had focus groups as well. Strikingly, most questions were unidimensional, brutally direct (many claimed each question was a theme, and some questions seemed inappropriate, using jargon), and operated more like structured interviews. There was no use of close-ended questions or rankings. Lacking was rapport building and easing into an interview, and the questions often cued the answers, especially in light of the relational samples. One could surmise apophenia flowed from the intersection of brief questions asked of friends and colleagues. Exemplars had a well-defined process to develop interview practices (e.g., expert review, piloting, existing literature review, cognitive interviews, and peer review, etc.).

Adequacy should be a central factor in data analysis, but many dissertations were questionable. Time in the field was brief, with most interviews 20-60 minutes, and most had no follow-up interviews or observations. Total time in hours in the field ranged between 6-12 hours (the briefest was 4 hours within a week, while the most was 25 hours direct interviews over a year), showing the opposite of prolonged engagement and persistent observation. Weighing evidence and determining adequacy needed evaluated (Freeman et al., 2007). Yet, if one mentioned saturation here, it was always reached. Shocking was how quick, down and dirty the total time in the field was. Imagine a dissertation titled “My 4 Hours Qualitative Study.”

The most involved dissertations embraced an iterative analysis and collection, though these studies were in a small minority. These studies had follow-up interviews and focus groups, with new questions to clarify previous results. Only one dissertation stated each interview shaped and improved the following interviews. One-and-done was the norm. Probing was not mentioned. Dissertations often ignored other sources of data (see Gehman et al., 2018), which should include the interactional nature of interviews (see Potter & Hepburn, 2005). Many claimed field notes and observations, but rarely did these data sources make it into the results sections. Time in the field should be a paramount concern, and a protocol which will be continuously monitored and updated in a cyclical fashion could aid researchers in probing beyond the surface. A rules-based, systematic approach improves research, but researchers

need to be open to searching for participants' meaning-making by listening and accepting the limitation field notes and interviews provided a partial view into a phenomenon which were often manipulated to be what the researcher desired (Fine, 1993; Gilligan & Eddy, 2017; Ross & Bibler Zaidi, 2019).

3.2.2 Coding Cycle

Most dissertations failed to move beyond general coding schemas, and the failure to link coding to category development with examples, definitions, and rules produced for readers an inability to understand the logic behind theory and theme development (see Gringeri et al., 2013; Mayring, 2019; Onwuegbuzie et al., 2016). "Strauss and Corbin (1990) perceive the coding paradigm as an obligatory element of a grounded theory: if the coding paradigm was not used in theory development, the theory would miss density and precision" (Vollstedt & Rezat, 2019, pp. 88). Coding was important, yet there was a ubiquitous rule in dissertations: General coding schemas strictly following experts were represented as the method, and few dissertations moved beyond the telling and not the showing.

There were two to three stages mentioned in most dissertations, but the what and how were missing. Grounded theory researchers produced the most detailed outlines, but few produced a detailed record of the mechanisms availed. A startling lack of creativity defined the research; little was mentioned beyond generalities or specificities, such as open coding or in vivo. Many incongruities or incomprehensible actions occurred: no coding mentioned, memos only after completion, hollow terms, or strictly frequentists. The idea of an evolutionary-devolutionary matrix was missing as well as the development of a codebook (Coker, 2021).

While qualitativist pretended to be qualitativist, most dissertations were reduced to quantitativist where something said instanter became something else devoid of context by being counted (a problem mentioned by van Manen, 2007). Few spoke about how to move from one cycle of coding to another or categories which developed relationship attributes beyond descriptives, presenting a miasma which lacked specificity and sensitivity (Gioia, 2021; Glaser & Strauss, 1967). Most dissertations lacked well-defined coding schemas, and the ones who did acted as if coding was a parametric procedure. The best dissertations used an iterative framework, moved from hollow terms to the concrete, and focused on "the tails," or anomalies, as much as the most common findings. Maxwell (1992) pointed out a problem: Data presented was merely an account which needed elucidated and validated. Hennink and Kaiser (2021) mentioned the need for code frequencies, batch comparisons, stopping criterion, counting, and code meaning to develop saturation, yet nary a one fully elucidated any procedures.

3.2.3 Themes and Theories

Themes and theories were the usual ending for most research in dissertations. Studies presented an average of six themes (range 3-20), while categories forming themes ranged from 0-60. How themes were developed was commonly mentioned in the dissertations but followed the findings of Brooks and Normore (2015): No one really knew how themes were developed or the connection with the literature and theory. For most researchers, themes

emerged, though there were several other major ways: clustered, constant comparison, generated, frequencies, and sorted. Others claimed grounded, apparent, refined, and patterns. Bizarre ways were researcher-identified (who else?), "not from the data" (from where, then?), grounded theory study claimed theories already in place, themes grounded in the literature, and conspicuous (did the data speak for itself?). Luborsky and Rubinstein (1995) suggested researchers should move beyond descriptions, give explanations for choices across the sample, and show alternatives and contradictory findings.

A troubling but common aspect: Qualitative research was quantitative research. Quantitizing was possible but not anywhere like Sandelowski et al. (2009) envisioned. Counting codes—often with exact counts of everything—meant a category lacked connections to major/minor themes, elements/dimensions, outliers, and context (Glaser and Strauss, 1967, called for properties). Rare in the dissertations were outliers, negative cases, exclusionary criteria, or multivocality, as every participant merged into one, which suggested the unlikelihood people across a setting were constant and unchanging (Mizzi, 2010; Peng et al., 2001; Sartori, 1970). A concern in frequentists was the decontextualization: Just because two participants stated something did not mean either meant the same idea or concluded with the same result (Milne & Oberle, 2005). Other examples: pure word count, adopting a bright line code frequency with no rationale (e.g., 4 codes was a theme, but why was 3 not?, etc.), and reaching a random number led to promotion to a theme, etc.

Most themes were underdeveloped, with many being more like a title with only one or a few words. Unlike quantitative research, all findings had a uniform distribution without any tails, and instead of explaining a percentage of a model (not uncommon for quantitative models to explain 20% or even far less and be considered significant, etc.), everything fit. Though there were many grounded theory studies, none utilized theoretic sampling consistent with any grounded theory strand and with the purpose to disconfirm (Gentles et al., 2015; Marks, 2007). Everyone found what they were looking for; a theme in all studies was finding one's ideas and positions were right and good, which could be due to concept creeping (similar to Marsteintredet & Malamud, 2020) or poorly constructing themes (Lucas & D'Enbeau, 2013). Peng et al. (2001) found a likely reason:

> But from where do theories come? It seems quite clear that cultural values must be an important source for theories: values guide our attention to what is good and important. Our views of what the world is like are shaped by what we think the would should be like" (pp. 25).

Experts might have claimed qualitative research should be chaotic, but everything channeled into an upside-down pyramid from the raw data, codes, categories, and themes. The best themes presented quotes with explanations and contextualization; some presented tables or many block quotes out of context and with no connections to themes. One unbelievable study produced long block quotes, yet there were no recordings. All models caricaturized what they purported to show (Tilman, 1991), with researchers commonly acting as if they were not in the findings (Fine, 1993). Relationships among categories were poorly defined, but everything was hierarchical and linear (there were no radial, familial, or processural, see

Gehman et al., 2018).

There were exemplars but uncommon. Some themes and theories were well developed, with dimensions, elements, and logic (as suggested by Morse et al., 2002). There were clear distinctions between other themes, and not every theme received the same weight. In the exemplars, there was a weaving of codes, categories, and the participants' voices to present a compelling story (some dissertations adhered to thick descriptions, as stated by Tracy, 2010, but most authors told more than they showed). While a few dissertations mentioned disconfirming evidence, no one explained the evaluation as possible anomalies (see Lewin et al., 2015; Lincoln & Guba, 1985) or the rhizomatic nature which should be common in most all studies.

Faith in interviews in all dissertations reigned supreme, though ethnographers have long known what was said and what one did should not be viewed as coherent and consistent (Jerolmack & Khan, 2014). The public persona, social desirability, and reciprocity between researcher-participant should all be concerns when reporting results, especially in light of relational sampling and a researcher's interest in a desired outcome. Grodal et al. (2021) mentioned two findings that could improve coding, categorization, and theming: Moving beyond forcing every bit of data into a hierarchical format and using negative cases and outliers to formulate theories and themes.

3.3 In the Field

Table 3 presents a framework for the reporting of findings, discussion, methods for validity and reliability, and writing. Though validity and reliability were presented at the end, all research should start with a plan to achieve validity and reliability in research. Developing rigor by establishing a verisimilitude of the phenomenon studied is paramount. Strategic thinking and ongoing evaluation necessitate each decision impacts all parts of the three frameworks. The following methods were examined in dissertations: validity and reliability, audits and member checks, saturation, triangulation, reflexivity, transferability and generalizability, and limitations.

Table 3. Comprehensive framework to report results.

Reporting the Results		
Findings	**10.2 Analysis: Clarity cont'd . . .**	**11.3 Audit Trail**
9.1 Audience a. purpose, b. relevance, c. resonance [c1. literature, c2. readers, c3. evocative, c4. generalizable], d. accessible, e. dissemination, f. transparent. [1, 4, 39-40, 42, 44, 46, 48]	f. structural, g. scope/linkage [g1. implications, g2. knowledge contributions, g3. meaningful/valuable, g4. internal / external], h. utility/insight [h1. setting, h2. population, h3. group, h4. empowering], i. inferences, j. interpretations [j1. not faulty, j2	a. replication, b. methodological awareness & integrity, c. limitations in use, d. reflexivity process, e. external, f. data access to readers.[2-3, 9, 11, 15-16, 22, 28, 30, 32, 44, 48, 53, 55, 61] **11.4 Confirmability/ consistency/ verification.**[11, 28, 42, 48, 58]

Discussion

10.1 Rigor (adequacy & appropriateness)

a. data rich,
b. supports findings,
c. adequate,
d. diverse,
e. depth/persistent involvement,
f. compare/contrast,
g. completeness [g1. definitive g2. comprehensive, g3. representativeness],
h. conceptual/theoretical, positioning,
i. credible [i1. trustworthy, i2. verisimilitude, i3., plausible],
j. data partly publicly available,
k. descriptive/concrete,
l. connected with empirical findings,
m. aligned/addressed and/or reformulated research question,
n. crystallization.[2-6, 8-9, 16, 18-19, 21, 23, 25, 27, 30-32, 35, 39, 40, 42, 44, 48, 62]

10.2 Analysis: Clarity

a. categorizing,
b. essential structure/storyline [b1. communicated/presented, b2. show not tell, b3. naturalistic generalization, b4. visual representation],
c. convincing,
d. corroboration [d1. grounded (referential), d2. interpretive, d3. extracts/sample, d4. hear/see range participants' voices, d5. context adequately/sufficiently described, d6. authentic/not cherry picked],
e. thick descriptions [e1. discrete, e2. extensive, e3. in depth, e4. immersive, e5. quotes, e6 erlebenis],
inadequate amount/variety of evidence, j3. negative cases (outliers, contrasts, negative), j4. discrepant analysis],
k. major/minor themes,
l. theory useful/fit and grab,
m. theory validation (empirical findings support theory),
n. internal coherence [n1. with approach, n2. with argument, n3. with findings, n4. across sample/study, n5. not overly uniform/complex, n6. all parts of study],
o. continuum of findings [o1. compared/contrasted, o2. extended previous findings, o3. no findings, o4. survey/topical, o5. integration/synthesis],
p. transferability [p1. value, p2. utility, p3. pragmatic validation, p4. within vs between group],
q. reliability,
r. summarize.[1-6, 8-9, 11, 13, 15-16, 18-28, 30-31, 36, 39-40, 42-44, 49, 51-53, 55, 57-58, 61-62]

10.3 Reflexivity

a. subjectivity/objectivity balance,
b. neutral,
c. clear.[27, 40, 52]

10.4 Limitations/strengths (in qualitative terms).[16, 18, 39, 49, 55, 61]

Validity and Reliability

11.1 Validity and reliability

11.2 Bias

a. bias [a1. cueing/directionality in questioning, a2. clarifying researcher, a3. consistency motif, a4. procedural validation, a5. reflexivity introduces bias, a6. social desirability, a7. informant bias, a8. confirming cases disproportionate weight].[9, 25, 29, 31]

11.5 Member Check

a. pre/during/post,
b. transcripts returned,
c. analyzing/coding,
d. external/internal consensus,
e. internal validity,
f. limitations.[2, 4, 9, 11, 16, 22, 30-31, 35, 48, 55, 58, 60]

11.6 Peer Debriefing

a. disinterested parties,
b. systematic,
c. review.[9, 11, 22, 30-31, 48, 55]

11.7 Dependability. [9, 18, 30, 31]

11.8 Legitimation.[23]

11.9 Saturation

a. data,
b. theoretical,
c. redundancy,
d. ideal.[48, 52, 55]

11.10 Triangulation

a. data,
b. researcher,
c. theoretical,
d. methodological,
e. complementarity/divergence
f. not just verification / validation can be dissonance / divergence / enrichment.[4]

Writing

12.1 Writing

a. creative risks,
b. persuasive [b1. logos, b2. ethos, b3. pathos],
c. voice,
d. cadence,
e. manuscript [e1. succinct, e2. organized, e3. headings, e4. edited, e5. formatted].[10, 46-47, 49]

Note. Appendix A lists references cited and can be used for further reading.

3.3.1 Validity and Reliability

Validity and reliability should be planned and negotiated throughout any study, but there was little evidence in the dissertations validity and reliability were anything more than *pro forma* meandering designed to pass a hurdle required to sanctify one's study. Other authors, such as Loh (2013), used different words to describe validity and reliability, such as trustworthiness, credibility, and verification, etc.; for purposes of conciseness, validity and reliability were used here, but others could replace the terms with their preferred usage and arrive at the same conclusions. Ross and Bibler Zaidi (2019) and Onwuegbuzie and Leech (2007) described different types of validity, but all researchers in the sample twisted validity to be a technique devoid of specificity.

There was the danger, seen in all analyses and validity/reliability methods: post hoc remodeling. Maybe validity and reliability could be dispensed with altogether, since everyone found their research was top notch. Previous research found the reverse butterfly effect where everything worked: If something was omitted or done, everything was improved. Everything, all the time, was a net positive and produced the desired results (Coker, 2022a).

The most common methods mentioned to produce valid and reliable results were member checks, thick descriptions, peer review, triangulation, saturation, and reflexivity. Uncommon methods, lacking explanation, included authenticity, removed biases, variety, and immersion. With few exceptions, few explicated the methods and instead focused on name-dropping (especially Lincoln & Guba, 1985); banal, self-serving, and self-righteous could describe the whole valid and reliable racket. Singularity in coherence was the goal, which filed every participant into a neat, connected category. Lack of perfection should not be a downside: 100% of the time, there were no problems or concerns. The rubber stamp strikes again. The kitchen sink approach to throwing a hodgepodge of methods seemed to be the major idea (Barusch et al., 2011) as opposed to a systematic research design (Oplatka, 2021). What was missing was that negative cases, dissonance, and incomplete findings (van Manen, 2007), coupled with designs that failed to produce all the leads and explore necessary factors, and should be the hallmarks of valid and reliable research.

There was a lack of systematicity, transparency, and formal attempts to question, qualify, and develop ideas in the data collection, analysis, and results. Some dissertations omitted validity and reliability, but one sentence or an announcement of the activity sans outcomes permeated most research. There were many questionable practices. Reflexivity was named without any results or challenges. One researcher asked a group of 4th grade students to discuss the analysis of a multiage group. No one ever disagreed, challenged, or changed. Peer review should produce a critical evaluation and debate (see V. Anderson, 2017), but all peer reviews produced head nods and total agreement. There should always be stubs, outliers, and new angles, and researchers should consider validity and reliability on a continuum which should be examined by the entire study, the individual factors, the context, and temporal factors (Yu & Ohlund, 2010). The activity was the outcome, though validity and reliability cannot be

reduced to a technique and might be stated though little effort or method used due to requirements (see Maxwell, 1992; B. Smith & McGannon, 2018).

3.3.2 Audits and Member Checks

Common techniques were audits and member checks to claim validity and reliability. There is a suggestion: Don't do them unless a method which improves and challenges one's results. Why? Audits never revealed any problems or concerns; everything was in balance. Member checks produced a rubber stamp; everyone, every time, checked complete agreement. There was a larger finding: Few really understand what audits or member checks even meant, and no one used either technique to improve a study.

Audits were popular in the dissertations, but most failed to mention any steps, and none used the original spirit promulgated by Halpern where there was a reevaluation, questioning, and negotiation of a study in its entirety (see Lincoln & Guba, 1985). There were differences between internal and external, but the results were strikingly unlike a financial audit: perfection. One researcher used a formal process and external checks, but everyone else stopped at the name-drop. A total mystery was what or how external auditors performed their functions, but their total agreement was reported as a badge of honor. Some claims were unintelligible, such as "thick description interviews" and "used peer debriefers" without any explanation.

Audits need reimagined by factor analysis as a multifaceted processual tool: internal/external, intensive/minimal, in situ/post hoc, and pro forma/analytical. While some claimed the value of well-conducted audit trails (see Carcary, 2020), the problem could be compared to the failed American company Enron: Audits cannot overcome poorly conducted studies with little fieldwork and confirmation bias. Confirmation bias was a problem in all disciplines, but McSweeney (2021) pointed out the problem of facticity could mean there was no reason to even conduct research—just state what one wanted. Friedman et al. (2017) explained the problem of confirmation bias:

> People get a rush from finding information that confirms that they are right; they would rather win an argument than discover the truth. People may have the ability to see flaws in their opponent's arguments. However, when it comes to their own opinions, that is when they are blind (pp. 110).

Of course, the same researchers would not apply such standards to their own beliefs. Researchers need to move beyond checking boxes and develop a systematic method to continuously check, challenge, and reformulate findings (Carcary, 2020; Coker, 2022a).

Member checks fell into two camps: transcript validation or discussion of themes/theories. No one ever had any problems or disagreements. Everyone. Like audits, nothing was learned except going through the motions, and member checks should be of little value in theory development (see Thomas, 2017). No one in the dissertations researched mentioned the value or possible ethical issues with member checks, especially overcoming the power differential, aggregating other participants' interviews, sharing with fellow participants, and ownership of interpretations (problems mentioned by Buchbinder, 2011; Candela, 2019; Karnieli-Miller et

al., 2009). Member checks seem to be from a bygone era before modern technology replaced scribbling down in situ or post hoc what was stated or observed (Varpio et al., 2017).

Within the sample, member checks lacked guidelines or discussions and explorations of what was said, what was meant, and the reflections afterwards. Some researchers suggested guidelines and procedures (Birt et al., 2016; Carlson, 2010; Chase, 2017; Harvey, 2015; Simpson & Quigley, 2016), but few (e.g., Birt et al., 2016; Karnieli-Miller et al., 2009; Kornbluh, 2015; Reilly, 2013) explored the problem of disagreements with previous transcripts. Reanalysis might lead to the destruction of previous data, does not necessarily negate previous findings, and might create an infinity loop of new material that needed considered within the new context. Carlson (2010) and Zahavi (2019) stated as much: Research will not reproduce a phenomenon as the experience was actually lived or even recounted at one point in time, as a person's context changes continuously.

3.3.3 Saturation

Half of the sample of dissertations researched claimed saturation. Saturation, like all components of practices to produce validity and reliability, was always reached if attempted and had questionable value (Low, 2019). The meanings of the kinds of saturation varied widely if stated at all: data, thematic, and/or theoretical (Guest et al., 2020). Only one in the sample mentioned how saturation was met (in grounded theory), and the lack of specificity was common in other fields (Guetterman, 2015; Varpio et al., 2017). There was a finding seemingly striking on the surface that all researchers probably did reach saturation, but the real question was the following: So what?

In the field, researchers often had homogeneous, relational samples, coded deductively (and thereby limiting what one tried to find), claimed one question was a theme, and spent little time exploring topics (mirroring findings by Hennink & Kaiser, 2021 and O'Reilly & Parker, 2013). With very little data and extremely underdeveloped themes, there was probably great similarity between participants of a sample who were friends or acquaintances with the researcher and knew what was expected. Unlike formal methods for saturation, such as with intermittent thematic formation, item tracking, or saturation tables (Coker, 2022a, Kerr et al., 2010, or Lowe et al., 2018), there were no mention how anyone made such claims (echoed by previous studies, such as Mason, 2010 and O'Reilly & Parker, 2013). Guest et al. (2020) provided a very different method to reach saturation, though one suspects saturation cannot be reduced to strictly a quantitative endeavor.

The recommendations of experts in qualitative research provided a reason why everyone probably reached a meaningless saturation: structural minimalism. Researchers mapped to saturation their simplistic, direct questions and not a phenomenon. Many experts recommended limiting codes, categories, and themes to an artificial number; intensity of contacts with prolonged engagement and persistent observation should also be an important component (Gentles et al., 2015; Onwuegbuzie & Leech, 2007). Structured minimalism, resulting in surface-level results, was the call of leading qualitative researchers, such as Creswell and Saldaña, to limit coding and categories to 30-50 codes and 3-7 categories (V. Elliott, 2018). There should be angularity and depth, but saturation should be considered by

degrees (O'Reilly & Parker, 2013; Saunders et al., 2018).

Angularity is exploring a topic from many perspectives, which does not necessarily mean a large sample. Depth is finding the stories and nuances about a topic. Finding thematic and theoretical saturation would need a process to investigate by developing factors and negative cases. Few dissertations appeared iterative, with the testing of ideas, codes, and themes. Some were incoherent in category and theme development, claiming a priori a certain number was needed or expressing arbitrary standards. Like the wave of the validity and reliability magic wand, all one had to do was state "nothing new." Topics which were limited and lacked thorough investigation resulted in surface-level saturation—question level or at the knowledge level, lacking contextualization (Weller et al., 2018, termed the problem saturation in salience).

A vignette explains the dilemma of depth and angle. There was a German Shepherd Dog named Kiki. When the owner entered the home after a long day of work, Kiki whined and ran up to the owner, wagging her tail. When the owner's son entered, she was shaking all over and jumping up and down, following him everywhere. When the son's friend entered, she loved him but had no respect, and jumped up on him and grabbed his arm with her mouth. Another friend who visited infrequently, saw the dog bark and hide under the table. He remarked, jokingly, "Every time I'm over, the dog is deathly afraid of you all." How the interviews were conducted and who were sampled could produce an answer to the "So what?" question posited earlier.

Depth was limited if everyone has or only spoke the same surface-level answers as one. Angularity results from gaining a diverse perspective from everyone seeing the same thing differently, as no two people will occupy the same context. Limited, one-question research areas and sparse-time immersion provided neither the breadth nor depth to offer the complexity of a phenomenon. There might be surface-level saturation, but the 360-degree view will be severely lacking. A researcher's focus, positionality, and methodology can also create saturation which would be of limited value and failed to differentiate coding versus meaning saturation (Aldiabat & Le Navenec, 2018). Large samples were not needed, but exploring the perspectives of the participants observing the same phenomenon—such as the dog hiding under the table—gave a representative image through dissonance. Furthermore, a one-and-done sampling might reveal important voices were missing. Just because everyone said the same thing did not mean everyone meant, saw, felt, and experienced the same thing. The attitude-behavior disconnect was an entirely different matter.

3.3.4 Triangulation

Triangulation was popular, being used in over half the dissertations as a method to ensure validity and reliability. Hammersley (2008) listed four uses of triangulation, but he pointed out the following:

> In other words, using data of different types can help us both to determine what interpretations of phenomena are more and less likely to be valid and to provide complementary information that illuminates different aspects of what we are studying (pp. 32-33).

Like all methods, if triangulation was attempted, it was a glowing success. All the dissertations used comparisons of internal data (most followed Creswell's definition and lacked the nuances of Denzin, see Fusch et al., 2018), though one also claimed theoretical triangulation. Some claimed triangulation with observations, though observations were either scant or nonexistent. The problem with triangulation was akin to the geographical replication: Successful triangulation required having other known spots; it was likely a researcher with the same positionality and methods would reproduce the same mistakes throughout the study (Makel & Plucker, 2014; Marks, 2007; Watt, 2007).

Triangulation suffered from common method variance, or the examination of a variable in the same way contaminates all research due to the perpetuation of systematic error (Marks, 2007; Podsakoff & Organ, 1986). Abdalla et al. (2018) provided a great overview about the lack of procedures or agreement, but the authors confused triangulation with trilateralism (claiming modern systems, such as GPS, utilized triangulation, though trilateralism was the standard). Triangulation required a known baseline while trilateralism needed a known value of an azimuth line and can be used in a sphere. All researchers in the present study looked internally, lacked a known value, failed to acknowledge working in a three-dimensional as opposed to two-dimensional space, and examined all variables in the same way. Researchers often located themselves as the known value and became the North star with which to measure the azimuth. Without a formal method to disrupt one's assumptions and biases with formal dysfluency methods, saturation and triangulation probably served to promulgate the problem of common method variance and could not be a substitute for a weak design (Hernandez & Preston, 2013; Thurmond, 2001).

Internal comparisons dominated triangulation usage in educational administration dissertations and served as a post hoc method to make claims of verification; the expansive definition of purposes and usage of different data sources, different researchers, different methods/methodologies, and different theories were lacking. Like the other methods listed, triangulation was a hollow term devoid of any methods and lacked challenges. One must know one's bearing and the angle to know the directionality as well as declination; merely claiming a singular study was triangulation stretched credulity. There was concern triangulation was more about the researcher, the desired results, and the theory than any method, with a confusion of the ontic and ontology. Most researchers presumed coherence should be the central goal of triangulation, but different data sets and methods should produce competing perspectives as the context changes (Barbour, 2001; Patton, 1999). Triangulation could be used to add to the richness of a phenomenon through identification of divergences.

3.3.5 Reflexivity

Perhaps no method was more poorly described in the dissertations than reflexivity, enabled by methodological illusions of post hoc remodeling of results (see Down et al., 2006; Fine, 1993). Besides the methodologies which rejected reflexivity, such as interpretative or hermeneutical phenomenology and autoethnography, everyone had the magic reflexivity wand which failed to show responsiveness and reactivity (purposes explained by Berger, 2015, and Morse et al., 2002). Reflexivity without method relies on the following dubious

position:

> Reflexivity as self-positioning and self-reporting, in depending on realistic self-awareness and honest disclosure, is a rather pre-Freudian idea, assuming, as it does, that all of our critical personal parameters are available to the consciousness, and that people present themselves with no ulterior motives. These assumptions appear to be unwarranted (Salzman, 2002, pp. 810).

Validity and reliability were claimed as guaranteed, though many omitted any method beyond journaling and peer debriefing. Journaling, or retreating into one's head, was all it took. If only eliminating biases were so simple. All one had to do was write a section of reflexivity, and voila, one neutralized any biases though the alignment between researcher and theory (Baumeister & Newman, 1994; Ditto & Lopez, 1992), which suggested a self-serving bias corrupted by where and what one searched.

How one could be an escape artists from oneself never materialized in any dissertation. Magical reflexivity and bracketing eliminated the researcher from the study as if true neutrality could be reached (Chan et al., 2013; B. Smith & McGannon, 2018). Researchers claimed to remove "experiential analysis," make biases "irrelevant," "compartmentalize," and "ensure" neutrality. There was never a mention beyond the superfluous of the hows and whats, especially with the problem every facet of research weighted confirming cases, a consistency motif, and followed the results researchers desired (Biemer, 2010; Podsakoff & Organ, 1986). Olmos-Vega et al. (2022) provided many possible methods, but no one listed anything besides someone did something. Kim (2015) called for reflexivity to be used as critique and the development of alternatives, but reflexivity in dissertations served a very different purpose. Unless working for dysfluency and disconfirmation, reflexivity was a.) probably a false method and b.) served to confirm and validate the researcher's positionality within the research (Jamie & Rathbone, 2021).

A paradigm shift was needed in reflexivity by considering cognitive sciences to structure and manage biases (see Dror in articles such as Kassin et al., 2013). Goffman (1978) pointed out the problem of impression management. There was the arrogance of self-questioning to create the appearance of muting the self to avoid the unimaginable quandary of autoethnography. Covert autoethnography was the attempt to keep the researcher's voice but hid the instrumentality (Coker, 2023). Connected to covert autoethnography, St. Pierre (2021) suggested a rationale: A researcher's value subverts actions to a positivism, rooted in the self. Values incongruent between researchers and participants probably hinder effective reflexivity (Catapano et al., 2019). A formalized method to reject confirmation bias (McSweeney, 2021) could be conducted by adopting all or some of the following: Consider the base rate, negative cases, adversarial collaboration, rejecting certainty, consider alternatives including the opposite, empirical measurement to counter overconfidence, framing, close inferences to the data versus one's politics, discomfort, outside sources, and formally writing the hypothesized end results before beginning, etc., could problematize a reflexivity reduced to a method which revealed little evidence of value (Friedman et al., 2017; Gringeri et al., 2013; Pillow, 2003; Tufford & Newman, 2012).

3.3.6 Transferability and Generalizability

If every other method mentioned was a positive, glowing success, transferability and generalizability were questioned. For most dissertations, transferability and generalizability were used interchangeably, seemed to be an afterthought despite all being required to have a value-driven purpose, and most claimed low transferability or generalizability. There was a haze, as how or what and to whom were lacking clarity (Onwuegbuzie & Leech, 2010, pointed out the problem of defining what characteristics produce transferability and utility; Firestone, 1993, stated there was a need for scope conditions, or what conditions were the findings applicable under). Transferability and generalizability lacked definition about sample to population, case to case, and/or analytical (Onwuegbuzie & Leech, 2007). A central question: If there was no transferability or generalizability, what was the point of the research? Porter (2007) answered the question: If a study had value, there should be a utility.

3.3.7 Limitations and Delimitations

The major limitation was probably unspoken: Educational administration doctoral programs, especially the Ed.D. programs, often have one or two courses required to complete a qualitative or quantitative dissertation (Lochmiller & Lester, 2017, called for improved preparation). The dissertation was, as a consequence, a pilot study, conducted by most students working full time. An overarching theme dominated limitations: Limitations were really delimitations, resulting in an *apologia* for qualitative research. Most researchers in the dissertation claimed sample problems (size, location, and selection, though as Hennink & Kaiser, 2021, found there were often no qualifications to such claims), method/methodology, and geographical diversity (Shaver & Norton, 1980, found the same long ago that educational studies have suffered from poor sampling methodologies). All points were delimitations chosen by the researcher while picking a relational sample, yet cast as the major limitations (Coker, 2022a). Some spoke about the issues common to all qualitative research by design, though some were bizarre: researcher's biases, participants' biases, self-reported data, and participants having different experiences. Less common limitations were lacking triangulation and little time to conduct the fieldwork.

Researchers in the dissertations struggled with the concept limitations were factors which hindered the research and were outside the researcher's control. One should ask during a design of research: Will the sample answer the research question? Is geographical diversity amenable to the purposes? Does the methodology selected answer the question? Jerolmack and Khan (2014) provided a delimitation which was an unstated limitation in the analysis of the dissertations: The ubiquitous use of interviews equated to action, which a century of research suggested what was said and what was done were generally two entirely different matters. Another salient issue was large samples were often inferior in qualitative research, as the time immersion would suffer (Micceri, 1989). The driver for the limitations/delimitations confusion probably stemmed from the requirement to have a limitations section and the misuse of qualitative research's intentional use of small *n* as a strength and not a weakness (Vasileiou et al., 2018, has found similar problems in peer reviewed articles).

3.4 Contradictions and Incongruities

Though not an aim, a metatheme connected all the phases of the dissertation: Frequent starts, stops, and revisions led to contradictions and incongruities. First, entire sections were repeated ad nauseum; authors should trust readers do not require the same section repeated sometimes three or four times. Secondly, lacking a coherent design, changes or decisions which caused a ripple were not reconciled. There were many examples: A researcher claimed 2 follow-up interviews, but then conducted 1 with a focus group; sampling changed in many studies, from purposive to snowball, etc.; criteria to develop semistructured interview changed between Chapters 3 and 4; bracketing would be done but then it was not; and transferability, triangulation, and saturation mentioned in Chapter 3 and never appeared. The lack of a coherent design meant many methods and methodologies made little sense between intentions and practices.

4. Discussion

Quality in qualitative research is more than simply listing Lincoln & Guba, Denzin, Merriam, Morgan, Glaser, Strauss, Charmaz, Patton, Moustkas, Creswell, and van Manen, etc. Cookie-cutter approaches and checklists can guide and improve studies, but such *pro forma* name-dropping cannot be the substitute for validity and reliability. Validity and reliability were baked into high-quality qualitative research through the following from the beginning: systematicity, unbiased attempts, representativeness and relevance, openness, utility, normotoxicity, and data-driven (SURROUND). Quality was neither citing nor solely evaluating post hoc confirmability, dependability, credibility, and trustworthiness or the many other iterations of validity and reliability. The SURROUND principles hinged on one simple question: Did the researcher tell the story well and completely of the participants? Every step, every action should translate into a valid and reliable study. Researchers cannot escape garbage in, garbage out (GIGO).

The SURROUND principles happen simultaneously, but the root was the S, or systematicity. Without systematicity, the other principles most probably happen helter-skelter. If a researcher, novice or otherwise, spent little time in the field and brashly and haphazardly examined the data with the intent to give the right answer (Gioia, 2021), one cannot save the results by claiming audits, member checks, and a plethora of many other sophistries. Flexible formulaic and strategic mapping in an iterative fashion using the frameworks should drive qualitative research. A key input of strategic mapping was actively considering alternatives, objectively evaluating practices, and predicting problems. (Strategic mapping is intentionally a gerund, highlighting an ongoing process.) The principles of SURROUND are as follows:

- **Systematicity**. Systematicity was more than dropping the moniker of a method and person championing it. The devil was in the details. Due to length requirements, many peer reviewed articles offered little evidence of systematicity. Systematicity was processual: deep immersion in designing, collecting, and analyzing data, with coherence in all phases of research (Aspers & Corte, 2019; Poucher et al., 2020). A key concept should be rooted in the

theoretical and paradigmatic nature of the phenomenon under investigation beyond one's personal preferences (Bird, 2014). To be systematic, there should be waves of analysis breaking down and reformulating the data in a loop to arrive at a conclusion which was always tentative (Barbour, 2001). A breakthrough should be the development of microprocesses with exacting definitions which incorporate explicit, tested methods for steps generally included in qualitative research, e.g., triangulation/crystallization, saturation, reflexivity, and positionality, etc.

- **Unbiased attempts**. Unbiased does not mean the elimination of biases but connotes three actions: identification, separation, and minimization. Usually an -ism was seen as a position and lacking objectivity. Quite the contrary, all data is data, but one must report the differences between the researcher's biases, introduced biases, and the findings while shedding the idea everything must be positivist and significant results (Greenwald, 1975; Maier et al., 2022; Roulston & Shelton, 2015). Active endeavors throughout the research process are required to find and minimize biases (elimination was ideal but a fool's errand). Dogmatism, with the reciprocity between researcher bias and participant bias, creates a research project where most everyone centers on the same end objective through doctrinal and relational means (Benge et al., 2012; Zmigrod, 2022).
- **Representativeness and Relevance**. Even systematic research can lack representativeness and relevance. Representativeness—rigor made visible--means there was sufficient and disperse coverage of the analysis to arrive at an individually, collectively, and contextually plausible answer to the research questions (Mays & Pope, 2000). Relevance plagues much of each discipline's research: Instead of standing on the shoulder of our predecessors, researchers were often blind and believed to have discovered something novel. "Incentives for surprising, innovative results are strong in science" (Nosek et al., 2012, para. 4) proved true in all disciplines. Theoretical and conceptual understandings placed the research into context and were more than name-dropping the popular theory of the day with blind adherence. There should be a tension among one's theoretical and conceptual framework, results, and analysis.
- **Openness**. In contrast to being unbiased, there was an omission of looking at extremes, opposites, and outliers where researchers see and find what they desire (Palaganas et al., 2017). Maxwell (1992) lamented a God's eye view, but researchers often were perched atop the world. The researcher must confront findings by looking for disconfirming evidence; such nuances might develop counterstories as well as explicating the multivocality and tension of the inherent diversity within every sample.
- **Utility**. Relevance was how the research fit into the previous findings and the field; in contrast, utility speaks to one's present research. There was one main question: How does the research contribute to the understanding and operation of the state of knowledge (Brooks & Normore, 2015)?

- **Normotoxicity**. Experts preached qualitative research was chaotic, nonlinear, and problematic. Yet, most researchers published results by smoothing all findings to a coherence in paradigms, theories, methodologies, and results (St. Pierre, 2021, citing Kuhn). While some claimed a distinction with quantitative methods, most qualitative research suffered from a lack of creativity and reduced all findings to three or so key findings. There was a pretend super-homogeneity (Walsh & Downe, 2005). Just as the quantitative side of research deals with overfitting, examine most extant research from most any researcher, and one sees the impossible phenomenon: Everything fits—no outliers or negatives to be found. What does not fit?
- **Data-Driven**. Whose findings were one reporting? The researcher's, the participants', or both? The intensity formula, based upon persistent observation and prolonged engagement, must also have a continuous reappraisal and challenge of emergent findings. There should be clarity and alterity, with weight on the participants as shaping the collection, analysis, and findings. Depth and angularity should connect data-driven with representativeness and reconstructions, borne of a systematic approach.

Research could be improved by using strategic mapping with the end in mind. Everything should involve defining delimitations and developing a negotiated coherence between and among all steps of the research process (Coker, 2022a). Part of the process should include shedding claims of perfection (Tracy, 2010) in favor of moving from satisficing to optimal; one does not have to be confessional, but making the dissertation the pilot will create limitations and problems. Mapping validity and reliability from the beginning (Hays et al., 2016), especially direct consideration of persistent observation and prolonged engagement, holds immense promise. Larger sample sizes were neither always better nor preferable in both qualitative and quantitative research (see Lortie-Forgues & Inglis 2019 for the problem in quantitative research, but the paradox was qualitative research rarely finds such uninformative effects). If researchers ignored directionality and relational sampling (by failing to recognize one's emotional and personal vulnerabilities, Down et al., 2006), one might as well skip Chapters 2-4 and proceed to Chapter 5.

"In doing phenomenological research, through the reflective methods of writing, the aim is not to create technical intellectual tools or prescriptive models for telling us what to do or how to do something" (van Manen, 2007, pp. 13). Qualitative research should be akin to negotiating an agreement, with an iterative process, false starts, failures, and methods which produce questionable value. Conceptual dilution of the methodologies and methods, often devoid of contextualization, harms all research. There was danger that theory led researchers to limit openness to contradictory data and ignoring questions of believability (Bowen, 2006; Freeman et al., 2007). Reducing what and how one operates to dropping names and terms with little erudition plagued much of research, from qualitative to quantitative.

Qualitative research must shed a predetermined narrative forcing results; hindsight bias, or an "illusion of causality" by forming what one wants to see because one works backwards, is a

major problem (Connelly & Clandinin, 1990, citing Crites, pp. 168). A concern within the relational sample paradigm was using gimmickry to find what one wanted and claiming validity and reliability without regard for reciprocity with the biases of the researcher, the participants, and expectations during interviews and/or observations (Benge et al., 2012; Harrison et al., 2001; Reilly, 2013). Concerns with being overly coherent and confirmation of one's worldviews have been expressed by others:

> But from where do theories come? It seems quite clear that cultural values must be an important source for theories: values guide our attention to what is good and important. Our views of what the world is like are shaped by what we think the world should be like (Peng et al., 2001, pp. 25).

A systematic way to consider the context, the phenomenon, and application to theory (Thorne, 2000, citing Morse) could disrupt the shaping of the world into one's schema, but the processes must be active resistance and acceptance of frayed findings as endemic. Unconscious psychological phenomena, positive suggestions, arousal, rejection of negative events, and false memories can cause biases in remembering and processing data (Makin, 2016; Shaw, 2020; Van Damme & Smets, 2014). Such considerations were largely missing from qualitative studies in education, but following other scientific fields could provide a model. As many philosophers and researchers lamented, oversimplification, superficiality, ignoring the time continuum, and illusions of truth were marshalled to create a positivist-post-positivist paradox.

5. Limitations

There were a number of limitations which could improve future research and answer new questions. Saturation was monitored with intermittent thematic formation, and some sections were more fully mapped than others. Sections which could profit from a larger sample or different methodology to elucidate the themes were methods/methodologies, ethics, and themes, but all were for different reasons. The methods/methodologies produced a salient theme, but there was much more variance between claims and actions; since the phenomenon was broad, the diversity of results could further elaborate the issue. Ethics turned out to have scant answers or evidence; since many dissertations lacked ethics sections, and the ones which did were superficial, surface-level saturation was reached which could be improved by thematic sampling. While there was a coherent answer to theme development in the dissertations, what was stated does not answer the actual process; a different methodology beyond archival research would be needed. A last issue, not addressed but of interest, was comparing Ed.D. and Ph.D. dissertations as well as examining specific preparations.

6. Conclusion

Perhaps most all research possesses one common characteristic: The chance of being false was much greater than being true (Friedman et al., 2017). Sartori (1970) stated a problem

plaguing quantitative research, often overlooked: Numerical rating scales were usually arbitrary. “Whatever the statistics show, it is fine to suggest reasons for your results, but discuss a range of potential explanations, not just favored ones” (Amrhein et al., 2019, para. 25). Qualitative research often developed its own scales and favored explanations in its research, but there were no tails amidst claims of perfect central distribution in all findings. How results can be perfectly uniform every time should be pause for concern, but researchers often searched to confirm their preconceived notions to prove the desired solutions by whatever means possible (Cox, 2012; Fanelli, 2009; Freeman et al., 2007). Bergen and Labonté (2020) detailed a prime reason for finding perfection: “Social desirability bias is problematic because it can lead to overestimation of the positive and diminished heterogeneity in responses, resulting in a questionable appearance of consensus” (para. 4). Social desirability was bidirectional, with field data being one aspect and the researcher and the cultural norms of the university being another.

There seemed to be a certainty in research: If one employs a qualitative technique, it worked and ensured validity and reliability. Moving away from illustrative theorizing (Gehman et al., 2018), liberation from univocality of self (Mizzi, 2010), and challenging claimed processes beyond a name (Morse, 2021) could provide direction to improve and sustain qualitative research. Most educational dissertations provided little value or utility, with little application to praxis (Hallinger, 2011). Researchers need to shake off the yoke of provincialism (Geertz, 1984) and accept fragmented understandings as the central theme of all research. Strategic mapping opens the possibilities for improvement for researchers to craft, monitor, and report qualitative studies.

References

Abdalla, M. M., Oliveira, L. G. L., Azevedo, C. E. F., & Gonzalez, R. K. (2018). Quality in qualitative organizational research: Types of triangulation as a methodological alternative. *Administração: Ensino e Pesquisa*, *19*(1), 66-98. https://doi.org/10.13058/raep.2018.v19n1.578

Aldiabat, K. M., & Le Navenec, C. (2018). Data saturation: The mysterious step in grounded theory method. *The Qualitative Report*, *23*(1), 245-261. https://doi.org/10.46743/2160-3715/2018.2994

Amalia, Y. S., Johnson-Mardones, D. F., Johnston-Parsons, M., Shen, W., Shin, Y. S. E., & Swanson, J. A. (2015). 1.4 Dialogue in narrative inquiry: Collaboration in doctoral study in the USA. In *International Handbook of Interpretation in Educational Research* (pp. 161-184). Springer.

Amankwaa, L. (2016). Creating protocols for trustworthiness in qualitative research. *Journal of Cultural Diversity, 23*(3), 121-127. Retrieved from https://tuckerpub.com/jcd.html

Amrhein V., Greenland S., McShane B. (2019). Scientists rise up against statistical significance. *Nature, 567*, 305–307. https://doi.org/10.1038/d41586-019-00857-9

Anderson, C. (2010). Presenting and evaluating qualitative research. *American Journal of*

Pharmaceutical Education, 74(8), 221-224. https://doi.org/10.5688/aj7408141

Anderson, V. (2017). Criteria for evaluating qualitative research. *Human Resource Development Quarterly*, 1-9. https://doi.org/10.1002/hrdq.21282

Anfara, V. A. Jr., Brown, K. M., & Mangione, T. L. (2002). Qualitative analysis on stage: Making the research process more public. *Educational Researcher, 31*(7), 28-38. https://doi.org/10.3102/0013189X031007028

Aref, H. A., Witry, M., Olufemi-Yusuf, D., & Guirguis, L. M. (2021). Ensuring quality qualitative research reporting in community pharmacy: A systematic literature review. *International Journal of Pharmacy Practice, 29*(5), 416-427. https://doi.org/10.1093/ijpp/riab027

Aspers, P., & Corte, U. (2019). What is qualitative in qualitative research. *Qualitative Sociology, 42*(2), 139-160. https://doi.org/10.1007/s11133-019-9413-7

Barbour, R. S. (2001). Checklists for improving rigour in qualitative research: A case of the tail wagging the dog? *BMJ, 322*(7294), 1115-1117. https://doi.org/10.1136/bmj.322.7294.1115

Barusch, A., Gringeri, C., & George, M. (2011). Rigor in qualitative social work research: A review of strategies used in published articles. *Social Work Research, 35*(1), 11-19. https://doi.org/10.1093/swr/35.1.11

Baumeister, R. F., & Newman, L. S. (1994). Self-regulation of cognitive inference and decision processes. *Personality and Social Psychology Bulletin, 20*(1), 3-19. https://doi.org/10.1177%2F0146167294201001

Benge, C. L., Onwuegbuzie, A. J., & Robbins, M. E. (2012). A model for presenting threats to legitimation at the planning and interpretation phases in the quantitative, qualitative, and mixed research components of a dissertation. *International Journal of Education, 4*(4), 65-124. http://dx.doi.org/10.5296/ije.v4i4.2360

Bergen, N., & Labonté, R. (2020). "Everything is perfect, and we have no problems": Detecting and limiting social desirability bias in qualitative research. *Qualitative Health Research, 30*(5), 783-792. https://doi.org/10.1177/1049732319889354

Berger, R. (2015). Now I see it, now I don't: Researcher's position and reflexivity in qualitative research. *Qualitative Research, 15*(2), 219-234. https://doi.org/10.1177/1468794112468475

Berkovich, I., & Grinshtain, Y. (2021). A review of rigor and ethics in qualitative educational administration, management, and leadership research articles published in 1999-2018. *Leadership and Policy in Schools*, 1-16. https://doi.org/10.1080/15700763.2021.1931349

Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly, 74*(5), 817-848. https://doi.org/10.1093/poq/nfq058

Bird, A. (2014). *Thomas Kuhn*. Routledge.

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member checking: A tool to enhance trustworthiness or merely a nod to validation? *Qualitative Health Research, 26*(13), 1802-1811. https://doi.org/10.1177/1049732316654870

Bowen, G. A. (2006). Grounded theory and sensitizing concepts. *International Journal of Qualitative Methods, 5*(3), 12-23. https://doi.org/10.1177%2F160940690600500304

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Brinkmann, S., & Kvale, S. (2005). Confronting the ethics of qualitative research. *Journal of Constructivist Psychology, 18*(2), 157-181. https://doi.org/10.1080/10720530590914789

Brooks, J. S., & Normore, A. H. (2015). Qualitative research and educational leadership: Essential dynamics to consider when designing and conducting studies. *International Journal of Educational Management, 29*(7), 788-806. https://doi.org/10.1108/IJEM-06-2015-0083

Buchbinder, E. (2011). Beyond checking: Experiences of the validation interview. *Qualitative Social Work, 10*(1), 106-122. https://doi.org/10.1177/1473325010370189

Caelli, K., Ray, L., & Mill, L. (2003). 'Clear as mud': Toward greater clarity in generic qualitative research. *International Journal of Qualitative Methods, 2*(2), 1-13. https://doi.org/10.1177%2F160940690300200201

Candela, A. G. (2019). Exploring the function of member checking. *The Qualitative Report, 24*(3), 619-628. https://doi.org/10.46743/2160-3715/2019.3726

Carcary, M. (2020). The research audit trail: Methodological guidance for application in practice. *Electronic Journal of Business Research Methods, 18*(2), 166-177. https://doi.org/10.34190/JBRM.18.2.008

Carlsen, B., & Glenton, C. (2011). What about N? A methodological study of sample-size reporting in focus group studies. *BMC Medical Research Methodology, 11*(1), 1-10. https://doi.org/10.1186/1471-2288-11-26

Carlson, J. A. (2010). Avoiding traps in member checking. *The Qualitative Report, 15*(5), 1102-1113. https://doi.org/10.46743/2160-3715/2010.1332

Catapano, R., Tormala, Z. L., & Rucker, D. D. (2019). Perspective taking and self-persuasion: Why "putting yourself in their shoes" reduces openness to attitude change. *Psychological Science, 30*(3), 424-435. https://doi.org/10.1177/0956797618822697

Chan, Z. C., Fung, Y. L., & Chien, W. T. (2013). Bracketing in phenomenology: Only undertaken in the data collection and analysis process. *The Qualitative Report, 18*(30), 1-9. https://doi.org/10.46743/2160-3715/2013.1486

Chase, E. (2017). Enhanced member checks: Reflections and insights from a participant-researcher collaboration. *The Qualitative Report, 22*(10), 2689-2703. https://doi.org/10.46743/2160-3715/2017.2957

Chenail, R. J. (2011). Ten steps for conceptualizing and conducting qualitative research studies in a pragmatically curious manner. *The Qualitative Report, 16*(6), 1713-1730. https://doi.org/10.46743/2160-3715/2011.1324

Coker, D. C. (2021). Making thematic analysis systematic: The seven deadly sins. *Journal of Studies in Education, 11*(3), 126-146. https://doi.org/10.5296/jse.v11i3.18882

Coker, D. C. (2022a). A thematic analysis of the structure of delimitations in the dissertation. *International Journal of Doctoral Studies, 17*, 141-159. https://doi.org/10.28945/4939

Coker, D. C. (2022b). Multiverses, myths, and the Big Bang: Problems in educational research. In V. Wang (Ed.), *Handbook of research on educational leadership and research methodology* (pp. 402-420). IGI Global.

Coker, D. C. (2023). Covert autoethnography. *2023 Proceedings of the International Symposium of Autoethnography and Narrative*. IAANI.

Connelly, F. M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational Researcher, 19*(5), 2-14. https://doi.org/10.3102/0013189X019005002

Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: The SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research, 22*(10), 1435-1443. https://doi.org/10.1177%2F1049732312452938

Cox, R. D. (2012). Teaching qualitative research to practitioner–researchers. *Theory Into Practice, 51*(2), 129-136. https://doi.org/10.1080/00405841.2012.662868

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

Cutcliffe, J. R., & McKenna, H. P. (2004). Expert qualitative researchers and the use of audit trails. *Journal of Advanced Nursing, 45*(2), 126-133. https://doi.org/10.1046/j.1365-2648.2003.02874.x

de Casterlé, B. D., Gastmans, C., Bryon, E., & Denier, Y. (2012). QUAGOL: A guide for qualitative data analysis. *International Journal of Nursing Studies, 49*(3), 360-371. https://doi.org/10.1016/j.ijnurstu.2011.09.012

De Jong, Y., van der Willik, E. M., Milders, J., Voorend, C. G. N., Morton, R. L., Dekker, F. W., ... & van Diepen, M. (2021). A meta-review demonstrates improved reporting quality of qualitative reviews following the publication of COREQ-and ENTREQ-checklists, regardless of modest uptake. *BMC Medical Research Methodology, 21*(1), 1-11. https://doi.org/10.1186/s12874-021-01363-1

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63*(4), 568–584. https://doi.org/10.1037/0022-3514.63.4.568

Dixon-Woods, M., Shaw, R. L., Agarwal, S., & Smith, J. A. (2004). The problem of appraising qualitative research. *BMJ Quality & Safety, 13*(3), 223-225.

http://dx.doi.org/10.1136/qshc.2003.008714

Down, S., Garrety, K., & Badham, R. (2006). Fear and loathing in the field: Emotional dissonance and identity work in ethnographic research. *M@n@gement, 9*(3), 95-115. https://doi.org/10.3917/mana.093.0095#xd_co_f=ZTk5NjZhNGYtNzgyOS00ZjY0LWJjMDAtYjNiNjY2NmRkNTVi~

Drisko, J. W. (1997). Strengthening qualitative studies and reports: Standards to promote academic integrity. *Journal of Social Work Education, 33*(1), 185-197. https://doi.org/10.1080/10437797.1997.10778862

Elliott, R., Fischer, C. T., & Rennie, D. L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology, 38*(3), 215-229. https://doi.org/10.1348/014466599162782

Elliott, V. (2018). Thinking about the coding process in qualitative data analysis. *The Qualitative Report, 23*(11), 2850-2861. https://doi.org/10.46743/2160-3715/2018.3560

Engel, J. D., & Kuzel, A. J. (1992). On the idea of what constitutes good qualitative inquiry. *Qualitative Health Research, 2*(4), 504-510. https://doi.org/10.1177%2F104973239200200409

Epp, A. M., & Otnes, C. C. (2021). High-quality qualitative research: Getting into gear. *Journal of Service Research, 24*(2), 163-167. https://doi.org/10.1177/1094670520961445

Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching*, 3rd ed., (pp. 119–161). Macmillan.

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PloS One, 4*(5), e5738. https://doi.org/10.1371/journal.pone.0005738

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods, 5*(1), 80-92. https://doi.org/10.1177/160940690600500107

Fine, G. A. (1993). Ten lies of ethnography: Moral dilemmas of field research. *Journal of Contemporary Ethnography, 22*(3), 267-294. https://doi.org/10.1177%2F089124193022003001

Firestone, W. A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher, 22*(4), 16-23. https://doi.org/10.3102/0013189X022004016

Flick, U. (1992). Triangulation revisited: Strategy of validation or alternative? *Journal for the Theory of Social Behaviour,* 22(2), 175–197. https://doi.org/10.1111/j.1468-914.1992.tb00215.x

Flick, U. (2019). *An introduction to qualitative research* (6th ed.). Sage.

Freeman, M., DeMarrais, K., Preissle, J., Roulston, K., & St. Pierre, E. A. (2007). Standards of evidence in qualitative research: An incitement to discourse. *Educational Researcher, 36*(1), 25-32. https://doi.org/10.3102%2F0013189X06298009

Frels, R. K., Sharma, B., Onwuegbuzie, A. J., Leech, N. L., & Stark, M. D. (2011). The use of a checklist and qualitative notebooks for an interactive process of teaching and learning qualitative research. *Journal of Effective Teaching, 11*(1), 62-79. Retrieved from http://www.uncw.edu/cte/et

Frias-Navarro, D., Pascual-Llobell, J., Pascual-Soler, M., Perezgonzalez, J., & Berrios-Riquelme, J. (2020). Replication crisis or an opportunity to improve scientific production? *European Journal of Education, 55*(4), 618-631. https://doi.org/10.1111/ejed.12417

Friedman, H. H., Fireworker, R. B., & Nagel, H. (2017). Biasing research: Can science be trusted? *Journal of Leadership, Accountability, and Ethics, 14*(2), 105-116. Retrieved from https://articlegateway.com/index.php/JLAE/index

Fusch, P., Fusch, G. E., & Ness, L. R. (2018). Denzin's paradigm shift: Revisiting triangulation in qualitative research. *Journal of Social Change, 10*(1), 19-32. https://doi.org/10.5590/JOSC.2018.10.1.02

Geertz, C. (1984). Distinguished lecture: Anti anti-relativism. *American Anthropologist, 86*(2), 263-278. Retrieved from https://www.americananthro.org/

Gehman, J., Glaser, V. L., Eisenhardt, K. M., Gioia, D., Langley, A., & Corley, K. G. (2018). Finding theory–method fit: A comparison of three qualitative approaches to theory building. *Journal of Management Inquiry, 27*(3), 284-300. https://doi.org/10.1177%2F1056492617706029

Gentles, S. J., Charles, C., Ploeg, J., & McKibbon, K. A. (2015). Sampling in qualitative research: Insights from an overview of the methods literature. *The Qualitative Report, 20*(11), 1772-1789. https://doi.org/10.46743/2160-3715/2015.2373

Gergen, K. J. (2014). Pursuing excellence in qualitative inquiry. *Qualitative Psychology, 1*, 49–60. http://dx.doi.org/10.1037/qup0000002

Gilligan, C., & Eddy, J. (2017). Listening as a path to psychological discovery: An introduction to the Listening Guide. *Perspectives on Medical Education, 6*(2), 76-81. https://doi.org/10.1007/s40037-017-0335-3

Gioia, D. (2021). A systematic methodology for doing qualitative research. *The Journal of Applied Behavioral Science, 57*(1), 20-29. https://doi.org/10.1177%2F0021886320982715

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Aldine.

Glesne, C., & Peshkin, A. (1992). *Becoming qualitative researchers.* White Plains.

Goertzen, M. J. (2017). Introduction to quantitative research and data. *Library Technology*

Reports, 53(4), 12-18. Retrieved from https://journals.ala.org/index.php/ltr

Goffman, E. (1978). *The presentation of self in everyday life.* Harmondsworth.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report, 8*(4), 597-607. https://doi.org/10.46743/2160-3715/2003.1870

Grbich, C. (2012). *Qualitative data analysis: An introduction*. sage.

Greenhalgh, T., Annandale, E., Ashcroft, R., Barlow, J., Black, N., Bleakley, A., ... & Ziebland, S. (2016). An open letter to The BMJ editors on qualitative research. *BMJ, 352*. https://doi.org/10.1136/bmj.i563

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82*(1), 1-20. https://psycnet.apa.org/doi/10.1037/h0076157

Gringeri, C., Barusch, A., & Cambron, C. (2013). Examining foundations of qualitative research: A review of social work dissertations, 2008–2010. *Journal of Social Work Education, 49*(4), 760-773. https://doi.org/10.1080/10437797.2013.812910

Grodal, S., Anteby, M., & Holm, A. L. (2021). Achieving rigor in qualitative analysis: The role of active categorization in theory building. *Academy of Management Review, 46*(3), 591-612. https://doi.org/10.5465/amr.2018.0482

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*, 59-82. http://dx.doi.org/10.1177/1525822X05279903

Guest, G., Namey, E., & Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. *PloS One*, 15(5), e0232076. https://doi.org/10.1371/journal.pone.0232076

Guetterman, T. C. (2015). Descriptions of sampling practices within five approaches to qualitative research in education and the health sciences. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 16*(2), Art. 25, http://nbn-resolving.de/urn:nbn:de:0114-fqs1502256.

Hallinger, P. (2011). A review of three decades of doctoral studies using the principal instructional management rating scale: A lens on methodological progress in educational leadership. *Educational Administration Quarterly, 47*(2), 271-306. https://doi.org/10.1177/0013161X10383412

Hammersley, M. (2007). The issue of quality in qualitative research. *International Journal of Research & Method in Education, 30*(3), 287-305. https://doi.org/10.1080/17437270701614782

Hammersley, M. (2008). Troubles with triangulation. In M. M. Bergman (Ed.), *Advances in mixed methods research* (pp. 22-36). Sage.

Hannes, K., Heyvaert, M., Slegers, K., Vandenbrande, S., & Van Nuland, M. (2015).

Exploring the potential for a consolidated standard for reporting guidelines for qualitative research: An argument Delphi approach. *International Journal of Qualitative Methods, 14*(4), 1609406915611528. https://doi.org/10.1177%2F1609406915611528

Harrison, J., MacGibbon, L., & Morton, M. (2001). Regimes of trustworthiness in qualitative research: The rigors of reciprocity. *Qualitative Inquiry, 7*(3), 323-345. https://doi.org/10.1177%2F107780040100700305

Harvey, L. (2015). Beyond member-checking: A dialogic approach to the research interview. *International Journal of Research & Method in Education, 38*(1), 23-38. https://doi.org/10.1080/1743727X.2014.914487

Hays, D. G., Wood, C., Dahl, H., & Kirk-Jenkins, A. (2016). Methodological rigor in Journal of Counseling & Development qualitative research articles: A 15-year review. *Journal of Counseling & Development, 94*(2), 172-183. https://doi.org/10.1002/jcad.12074

Hennink, M., & Kaiser, B. N. (2021). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, 114523. https://doi.org/10.1016/j.socscimed.2021.114523

Hernandez, I., & Preston, J. L. (2013). Disfluency disrupts the confirmation bias. *Journal of Experimental Social Psychology, 49*(1), 178-182. https://doi.org/10.1016/j.jesp.2012.08.010

Howe, K., & Eisenhart, M. (1990). Standards for qualitative (and quantitative) research: A prolegomenon. *Educational Researcher, 19*(4), 2-9. https://doi.org/10.3102%2F0013189X019004002

Jamie, K., & Rathbone, A. P. (2021). Using theory and reflexivity to preserve methodological rigour of data collection in qualitative research. *Research Methods in Medicine & Health Sciences*, 26320843211061302. https://doi.org/10.1177%2F26320843211061302

Jerolmack, C., & Khan, S. (2014). Talk is cheap: Ethnography and the attitudinal fallacy. *Sociological Methods & Research, 43*(2), 178-209. https://doi.org/10.1177/0049124114523396

Karnieli-Miller, O., Strier, R., & Pessach, L. (2009). Power relations in qualitative research. *Qualitative Health Research, 19*(2), 279-289. https://doi.org/10.1177/1049732308329306

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition, 2*(1), 42-52. https://doi.org/10.1016/j.jarmac.2013.01.001

Kennedy, M. M. (1979). Generalizing from single case studies. *Evaluation Quarterly, 3*(4), 661-678. https://doi.org/10.1177/0193841X7900300409

Kerasovitis, K. (2020). Post qualitative research-reality through the antihierarchical assemblage of non-calculation. *Qualitative Report, 25*(13), 56-70. https://doi.org/10.46743/2160-3715/2020.4756

Kim, J. H. (2015). *Understanding narrative inquiry: The crafting and analysis of stories as*

research. Sage Publications.

Kline, W. B. (2008). Developing and submitting credible qualitative manuscripts. *Counselor Education and Supervision, 47*(4), 210-217. https://doi.org/10.1002/j.1556-6978.2008.tb00052.x

Knapp, M. S. (2017). The practice of designing qualitative research on educational leadership: Notes for emerging scholars and practitioner-scholars. *Journal of Research on Leadership Education, 12*(1), 26-50. https://doi.org/10.1177/1942775116647365

Kornbluh, M. (2015). Combatting challenges to establishing trustworthiness in qualitative research. *Qualitative Research in Psychology, 12*(4), 397-414. https://doi.org/10.1080/14780887.2015.1021941

Lester, J. N., Cho, Y., & Lochmiller, C. R. (2020). Learning to do qualitative data analysis: A starting point. *Human Resource Development Review*, *19*(1), 94-106. https://doi.org/10.1177/1534484320903890

Levitt, H. M., Motulsky, S. L., Wertz, F. J., Morrow, S. L., & Ponterotto, J. G. (2017). Recommendations for designing and reviewing qualitative research in psychology: Promoting methodological integrity. *Qualitative Psychology, 4*(1), 2-22. https://doi.org/10.1037/qup0000082

Lewin, S., Glenton, C., Munthe-Kaas, H., Carlsen, B., Colvin, C. J., Gülmezoglu, M., ... & Rashidian, A. (2015). Using qualitative evidence in decision making for health and social interventions: An approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). *PLoS Medicine, 12*(10), e1001895. https://doi.org/10.1371/journal.pmed.1001895

Liao, H., & Hitchcock, J. (2018). Reported credibility techniques in higher education evaluation studies that use qualitative methods: A research synthesis. *Evaluation and Program Planning, 68,* 157-165. https://doi.org/10.1016/j.evalprogplan.2018.03.005

Lietz, C. A., & Zayas, L. E. (2010). Evaluating qualitative research for social work practitioners. *Advances in Social Work, 11*(2),188-202. https://doi.org/10.18060/589

Lincoln, Y. S. (2002, November). On the nature of qualitative evidence. Paper presented at the annual meeting of the Association for the Study of Higher Education, Sacramento, CA.

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

Lochmiller, C. R., & Lester, J. N. (2017). Conceptualizing practitioner-scholarship for educational leadership research and practice. *Journal of Research on Leadership Education, 12*(1), 3-25.https://doi.org/10.1177/1942775116668525

Loh, J. (2013). Inquiry into issues of trustworthiness and quality in narrative studies: A perspective. *The Qualitative Report, 18*(33), 1-15. https://doi.org/10.46743/2160-3715/2013.1477

Long, H. A., French, D. P., & Brooks, J. M. (2020). Optimising the value of the critical

appraisal skills programme (CASP) tool for quality appraisal in qualitative evidence synthesis. *Research Methods in Medicine & Health Sciences, 1*(1), 31-42. https://doi.org/10.1177%2F2632084320947559

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher, 48*(3), 158-166. https://doi.org/10.3102/0013189X19832850

Low, J. (2019). A pragmatic definition of the concept of theoretical saturation. *Sociological Focus, 52*(2), 131-139. https://doi.org/10.1080/00380237.2018.1544514

Lowe, A., Norris, A. C., Farris, A. J., & Babbage, D. R. (2018). Quantifying thematic saturation in qualitative data analysis. *Field Methods, 30*(3), 191-207. https://doi.org/10.1177%2F1525822X17749386

Luborsky, M. R., & Rubinstein, R. L. (1995). Sampling in qualitative research: Rationale, issues, and methods. *Research on Aging, 17*(1), 89-113. https://doi.org/10.1177%2F0164027595171005

Lucas, K., & D'Enbeau, S. (2013). Moving beyond themes: Reimagining the qualitative analysis curriculum. *Qualitative Communication Research, 2*(2), 213-227. https://doi.org/10.1525/qcr.2013.2.2.213

Maier, M., Bartoš, F., Stanley, T. D., Shanks, D. R., Harris, A. J., & Wagenmakers, E. J. (2022). No evidence for nudging after adjusting for publication bias. *Proceedings of the National Academy of Sciences, 119*(31), e2200300119. https://doi.org/10.1073/pnas.2200300119

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher, 43*(6), 304-316. https://doi.org/10.3102%2F0013189X14545513

Makin, S. (2016, September 9). What happens in the brain when we misremember. *Scientific American.* Retrieved from https://www.scientificamerican.com/article/what-happens-in-the-brain-when-we-misremember/

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research, 26*(13), 1753-1760. https://doi.org/10.1177/1049732315617444

Marks, G. (2007). Introduction: Triangulation and the square-root law. *Electoral Studies, 26*(1), 1-10. https://doi.org/10.1016/j.electstud.2006.03.009

Marsteintredet, L., & Malamud, A. (2020). Coup with adjectives: Conceptual stretching or innovation in comparative research? *Political Studies, 68*(4), 1014-1035. https://doi.org/10.1177%2F0032321719888857

Mason, M. (2010). Sample size and saturation in PhD studies using qualitative interviews.

Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 11(3). https://doi.org/10.17169/fqs-11.3.1428

Maxwell, J. (1992). Understanding and validity in qualitative research. *Harvard Educational Review, 62*(3), 279-301. https://doi.org/10.17763/haer.62.3.8323320856251826

Mayring, P. (2019). Qualitative content analysis: Demarcation, varieties, developments. *Forum: Qualitative Social Research, 20*(3), 1-26. https://doi.org/10.17169/fqs-20.3.3343

Mays, N., & Pope, C. (2000). Assessing quality in qualitative research. *BMJ, 320*(7226), 50-52. https://doi.org/10.1136/bmj.320.7226.50

McCaslin, M. L., & Scott, K. W. (2003). The five-question method for framing a qualitative research study. *The Qualitative Report, 8*(3), 447-461. https://doi.org/10.46743/2160-3715/2003.1880

McSweeney, B. (2021). Fooling ourselves and others: confirmation bias and the trustworthiness of qualitative research–Part 2 (cross-examining the dismissals). *Journal of Organizational Change Management, 34*(5), 841-859. https://doi.org/10.1108/JOCM-04-2021-0118

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation.* Jossey-Bass.

Meyrick, J. (2006). What is good qualitative research? A first step towards a comprehensive approach to judging rigour/quality. *Journal of Health Psychology, 11*(5), 799-808. https://doi.org/10.1177%2F1359105306066643

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*(1), 156. https://psycnet.apa.org/doi/10.1037/0033-2909.105.1.156

Milne, J., & Oberle, K. (2005). Enhancing rigor in qualitative description. *Journal of Wound Ostomy & Continence Nursing, 32*(6), 413-420. Retrieved from https://journals.lww.com/jwocnonline/

Mitchell, K. M., & Clark, A. M. (2018). Five steps to writing more engaging qualitative research. *International Journal of Qualitative Methods, 17*(1), 1-3. https://doi.org/10.1177%2F1609406918757613

Mizzi, R. (2010). Unraveling researcher subjectivity through multivocality in autoethnography. *Journal of Research Practice, 6*(1), M3. Retrieved from https://jrp.icaap.org/index.php/jrp/article/

Morse, J. M. (2015). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research, 25*(9), 1212-1222. https://doi.org/10.1177%2F1049732315588501

Morse, J. M. (2021). Why the qualitative health research (QHR) review process does not use checklists. *Qualitative Health Research, 31*(5), 819-821.

https://doi.org/10.1177%2F1049732321994114

Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods, 1*(2), 13-22. https://doi.org/10.1177%2F160940690200100202

Munthe-Kaas, H. M., Glenton, C., Booth, A., Noyes, J., & Lewin, S. (2019). Systematic mapping of existing tools to appraise methodological strengths and limitations of qualitative research: First stage in the development of the CAMELOT tool. *BMC Medical Research Methodology, 19*(1), 1-13. https://doi.org/10.1186/s12874-019-0728-6

Nassaji, H. (2020). Good qualitative research. *Language Teaching Research, 24*(4), 427-431. https://doi.org/10.1177%2F1362168820941288

错误!未找到引用源。National Institute for Health and Care Excellence. (2012). Methods for the Development of NICE Public Health Guidance [Internet]. Retrieved from https://nice.org,uk/

Norris, N. (1997). Error, bias and validity in qualitative research. *Educational Action Research, 5*(1), 172-176. https://doi.org/10.1080/09650799700200020

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*(6), 615-631. https://doi.org/10.1177/1745691612459058

O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., & Cook, D. A. (2014). Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine, 89*(9), 1245-1251. https://doi.org/10.1097/ACM.0000000000000388

Olmos-Vega, F. M., Stalmeijer, R. E., Varpio, L., & Kahlke, R. (2022). A practical guide to reflexivity in qualitative research: AMEE Guide No. 149. *Medical Teacher*, 1-11. https://doi.org/10.1080/0142159X.2022.2057287

Onwuegbuzie, A. J., Frels, R. K., & Hwang, E. (2016). Mapping Saldaña's coding methods onto the literature review process. *Journal of Educational Issues, 2*(1), 130-150. http://dx.doi.org/10.5296/jei.v2i1.8931

Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron? *Quality & Quantity, 41*(2), 233-249. https://doi.org/10.1007/s11135-006-9000-3

Onwuegbuzie, A. J., & Leech, N. L. (2010). Generalization practices in qualitative research: A mixed methods case study. *Quality & Quantity, 44*(5), 881-892. https://doi.org/10.1007/s11135-009-9241-z

Oplatka, I. (2021). Eleven pitfalls in qualitative research: Some perils every emerging scholar and doctoral student should be aware of! *The Qualitative Report, 26*(6), 1881-1890. https://doi.org/10.46743/2160-3715/2021.4783

Orb, A., Eisenhauer, L., & Wynaden, D. (2001). Ethics in qualitative research. *Journal of Nursing Scholarship, 33*(1), 93-96. https://doi.org/10.1111/j.1547-5069.2001.00093.x

O'Reilly, M., & Parker, N. (2013). 'Unsatisfactory saturation': A critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research, 13*(2), 190-197. https://doi.org/10.1177%2F1468794112446106

Palaganas, E. C., Sanchez, M. C., Molintas, V. P., & Caricativo, R. D. (2017). Reflexivity in qualitative research: A journey of learning. *The Qualitative Report, 22*(2), 426-438. https://doi.org/10.46743/2160-3715/2017.2552

Parker, I. (2004). Criteria for qualitative research in psychology. *Qualitative Research in Psychology, 1*(2), 95-106. https://doi.org/10.1191/1478088704qp010oa

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research, 34*(5 Pt 2), 1189-1208. Retrieved from https://www.hsr.org/

Patton, M. Q. (2003). Qualitative evaluation checklist. *Evaluation Checklists Project, 21*, 1-13. Retrieved from www.wmich.edu/evalctr/checklists

Peng, K., Ames, D., & Knowles, E. D. (2001). Culture and human inference: Perspectives from three traditions. In D. Masumoto (Ed.), *Handbook of culture and psychology*, 243-263. Oxford University Press

Perry, T., Morris, R., & Lea, R. (2022). A decade of replication study in education? A mapping review (2011–2020). *Educational Research and Evaluation, 27*(1-2), 12-34.

https://doi.org/10.1080/13803611.2021.2022315

Pillow, W. (2003). Confession, catharsis, or cure? Rethinking the uses of reflexivity as methodological power in qualitative research. *International Journal of Qualitative Studies in Education, 16*(2), 175-196. https://doi.org/10.1080/0951839032000060635

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*(4), 531-544. https://doi.org/10.1177%2F014920638601200408

Porritt, K., Gomersall, J., & Lockwood, C. (2014). JBI's systematic reviews: Study selection and critical appraisal. *AJN The American Journal of Nursing, 114*(6), 47-52. https://doi.org/10.1097/01.NAJ.0000450430.97383.64

Porter, S. (2007). Validity, trustworthiness and rigour: Reasserting realism in qualitative research. *Journal of Advanced Nursing, 60*(1), 79-86. https://doi.org/10.1111/j.1365-2648.2007.04360.x

Potter, J., & Hepburn, A. (2005). Qualitative interviews in psychology: Problems and possibilities. *Qualitative Research in Psychology, 2*(4), 281-307. https://doi.org/10.1191/1478088705qp045oa

Poucher, Z. A., Tamminen, K. A., Caron, J. G., & Sweet, S. N. (2020). Thinking through and designing qualitative research studies: A focused mapping review of 30 years of qualitative research in sport psychology. *International Review of Sport and Exercise Psychology, 13*(1), 163-186. https://doi.org/10.1080/1750984X.2019.1656276

Reilly, R. C. (2013). Found poems, member checking and crises of representation. *The Qualitative Report, 18*(30), 1-18. https://doi.org/10.46743/2160-3715/2013.1534

Richards, H. M., & Schwartz, L. J. (2002). Ethics of qualitative research: Are there special issues for health services research? *Family Practice, 19*(2), 135-139. https://doi.org/10.1093/fampra/19.2.135

Rocco, T. S. (2010). Criteria for evaluating qualitative studies. *Human Resource Development International, 4*, 375-378. https://doi.org/10.1080/13678868.2010.501959

Ross, P. T., & Bibler Zaidi, N. L. (2019). Limited by our limitations. *Perspectives on Medical Education, 8*(4), 261-264. https://doi.org/10.1007%2Fs40037-019-00530-x

Roulston, K., & Shelton, S. A. (2015). Reconceptualizing bias in teaching qualitative research methods. *Qualitative Inquiry, 21*(4), 332-342. https://doi.org/10.1177/1077800414563803

Salzman, P. C. (2002). On reflexivity. *American Anthropologist, 104*(3), 805-811. https://doi.org/10.1525/aa.2002.104.3.805

Sandelowski, M., & Barroso, J. (2003). Classifying the findings in qualitative studies. *Qualitative Health Research, 1*3(7), 905-923. https://doi.org/10.1177%2F1049732303253488

Sandelowski, M., Voils, C. I., & Knafl, G. (2009). On quantitizing. *Journal of Mixed Methods Research, 3*(3), 208-222. https://doi.org/10.1177%2F1558689809334210

Sartori, G. (1970). Concept misformation in comparative politics. *American Political Science Review, 64*(4), 1033-1053. https://doi.org/10.2307/1958356

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., ... & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity, 52*(4), 1893-1907. https://doi.org/10.1007/s11135-017-0574-8

Shaver, J. P., & Norton, R. S. (1980). Randomness and replication in ten years of the American Educational Research Journal. *Educational Researcher, 9*(1), 9-15. https://doi.org/10.3102%2F0013189X009001009

Shaw, J. (2020). Do false memories look real? Evidence that people struggle to identify rich false memories of committing crime and other emotional events. *Frontiers in Psychology, 11*, 650. https://doi.org/10.3389/fpsyg.2020.00650

Sikes, P. (2000). 'Truth' and 'lies' revisited. *British Educational Research Journal, 26*(2), 257-270. https://doi.org/10.1080/01411920050000980

Simpson, A., & Quigley, C. F. (2016). Member checking process with adolescent students: Not just reading a transcript. *The Qualitative Report, 21*(2), 376-392. https://doi.org/10.46743/2160-3715/2016.2386

Smith, B., & McGannon, K. R. (2018). Developing rigor in qualitative research: Problems and opportunities within sport and exercise psychology. *International Review of Sport and*

Exercise Psychology, 11(1), 101-121. https://doi.org/10.1080/1750984X.2017.1317357

Smith, G. D., Gelling, L., Haigh, C., Barnason, S., Allan, H. T., & Jackson, D. (2018). The position of reporting guidelines in qualitative nursing research. *Journal of Clinical Nursing, 27*(5-6), 889-891. http://dx.doi.org/10.1111/jocn.14213

Spencer, L., Ritchie, J., Lewis, J., & Dillon, L. (2004). Quality in qualitative evaluation: A framework for assessing research evidence. Government's Chief Social Researcher's Office.

Stenfors, T., Kajamaa, A., & Bennett, D. (2020). How to… assess the quality of qualitative research. *The Clinical Teacher, 17*(6), 596-599. https://doi.org/10.1111/tct.13242

Stiles, W. B. (1993). Quality control in qualitative research. *Clinical Psychology Review, 13*, 593–618. http://dx.doi.org/10.1016/0272-7358(93)90048-Q

St. Pierre, E. A. (2021). The lure of the new and the hold of the dogmatic. *Qualitative Inquiry, 27*(5), 480-490. https://doi.org/10.1177/1077800420939133

Taylor, E.W., Beck, J., & Ainsworth, E. (2001). Publishing qualitative adult education research: A peer review perspective. *Studies in the Education of Adults, 33*(2), 163–79. https://doi.org/10.1080/02660830.2001.11661452

Thomas, D. R. (2017). Feedback from research participants: Are member checks useful in qualitative research? *Qualitative Research in Psychology, 14*(1), 23-41. https://doi.org/10.1080/14780887.2016.1219435

Thorne, S. (2000). Data analysis in qualitative research. *Evidence-Based Nursing, 3*(3), 68-70. http://dx.doi.org/10.1136/ebn.3.3.68

Thurmond, V. A. (2001). The point of triangulation. *Journal of Nursing Scholarship, 33*(3), 253-258. https://doi.org/10.1111/j.1547-5069.2001.00253.x

Tilman, D. (1991). The schism between theory and ardent empiricism: A reply to Shipley and Peters. *The American Naturalist, 138*(5), 1283-1286. https://doi.org/10.1086/285285

Tong, A., Flemming, K., McInnes, E., Oliver, S., & Craig, J. (2012). Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology, 12*(1), 1-8. https://doi.org/10.1186/1471-2288-12-181

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care, 19*(6), 349-357. https://doi.org/10.1093/intqhc/mzm042

Tooley, J., & Darby, D. (1998). *Educational research: A critique: A survey of published educational research.* London: Office for Standards in Education.

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry, 16*(10), 837-851. https://doi.org/10.1177%2F1077800410383121

Treharne, G. J., & Riggs, D. W. (2014). Ensuring quality in qualitative research. In P.

Rohleder & A. Lyons (Eds.), *Qualitative research in clinical and health psychology* (pp. 57-73). Palgrave.

Tufford, L., & Newman, P. (2012). Bracketing in qualitative research. *Qualitative Social Work, 11*(1), 80-96. https://doi.org/10.1177%2F1473325010368316

Ullrich, C., Stürmlinger, A., Wensing, M., & Krug, K. (2020). Qualitative research methods in medical dissertations: An observational methodological study on prevalence and reporting quality of dissertation abstracts in a German university. *BMC Medical Research Methodology, 20*(1), 1-9. https://doi.org/10.1186/s12874-020-01186-6

Van Damme, I., & Smets, K. (2014). The power of emotion versus the power of suggestion: Memory for emotional events in the misinformation paradigm. *Emotion, 14*(2), 310-320. https://psycnet.apa.org/doi/10.1037/a0034629

van Manen, M. (2007). Phenomenology of practice. *Phenomenology & Practice, 1*(1), 11-30. Retrieved from https://journals.library.ualberta.ca/pandpr/index.php/pandpr/

Varpio, L., Ajjawi, R., Monrouxe, L. V., O'Brien, B. C., & Rees, C. E. (2017). Shedding the cobra effect: Problematising thematic emergence, triangulation, saturation and member checking. *Medical Education, 51*(1), 40-50. https://doi.org/10.1111/medu.13124

Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology, 18*(1), 1-18. https://doi.org/10.1186/s12874-018-0594-7

Vollstedt, M., & Rezat, S. (2019). An introduction to grounded theory with a special focus on axial coding and the coding paradigm. In G. Kaiser & N. Presmeg (Eds.), *Compendium for Early Career Researchers in Mathematics Education* (pp. 81-100). Springer. https://doi.org/10.1007/978-3-030-15636-7

von Unger, H. (2016). Reflexivity beyond regulations: Teaching research ethics and qualitative methods in Germany. *Qualitative Inquiry, 22*(2), 87-98. https://doi.org/10.1177/1077800415620220

Walsh, D., & Downe, S. (2005). Meta-synthesis method for qualitative research: A literature review. *Journal of Advanced Nursing, 50*(2), 204-211. https://doi.org/10.1111/j.1365-2648.2005.03380.x

Walther, J., Sochacka, N. W., & Kellam, N. N. (2012). Challenges to ensuring quality in qualitative research: A procedural view. In *2012 ASEE Annual Conference & Exposition* (pp. 25.298.1-17). https://doi.org/10.18260/1-2-21056

Watkins, D. C. (2017). Rapid and rigorous qualitative data analysis: The "RADaR" technique for applied research. *International Journal of Qualitative Methods, 16*(1), 1609406917712131. https://doi.org/10.1177%2F1609406917712131

Watt, D. (2007). On becoming a qualitative researcher: The value of reflexivity. *Qualitative*

Report, 12(1), 82-101. Retrieved from http://www.nova.edu/ssss/QR/QR12-1/watt.pdf

Weller, S. C., Vickers, B., Bernard, H. R., Blackburn, A. M., Borgatti, S., Gravlee, C. C., & Johnson, J. C. (2018). Open-ended interview questions and saturation. *PloS One, 13*(6), e0198606. https://doi.org/10.1371/journal.pone.0198606

Whittemore, R., Chase, S. K., & Mandle, C. L. (2001). Validity in qualitative research. *Qualitative Health Research, 11*(4), 522-537. https://doi.org/10.1177%2F104973201129119299

Williams, E. N., & Morrow, S. L. (2009). Achieving trustworthiness in qualitative research: A pan-paradigmatic perspective. *Psychotherapy Research, 19*(4-5), 576-582. https://doi.org/10.1080/10503300802702113

Yadav, D. (2021). Criteria for good qualitative research: A comprehensive review. *The Asia-Pacific Education Researcher*, 1-11. https://doi.org/10.1007/s40299-021-00619-0

Yonge, O., & Stewin, L. (1988). Reliability and validity: Misnomers for qualitative research. *Canadian Journal of Nursing Research Archive*, 61-68. Retrieved from https://cjnr.archive.mcgill.ca/article/view/1000

Yu, C. H., & Ohlund, B. (2010). Threats to validity of research design. Retrieved from http://www.creative-wisdom.com/teaching/WBI/threat.shtml

Zahavi, D. (2019). Getting it quite wrong: Van Manen and Smith on phenomenology. *Qualitative Health Research, 29*(6), 900-907. https://doi.org/10.1177/1049732318817547

Zmigrod, L. (2022). A psychology of ideology: Unpacking the psychological structure of ideological thinking. *Perspectives on Psychological Science, 17*(4), 1072-1092. https://doi.org/10.1177/17456916211044140

Appendix A

Authors cited in Tables 1, 2, and 3.

1. Lewin et al., 2015	21. Patton, 2003	42. Parker, 2004
2. Dixon et al., 2004	22. Lietz & Zayas, 2010	43. Stiles, 1993
3. Spencer et al., 2004	23. Frels et al., 2011	44. Tracy, 2010
4. Merriam, 2009	24. Goertzen, 2017	45. McCaslin & Scott, 2003
5. de Casterlé et al., 2012	25. C. Anderson, 2010	46. Mitchell & Clark, 2018
6. Meyrick, 2006	26. Erickson, 1986	47. Chenail, 2011
7. Cooke et al., 2012	27. Lincoln, 2002	48. V. Anderson, 2017
8. Sandelowski & Barroso, 2003	28. Cutcliffe & McKenna, 2004	49. Rocco, 2010
9. Morse, 2015	29. Podsakoff & Organ, 1986	50. Caelli et al., 2003
10. Glaser Dissertation Timeline,	30. Treharne & Riggs, 2014	51. Taylor et al., 2001
11. Morse et al., 2002	31. Walther et al., 2012	52. Williams & Morrow, 2009
12. Bowen, 2006	32. Aref et al., 2021	53. Levitt et al., 2017
13. Biemer, 2010	33. Jamie & Rathbone, 2021	54. Drisko, 1997
14. De Jong et al., 2021; Tong et al., 2012	34. Watkins, 2017	55. Barusch et al., 2011
15. G. Smith et al., 2018	35. Freeman et al., 2007	56. Nassaji, 2020
16. O'Brien et al., 2014	36. Howe & Eisenhart, 1990	57. Epp & Otnes, 2021
17. Coker, 2022a	37. Milne & Oberle, 2005	58. Tong et al., 2007
18. National Institute for Health and Care Excellence, 2012	38. Amankwaa, 2016	59. Hammersley, 2007
19. Porritt et al., 2014	39. Kline, 2008	60. Morse, 2021
20. Long et al., 2020	40. R. Elliott et al., 1999	61. Munthe-Kaas et al., 2019
	41. Gergen, 2014	62. Whittemore et al., 2001