



Beitrag aus Zeitschrift für Evaluation

17. Jg. · Heft 1 · 2018

ISSN 1619-5515
© Waxmann Verlag GmbH 2018
Steinfurter Str. 555
48159 Münster
www.waxmann.com

Gütekriterien qualitativer Evaluationsforschung

Philipp Mayring¹

Zusammenfassung: Auf der Suche nach Standards, Maßstäben, Gütekriterien für qualitativ orientierte Evaluationsprojekte wird zunächst Evaluation als Evaluationsforschung definiert, die sich also auch Forschungsstandards unterwerfen muss. Es wird in Frage gestellt, ob sich qualitative von quantitativer Forschung eindeutig abheben lässt und somit völlig eigene Standards für qualitative Evaluation anzuwenden seien. Schließlich werden Gütekriterien für Evaluationsforschung, die auch für qualitative Projekte anwendbar sind, auf mehreren Ebenen diskutiert: Allgemeine Kriterien, die sich am Forschungsablauf orientierten (Fragestellung – Theorieeinbettung – Design/Methoden – Ergebnisse – Interpretation/Konsequenzen), aus quantitativer Forschung kommende Standards (Objektivität, Reliabilität, Validität), Standards, die sich konkret auf Evaluation beziehen (DeGEval-Standards, Standards für einzelne qualitative Evaluationsdesigns sowie konkrete Methodenstandards). Es wird dafür plädiert, sich auf ein möglichst breites Spektrum von Standards zu beziehen.

Schlagwörter: Gütekriterien, qualitative Forschung, Evaluation, Mixed Methods, Standards

Quality Standards for Qualitative Evaluation Research

Abstract: Looking for quality criteria or standards for qualitative evaluation the paper defines evaluation as evaluation research, and therefore qualitative evaluation has to meet standards for good empirical research. The differentiation between qualitative and quantitative evaluation seems to be difficult and less useful, so we have to look at general research criteria instead of developing specific qualitative standards. Those standards are discussed at different levels: standards oriented at the steps of scientific analysis (research question – theoretical basis – design/methods – results – discussion/consequences); general standards like objectivity, reliability, validity; standards for evaluation (DeGEval-standards, standards for qualitative evaluation designs); standards for specific methods. The paper argues for a broad reference to quality criteria in qualitative evaluation.

Keywords: Quality Criteria, Qualitative Research, Evaluation, Mixed Methods, Standards

1 Alpen-Adria Universität Klagenfurt

1. Begriffliche Vorbemerkungen: Evaluation versus Evaluationsforschung?

Bevor versucht wird, Qualitätsstandards für qualitativ orientierte Evaluationsforschung zu diskutieren, sind begriffliche Festlegungen notwendig. Bisweilen wird nämlich ein Unterschied zwischen Evaluation und Evaluationsforschung gemacht. Evaluation sei die Durchführung von Bewertungsmaßnahmen in der Praxis und würde nicht harten Wissenschaftsstandards genügen müssen, nur Evaluationsforschung müsse sich an gängige Gütekriterien halten. Demgegenüber soll hier von einer Definition von Evaluation ausgegangen werden, die immer wissenschaftlich fundiertes Vorgehen impliziert. Döring und Bortz (2015) setzen in ihrem Lehrbuch ebenfalls Evaluation mit wissenschaftlicher Evaluation gleich und beschreiben ihr Vorgehen als „auf der Basis von Forschungsergebnissen in einem wissenschaftlichen Prozess durch sozialwissenschaftlich qualifizierte Evaluationsfachleute“ (Döring/Bortz 2015: 977) durchgeführt. In eine ähnliche Richtung weisen Spiel/Grading/Lüftenegger (2010) und Wottawa/Thierau (2003). Nimmt man diese Begriffsbestimmung nicht auf, so besteht die Gefahr, Evaluation nicht mehr von wertenden Stellungnahmen abgrenzen zu können. Hier scheint auch keine Zwischenposition möglich (Evaluation eher oder teilweise an Wissenschaft orientiert), denn das wissenschaftliche Vorgehen ist durch klare Kriterien gekennzeichnet, die erfüllt oder nicht erfüllt werden.

Auch eine Unterscheidung von Evaluationsforschung als wissenschaftliche Tätigkeit von Evaluation als Praxis oder Anwendung mit weniger starken Anforderungen kann nicht genügen, da Praxisanwendungen immer wissenschaftlich begründet sein müssen, gerade im Bereich der Evaluation, in dem die Konsequenzen von falschen Ergebnissen oder Fehleinschätzungen fatal sein können.

Dies mag zunächst als Bürde, als Last, als Zusatzaufgabe erscheinen gegenüber einem einfachen Evaluationsverständnis. Die Vorteile für Evaluation sind aber zweierlei. Zum einen gewinnt die Evaluation an Genauigkeit und Verlässlichkeit. Zum anderen profitiert sie von dem grundlegenden Ansatz wissenschaftlichen Vorgehens: Konkrete Beobachtungen, empirisches Material wird gesammelt und ausgewertet, um allgemeinere Schlussfolgerungen abzuleiten, die Ergebnisse also zu Theorie zu generalisieren. Diese Theoriebausteine, zum Beispiel allgemeine Variablenzusammenhänge, lassen sich auf zukünftige Situationen anwenden, ohne dass eine erneute Untersuchung nötig wäre (wenn die Äquivalenz der Situationen sichergestellt ist). Es wird eine Wissensbasis geschaffen, die auf konkrete Probleme angewandt werden kann. Wenn die Ergebnisse einer Evaluation (z.B. einer Maßnahme betrieblicher Gesundheitsförderung) also wirklich wissenschaftlich fundiert zustande gekommen sind, lassen sie sich auf ähnliche Betriebe generalisieren; eine erneute Evaluation in einem weiteren Betrieb erscheint nicht mehr unbedingt notwendig. Deshalb sollte gute Evaluation(-sforschung) als ersten Schritt immer eine Analyse des Stands der Forschung umfassen und prüfen, ob zum Evaluationsgegenstand bereits verlässliche Evaluationsergebnisse vorliegen, an die angeknüpft werden kann. Auf diese Weise wird Evaluation dann auch wieder sparsam.

Wenn man also Evaluation auf diesem Hintergrund als Bewertung von Konzepten oder Maßnahmen mit empirisch-wissenschaftlichen Methoden versteht, so folgt daraus, dass sich Evaluation auch den wissenschaftlichen Gütekriterien unterwerfen muss. Dies soll in diesem Beitrag weiter untersucht werden.

2. Die Qualitativ-quantitativ-Kontroverse in der Forschungsmethodik

Wenn nun festgelegt wurde, dass sich Evaluation den Wissenschaftsstandards empirischer Forschung zu stellen hat, erhebt sich das zweite Problem, auf welche Art von Wissenschaft sie sich beziehen soll. Die Kontroverse zwischen qualitativer und quantitativer Forschungsmethodik, mit jeweils sehr unterschiedlichen Standards und Gütekriterien, reicht weit zurück und ist immer noch virulent.

Ein Blick in die Wissenschaftsgeschichte zeigt eine stetige Pendelbewegung (vgl. Mayring 2016). Am Beginn stehen dabei eher qualitativ orientierte Ansätze. Das bis ins Mittelalter bindende Wissenschaftsverständnis geht auf Aristoteles zurück, der ein qualitativ-beschreibendes, kategorisierendes Vorgehen entwickelte, induktives Schließen auf Verallgemeinerungen zulassend, Prozesse (Werden und Vergehen) besonders betonend, alles Elemente, die wir heute eher dem qualitativen Wissenschaftsparadigma zuordnen würden. Quantitatives Denken hat sich dann zu Beginn der Neuzeit (Galileo Galilei, Descartes, Newton) entwickelt, gekennzeichnet durch die Suche nach raum-zeitlich unabhängigen Naturgesetzen, durch Messung und Experiment. Aus der Kritik am naturwissenschaftlich-mechanistischen Weltbild Descartes hat Gianbattista Vico in der ersten Hälfte des 18. Jahrhunderts ein kulturwissenschaftliches Gegenkonzept („Scienza Nuova“) entwickelt, das den Untersuchungsgegenstand in seinen historischen und kulturellen Kontext einordnet. Vor allem im 19. Jahrhundert haben sich hermeneutische Ansätze entwickelt, die statt gesetzesbezogener Erklärungen nach einem Verstehen des Gegenstandes in einem Wechselspiel zwischen Interpret und Interpretiertem suchten. Bei der Etablierung humanwissenschaftlicher Einzeldisziplinen zu Beginn des 20. Jahrhunderts (erste Universitätsinstitute für Soziologie, Psychologie, usw.) kamen dann wieder quantitative Ansätze auf, wie die großangelegte Umfrageforschung in der Soziologie oder psychologische Experimente. Moderne qualitative Ansätze haben sich vor allem in der Nachkriegszeit etabliert. Ihre Akzeptanz als gleichwertige wissenschaftliche Verfahren zeigt sich jedoch sehr unterschiedlich in den Einzeldisziplinen, von geringer Akzeptanz in Psychologie oder Medizin bis zu Gleichwertigkeit in Soziologie oder Erziehungswissenschaft.

So hat uns diese Entwicklung in eine stark kontroverse Situation in den Humanwissenschaften geführt. Hier ist von zwei unterschiedlichen Kulturen wissenschaftlichen Vorgehens (vgl. Snow 1998), ja sogar von „science wars“ (Ross 1996; Bammé 2004) die Rede. Die gängigen Methodenhand- und Lehrbücher lassen sich meist eindeutig zuordnen. Die quantitativ orientierten Lehrbücher (vgl. z.B. Eid/Gollwitzer/Schmitt 2015; Bierhoff/Petermann 2015) setzen Forschung mit quanti-

tativem Hypothesentesten mit Variablen, die durch standardisierte Testverfahren erhoben wurden, gleich. Qualitativ orientierte Methodenhandbücher (vgl. Denzin/Lincoln 2011; Flick 2005) gehen von einem unterschiedlichen Wissenschaftsverständnis aus und beschreiben völlig andere Erhebungs- und Auswertungsmethoden. Etliche neuere Methodenliteratur versucht, auf beides einzugehen (vgl. z.B. Baur/Blasius 2014; Hussy/Schreier/Echterhoff 2013), stellt sie aber mehr oder weniger unabhängig dar, jeweils ihrer eigenen (völlig unterschiedlichen) Logik folgend. Eine wirkliche Integration gelingt nicht. Ob eine solche Integration im Rahmen der in letzter Zeit viel diskutierten Mixed-Methods-Ansätze (vgl. Mayring 2012; Mayring et al. 2007) völlig umgesetzt ist, erscheint ebenso fraglich. Hier wird meist vorgeschlagen, die quantitativen Teile des Mixed-Methods-Projektes nach quantitativen Gütekriterien durchzuführen und die qualitativen Teile nach qualitativen Kriterien (vgl. z.B. Flick 2011). Onwuegbuzie/Johnson (2006) diskutieren verschiedene Legitimationstypen für Mixed Methods. Auch Dellinger/Leech (2007) betonen neben traditionellen Gütekriterien für quantitative Analyseschritte und alternativen Gütekriterien für qualitative Analyseschritte die Notwendigkeit eigener Gütekriterien für die Kombination (Angemessenheit und Transparenz des Mixed-Methods-Konzeptes), bleiben damit aber recht vage. Auf der Suche nach klaren, einheitlichen Kriterien für angemessenes Forschen bleiben wir also auch hier ratlos. So soll im Folgenden der Frage der angemessenen Gütekriterien näher nachgegangen werden.

3. Die Diskussion um Standards empirischer Sozialforschung

Auf der Suche nach Mindestanforderungen an Methoden qualitativ orientierter Evaluationsforschung könnte man nun die Spur der qualitativen Forschung verfolgen und sich an dort diskutierten Gütekriterien orientieren. Gerade aber angesichts der eben dargestellten Methodenkontroverse erscheint dies gefährlich. Personen, die sich nicht dezidiert dem qualitativen ‚Lager‘ zuordnen, würden die Forschungsergebnisse vorschnell ablehnen, und oft sind dies auch die Auftraggeber(innen) von Evaluation. Die im Rahmen von Evaluation diskutierte Forderung nach Evidenzbasierung wird oft als quantitative Evidenz, einem experimentellen Paradigma mit Interventions- und Kontrollgruppen folgend verstanden. Qualitative Evaluation, die sich davon prinzipiell absetzt, gerät in die Gefahr, als unwissenschaftlich abgetan zu werden. Deshalb möchte ich hier Ansätze vorstellen, die nach gemeinsamen Qualitätsmaßstäben für qualitativ und quantitativ orientiertes Vorgehen suchen und damit auch für Mixed-Methods-Ansätze relevant sind.

Dies erscheint umso wichtiger, als eine klare Abgrenzung qualitativer von quantitativer Forschung schwierig erscheint. Auch in qualitativer Forschung wird mit Zahlen, mit numerischen Begriffen und Relationen gearbeitet. Das Auftreten untersuchter Phänomene wird als zentral (= häufig) bezeichnet, es wird mit wenn auch kleinen Stichproben gearbeitet, Zusammenhänge werden untersucht, also quantitative Begrifflichkeiten, was eine klare Abgrenzung schwierig macht. Weiterhin wird quantitatives Forschen als linear konzipiert, von der Forschungsfrage zu ihrer Be-

antwortung in den Ergebnissen im Gegensatz zum zirkulären Vorgehen qualitativer Forschung (vgl. z.B. Flick 2005), bei der von einem offeneren Forschungsthema durch empirische Ergebnisse die Forschungsfrage weiter präzisiert wird und erneute empirische Erhebungen folgen. Dabei wird übersehen, dass auch hier eine Beantwortung der (revidierten) Forschungsfrage im Sinne eines linearen Prozesses folgt. Gerade in Evaluationsforschung muss ja am Ende eine Beantwortung der Bewertungsfrage stehen und nicht ein ergebnisloses Kreisen um den Gegenstand (vgl. dazu Mayring 2010). Oft wird quantitatives Forschen als variablenbezogen im Gegensatz zur Subjektbezogenheit qualitativer Forschung bezeichnet (vgl. z.B. Husy/Schreier/Echterhoff 2013). Dabei wird übersehen, dass quantitative Forschung natürlich ebenso subjektbezogen sein möchte, eine möglich genaue und adäquate Erfassung der Probanden anstrebt. Auf der anderen Seite spielen in qualitativer Forschung Variablen ebenso eine Rolle, wenngleich auf einem niedrigeren Skalenniveau. Wenn in einer klassischen qualitativen Studie (vgl. Girtler 1980) die Auswirkungen von Obdachlosigkeit auf die Betroffenen untersucht werden, so ist dies auch als ein Variablenzusammenhang (Obdachlosigkeit und subjektives Befinden) zu verstehen. Gerade Evaluationsforschung bezieht sich immer auch auf eine Bewertungsvariable.

Da die Abgrenzung qualitativ – quantitativ nicht so einfach möglich ist, werden immer wieder Versuche unternommen, die Gemeinsamkeiten zu betonen und übergreifende Standards zu entwickeln. Sehr einflussreich waren hier King/Keohane/Verba (1994). Sie bezeichnen qualitativ und quantitativ zwar als unterscheidbare Forschungsstile, die aber der gleichen Logik des Ableitens allgemeiner Schlussfolgerungen, kausaler Zusammenhänge folgen. Auch explorative oder deskriptive Analyseschritte haben das Ziel, zu Generalisierungen zu führen. Gerade in Evaluationsforschung ist dies unmittelbar einsichtig, da auch qualitative Evaluation schlussendlich die Frage der Wirksamkeit der untersuchten Intervention und damit eine Kausalbeziehung beantworten soll. Somit wären auch Gütestandards ähnlich zu konzipieren.

Im Folgenden möchte ich Ansätze für Gütekriterien für qualitative wie auch quantitative Evaluationsforschung auf drei Ebenen untersuchen: auf der Ebene allgemeinen wissenschaftlichen Vorgehens, auch im Sinne eines gemeinsamen Ablaufmodelles, auf der Ebene übergeordneter Kriterien wie Objektivität, Reliabilität und Validität sowie auf der Ebene konkreter methodenspezifischer Anforderungen, bezogen auf Evaluation generell, auf spezifische Evaluationsdesigns sowie konkreter Erhebungs- und Auswertungsmethoden.

3.1 Allgemeine Standards, auf den Ablauf wissenschaftlichen Arbeitens bezogen

Ein international einflussreiches, im deutschsprachigen Raum viel zu wenig diskutiertes Modell stammt vom US-amerikanischen National Research Council NRC, der zentralen amerikanischen Forschungsförderungsinstitution. Hier kam es zu Beschwerden, warum vom NRC nur selten Forschungsprojekte mit qualitativer Me-

thodik gefördert werden. Eine Kommission hat daraufhin Wissenschaftskriterien ausgearbeitet, die für qualitative (aber auch für quantitative) Projekte zentral bei der Begutachtung sein sollen. Diese Prinzipien, die sich auf den Ablauf eines Forschungsprojektes beziehen, Schritte im wissenschaftlichen Arbeiten darstellen, sind (vgl. National Research Council 2002: 52):

- Pose significant questions that can be investigated empirically.
- Link research to relevant theory.
- Use methods that permit direct investigation of the question.
- Provide a coherent and explicit chain of reasoning.
- Replicate and generalize across studies.
- Disclose research to encourage professional scrutiny and critique.

Mit dem ersten Standard ist gemeint, dass das Projekt nicht nur eine vage Thematik als Ausgangspunkt haben sollte, sondern eine klare Forschungsfrage. Ein Thema aus verschiedenen Blickwinkeln zu bearbeiten fällt eher in den Bereich journalistischer Tätigkeit. Auch der zweite Punkt des Anknüpfens an Theorie ist nicht selbstverständlich in qualitativer Forschung, wird dort oft mit rein deduktivem, hypothesengesteuertem Vorgehen assoziiert. Gemeint ist hier aber, dass auch Ergebnisse qualitativer Forschung an den bisherigen Wissensstand anknüpfen und diesen erweitern sollten, im Sinne eines allgemeinen Theoriebegriffes als Summe des bisherigen Gegenstandswissens. Die genaue Begründung und Beschreibung der Methodik, die sich auf den Untersuchungsgegenstand bezieht, stellt einen weiteren Schwachpunkt vieler qualitativ orientierter Forschungsanträge dar. Der vierte Punkt in der Liste der Standards des NRC bezieht sich auf die Forschungslogik. Dies bedeutet vor allem die Begründung und Ausarbeitung eines Forschungsdesigns, das sich auf die zu untersuchende Fragestellung klar bezieht. Weiterhin haben qualitativ orientierte Studien oft das Problem kleiner, wenig repräsentativer Stichproben. Hier sind Replikationen und Argumentationen der Generalisierung der Forschungsergebnisse besonders wichtig. Schließlich ist die Offenlegung, kritische Diskussion und Veröffentlichung aller Materialien und Analyseschritte zentrales Gütekriterium.

Im Folgenden sind solche Bemühungen um Wissenschaftsstandards für qualitative Forschung, die sich auf die zentralen Schritte des Forschungsprozesses beziehen, vielfach aufgegriffen worden. So hat die amerikanische Erziehungswissenschaftliche Gesellschaft (American Educational Research Association AERA), selbst auch Herausgeberin mehrerer zentraler erziehungswissenschaftlicher Zeitschriften mit ausführlichen Reviewprozessen, eigene Kriterien für qualitativ orientierte Forschungsprojekte entwickelt und veröffentlicht (vgl. AERA 2006), die auch für quantitative Forschungsprozesse Gültigkeit besitzen. Dazu zählen:

- Klare Formulierung des Forschungsproblems, Anknüpfen an den theorie- und praxisbezogenen Wissensstand, Dokumentation der diesbezüglichen Literaturrecherche mit kritischer Einschätzung des Stands der Forschung, Darlegung des theoretischen und methodologischen Bezugsrahmens
- Ausarbeitung eines Forschungsdesigns, Darlegung der Logik der Untersuchung
- Ausarbeitung der empirischen Basis, Festlegung der Studieneinheiten, Begründung und Auswahlstrategie der Stichproben, Materialsammlungsbeschreibung

- Entwicklung und Begründung der Materialstrukturen, gegebenenfalls der Klassifikationsschemata, gegebenenfalls der Transkription sprachlichen Materials und die Analyse der Aussagekraft der erhobenen empirischen Materialien
- Beschreibung und Begründung des Auswertungskonzeptes, der konkreten Auswertungsprozeduren, der abgeleiteten Schlussfolgerungen und Ergebnispräsentationen
- Argumente zur Generalisierung der Forschungsergebnisse
- Forschungsethische Rahmenbedingungen und Prozeduren

Wie wir sehen, wird in diesen Standardkonzeptionen nicht auf den typischen quantitativen Analyseschritten wie eine Hypothesenentwicklung vorab, eine quantitative Datenerhebung mithilfe standardisierter Messinstrumente, Verwendung vollrepräsentativer Stichproben, ein experimentelles Design, eine statistische Auswertung bestanden und damit das Tor für qualitativ orientierte Projekte geöffnet. Damit ist meiner Meinung nach auch eine wirkliche Integration qualitativer und quantitativer Forschung möglich, damit wird der Gegensatz qualitativ vs. quantitativ obsolet.

Solche Standards qualitativer Forschung, wie sie dann auch in anderen Bereichen (z.B. Pflegeforschung) aufgegriffen wurden, sind auf Seiten der qualitativen Forschung nicht immer begrüßt worden. So gelten sie für Norman Denzin in seinem provokanten qualitativen Manifest (Untertitel: „A call to Arms“, Denzin 2010) als unzulässige Einschränkungen gegenstandsspezifischer Offenheit und Variabilität, als Versuche neoliberaler Politik, sich qualitative Forschung zunutze zu machen, als Zurückdrängung kritisch-emanzipatorischer Forschung. Demgegenüber denke ich aber, dass die Kriterien so offen formuliert sind, dass auch explorative oder kritische Forschung sie akzeptieren kann und so überhaupt die Möglichkeit bekommt, Forschungsgelder und Publikationsmöglichkeiten zu bekommen.

Ich habe an anderer Stelle (vgl. Mayring 2012) solche Kriterien noch stärker auf ein allgemeines Ablaufmodell für empirische Sozialforschung (sei sie qualitativ, quantitativ oder Mixed Methods) bezogen und ausgearbeitet und möchte sie hier genauer auf Evaluationsforschung beziehen. Die einzelnen Schritte sind (vgl. Mayring 2012: 298):

1. Konkrete Fragestellung (verallgemeinerte Variablenzusammenhänge), Praxisrelevanz, ggf. Hypothesen, Formulierung des Forscher(innen)standpunkts)
2. Theorieeinordnung (Stand der Forschung, Theorieansatz bzw. integrierte -ansätze (Disziplintheorie, Bereichstheorie, Wissenschaftstheorie inkl. Begründung für Mixed Methods), Vorverständnis bei Interpretationen)
3. Ableitung und Ausarbeitung des Designs (explorativ, deskriptiv, zusammenhangs-, kausalanalytisch, mixed)
4. Klare Stichproben/Materialbestimmung, Samplingstrategie, auch kleine Stichproben, auch unterschiedliche Stichproben
5. Begründete und pilotgetestete (ggf. in der gleichen Studie) Methoden (Erhebung qualitativ und/oder quantitativ, Auswertung textanalytisch und/oder statistisch)
6. Durchführung der Studie(n), bei Auftauchen neuer Aspekte Modifizierung der Fragestellung und neues Durchlaufen der Ablaufschritte
7. Klare Ergebnisdarstellung und Beantwortung der Fragestellung

8. Gütekriterien (Objektivität bzw. Diskussion der Subjekt-Objekt-Beziehung, Validität/Gegenstandsnahe; Reliabilität/Genauigkeit; methodenspezifische Gütekriterien)
9. Generalisierungsargumente, Theoriekonsequenzen
10. Praxiskonsequenzen

Was heißt das nun für qualitativ orientierte Evaluationsprojekte? Das nun Folgende bezieht sich auf die oben genannten zehn Punkte.

Der erste Schritt stellt ein wichtiges Gütekriterium im Sinne des oben skizzierten Verständnisses von Evaluation als Evaluationsforschung dar. Der Ausgangspunkt des Projektes ist zwar in der Regel der Auftrag zur Bewertung einer konkreten Maßnahme oder Intervention; die wissenschaftliche Fragestellung lautet jedoch: Wie sind Maßnahmen vom Typ x (Variable 1) nach den Kriterien a folgende (Variable 2 ff) einzuschätzen? Wenn solche Maßnahmen in ähnlichen Kontexten bereits genügend evaluiert sind, braucht es kein eigenes Evaluationsforschungsprojekt. Die Recherche, ob solche Evaluationen bereits vorliegen, ist dann ein wichtiger zweiter Schritt (Stand der Forschung). Wenn bei der Formulierung der Evaluationsfragestellung bereits sinnvolle Vermutungen über die möglichen Ergebnisse formulierbar sind, so sollte dies in Form von Hypothesen festgehalten werden. Die Funktion von vorab formulierten Hypothesen ist, dass im Forschungsprozess besonders intensiv nach einer Widerlegung der Hypothesen gesucht werden sollte, also im Sinne der Wissenschaftstheorie Poppers eine Falsifikation versucht werden sollte. Dies stärkt die Aussagekraft der Ergebnisse entscheidend. Im Rahmen qualitativ orientierter Projekte wird die Beziehung der Forscherpersonen zum Gegenstand nicht als simple Realitätsabbildung, sondern als Interaktion begriffen (siehe dazu unten im Abschnitt zu Objektivität). Dies, zunächst eine Quelle von Subjektivität, kann kontrolliert werden, indem der Standpunkt der Forscherinnen und Forscher vorab formuliert wird.

Im zweiten Schritt geschieht nun die Theorieeinbettung, das Anknüpfen an den Stand der Forschung, in unserem Falle der Evaluationen zu ähnlichen Gegenständen. Hier stoßen wir schnell auf Konzepte von Mixed Methods, da oft qualitative und quantitative Evidenz zum Evaluationsgegenstand vorliegt. Sowohl die Einbettung in solche Projekte als auch die Verwendung von Mixed Methods im eigenen Evaluationsprojekt erfordert eine gesonderte Begründung der Methodenkombination oder Methodenintegration (vgl. dazu Flick 2011). In der Evaluationsforschung sind weiterhin in den letzten Jahren unterschiedliche Evaluationstheorien diskutiert worden (z.B. konstruktivistische Evaluation, realistische Evaluation, experimentelle Evaluation). Ein wichtiges Gütekriterium wäre hier eine explizite Zuordnung zu einer dieser Evaluationstheorien. Sobald Konzepte aus einzelnen Disziplinen verwendet werden (z.B. Psychoanalyse aus der Psychologie oder Systemtheorie aus der Soziologie) sollte dieser Bezug klargemacht werden und begründet werden. Sobald einzelne Konzepte oder Variablen verwendet werden, ist auch hier der Theoriebezug notwendig (z.B. Stresstheorie bei Belastungserhebung, Zufriedenheitstheorien). Im Rahmen qualitativer Forschung spielen Interpretationen von Textmaterial (z.B. aus Interviews) eine besondere Rolle. Hier ist die Formulierung des Vorverständnisses der Interpreten im Sinne eines hermeneutischen Zirkels wichtig (vgl. Mayring 2016), und auch dies gehört zur Theorieeinbettung der Forschungsfrage.

Der nächste Schritt im Forschungsprozess ist die Festlegung und Begründung des Forschungsdesigns. Prinzipiell sind hier explorative, deskriptive, zusammenhangsanalytische oder kausalanalytische Designs möglich (vgl. Mayring 2010). Bei Evaluationsprojekten handelt es sich jedoch in der Regel um kausalanalytische Untersuchungsanlagen: Hat die Intervention einen kausalen Einfluss auf die Zielvariablen? Dafür sind aber verschiedene speziellere Evaluationsdesigns entwickelt worden. Zumindest formative Evaluation, summative Evaluation, Selbstevaluation, konstruktivistische Evaluation und experimentelle Evaluation lassen sich unterscheiden. Der klare und begründete Bezug zu einem solchen Evaluationsdesign und die konkrete Ausarbeitung stellen einen weiteren Gütestandard dar.

Die Begründung der Stichprobe stellt in qualitativ orientierten Projekten oft ein Problem dar, da oft mit kleinen, unsystematisch ausgewählten Stichproben gearbeitet wird. Hier muss man sich klarmachen, dass auch quantitative Studien selten mit voll repräsentativen Zufallsstichproben arbeiten, da diese sehr aufwändig sind und oft gar nicht zu bewerkstelligen sind. Es kommt also darauf an, die Stichprobe oder Materialauswahl (beispielsweise bei Dokumentenanalysen) argumentativ zu begründen. Für qualitative Forschung sind hier Samplingstrategien vorgeschlagen worden, wie theoretische Stichprobenziehung, qualitative Stichprobenpläne (vgl. Hussey/Schreier/Echterhoff 2013), die die Verallgemeinerbarkeit entscheidend verbessern können.

Die Methodenbegründung spielt in qualitativ orientierten Projekten eine besondere Rolle, da hier in der Regel nicht auf standardisierte, lang erprobte Verfahren zurückgegriffen wird, sondern das Erhebungsinstrumentarium (z.B. Interviewleitfäden) und Auswertungsinstrumentarium (z.B. Kodierpläne) im Projekt selbst entwickelt werden. Entscheidend für die Qualität der Studie ist dann aber, dass die neu entwickelten Verfahren wenigstens in kleinem Rahmen (z.B. Pilotinterviews) getestet werden.

Der nächste Ablaufschritt ist die Durchführung der Studie. Von Vertretern qualitativer Forschung ist oft kritisiert worden, dass quantitative Forschung linear vorgehe, also nur Material und Arbeitsschritte zulasse, die direkt auf die vorab formulierte Fragestellung inkl. Hypothesen bezogen sind (vgl. Flick 2005). Qualitative Forschung dagegen würde zirkulär vorgehen, aufgrund erster Erfahrungen im Feld die Fragestellung modifizieren, bei neu auftauchenden Aspekten die Fragestellung erweitern und wieder ins Untersuchungsfeld gehen. Dies scheint mir eine zu apodiktische Gegenüberstellung zu sein, denn auch qualitative Forschung will schlussendlich (linear) die Forschungsfrage beantworten. Wichtig wäre aber, auf neu auftretende Fragen im Sinne der Hauptfragestellung einzugehen, was dann aber unter Umständen erneute Recherchen zu Stand der Forschung, Theoriehintergrund und Methodenkonzept erfordert.

Auch von qualitativen Evaluationsprojekten erwarten wir uns schließlich klare Ergebnisse zu ausgewählten Qualitätsindikatoren des Evaluationsgegenstandes. Sich hier nur auf offene Prozessbeschreibungen zurückzuziehen erscheint defizitär.

Schließlich sollten auch qualitative Evaluationsprojekte am Ende sorgfältig Gütekriterien auf das eigene Projekt anwenden, Stärken und Schwächen des methodischen Vorgehens offen diskutieren. Da wir heute davon ausgehen müssen, dass mit

einem Projekt die Forschungsfrage niemals vollständig und abschließend bearbeitet werden kann, stellt das Aufzeigen der Desiderata der eigenen Arbeit, das Aufweisen weiterer Untersuchungsmöglichkeiten der Evaluationsfragestellung einen besonderen wissenschaftlichen Wert dar, auch wenn das Auftraggebern von Evaluation nicht immer vermittelbar ist.

Gerade qualitative Evaluationsprojekte haben besonderen Bedarf, die Verallgemeinerbarkeit ihrer Ergebnisse gesondert zu begründen, da sie weniger mit voll repräsentativen Stichproben arbeiten. Hier kann es auch darum gehen, die Grenzen der Verallgemeinerbarkeit (Limitationen, vgl. Steinke 1999) aufzuzeigen. Die Generalisierung der Forschungsergebnisse bedeutet nun auch eine erneute Theorieeinbettung und Weiterentwicklung des Theoriestands.

Sind konzise Ergebnisse durch den Forschungsprozess erzielt worden, so ist gerade für Evaluationsforschung die Ableitung von Praxiskonsequenzen zentral. In aller Regel sind nun Verbesserungsvorschläge und Modifikationen der untersuchten Maßnahme möglich.

Auf die zehn Schritte des Forschungsprozesses lassen sich also Gütekriterien beziehen, die für qualitative und quantitative Evaluationsforschung gleichermaßen Relevanz besitzen und damit zu einer Überwindung des Gegensatzes, besonders aber auch zur Schärfung qualitativer Evaluationsansätze beitragen können. Daneben werden in der Methodendiskussion aber auch Kriterien diskutiert, die sich nicht auf einzelne Forschungsschritte beziehen, sondern als prozessübergreifend angesehen werden können.

3.2 Prozessübergreifende Gütekriterien für qualitative Evaluationsforschung

Auf der Suche nach allgemeinen Gütekriterien für empirische Forschung, die nicht an ein Ablaufmodell gebunden sind, stößt man unweigerlich auf die Konzepte von Objektivität, Reliabilität und Validität (vgl. z.B. Döring/Bortz 2015). Die Konzepte stammen aus der quantitativen Forschung und wurden zunächst für die Bewertung standardisierter Testverfahren verwendet: Sie sollen zu gleichen Messwerten führen unabhängig von der testdurchführenden Person (Objektivität), sie sollen zu exakten Messungen führen (Reliabilität) und auch das erfassen, was sie vorgeben (Validität). Diese Kriterien sind jedoch nicht einfach auf qualitative Forschung übertragbar. Validität wird traditionell dadurch überprüft, dass ein Test, von dem man bereits weiß, dass er die Zielvariable erfasst, zusammen mit dem zu überprüfenden neuen Test auf eine Stichprobe anwendet und die Übereinstimmung (korrelativ) bestimmt (kriteriumsbezogene Validität). Da in qualitativen Projekten in der Regel die Instrumente gegenstandsbezogen entwickelt werden und keine Vergleichsinstrumente vorliegen, muss dies hier scheitern. Zur Reliabilitätsüberprüfung sind im testtheoretischen Kontext zwei Verfahren entwickelt worden, die die Genauigkeit über die Stabilität des Instruments (zwei Testungen derselben Personen in kurzem Abstand werden korreliert; Retest-Reliabilität) und der internen Konsistenz des Instruments (es wird untersucht, ob die einzelnen Testitems alle in dieselbe Richtung weisen) überprüfen.

Auch diese Verfahren kommen für qualitativ orientierte Projekte eher nicht in Frage. Ein Interview lässt sich nicht ein zweites Mal in der exakt gleichen Weise durchführen (die Bereitschaft der Befragten würde auch ein Problem sein). Auch sind die Interviewfragen nicht so angelegt, identische Aspekte zu erfassen, sondern betreffen gerade unterschiedliche Bereiche. Mit dem Gütekriterium der Objektivität haben Vertreter(innen) qualitativer Forschung die größten Probleme (vgl. Flick 2006). Eine Erfassung der sozialen Realität völlig unabhängig von der Person des Untersuchenden wird nicht als möglich erachtet, Vorannahmen, Referenzrahmen der Forscherinnen und Forscher beeinflussen die Ergebnisse. Solche Argumente haben dazu geführt, dass oft die klassischen Gütekriterien für qualitative Forschung abgelehnt werden und eigene Kriterien formuliert werden. Es gibt allerdings auch Versuche, sich an den drei Gütekriterien anzulehnen und Standards zu formulieren, die in die gleiche Richtung weisen, aber die Besonderheiten qualitativer Forschung berücksichtigen (vgl. Steinke 1999). Statt exakter Replikation wird die Genauigkeit der Analyse über die intersubjektive Nachvollziehbarkeit, die Dokumentation aller Verfahren ausgewiesen. Zur Validierung der Analyse sind eigene Verfahren entwickelt worden wie die Diskussion der Interpretationsergebnisse mit den untersuchten Personen (kommunikative Validierung). Auch liegen hier eher die Stärken qualitativer Forschung, wenn über die Nähe zum Gegenstand und die Alltagsorientierung (statt Laborforschung) eine gültige Erfassung des Gegenstands befördert wird. Aber auch am Objektivitätskriterium lässt sich in qualitativen Projekten anknüpfen. Eine völlige Ablehnung würde ja bedeuten, dass das Evaluationsergebnis relativiert werden müsste. Eine Lösung besteht darin, dass die Interaktion zwischen Forschenden und ihrem Gegenstand offengelegt, analysiert, reflektiert wird und so mögliche Verzerrungen greifbar werden. Die klassischen Gütekriterien der quantitativen Forschung lassen sich also im weitesten Sinne auch auf qualitative Evaluationsprojekte beziehen, wenn auch in eher unspezifischer, sehr offener Form. Allein sind sie sicherlich nicht ausreichend zur Qualitätseinschätzung des Forschungsprozesses. Sie können Orientierungspunkt sein, müssen aber ergänzt werden.

Einen spezifischeren Weg, aber wieder übergreifend und eher forschungsprozessunabhängig, hat die Gesellschaft für Evaluation DeGEval eingeschlagen, indem sie Standards für Evaluation formuliert hat, ausgehend von den in den USA in den 1990er Jahren entwickelten Standards for Educational Evaluation, und mehrfach überarbeitet sowie von den Mitgliedern der DeGEval beschlossen hat (vgl. DeGEval 2017). Diese Standards mit den vier Hauptkriterien der Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit, operationalisieren über 25 spezielle Kriterien mit ausführlicher Erläuterung, sind von qualitativer Seite zwar als immer noch vage und allgemein kritisiert worden (vgl. Flick 2006), gehen aber explizit auch auf qualitative Evaluationsforschung ein. Hervorzuheben wären hier unter den Nützlichkeitsstandards der Bezug zu den Beteiligten und Betroffenen (N1) sowie die Klärung der Evaluationszwecke (N2), was auf partizipative Ansatzpunkte, in qualitativer Evaluation besonders wichtig, hinweist. Unter den Fairnesskriterien wird die unparteiische Durchführung und Berichterstattung (F4) betont, die vor allem bei interner Evaluation und Selbstevaluation bedeutsam ist, wenngleich sie nie völlig erreichbar sein wird. Unter den Genauigkeitskriterien wird explizit auf Validität und Reliabi-

lität (G3) hingewiesen, hier aber in einer offeneren Form: „Die eingesetzten, qualitativen wie quantitativen Instrumente ... sollen konsistente, reproduzierbare sowie intersubjektiv nachvollziehbare ... Informationen beisteuern“ (DeGEval 2017: 24). Auch im Weiteren wird auf eine angemessene Analyse qualitativer und quantitativer Informationen (G7) abgehoben. „Sowohl qualitative als auch quantitative Verfahren können fruchtbare Daten und Informationen erbringen“ (DeGEval 2017: 25). Die Standards für Evaluation der DeGEval gehen also explizit auf qualitative Evaluationsforschung ein, versuchen dabei aber, wie auch hier vorgeschlagen, nicht völlig eigenständige Qualitätsmaßstäbe anzulegen.

3.3 Methodenspezifische Gütekriterien

Ein letzter Ansatz in der Diskussion über Gütekriterien für qualitative Evaluationsforschung besteht darin, methodenspezifische Maßstäbe zu formulieren. In quantitativer Forschung existieren solche Kriterien weithin, beispielsweise für experimentelle Designs die interne und externe experimentelle Validität (vgl. Döring/Bortz 2015). Wie könnten sie für qualitative Evaluation aussehen?

Hier wäre zu unterscheiden zwischen Evaluationsdesigns und Evaluationsmethoden im engeren Sinne. Typische qualitative Evaluationsdesigns sind formative Evaluation, Selbstevaluation und konstruktivistische Evaluation. Es geht nun darum, zu definieren, was zentrale unverzichtbare methodische Prinzipien dieser unterschiedlichen Designs sind. So wäre ein Gütekriterium für formative Evaluation, dass auch wirklich aufgrund erster Evaluationsergebnisse verändernd eingegriffen wird und wiederum die Wirkung überprüft wird. Es muss also mehrere empirisch überprüfte Programmversionen geben. Für Selbstevaluation wäre entscheidend, dass auch die Instrumente der Evaluation von den Beteiligten, zwar ggf. unter externer Beratung, aber doch selbst bestimmt werden und über die Art der Ergebnisberichterstattung selbst entschieden wird. Bei konstruktivistischer Evaluation ist die Identifizierung und Beteiligung aller relevanten Stakeholder sowie die professionelle Moderation der Qualitätsdiskurse unter diesen Gruppen entscheidend.

Dies lässt sich fortsetzen für die konkreten Erhebungs- und Auswertungsmethoden, die in der qualitativen Evaluation eingesetzt werden. Werden offene Interviews eingesetzt, so wäre entscheidend, ob der Interviewleitfaden theoriegeleitet entwickelt wurde und die Interviews pilotgetestet wurden. Werden Feldbeobachtungen vorgenommen, so ist auch hier eine Beobachtungsanleitung sowie eine Beobachterschulung notwendig. Für Qualitative Inhaltsanalyse sind die Festlegung inhaltsanalytischer Regeln (Kategoriendefinition bzw. Kodierleitfaden) sowie die Überprüfung von Intra- und/oder Interkoder-Übereinstimmung Gütekriterien.

4. Schlussbemerkung

Wir haben also eine Reihe von Gütekriterien auf den unterschiedlichsten Ebenen des Forschungsprozesses kennengelernt, die für qualitative Evaluation relevant sein können. Für manche Vertreter(innen) qualitativer Forschung ist die letztgenannte Kategorie der methodenspezifischen Kriterien am wichtigsten. So betont Flick (2006) die Methodenangemessenheit von Gütekriterien. Besser als allgemeine Standards oder Checklisten seien spezifische Maßstäbe für die jeweiligen Forschungsschritte. Das würde aber bedeuten, dass zwischen unterschiedlichen Projekten oder in Mixed-Methods-Ansätzen unterschiedliche Maßstäbe angewandt würden. Hierin sehe ich eine Gefahr. Die Diskussion von Gütekriterien sollte, wie oben dargestellt, ein fester Bestandteil im Forschungsprozess sein und am Ende des Projekts zu einer einheitlichen Einschätzung des Werts der Ergebnisse führen, die dann auch zwischen verschiedenen Evaluationsprojekten Vergleiche zulässt. So sollte hier auf möglichst vielen Ebenen argumentiert werden, auf der Ebene der einzelnen Schritte im Forschungsprozess, auf übergeordneter Ebene orientiert an Objektivität, Reliabilität und Validität, an Evaluationsstandards orientiert sowie auf die einzelnen Evaluationsdesigns, Erhebungs- und Auswertungsmethoden bezogen. Hier ist argumentiert worden, dass solche Kriterien für quantitative und qualitative Forschung vergleichbar sein müssen. Nur dann lassen sich Ergebnisse aus unterschiedlichen Projekten zusammenführen, lässt sich der Stand der Forschung in spezifischen Feldern zusammenfassen.

Wir gelangen so zu einem Qualitätsprofil, das Stärken, aber auch Schwächen des jeweiligen Evaluationsforschungsprojektes aufzeigt und so die Möglichkeit bietet, die Verallgemeinerbarkeit der Ergebnisse auf andere Praxiskontexte spezifisch abzuwägen, aber auch weitere Forschungen zur Klärung von Schwachpunkten anzuregen.

5. Literatur

- American Educational Research Association AERA (Hg.) (2006): Standards for Reporting on Empirical Social Science Research in AERA Publications. In: Educational Researcher, 35 (6), S. 33-40.
- Bammé, Arno (2004): Science Wars: Von der akademischen zur postakademischen Wissenschaft. Frankfurt: Campus.
- Baur, Nina/Blasius, Jörg (Hg.) (2014): Handbuch der Methoden empirischer Sozialforschung. Wiesbaden: Springer VS.
- Bierhoff, Hans-Werner/Petermann, Franz (2015): Forschungsmethoden der Psychologie. Göttingen: Hogrefe.
- DeGEval – Gesellschaft für Evaluation (2017): Standards für Evaluation (DeGEval-Standards). Verfügbar unter: www.degeval.de/publikationen/standards-fuer-evaluation [15.08.2017].
- Dellinger, Amy B./Leech, Nancy L. (2007): Toward a Unified Validation Framework in Mixed Methods Research. In: Journal of Mixed Methods Research, 1 (4), S. 309-332.
- Denzin, Norman K. (2010): The Qualitative Manifesto. A Call to Arms. Walnut Creek, CA: Left Coast Press.
- Denzin, Norman K./Lincoln, Yvonne (2011): The Sage Handbook of Qualitative Research. London: Sage (5. Aufl.).

- Döring, Nicola/Bortz, Jürgen (2015): Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. Berlin: Springer (5. Aufl.).
- Eid, Michael/Gollwitzer, Mario/Schmitt, Manfred (2015): Statistik und Forschungsmethoden. Weinheim: Beltz (4. Aufl.).
- Flick, Uwe (2005): Qualitative Sozialforschung: Ein Handbuch. Reinbek: Rowohlt.
- Flick, Uwe (2006): Qualität in der qualitativen Evaluationsforschung. In: Flick, Uwe (Hg.): Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Reinbek: Rowohlt. S. 424-443.
- Flick, Uwe (2011): Triangulation. Eine Einführung. Wiesbaden: Springer VS (3. Aufl.).
- Girtler, Roland (1980): Vagabunden in der Großstadt. Teilnehmende Beobachtung in der Lebenswelt der „Sandler“ Wiens. Stuttgart: Enke.
- Hussy, Walter/Schreier, Margrit/Echterhoff, Gerald (2013): Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor. Berlin: Springer (2. Aufl.).
- King, Gary/Keohane, Robert O./Verba, Sidney (1994): Designing Social Inquiry. Scientific Inference in Qualitative Research. Princeton: University Press.
- Mayring, Philipp (2010). Design. In: Mey, Günter/Mruck, Katja (Hg.): Handbuch qualitative Forschung in der Psychologie. Wiesbaden: Springer VS, S. 225-237.
- Mayring, Philipp (2012): Mixed Methods – ein Plädoyer für gemeinsame Forschungsstandards qualitativer und quantitativer Methoden. In: Gläser-Zikuda, Michaela/Seidl, Tina/Rohlf, Carsten/Gröschner, Alexander/ Ziegelbauer, Sascha (Hg.): Mixed Methods in der empirischen Bildungsforschung. Münster: Waxmann, S. 287-300.
- Mayring, Philipp (2016): Einführung in die Qualitative Sozialforschung. Weinheim: Beltz (6. Aufl.).
- Mayring, Philipp/Huber, Günter L./Gürtler, Leo/Kiegelmann, Mechthild (Hg.) (2007): Mixed Methodology in Psychological Research. Rotterdam: Sense Publisher.
- National Research Council NRC (2002): Scientific Research in Education. Committee on Scientific Principles for Educational Research. Washington, DC: National Academy Press.
- Onwuegbuzie, Anthony/Johnson, R. Burke (2006): The Validity Issue in Mixed Research. In: Research in the Schools, 13 (1), S. 48-63.
- Ross, Andrews (Hg.) (1996): Science Wars. Durham: Duke University Press.
- Snow, Charles P. (1998): The two Cultures. Cambridge: University Press.
- Spiel, Christiane/Grading, Petra/Lüftenegger, Marko (2010): Grundlagen der Evaluationsforschung. In: Holling, Heinz/Schmitz, Bernhard (Hg.): Handbuch Statistik, Methoden und Evaluation. Göttingen: Hogrefe, S. 223-232.
- Steinke, Ines (1999): Kriterien qualitativer Forschung. Ansätze zur Bewertung qualitativ-empirischer Sozialforschung. Weinheim: Juventa.
- Wottawa, Heinrich/Thierau, Heike (2003): Lehrbuch Evaluation. Bern: Huber (3. Aufl.).