

Uwe Flick

Standards, Kriterien, Strategien –

Zur Diskussion über Qualität qualitativer Sozialforschung

Standards, Criteria, Strategies –

Discussing the Quality of Qualitative Research

Zusammenfassung:

Der Kontext der Frage nach der Qualität qualitativer Forschung wird kurz umrissen. Hier werden insbesondere Diversifizierungen qualitativer Forschung (sprachraumbezogen, schulenspezifisch, unterschiedliche Anwendungsfelder und Disziplinen) behandelt, bevor drei Ansätze der Auseinandersetzung mit der Frage der Qualität qualitativer Forschung an Beispielen aus der Diskussion behandelt werden: Die Formulierung von Standards, Kriterien und Strategien. Für diese Ansätze werden als Probleme die Anwendbarkeit auf qualitative Forschung insgesamt oder auf bestimmte Bereiche diskutiert. Die Frage der Festlegung von Grenzwerten als ein Teil der Formulierung von Kriterien und Standards wird ebenso behandelt wie die Akzeptanz von Guidelines und Checklists. Als Alternative wird die Verfolgung von Strategien der Geltungsbe gründung (analytische Induktion, Triangulation, Qualitätsmanagement) diskutiert und vorgeschlagen, empirische Methodenfor schung zur Entwicklung von Antworten auf die Qualitätsfrage voranzutreiben, die sowohl die Forschungspraxis als auch die Perspektive der Abnehmer von Forschungsergebnissen berücksichtigt.

Schlagworte: Qualität qualitativer Forschung, Triangulation, Standards qualitativer Forschung, Gütekriterien

Abstract:

The context of the question of quality in qualitative research is briefly outlined. The main emphasis here is on the diversification of qualitative research (in different language contexts, in schools of research, in different fields of application and different disciplines). Using examples from ongoing debates, the discussion then focuses on three approaches to the question of quality in qualitative research: standards, criteria and strategies. The problems addressed concerning the question of whether or not these approaches can be used in qualitative research in general, and in specific fields in particular. The question of establishing benchmarks as part of the process of defining criteria and standards is discussed, as well as the acceptance of guidelines and checklists. As an alternative, the use of strategies for enhancing the quality of qualitative research (analytic induction, triangulation, quality management) is proposed, and it is suggested that researchers should be encouraged to engage in empirical methods research in order to provide answers to the question of quality while taking into account, both, research practices and the perspectives of the addressees of research findings.

Keywords: quality of qualitative research, triangulation, standards in qualitative research, quality criteria

1. Einleitung

Die Frage, wie die Qualität qualitativer Forschung bestimmt werden soll bzw. kann, wird gestellt, seitdem es qualitative Forschung gibt. Sie erfreut sich aktuell, immer noch und wieder, einer besonderen Aufmerksamkeit, wie etwa das vorliegende Schwerpunktheft unterstreicht. Die Antworten auf diese Frage sind bislang allerdings noch nicht gefunden – bzw. nicht in einer allgemein akzeptierten Form. Beiträge zu der entsprechenden Diskussion liegen etwa in Versuchen, Güte-Kriterien zu formulieren (vgl. Seale 1999 oder Steinke 1999), diese einzufordern (etwa die Initiative für ein Rundgespräch der DFG zu diesem Thema – vgl. Helsper et al. 2001 für die Resultate) oder in der mehr oder minder lakonischen Feststellung, die Antworten auf entsprechende Fragen seien noch nicht gefunden (vgl. Lüders 2000 oder 2006a). Dass solche Einschätzungen trotz der vielfältigen Versuche, das Problem anzugehen, noch eine gewisse Berechtigung haben, liegt auch in der Natur der Sache – der besonderen Situation, in der sich qualitative Forschung befindet – begründet. Im folgenden Beitrag soll ein Überblick über die entsprechende Diskussion gegeben werden, der die unterschiedlichen Wege der Beantwortung unserer Ausgangsfrage vergleichend gegenüberstellt.

2. Was ist qualitative Forschung und wovon reden wir?

Bevor wir in die eigentliche Diskussion einsteigen, sind noch ein paar Vorbemerkungen dazu notwendig, was hier der Bezugspunkt sein soll. Qualitative Forschung hat sich in verschiedenen Kontexten entwickelt. Hier sind einerseits theoretische und methodologische Schulen zu unterscheiden, die jeweils bestimmte Grundannahmen, Forschungsinteressen und – in der Regel, aber nicht immer daraus resultierend – Methoden(-präferenzen) kennzeichnen bzw. unterscheiden. So ist der ursprünglich in den USA entstandene Ansatz der gegenstandsbegründeten Theoriebildung (grounded theory) im englischen aber auch im deutschen Sprachraum als eigener Ansatz zu verzeichnen, dessen Interesse sich in der Regel auf die Entwicklung von Theorien über einen bestimmten Gegenstand aus empirischem Material bzw. aus dessen Analyse konzentriert. Gleiches lässt sich über die Biographieforschung sagen, die diesseits und jenseits der Sprachgrenze(n) die Analyse von Lebensgeschichten auf theoretisch relevante Verdichtungen orientiert ist. Daneben gibt es Ansätze, oder Schulen, die spezifisch für bestimmte Kontexte sind, dort jedoch eine zentrale Rolle spielen, während sie in anderen Kontexten kaum rezipiert werden bzw. diese Rezeption dort auch nicht suchen. Beispiele sind hier etwa die objektive Hermeneutik oder die hermeneutische Wissenssoziologie, die ihre Wirkung (und Publikationsaktivitäten) fast ausschließlich im deutschen Sprachraum entfalten (vgl. Reichertz 2000, vgl. auch ZBBS-Heft 2/2004). Ähnliches gilt für die im englischen Sprachraum sich differenzierenden Formen der Diskursanalyse, die etwa in England eine starke Dominanz in der Diskussion entwickelt haben, hierzulande jedoch kaum rezipiert werden (auch da der Begriff der Diskursanalyse hier mit anderen Wurzeln assoziiert ist). Das

heißt, die Diskussion über qualitative Forschung ist durch unterschiedliche Differenzierungen gekennzeichnet – Schulen auf der einen Seite, sprachraumbezogene Schwerpunkte und Unterschiede auf der anderen Seite (vgl. hierzu auch Flick 2005 und Knoblauch/Flick/Maeder 2005 für Überblicke). Dazu kommen noch (mindestens) zwei weitere Differenzierungen. Zunächst einmal sind disziplinspezifische Entwicklungen zu verzeichnen. Der Diskurs in der Erziehungswissenschaft (über qualitative Forschung) entwickelt sich in mehr oder minder enger Verzahnung (oder mehr oder minder unabhängig) von dem in der Soziologie oder in der Psychologie. Beispiel hierfür ist das erwähnte DFG-Rundgespräch, das sich auf die Erziehungswissenschaft beschränkt hat oder das von Friebertshäuser und Prengel (1997) herausgegebene Handbuch mit der gleichen Fokussierung.

Ebenso relevant (auch für unsere Fragestellung im vorliegenden Beitrag) wird aber in den letzten Jahren die Differenzierung der unterschiedlichen Anwendungsfelder qualitativer Forschung. Zu nennen sind hier Bereiche wie die Gesundheitsforschung (vgl. NIH 2001 bzw. Schaeffer/Müller-Mundt 2002), qualitative Management- und Organisationsforschung (Cassel/Symon 2004) oder die qualitative Evaluationsforschung (Flick 2006a). In diesen Feldern beginnt sich die methodische Diskussion über qualitative Forschung und mehr noch über ‚gute‘ qualitative Forschung langsam zu verselbständigen. Dies hat auch mit den Rahmenbedingungen zu tun, unter denen qualitative Forschung hier durchgeführt wird. In der Regel handelt es sich hier um Auftragsforschung, die mit spezifischen Erwartungen hinsichtlich der Ergebnisse und vor allem ihrer praktischen Relevanz verknüpft ist und häufig unter anderen Rahmenbedingungen realisiert werden muss als qualitative Grundlagen- bzw. Qualifikationsforschung. Zu nennen ist hier etwa der zeitliche Rahmen, dessen Folgen sich etwa an der Diskussion über die Legitimität von „Abkürzungsstrategien“ (vgl. Lüders 2000, Flick 2000) bei der Verwendung qualitativer Methoden in solchen Kontexten festmachen lässt oder auch an der Frage der Überzeugung von – außerwissenschaftlichen – Zielgruppen mit den gefundenen Ergebnissen (vgl. hierzu Lüders 2006b; Mensching 2006).

Aus dieser knappen Skizzierung der Diversifizierung qualitativer Forschung – die keinerlei Anspruch auf Vollständigkeit erhebt, sondern exemplarisch bzw. selektiv angelegt ist – lässt sich ein Dilemma ableiten, in dem die hier interessierende Diskussion über die Qualität qualitativer Forschung steckt. Die Frage nach der angemessenen Bestimmung, Sicherung oder Verbesserung dieser Qualität stellt sich über alle der genannten Bereiche hinweg. Die Lösungswege, die dabei beschritten werden, unterscheiden sich ebenso wie die Klärungsnotwendigkeiten und die gefundenen bzw. vorgeschlagenen Lösungen. Daraus lässt sich die Frage ableiten, ob es erwartbar (und sinnvoll) ist, über die verschiedenen Bereiche und Kontexte hinweg eine gültige Antwort auf die Qualitätsfrage zu finden oder nicht. Kaum umstritten ist, dass qualitative Forschung eine Antwort auf diese Frage finden muss. Jedoch herrscht wenig Einigkeit darüber, wie diese Antwort aussehen soll: Liegt sie darin, Gütekriterien zu formulieren, die idealerweise Grenzwerte oder Benchmarks zur Unterscheidung von guter und weniger guter Forschung ‚mitliefern‘? Dann lautet die erste Frage, welche Kriterien hierfür geeignet sind, und die zweite, ob sie für ‚die‘ qualitative Forschung gültig sein sollen oder für bestimmte Richtungen in der qualitativen Forschung. Wenn Kriterien, sollen sie dann auf eine grounded-theory-Studie gleichermaßen anwendbar sein wie auf eine Untersuchung, die auf Fallrekonstruktionen im Sinne der Objektiven Hermeneutik basiert – oder auch auf eine Fallstudie zur Evaluation

einer Institution? Oder stellt sich die Frage der Qualität bei qualitativer Forschung grundsätzlich anders – jenseits von Kriterien? Dann stellt sich die Frage, was an die Stelle von Kriterien treten soll und kann. Im Folgenden sollen verschiedene Auseinandersetzungen mit der Qualitätsfrage in der qualitativen Forschung diskutiert werden.

3. Standards nicht-standardisierter Forschung

Einen interessanten Vorschlag hat kürzlich Ralf Bohnsack (2005) vorgelegt. Hier wird die Beantwortung unserer Leitfrage daran festgemacht, in wie weit sich Standards nicht-standardisierter Forschung identifizieren lassen bzw. herausgebildet haben. Bohnsack verdeutlicht einerseits, dass sich die Standards bei nicht-standardisierter Forschung nicht ‚am grünen Tisch‘ (des Methodikers) entwickeln lassen, sondern quasi im Nachgang aus der Rekonstruktion nicht-standardisierter bzw. qualitativer Forschungspraxis ableiten und explizieren lassen. Gleiches wird auch für die Methoden qualitativer Forschung selbst festgestellt („These 1: Die Methoden und Standards qualitativer Forschung werden auf der Grundlage einer empirischen Rekonstruktion der Forschungspraxis entwickelt“ – Bohnsack 2005, S. 65). Dass die vorliegenden Methoden der qualitativen Forschung sich aus konkreten Forschungsfragen und -projekten entwickelt haben, lässt sich gut nachzeichnen. Mittlerweile haben sich allerdings eine ganze Reihe von mehr oder minder kanonisierten Methoden in der qualitativen (bzw. nicht-standardisierten oder rekonstruktiven) Forschung entwickelt und etabliert, so dass Forscherinnen und Forscher heute häufig eher vor der Frage stehen, welche davon sie anwenden wollen für die Beantwortung ihrer Forschungsfragen und welche methodischen Neuentwicklungen aus der Praxis eher die Ausnahme sind. Hier stellt sich dann die Frage, worin sich gute von weniger guter Anwendung bestimmter Methoden unterscheiden lässt. Standards in der nicht-standardisierten Forschung stellen nach Bohnsack Standards zweiten Grades dar, die aus der Auseinandersetzung mit den natürlichen Standards (ersten Grades) entwickelt werden sollen. Folgt man diesem Ansatz und der darauf bezogenen Argumentation bei Bohnsack, so lassen sich Standards qualitativer Forschung aus der Analyse alltäglicher Standards der Kommunikation entwickeln und darüber die Kriterien Gültigkeit und Zuverlässigkeit in der qualitativen Forschung rekonstruieren (S. 76). Dabei wird argumentiert, dass die wesentliche Bezugsebene für die Formulierung von Standards die methodologische und theoretische Begründung des jeweiligen Vorgehens ist. Weiterhin wird ausgeführt, dass bei qualitativen Methoden zwischen offenen und rekonstruktiven Verfahren unterschieden werden sollte, wobei nur letztere den von Bohnsack entwickelten Qualitätsstandards entsprechen (vgl. These 7, S. 74).

Der Ansatz von Bohnsack liefert eine ganze Reihe von theoretisch und methodologisch aufschlussreichen Vorschlägen für eine metatheoretische Fundierung der Diskussion über die Qualität qualitativer Forschung. Allerdings bleiben verschiedene Fragen offen. Zunächst einmal die Frage, ob die Formulierung von Standards in einem derart heterogenen Feld wie der qualitativen Forschung (überhaupt bzw. schon zum gegenwärtigen Zeitpunkt) realisiert werden kann – wenn noch nicht einmal Einigkeit über die Bezeichnungen (qualitativ, interpreta-

tiv, rekonstruktiv) und Zugehörigkeiten zum Feld besteht. Zweitens laufen Formulierungen von Standards in der Regel Gefahr, Standardisierung (von Vorgehensweisen und Prozeduren) mit sich zu bringen – was den Ansatz nicht-standardisierter Forschung zumindest in einen Widerspruch verwickelt. Drittens, und das ist in unserem Kontext das entscheidende Argument, wird über den Weg der Formulierung von Standards, den Bohnsack einschlägt, die Frage der Geltungsbegründung von der Ebene der Qualitätsbestimmung praktischer Vorgehensweisen im Feld auf die Ebene der Angemessenheit ganzer Forschungsprogramme verlagert. Wendet man Bohnsacks Vorschlag an, weiß man zwar, dass bestimmte Ansätze – rekonstruktive Verfahren – den (?) Standards qualitativer Forschung entsprechen, andere – offene Verfahren – dagegen nicht. Weniger hilfreich sind diese Vorschläge dagegen bei der Suche nach Antworten auf die Frage, wonach die konkreten Anwendungen und Verfahrensweisen in einem Forschungsprojekt oder Artikel zu bewerten sind. In diese Richtung zielen die im folgenden Abschnitt behandelten Ansätze.

4. Kriterien qualitativer Forschung

4.1 Verwendung der klassischen Kriterien

Lange Zeit wurde vorgeschlagen, die klassischen Kriterien der empirischen Sozialforschung – Reliabilität, Validität und Objektivität auch auf qualitative Forschung anzuwenden oder sie für diesen Gegenstand zu modifizieren. Kirk und Miller (1986) widmen sich Reliabilität und Validität in dieser Hinsicht. Zum einen wird dabei deutlich, dass die Reliabilität von Daten und Verfahren im traditionellen Sinne – als die Stabilität von Daten und Ergebnissen bei mehreren Erhebungen – für die Bewertung qualitativer Daten eher ungeeignet ist: Die identische Wiederholung einer Erzählung bei wiederholten narrativen Interviews ist eher ein Hinweis auf eine ‚zurechtgelegte‘ Version als auf die Verlässlichkeit des Erzählten. Validität (vgl. Kvale 1995) wird ebenfalls häufiger für die qualitative Forschung diskutiert. Kirk und Miller (1986, S. 21) fassen die Frage der Validität darin zusammen, ob „der Forscher sieht, was er (...) zu sehen meint“. Hier ergeben sich ebenfalls Probleme bei der unmittelbaren Anwendung klassischer Validitätskonzeptionen. Interne Validität wird etwa erhöht bzw. sichergestellt, indem ausgeschlossen werden soll, dass andere als die in der Untersuchungshypothese enthaltenen Variablen den beobachteten Zusammenhang bestimmen (z. B. Bortz/Döring 2001, S. 53). In diesem Verständnis liegen bereits die Probleme bei der Übertragung auf qualitative Forschung begründet: Interne Validität soll durch eine möglichst umfassende Kontrolle der Kontextbedingungen in der Untersuchung erhöht werden. Zu diesem Zweck wird die weitgehende Standardisierung der Erhebungs- bzw. Auswertungssituation angestrebt. Der dafür notwendige Grad an Standardisierung ist jedoch mit dem größten Teil der gängigen qualitativen Methoden nicht kompatibel bzw. stellt ihre eigentlichen Stärken in Frage. Ähnlich lässt sich für die anderen Formen der Validität aufzeigen, warum sie

nicht direkt auf qualitative Forschung übertragen werden können (vgl. Steinke 1999).

Das dritte Kriterium aus dem Kanon der quantitativen Forschung ist die Objektivität. Hier finden sich kaum Versuche, dieses Kriterium auf qualitative Forschung anzuwenden. Eine Ausnahme ist die Arbeit von Madill et al. (2000). Darin wird jedoch Objektivität ausschließlich an der Analyse qualitativer Daten festgemacht und mit der Frage, ob zwei Forscher zu gleichen Ergebnissen bei der Analyse vorliegender qualitativer Daten kommen und damit mit der „Konsistenz der Bedeutung durch die Triangulation der Ergebnisse zweier unabhängiger Forscher“ (ebd., S. 17) gleichgesetzt. Insgesamt findet sich zwar gelegentlich der Anspruch, qualitative Forschung müsse sich zumindest den Fragen stellen, die mit Konzepten wie Reliabilität und Validität (z.B. bei Morse 1999, S. 717) oder Objektivität (Madill et al. 2000) verknüpft sind. In der Umsetzung überwiegt jedoch die Modifikation oder Reformulierung der Konzepte.

4.2 Reformulierung herkömmlicher Kriterien

Reliabilität

Vorschläge zur Reformulierung des Reliabilitätskonzeptes im Sinne einer stärker prozeduralen Konzeption zielen darauf ab, das Zustandekommen der Daten dahingehend zu explizieren, dass überprüfbar wird, was Aussage noch des jeweiligen Subjekts ist und wo die Interpretation des Forschers schon begonnen hat. Hierzu gehören etwa exakte und einheitliche Vorgaben, wie Interviews oder Gespräche transkribiert werden sollen (vgl. hierzu Kowal/O'Connell 2000) oder die Kennzeichnung von wörtlich wiedergegebenen Aussagen in Feldnotizen in Abhebung von Zusammenfassungen oder Paraphrasen durch den Forscher. Schließlich soll sich die Reliabilität im gesamten Prozess durch dessen reflexive Dokumentation erhöhen (vgl. hierzu auch Seale 1999).

Validität

Als eine Reformulierung der Validitätsbestimmung wird eine Analyse der Interviewsituation ausgehend von Habermas' Theorie des kommunikativen Handelns (1981) vorgeschlagen (Legewie 1987). Geltungsansprüche im Interview sind zu differenzieren in den Inhalt des Gesagten, die Angemessenheit der Beziehung und die aufrechte Selbstdarstellung des Interviewpartners. Validierung erfolgt über eine Analyse der Interviewsituation auf Auffälligkeiten und Verzerrungen und auf das Vorliegen eines Arbeitsbündnisses und einer nicht-strategischen Kommunikation. Ein Problem bei diesem Ansatz ist die (zumindest implizite) Annahme einer ‚richtigen‘ bzw. ‚gültigen‘ Version der Erzählung, wodurch sich die Validitätsfrage auf die Bestimmung von Abweichungen von dieser Version bzw. auf die Identifizierung von Hinweisen auf potentielle Abweichungen („Verzerrungen“) reduzieren lässt.

Kommunikative Validierung

Die Zustimmung der Untersuchungsteilnehmer – als kommunikative Validierung oder member checks bezeichnet – nach Abschluss des Interviews wird als eine weitere Form der Validierung diskutiert (für allgemeinere Diskussionen vgl. Terhart 1995, S. 388 ff.). Für eine allgemeinere Anwendung solcher Strategien sind zwei Fragen noch nicht befriedigend beantwortet: (1) Wie ist das methodische Vorgehen bei der kommunikativen Validierung zu gestalten, damit es den untersuchten Sachverhalten und der Sicht der Subjekte tatsächlich gerecht wird? (2) Wie lässt sich jenseits der Zustimmung der Subjekte die Frage der Geltungsbegründung weitergehend beantworten? Hierzu sind andere Qualitätsprüfungen notwendig, die kommunikative Validierungen ergänzen (vgl. als Überblick Flick 1987).

Prozedurale Validierung

Mishler (1990) setzt am Prozess der Validierung an (statt am Zustand der Validität) und definiert „Validierung als soziale Konstruktion von Wissen“ (ebd., S. 417), durch die wir „Behauptungen über die ‚Vertrauenswürdigkeit‘ berichteter Beobachtungen, Interpretationen und Verallgemeinerungen aufstellen und diese bewerten“ (ebd., S. 419). Schließlich umgeht „Validierung, verstanden als der soziale Diskurs, durch den Vertrauenswürdigkeit hergestellt wird, solche vertrauten Konventionen wie Reliabilität, Falsifikation und Objektivität“. Als empirische Basis für diesen Diskurs und die Konstruktion von Vertrauenswürdigkeit erörtert Mishler die Verwendung von Beispielen aus narrativen Studien.

Altheide und Johnson (1998, S. 291f.) formulieren schließlich das Konzept der „Validität-als-reflexive-Erklärung“, das Forscher, den Gegenstand und den Prozess der Sinnfindung in Beziehung setzt und Validität am Prozess der Forschung und den verschiedenen Beziehungen darin festmacht. Hier werden Beziehungen zwischen den Beobachtungen und weiteren Kontexten, zwischen dem Beobachter, Beobachteten und Setting, Perspektiven (des Mitglieds oder Forschers) und der Art der Darstellung zum Ansatzpunkt für Validierungen. Damit wird die Validierung unter der Perspektive des gesamten Forschungsprozesses und der beteiligten Faktoren behandelt. Die Vorschläge bleiben dabei jedoch eher auf der Ebene der Programmatik, als dass konkrete Kriterien oder Anhaltspunkte formuliert werden, anhand derer sich einzelne Studien oder Bestandteile davon beurteilen lassen. Die Versuche der Verwendung oder Reformulierung von Validität und Validierung haben insgesamt mit verschiedenen Problemen zu kämpfen: Formale Analysen des Zustandekommens von Daten in der Interviewsituation beispielsweise sagen noch nichts über Inhalte und ihre angemessene Behandlung im weiteren Verlauf der Forschung aus. Das Konzept der kommunikativen Validierung oder Member Checks ist mit dem Problem konfrontiert, dass Zustimmung dort als Kriterium schwierig ist, wo die Sicht des Subjekts systematisch überschritten wird – in Interpretationen, die ins soziale oder psychische Unbewusste vordringen wollen oder sich gerade aus der Unterschiedlichkeit verschiedener subjektiver Sichtweisen ableiten. Aus diesem Grund wird dieses Kriterium von verschiedenen Seiten immer wieder vehement in Frage gestellt. Die behandelten Reformulierungen des Validitätskonzepts zeichnen sich insgesamt durch eine gewisse Unschärfe aus, die der Forschungspraxis durch ihre generelle Problematisierung

und Programmatik nicht unbedingt eine Lösung für die Frage der Geltungsbe-gründung anbietet. Als gemeinsame Tendenz bleibt jedoch eine Verlagerung von Validität zur Validierung und von der Beurteilung des einzelnen Forschungsschritts oder -bestandteils zur Herstellung von Transparenz über den Forschungsprozess festzuhalten.

Die Anwendung klassischer Kriterien auf qualitative Forschung wird seit längerem in Frage gestellt, da „das ‚Wirklichkeitsverständnis‘“ beider Forschungsrichtungen dafür „zu unterschiedlich“ (Lüders/Reichert 1986, S. 97) sei. Ähnliche Vorbehalte finden sich schon bei Glaser und Strauss (1979, S. 92), die „bezweifeln, ob der Kanon quantitativer Sozialforschung als Kriterium (...) auf qualitative Forschung (...) anwendbar ist. Die Beurteilungskriterien sollten vielmehr auf einer Einschätzung der allgemeinen Merkmale qualitativer Sozialforschung beruhen – der Art der Datensammlung (...), der Analyse und Darstellung und der (...) Weise, in der qualitative Analysen gelesen werden.“ Aus dieser Skepsis resultieren im Lauf der Zeit eine Reihe von Versuchen, „methodenangemessene Kriterien“ (Flick 1987) zu entwickeln und diese an die Stelle von Kriterien wie Validität und Reliabilität zu setzen.

4.3 Formulierung alternativer, methodenangemessener Kriterien

Die dritte Variante der Beantwortung der Frage nach der Bewertung qualitativer Forschung ist entsprechend die Suche nach alternativen, methodenangemessenen Kriterien. Dabei ist der Gedanke leitend, dass die Frage nach der Qualität grundsätzlich durch die Formulierung und Anwendung von Kriterien beantwortet werden kann und sollte, dass jedoch die klassischen Kriterien an den Charakteristika qualitativer Forschung und Methoden vorbeizielten.

Lincoln und Guba (1985) propagieren Vertrauenswürdigkeit, Glaubwürdigkeit, Übertragbarkeit, Zuverlässigkeit und Bestätigbarkeit als Kriterien qualitativer Forschung, wobei das erstgenannte zum zentralen Kriterium wird. Um die Glaubwürdigkeit qualitativer Forschung, Daten und Ergebnisse zu erhöhen, skizzieren sie verschiedene Strategien. Dazu zählen neben einem „verlängerten Engagement“ im Feld, „ausdauernden Beobachtungen“ und der Triangulation (s. u.) verschiedener Methoden, Forscher und Datensorten auch „Peer debriefing“ (regelmäßige Besprechungen mit anderen Forschern zur Aufdeckung blinder Flecke sowie die Analyse abweichender Fälle (s. u.) und die Überprüfung der Angemessenheit von Interpretationen und Member checks im Sinne der kommunikativen Validierung von Daten und Interpretationen).

Damit sind verschiedene Ansatzpunkte für die Sicherung und Überprüfung von Qualität im qualitativen Forschungsprozess aufgezeigt. Auf diesem Weg lassen sich Vorgehen und Durchführung im Prozess der Forschung offen legen und beurteilen. Unter dem Blickwinkel der produzierten Erkenntnisse lassen sich die Fragen, die ein solcher Prozess der Überprüfung beantworten soll, nach Huberman und Miles (1998, S. 202) allgemeiner zusammenfassen. Sie richten sich auf die Begründetheit der Erkenntnisse in den Daten und der Schlüsse, sowie die Angemessenheit der Kategorienstruktur. Sie sollen prüfen, ob Forschungsentscheidungen gerechtfertigt waren und ob Strategien zur Erhöhung der Glaubwürdigkeit angewendet wurden. Dabei sind zwar die Ergebnisse Ausgangspunkt der Bewertung der Forschung, die zu ihnen geführt hat; jedoch wird diese Frage

in der Verbindung einer ergebnisorientierten Sichtweise mit einem prozessorientierten Herangehen zu beantworten gesucht.

Die bislang skizzierten Strategien zielen auf die Formulierung von Kriterien ab, die analog zu den in der quantitativen Forschung etablierten Kriterien in der qualitativen Forschung eingesetzt werden können.

Steinke (1999) schlägt in einem ersten Versuch im deutschen Sprachraum sieben Kriterien vor: Intersubjektive Nachvollziehbarkeit des Prozesses, der zu Ergebnissen geführt hat; Indikation bzw. Gegenstandsangemessenheit des Vorgehens; empirische Verankerung von Theoriebildung und -prüfung; Limitation, d.h. die Benennung der Reichweite von Ergebnissen; reflektierte Subjektivität; Kohärenz der Theorie und Relevanz von Fragestellung und Theoriebildung.

In den hier kurz vorgestellten Vorschlägen tauchen jeweils verschiedene Probleme auf. Einerseits ist es bei diesen Kriterien – anders als bei der Reliabilitätsbestimmung in der quantitativen Forschung – schwierig, Grenzwerte oder Punkte zu definieren, die zwischen guter und schlechter Forschung unterscheiden: Im Beispiel der Glaubwürdigkeit werden von Lincoln und Guba lediglich Strategien formuliert, wie diese hergestellt bzw. erhöht werden kann. Der Forscher, der diese zur Sicherung von Qualität und Glaubwürdigkeit auf seine Forschung anwenden möchte, ist ebenso mit der Frage allein gelassen wie der Leser, der einen Forschungsbericht anhand dieses Kriteriums bewerten möchte: Welche Resultate müssen Peer debriefing und/oder Member checks bringen, damit sie ein Indikator für die Glaubwürdigkeit der damit überprüften Forschung sind? Müssen *alle* dabei Befragten zu einheitlichen Einschätzungen kommen – etwa was die Plausibilität der Resultate angeht – oder reicht es, wenn die *Mehrheit* oder bestimmte Personen diese Plausibilität bestätigt? Ist etwa die Bestätigung seitens bestimmter Personen anders zu gewichten als die Ablehnung durch die anderen Befragten? Zum Problem wird dies, da ohne die Angabe von Grenzwerten die Idee der Kriterien häufig zu gut gemeinten Absichtserklärungen verkommt (vgl. auch Lüders 2000). Andererseits sind all diese Vorschläge jeweils vor dem Hintergrund eines bestimmten Ansatzes formuliert und in ihrer Anwendung auf andere Ansätze eher begrenzt (vgl. hierzu auch Lüders 2003).

5. Guidelines, Checklists, Kriterienkataloge

Die Frage der Bewertung qualitativer Forschung wird aktuell in drei Kontexten gestellt: Zum einen stellt sie sich dem Forscher, der sein Vorgehen und seine Ergebnisse überprüfen und absichern möchte; weiterhin stellt sie sich dem Abnehmer der Forschung – dem Leser von Veröffentlichungen oder dem Auftraggeber, der das Vorgestellte einschätzen und bewerten möchte; schließlich wird sie bei der Begutachtung qualitativer Forschung, bei der Beurteilung von Forschungsanträgen und zunehmend auch bei der von Manuskripten im peer review von Zeitschriften aufgeworfen. Gerade in Hinblick auf den letzten Kontext ist eine wachsende Zahl von Guidelines zur Bewertung von Forschungspapieren (Artikeln, Anträgen etc.) in verschiedenen Anwendungsfeldern zu verzeichnen.

5.1 Gesundheitsforschung

Seale (1999, S. 189-192) gibt einen Kriterienkatalog der British Sociological Association of Medical Sociology wieder, der einen Fragenkatalog zu 20 Bereichen von der Fragestellung über das Vorgehen bei Sampling, Erhebung, Analyse und Darstellung sowie Ethik umfasst. Die vorgestellten Leitfragen sind sicherlich hilfreich, bei der Beantwortung ist der Anwender des Kataloges jedoch auf seine eigenen ggf. impliziten Kriterien angewiesen, um bspw. im Bereich 19 (Are the results credible and appropriate?) die Frage „do they address the research question(s)“? (ebd., S. 192) zu beantworten.

Einen weiteren Katalog legten die National Institutes of Health, Office of Behavioral and Social Sciences (NIH 2001) für den Bereich Public Health vor. Hierbei wird vor allem auf Designfragen Wert gelegt, zu denen auch Fragen der Erhebung und Analyse gerechnet werden, sowie auf die Kombination von qualitativer und quantitativer Forschung. Die Erläuterung der relevanten Teile eines Forschungsantrags und der Fragen, die dabei zu beachten sind, wird durch eine Checklist ergänzt, die Items enthält wie bspw.: „Data collection procedures are fully explained“ (ebd., S. 16). Ziel ist, den eingeführten Katalog für quantitative Projekte auf die Besonderheit qualitativer Anträge hin zu modifizieren.

Für den Bereich der klinischen Psychologie haben Elliot/Fischer/Rennie (1999) einen Katalog von Guidelines für die Publikation qualitativer Studien vorgelegt. Dieser enthält zwei Bereiche, von denen der erste für qualitative und quantitative Forschung gleichermaßen gelten soll, der zweite dagegen auf die Besonderheiten qualitativer Forschung zugeschnitten ist. Im ersten Teil werden Fragen der Angemessenheit der Methodenwahl oder der ausreichenden Spezifikation der Methoden behandelt, während im zweiten Teil Fragen der ausreichenden Verankerung von Aussagen in Beispielen, die Kohärenz von Ergebnissen oder die Anwendung von „credibility checks“ (z.B. Member checks oder Peer debriefing, Triangulation etc. – ebd., S. 229) behandelt werden. Wie die heftige Reaktion von Reicher (2000) verdeutlicht, sind diese Guidelines trotz ihrer relativ allgemeinen Formulierung nicht unbedingt konsensfähig und auf die eingangs kurz skizzierte Bandbreite qualitativer Forschung übertragbar.

5.2 Qualitative Evaluationsforschung

Speziell für den Kontext der qualitativen Evaluationsforschung werden in den letzten Jahren verstärkt Checklists, Frameworks und Kriterienkataloge entwickelt (vgl. auch Flick 2006b). So haben Spencer/Ritchie/Lewis/Dillon (2003) einen „Framework for assessing research evidence“ für diesen Bereich vorgelegt. Dieser basiert auf Literaturanalysen und 29 Experteninterviews mit Auftraggebern, Abnehmern, Forschern und Praktikern, die jeweils in Evaluationen involviert waren. Der Rahmen orientiert sich an vier leitenden Prinzipien:

„Research should be

1. contributory in advancing wider knowledge or understanding;
2. defensible in design by providing a research strategy which can address the evaluation questions posed;
3. rigorous in conduct through the systematic and transparent collection, analysis and

- interpretation of qualitative data;
4. credible in claim through offering well-founded and plausible arguments about the significance of the data generated" (S. 6).

Zur Bewertung konkreter Vorhaben haben sie insgesamt 18 Fragen formuliert, die sich sieben Bereichen bzw. Schritten des Forschungsprozesses zuordnen lassen: Die Fragen beziehen sich auf die Ergebnisse (z.B.: How credible are the findings? How has knowledge/understanding been extended by the research?), auf Designs, Sampling (z. B.: How well defended is the sample design/target selection of cases/documents?), Datensammlung und -analyse (z. B.: How well has the approach to and formulation of the analysis been conveyed? Contexts of data sources – how well are the retained and portrayed?). Die Darstellung der Ergebnisse sowie Reflexivität und Neutralität des Forschers werden ebenso in Fragen gefasst. Während diese Fragen eher eine orientierende Funktion haben, wird das Framework in den mitgelieferten Quality Indicators konkretisiert. Davon werden insgesamt 88 formuliert, zu den einzelnen Fragen unterschiedlich viele. Zur Frage der Glaubwürdigkeit findet sich z. B. der Indikator „Findings/conclusions make sense/have a coherent logic“, zur Sample Composition der Indikator „Detailed profile of achieved sample/case coverage“, jeweils neben weiteren Indikatoren für beide Fragen (vgl. Spencer et al. 2003, S. 22 ff.).

Dieser Fragenkatalog wurde im Auftrag des „Research Cabinet“ bzw. „the UK Government's Office of the Chief Social Researcher“ (Kushner 2005, S. 111) erstellt und an Ministerien bzw. alle Government Departments verschickt. Er soll diesen bei der Vergabe und vor allem Bewertung von Evaluationen, die auf qualitativen Methoden basieren, eine Orientierung bieten. Vor allem wegen dieser Verbreitung und Funktion wird das Framework etwa von Kushner (2005) mit besonderer Aufmerksamkeit und kritisch betrachtet.

Dabei werden verschiedene Vorteile gesehen, etwa dass es Klarheit in den Bereich der Evaluation bringe, dass es ein Beitrag zur methodischen Weiterentwicklung qualitativer Forschung sei und Evaluatoren vor unsinnigen Verträgen bzw. Aufträgen retten würde (ebd., S. 115 f.). Gewichtiger sind jedoch die formulierten Vorbehalte. Grundsätzlich kritisiert Kushner (2005), dass mit der Formulierung und Verteilung solcher Bewertungsinstrumente die Verantwortung für die Qualität von den Evaluatoren zu den Auftraggebern (häufig Verwaltungen) verlagert würde: „(...) it places on government an unreasonable responsibility to manage and guarantee the independence of an evaluation. Independence is conventionally guaranteed by the principle that an evaluation externally is ‚sponsored‘ and not ‚bought‘ by government“ (ebd., S. 116).

Weiterhin kritisiert Kushner, dass die Ausführungen im Rahmen des Framework zu sehr auf die erkenntnistheoretische und method(olog)ische Literatur gestützt würden, kaum jedoch auf die „politics of inquiry“ oder den Auftrag von Evaluatoren abzielen. Gerade die Konfrontation mit vielfältigen und häufig konfligierenden Absichten würde zu wenig berücksichtigt (ebd., S. 116). Insgesamt sei das Framework zu sehr an Fragen angewandter Sozialforschung und zu wenig an den Besonderheiten qualitativer Evaluation orientiert (dass beides unterschiedlich zu sehen ist, hält auch Lüders 2006b fest). Ebenso würde mehr die erkenntnistheoretische als die praktische Dimension der Evaluation behandelt und „policy evaluation“ mit Programmevaluation gleichgesetzt. An diesem umfangreichen Vorschlag für einen Bewertungskatalog zur Beurteilung qualitativer Evaluationen und der detaillierten Kritik daran lässt sich die Problematik der

Qualitätsbestimmung bei qualitativen Studien erneut zeigen. Auch hier sind die Fragen aufgeworfen, ob solche Kataloge für alle Formen von Evaluationen gelten können (Policy- vs. Programmevaluation), wie sich der eher grundsätzliche Diskurs über Qualität auf eine handhabbare Anleitung in der Forschungspraxis herunterbrechen lässt, die gleichzeitig noch den Besonderheiten von Evaluation gegenüber Forschung Rechnung trägt. Schließlich, und das hat Kushner in ihrer Kritik nicht weiter berücksichtigt, sind die Fragen (appraisal questions) und quality indicators, die vorgeschlagen werden, eher Leitfragen zur Orientierung als Hinweise, wo die Grenze zwischen guter und schlechter Evaluation verläuft. Kriterien in der quantitativen Forschung beinhalten solche Grenzwerte in der Regel – eine bestimmter Wert der Übereinstimmung zwischen Kodierern bei der Überprüfung der Interrater-Reliabilität muss schon gegeben sein, damit das Kriterium erfüllt ist. Und schließlich ist auch hier die Frage zu stellen, welchen Stellenwert das Framework bzw. die ergänzenden Ausführungen für qualitative Forschung außerhalb der Evaluation haben.

5.3 Management- und Organisationsforschung

Für diesen Bereich haben Cassell/Buehring/Symon/Johnsop/Bishop (2005) ein Projekt zu „Benchmarking good practice in qualitative management research“ durchgeführt, das auf der Grundlage von Interviews mit Abnehmern solcher Forschung deren implizite und explizite Standards der Bewertung qualitativer Forschung herausgearbeitet hat. Dabei werden die Konzepte der Befragten einerseits davon abhängig gemacht, was qualitative Forschung ausmacht, andererseits daran, woran die Glaubwürdigkeit ihrer Ergebnisse bestimmt wird. Zu letzterem finden sich Aspekte wie die Quantifizierbarkeit von Ergebnissen, die rigorose Anwendung von Methoden, das Vorhandensein „technischer skills“ bei der Durchführung und die Möglichkeit, praktisch relevante Schlussfolgerungen aus den Resultaten zu ziehen u. a. „Good practice“ in qualitativer Forschung wird demnach v. a. an Aspekten gemessen, wie dem Forschungsdesign (warum wurde welche Methode gewählt, wie war das Sampling konzipiert) und der Anwendung von Kombinationen von Methoden (v. a. qualitativ und quantitativ). Daneben wird die Analyse und Reflexivität hinsichtlich des gewählten Vorgehens und die Aufbereitung und Verbreitung der Ergebnisse zu einem Anhaltspunkt. Bei der Bestimmung der Qualität qualitativer Managementforschung steht die Frage im Vordergrund, welche „contribution“ von den Ergebnissen geleistet wird: liefern sie neue Erkenntnisse, praktische Konsequenzen oder die Entdeckung neuer Problemlagen als Resultat? Daneben werden aber auch Faktoren wie „technical accomplishment“ in der Anwendung von Methoden, oder die Frage, inwieweit Kriterien bei der Durchführung eines Projektes eine Rolle gespielt haben, als relevant erachtet.

Diese Studie zeigt vor allem, welche impliziten und expliziten Kriterien die Abnehmer qualitativer Forschungsergebnisse in einem speziellen Bereich anwenden. Man kann sie einerseits als Beitrag zur sozialwissenschaftlichen Verwendungsforschung lesen – der wieder einmal zeigt, dass die Praxis möglicherweise ganz andere Maßstäbe und Kriterien bei der Bewertung von Forschungsergebnissen und Vorgehensweisen anlegt als die Forscher selbst (vgl. auch Lüders 2006a).

Andererseits zeigen sie die Notwendigkeit auf, von der Seite der Forscher selbst zu definieren, was gute qualitative Forschung ist, da sich die Bewertung qualitativer Forschung ansonsten (noch stärker) verselbstständigt bzw. von außen vorgenommen wird. Schließlich zeigt dieses Projekt die Schwierigkeiten, die sich ergeben, wenn man die Standards von Forschung aus den Standards im Feld bzw. in der Praxis rekonstruieren will (vgl. hierzu Bohnsack 2005, s. o.), da es dieser Gruppe zumindest nicht gelungen ist, aus ihren Analysen überzeugende Kriterien, Standards oder einen Beurteilungskatalog für Forschungsprojekte und -ergebnisse zu entwickeln.

Die hier kurz skizzierten Beurteilungskataloge stellen weniger eine abschließende Antwort auf die Frage nach den angemessenen Kriterien zur Bewertung qualitativer Forschung dar. Sie verdeutlichen vielmehr die Brisanz der Frage der Bewertung und zeigen, dass diese zunehmend konkretisiert gestellt wird und dass ggf. auch Antworten von außen an die qualitative Forschung herangetragen werden, wenn sie diese nicht selber formuliert.

6. Die besondere Problematik der Qualität qualitativer Forschung

Die hier behandelten Kriterienkataloge und die Reaktionen, die sie provozieren, zeigen aber auch die spezielle Problematik qualitativer Forschung im Vergleich zu quantitativer Forschung. Diese Problematik lässt es zunehmend unwahrscheinlich erscheinen, dass die Antwort auf die Qualitätsfrage in einem eigenen Katalog von Qualitätskriterien liegen wird, der komplementär zu den ‚quantitativen‘ Kriterien eingesetzt werden kann: Einerseits wird bezweifelt, dass es sinnvoll ist, von ‚der‘ qualitativen Forschung zu sprechen, auch wenn es eine ganze Reihe von verbindenden Kennzeichen gibt (vgl. Flick 2002, Schütze in diesem Heft). Autoren wie Reicher (2000) oder auch Madill et al. (2000) stellen in Frage, ob es einheitliche Kriterien oder Bewertungsansätze für alle Formen qualitativer Forschung geben kann. Sie treffen eine Unterscheidung vor allem zwischen realistischen und (radikal) konstruktivistischen (Madill et al. 2000) oder zwischen „experientiellen“ und diskursiven Methoden (Reicher 2000). Diese von der Diskussion in England geprägten Unterscheidungen können jedoch in ihren Konsequenzen auch auf die deutschsprachige Diskussion übertragen werden: Sollen qualitative Forschungsprojekte, die sich mit Interviews auf den Weg zur Entwicklung einer „grounded theory“ begeben, nach denselben Maßstäben bewertet werden wie Untersuchungen, die mittels objektiver Hermeneutik fallrekonstruktive Forschung betreiben? Oder weiter gedacht: Welchen Stellenwert können die vor einem spezifischen theoretischen und Diskurshintergrund entstandenen Kriterien von bspw. Lincoln/Guba (1985) für die Bewertung der vor einem ganz anderen Hintergrund operierenden hermeneutischen Verfahren in der deutschen Diskussion haben? Selbst wo ein kontextübergreifender Trend wie der zur Ethnographie festzustellen ist, bleibt zu bezweifeln, dass sich hier ein gemeinsames Verständnis von Qualität etablieren wird, wenn man etwa die Vehemenz berücksichtigt, mit der etwa Hirschauer (2001) den Ansatz von Denzin in Frage stellt.

Das heißt eine Seite der speziellen Problematik der qualitativen Forschung in der Beantwortung der Qualitätsfrage ist das Spannungsverhältnis zwischen qua-

litativer ‚Einheitswissenschaft‘ und Methoden- oder Schulenvielfalt (vgl. hierzu auch Flick 2005), der man mit einheitlichen oder (nur mit) differenzierenden Kriterien oder Bewertungsansätzen gerecht werden kann. Es soll dabei nicht unterstellt werden, dass in der quantitativen Forschung nicht auch unterschiedliche Ansätze und Richtung zu verzeichnen wären, jedoch hat diese Vielfalt bislang nicht zu einer schulenspezifischen Infragestellung der Kriterien Reliabilität, Validität und Objektivität geführt.

Die andere Seite der Problematik ist, dass die Qualität qualitativer Forschung jenseits dessen liegt, was in eindeutige Kriterien gefasst werden kann (vgl. hierzu Flick 2002, Kap. 22). Yardley (2000) diskutiert in diesem Kontext „dilemmas in qualitative research“. Wie lässt sich bei einer explorativen Studie etwa bewerten, was den tatsächlichen Gewinn an neuem Wissen darstellt? Wie lässt sich bewerten, ob die verwendeten Methoden dem untersuchten Feld und der Fragestellung angemessen waren? Kann man die Originalität im methodischen und im Feld-Zugang beurteilen? Auf welche Weise kann man die Kreativität im Zugang zum und im Umgang mit dem Material bewerten? Wie lässt sich das Verhältnis von Einzelschritt und Gesamtprozess beurteilen? Die meisten der oben behandelten Bewertungsansätze versuchen, die Qualitätsfrage auf den einzelnen Schritt im Forschungsprozess herunterzubrechen: Madill et al (2000) bspw. betrachten die Frage der Objektivität und Reliabilität ausschließlich an der Übereinstimmung der Interpretationen unterschiedlicher Forscher, ohne die anderen Schritte des Forschungsprozesses dabei zu berücksichtigen.

7. Strategien der Geltungsbegründung

Eine dritte Alternative – neben der Formulierung von Standards oder Kriterien – ist die Entwicklung und Anwendung von Strategien der Geltungsbegründung, um darüber die Qualität von Daten und Erkenntnissen zu bestimmen. Damit wird die Qualitätsfrage auf die Bewertung des Forschungsprozesses als Ganzen erweitert. Darauf soll abschließend eingegangen werden.

7.1 Analytische Induktion

Die analytische Induktion setzt am abweichenden Fall an. Darunter ist nach Bühler-Niederberger 1985, S. 476) zu verstehen: „Analytische Induktion ist eine Methode systematisierter Ereignisinterpretation, die sowohl den Prozeß der Genese wie auch der Prüfung von Hypothesen umfaßt. Ihr entscheidendes Instrument ist die Analyse der Ausnahme, des von der Hypothese abweichenden Falls.“ Dieses Verfahren setzt nach der Entwicklung einer vorläufigen Theorie (bzw. eines Musters, Modells etc.) an der Suche nach und Analyse von abweichenden Fällen (oder Gruppen) an. Dabei ist die analytische Induktion vor allem an der Absicherung gewonnener Theorien und Erkenntnisse durch die Analyse bzw. Integration abweichender Fälle orientiert. Als „Analyse negativer Fälle“ greifen Lincoln und Guba (1985) das Konzept auf. Anknüpfungen ergeben sich zu Fragen der Verallgemeinerung von Fallstudien, jedoch hat die analytische Induktion ihren eigenen Stellenwert als Prüfverfahren für Analysen.

7.3 Triangulation

Als Strategie der Geltungsbegründung qualitativer Forschung wird seit langem die ursprünglich von Denzin (1970/1989) vorgeschlagene Triangulation diskutiert, die zwischenzeitlich weiter elaboriert und differenziert wurde (vgl. Flick 2004 für einen Überblick). Dabei geht es im wesentlichen darum, die Erkenntnismöglichkeiten qualitativer Forschung durch zusätzliche Perspektiven zu erweitern – durch die Verwendung verschiedener Datensorten, die Einbeziehung unterschiedlicher Theorien und v. a. verschiedener Methoden. In seinen aktuelleren Varianten ist der Ansatz der Triangulation um eine fundiertere Verknüpfung unterschiedlicher qualitativer Methoden oder qualitativer mit quantitativen Methoden bemüht. Angesichts der eher pragmatisch ausgerichteten Diskussionen um eine Verknüpfung qualitativer und quantitativer Methoden in „Mixed Method Designs“ (vgl. Tashakkori/Teddlie 2003) erscheint eine solche fundierte Diskussion über die Verknüpfung und Verknüpfbarkeit von Methoden notwendig. Zu Qualitätsdiskussion in der qualitativen Forschung kann die Triangulation insofern einen Beitrag leisten, als sie Erkenntnisse verbreitern kann und der Widersprüchlichkeit und Vielschichtigkeit untersuchter Phänomene durch die unterschiedlichen Aspekte gerecht werden kann, die unterschiedliche methodische Zugänge aufzeigen. Die wechselseitige Bestätigung der Ergebnisse unterschiedlicher methodischer Zugänge im Sinne einer wechselseitigen Validierung, die Denzin propagiert hatte, hat sich als nicht realisierbare und theoretisch nicht haltbare Zielsetzung herausgestellt (vgl. hierzu auch Denzin 1989 und Flick 2004).

7.3 Qualitätsmanagement in der qualitativen Forschung

Eingangs war die Rede davon, dass die Standards qualitativer Forschung aus der Forschungspraxis rekonstruiert werden sollten (vgl. Bohnsack 2005). In gewisser Weise einen Schritt weiter geht das Konzept des Qualitätsmanagements in der qualitativen Forschung (vgl. Flick 2002, Kap. 22), wobei es stärker in der Forschung selbst verankert ist. Die Diskussion zum Qualitätsmanagement wird im Bereich der industriellen Produktion, Dienstleistungen oder im Gesundheitswesen seit längerem geführt (Kamiske/Brauer 1995). Dieser Ansatz lässt sich auf die qualitative Forschung übertragen, um eine Diskussion über Qualität in der Forschung voranzutreiben. Über das Konzept des Auditing ergeben sich bereits erste Anknüpfungspunkte. So wird für die Überprüfung der Verlässlichkeit qualitativer Daten von Lincoln/Guba (1985) ein Prozess des „auditing“ vorgeschlagen, der am Vorgang der Buchprüfung im Finanzwesen orientiert ist. Dafür wird ein „Überprüfungspfad“ (auditing trail) skizziert: Ein Auditing trail erfasst

- die Rohdaten, ihre Erhebung und Aufzeichnung;
- Datenreduktion und Ergebnisse von Synthesen durch Zusammenfassung, theoretische Notizen, Memos etc., Summaries, Kurzdarstellungen von Fällen etc.;
- Datenrekonstruktionen und Ergebnisse von Synthesen anhand der Struktur entwickelter und verwendeter Kategorien (Themen, Definitionen, Beziehungen), Erkenntnisse (Interpretationen und Schlüsse) sowie die erstellten Be-

richte mit ihren Integrationen von Konzepten und den Bezügen zu existierender Literatur;

- Prozessnotizen, d. h. methodologische Notizen und Entscheidungen auch hinsichtlich der Herstellung von Vertrauens- und Glaubwürdigkeit der Erkenntnisse;
- Materialien in Bezug auf Absichten und Anordnungen wie die Forschungskonzeption, persönliche Aufzeichnungen und Erwartungen der Beteiligten;
- Informationen über die Entwicklung der Instrumente einschließlich der Pilotversionen und vorläufigen Plänen (vgl. Lincoln/Guba 1985: 320-321).

Damit ist bereits die Prozessperspektive angelegt, die alle relevanten Schritte des Forschungsprozesses umfasst, der zu den Daten und ihrer Interpretation geführt hat. Im Kontext des Qualitätsmanagements ist ein Audit „(...) die systematische, unabhängige Untersuchung einer Aktivität und deren Ergebnisse, durch die Vorhandensein und sachgerechte Anwendung spezifizierter Anforderungen beurteilt und dokumentiert werden“ (Kamiske/Brauer 1995: 5). Insbesondere das „Verfahrensaudit“ ist für die Forschung interessant. Ein Verfahrensaudit soll sicherstellen, „daß die vorgegebenen Anforderungen eingehalten werden und für die jeweilige Anwendung zweckmäßig sind. (...) Vorrang hat immer das nachhaltige Abstellen von Fehlerursachen, nicht die einfache Fehleraufdeckung“ (Kamiske/Brauer 1995, S. 8). Solche Qualitätsbestimmungen werden nicht abstrakt – etwa an bestimmten Methoden per se – vorgenommen, sondern mit Blick auf die *Kundenorientierung* und die *Mitarbeiterorientierung* (Kamiske/Brauer 1995, S. 95f., 110f.). Dabei ergibt sich die Frage, wer eigentlich die Kunden sozialwissenschaftlicher Forschung sind. Im Qualitätsmanagement wird zwischen internen und externen Kunden unterschieden. Während letztere die Abnehmer des jeweiligen Produktes sind, gehören zu den ersteren die Beteiligten an der Herstellung im weiteren Sinn (z. B. Mitarbeiter anderer Abteilungen). Für die Forschung lässt sich diese Unterteilung übersetzen in diejenigen, für die das Ergebnis nach außen produziert wird (Auftraggeber, Gutachter etc. als externe Kunden), und diejenigen, für die und an denen das jeweilige Ergebnis zu erzielen gesucht wird (Interviewpartner, untersuchte Institutionen etc. als interne Kunden). Zur Überprüfung lassen sich beide Aspekte explizit analysieren: Inwieweit ist die Untersuchung so verlaufen, dass sie die Fragestellung beantwortet (externe Kundenorientierung) und den Perspektiven der Beteiligten ausreichend Raum lässt (interne Kundenorientierung)?

Die Mitarbeiterorientierung will berücksichtigen, dass „Qualität unter Anwendung geeigneter Techniken, aber auf der Basis einer entsprechenden Geisteshaltung entsteht“, wobei die „Übertragung von (Qualitäts-)Verantwortung auf die Mitarbeiter durch die Einführung von Selbstprüfung anstelle von Fremdkontrolle“ (Kamiske/Brauer 1995, S. 110f.) ein weiterer Ansatzpunkt ist. Entsprechend bezeichnet Qualitätsmanagement „Tätigkeiten (...), die die Qualitätspolitik, die Ziele und Verantwortlichkeiten festlegen sowie diese durch Mittel wie Qualitätsplanung, Qualitätslenkung, Qualitätssicherung/Qualitätsmanagement-Darlegung und Qualitätsverbesserung verwirklichen“ (ISO 1994; zit. nach Kamiske/Brauer 1995, S. 149).

Qualität im qualitativen Forschungsprozess lässt sich nur realisieren, wenn sie mit den beteiligten Forschern gemeinsam hergestellt und überprüft wird. Zunächst wird festgelegt, was eigentlich unter Qualität in diesem Zusammenhang

zu verstehen ist und verstanden wird. Dabei lassen sich folgende Leitgedanken für das Qualitätsmanagement in der qualitativen Forschung festhalten:

- eine möglichst klare Festlegung der zu erreichenden Ziele und einzuhaltenden Standards des Projekts. Daran müssen alle Forscher und Mitarbeiter beteiligt werden;
- eine Festlegung, wie diese Ziele und Standards und allgemeiner die angestrebte Qualität zu erreichen sind; damit sind eine Einigung über die Weise der Anwendung bestimmter Methoden und ihre Umsetzung, etwa durch gemeinsame Interviewtrainings und deren Auswertung, Voraussetzungen für Qualität im Forschungsprozess;
- die klare Festlegung der Verantwortlichkeiten für die Herstellung von Qualität im Forschungsprozess und
- die Transparenz der Beurteilung und Sicherstellung der Qualität im Prozess.

Die Bestimmung dessen, was Qualität ist, deren Herstellung und Sicherstellung im Prozess und die Erfahrung, dass Qualität sich nur in der Kombination von Methoden und einer entsprechenden Haltung realisieren lässt, sind Anknüpfungspunkte zur Diskussion um Qualitätsmanagement in der sozialwissenschaftlichen Forschung. Im Unterschied zu anderen Ansätzen der Qualitätsprüfung in der qualitativen Forschung klären beim Qualitätsmanagement zunächst allen Beteiligten, was unter Qualität verstanden wird, welche Qualitätsziele sich daraus ableiten lassen und wie diese jeweils zu erreichen sind. Hier wird der Gedanke aufgegeben, Forschungsqualität ließe sich allgemein, abstrakt und von außen bestimmen, zugunsten einer gemeinsamen Klärung des Qualitätskonzeptes und seiner Umsetzung (vgl. hierzu ausführlicher Flick 2002, Kap. 22).

7.4 Indikation qualitativer Forschung

Eine weitere Strategie in diesem Zusammenhang ist die Klärung der Indikationsfrage – ähnlich wie dies in der Medizin und Psychotherapie für die Eignung von Behandlungsmethoden bei bestimmten Problemen und Personengruppen geprüft wird. Auf den hier behandelten Kontext übertragen meint das die Frage, warum eigentlich bestimmte – und nicht andere – Methoden für die konkrete Untersuchung verwendet wurden. Nicht nur in qualitativer Forschung, sondern in empirischer Forschung generell geben Lehrbücher kaum eine Hilfestellung für die Entscheidung, wann eine bestimmte Methode für eine Untersuchung ausgewählt werden sollte. Die meisten dieser Bücher behandeln die einzelnen Methoden oder Forschungsdesigns separat, wenn sie ihre Eigenschaften und Probleme beschreiben. In den meisten Fällen gelangen sie nicht zu einer vergleichenden Darstellung verschiedener methodischer Alternativen oder zur Formulierung von Ansatzpunkten dafür, wie eine spezielle (und nicht eine andere) Methode für einen Forschungsgegenstand ausgewählt werden sollte. Entsprechend ist es für die qualitative Forschung notwendig, die Frage der Indikation weiter zu klären. Dabei heißen die relevanten Fragen: Wann sind welche qualitativen Methoden angemessen – für welchen Gegenstand? Für welche Fragestellung? Für welche Untersuchungsgruppe (Population) oder welches Untersuchungsfeld etc.? Wann sind quantitative Methoden oder eine Kombination von quantitativen und qualitativen Methoden indiziert? Die Klärung dieser Frage wird hier jedoch – anders als

bei Steinke (1999) – nicht als ein Kriterium sondern als eine Strategie der Geltungsbegründung verstanden (vgl. auch hierzu ausführlicher Flick 2002, Kap. 22).

8. Fazit und Ausblick

In diesem Beitrag wurde die Problematik der Beantwortung der Frage nach der Qualität qualitativer Forschung diskutiert. Die drei Ansätze, die hier in aller Kürze behandelt wurden (vgl. ausführlicher Flick 2007), sind als Wege zu einer solchen Antwort zu sehen, von denen keiner bislang eine umfassende Antwort geliefert hat. Angesichts der eingangs skizzierten Diversifizierung des Feldes bleibt auch abzuwarten, ob es zu einer einheitlichen Formulierung von Standards und Kriterien für qualitative Forschung insgesamt kommen kann und ob dies wünschenswert ist. Entsprechende Versuche werden in der Regel von verschiedenen Seiten als nicht praktikabel zurückgewiesen, da sie den Besonderheiten spezifischer Ansätze nicht angemessen Rechnung tragen. Dies wird für Kriterien und Guidelines diskutiert, stellt sich aber noch mehr für die Formulierung von Standards. Bei der Formulierung von Kriterien kommt noch eine zweite Problematik hinzu. Es erweist sich als schwierig, Grenzwerte oder Benchmarks zu formulieren, anhand derer zwischen guter und weniger guter Forschung unterschieden werden kann, so wie dies etwa für die quantitative Forschung und die dort verwendeten Kriterien praktikabel ist. Auch wenn die hier vorgestellten Beispiele einer empirischen Ermittlung von Qualitätsmerkmalen qualitativer Forschung (vgl. Abschnitt 5.2 und 5.3) noch nicht wirklich überzeugen können, ist damit ein Weg skizziert, der auch für andere Bereiche aufschlussreich sein könnte: durch empirische Analysen der Forschungspraxis und die Bewertung von qualitativer Forschung und ihrer Ergebnisse aus der Sicht ihrer Abnehmer Anhaltspunkte zu entwickeln, wonach die Qualität qualitativer Forschung bestimmt werden kann. Strategien wie analytische Induktion, Triangulation und Qualitätsmanagement können schließlich die Qualitätsentwicklung in den Forschungsprozess hineinragen und dort konkret und handhabbar machen.

Literatur

- Altheide D. L./Johnson, J. M. (1998): Criteria for Assessing Interpretive Validity in Qualitative Research. In: Denzin, N./Lincoln, Y. (Eds.): *Collecting and Interpreting Qualitative Materials* London/New Dehli, S. 293-312.
- Bohnsack, R. (2005): Standards nicht-standardisierter Forschung in den Erziehungs- und Sozialwissenschaften. In: *Zeitschrift für Erziehungswissenschaft*, 8, Beiheft 4/2005, S. 63-81.
- Bortz, J./Döring, N. (2001): *Forschungsmethoden und Evaluation für Sozialwissenschaftler* (3. Aufl.). Berlin u. a.
- Bühler-Niederberger, D. (1985): Analytische Induktion als Verfahren qualitativer Methodologie. In: *Zeitschrift für Soziologie*, 14 (1985), S. 475-485.

- Cassell, C./Symon, G. (eds) (2004): *Essential Guide to Qualitative Methods in Organizational Research*. London.
- Cassell, C., Buehring, A., Symon, G., Johnson, P. and Bishop, V. (2005): *Qualitative Management Research: A Thematic Analysis of Interviews with Stakeholders in the Field*.
- Denzin, N. K. (1978): *The Research Act* (2. Aufl.). Chicago 1978 (3. Aufl.: Englewood Cliffs, N. J.: Prentice Hall 1989).
- Elliot, R./Fischer, C. T./Rennie, D. L. (1999): Evolving Guidelines for publication of qualitative research studies in psychology and related fields. In: *British Journal of Clinical Psychology* (38), S. 215-229.
- Flick, U. (1987): Methodenangemessene Gütekriterien in der qualitativ-interpretativen Forschung. In: Bergold, J. B./ Flick, U. (Hrsg.): *Ein-Sichten. Zugänge zur Sicht des Subjekts mittels qualitativer Forschung*. Tübingen, S. 246-263.
- Flick U. (2000): Triangulation in der qualitativen Forschung In: Flick, U./Kardorff, E. v./Steinke, I. (Hrsg.): *Qualitative Forschung – Ein Handbuch*. Reinbek, S. 309-319.
- Flick U. (2000): *Qualitative Sozialforschung – Eine Einführung*. Reinbek 2002.
- Flick, U (2004): *Triangulation – Methodologie und Anwendung*. Wiesbaden.
- Flick, U. (2005): *Qualitative Research in Germany and the US - State of the Art, Differences and Developments*. *Forum Qualitative Sozialforschung* 6, No. 3, <http://www.qualitative-research.net/fqs/fqs-e/inhalt3-05-e.htm>.
- Flick, U. (Hrsg.) (2006a): *Qualitative Evaluationsforschung – Konzepte, Methoden, Anwendungen*. Reinbek.
- Flick, U. (2006b): Qualität in der qualitativen Evaluationsforschung. In: Flick, U. (Hrsg.): *Qualitative Evaluationsforschung – Konzepte, Methoden, Anwendungen*. Reinbek, S. 424-443.
- Flick, U. (2007): *How to Evaluate Qualitative Research*. (Vol. 8 of the *Qualitative Research Kit*). London/Dehli.
- Friebertshäuser, B./Prenzel, A. (Hrsg.) (1997): *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft*. Weinheim/München.
- Glaser, B. G./Strauss, A. L. (1979): Die Entdeckung gegenstandsbegründeter Theorie: Eine Grundstrategie qualitativer Forschung. In: Hopf, C./Weingarten, E. (Hrsg.): *Qualitative Sozialforschung*. Stuttgart, S. 91-112.
- Habermas J. (1981): *Theorie des kommunikativen Handelns* (2 Bd.). Frankfurt.
- Helsper, W./Herwartz-Emden, L./Terhart, E. (2001): Qualität qualitativer Forschung in der Erziehungswissenschaft. In: *Zeitschrift für Pädagogik*, (47), S. 251-269
- Hirschauer, S. (2001): Rezension zu Flick/Kardorff/Steinke (Hg.) (2000): *Qualitative Forschung – ein Handbuch*. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, (53), S. 809-81.
- Huberman, A. M./Miles, M. B. (1998): *Data Management and Analysis Methods*. In: N. Denzin & Y. S. Lincoln (eds.): *Collecting and Interpreting Qualitative Materials*. London/Thousand Oaks/New Dehli, S. 179-211.
- Kamiske, G. F./Brauer, J. P. (1995): *Qualitätsmanagement von A bis Z - Erläuterungen moderner Begriffe des Qualitätsmanagements* (2. Aufl.). München.
- Kirk J./Miller Marc L. (1986): *Reliability and Validity in qualitative Research*. Beverley Hills.
- Kowall, S./O'Connell, D. C. (2000): Zur Transkription von Gesprächen. In: Flick, U./Kardorff, E. v./Steinke, I. (Hrsg.): *Qualitative Forschung – Ein Handbuch*. Reinbek, S. 437-447.
- Knoblauch, H./ Flick, U./Maeder, C. (eds.) (2005): The State of the Art of Qualitative Research in Europe. In: *Forum Qualitative Sozialforschung* 6, No. 3 (<http://www.qualitative-research.net/fqs/fqs-e/inhalt3-05-e.htm>).
- Kushner, S. (2005): Qualitative Control – A Review of the Framework for Assessing Qualitative Evaluation. In: *Evaluation*, 11, S. 111-122.
- Kvale, S. (1995): Validierung: Von der Beobachtung zu Kommunikation und Handeln. In: Flick, U./Kardorff, E.v./Keupp, H./Rosenstiel, L.v./Wolff, S. (Hrsg.): *Handbuch Qualitative Sozialforschung* (2. Aufl.). München, S. 427-432.

- Legewie, H. (1987): Interpretation und Validierung biographischer Interviews. In: Jüttemann, G./Thomae, H. (Hrsg.): Biographie und Psychologie. Berlin, S. 138-150.
- Lincoln, Y. S./Guba, E. G. (1985): *Naturalistic Inquiry*. London, New Dehli.
- Lüders, C. (2000): Herausforderungen qualitativer Forschung In: Flick, U./ Kardorff, E. v./Steinke, I. (Hrsg.): *Qualitative Forschung – Ein Handbuch*. Reinbek, S. 632-643.
- Lüders, C. (2003): Gütekriterien. In: Bohnsack, R./Marotzki, W./Meuser, M. (Hrsg.): *Hauptbegriffe Qualitativer Sozialforschung*. Opladen, S. 80-82.
- Lüders, C. (2006a): Qualitative Evaluationsforschung – Was heißt denn hier Forschung? In: Flick, U. (Hrsg.): *Qualitative Evaluationsforschung – Konzepte, Methoden, Anwendungen*. Reinbek, S. 33-62.
- Lüders, C. (2006b): Qualitative Daten als Grundlage der Politikberatung. In: Flick, U. (Hrsg.): *Qualitative Evaluationsforschung – Konzepte, Methoden, Anwendungen*. Reinbek, S. 444-462.
- Lüders, C./Reichert, J. (1986): Wissenschaftliche Praxis ist, wenn alles funktioniert und keiner weiß warum. Bemerkungen zur Entwicklung qualitativer Sozialforschung. In: *Sozialwissenschaftliche Literaturreisenschau*, (12), S. 90-102.
- Madill, A./Jordan, A./Shirley, C. (2000): Objectivity and reliability in qualitative analysis: Realist, contextualist and radical constructionist epistemologies. In: *British Journal of Psychology*, (91), S. 1-20.
- Mensching, A. (2006): Zwischen Überforderung und Banalisierung – zu den Schwierigkeiten der Vermittlungsarbeit im Rahmen der qualitativen Evaluationsforschung. In: Flick, U. (Hrsg.): *Qualitative Evaluationsforschung – Konzepte, Methoden, Anwendungen*. Reinbek, S. 339-362.
- Mishler, E. G. (1986): *Research Interviewing. Context and Narrative*. Cambridge, Ma.
- Morse, J. M. (1999): Myth #93: Reliability and Validity are not relevant for qualitative inquiry – Editorial. In: *Qualitative Health Research*, (9), S. 717-718.
- NIH-Office of Behavioral and Social Sciences Research des National Institutes of Health (ed.) (2001): *Qualitative Methods in Health Research – Opportunities and Considerations In Application and Review*.
- Reicher, S. (2000): Against methodolatry: Some comments on Elliot, Fischer, and Rennie. In: *British Journal of Clinical Psychology*, (39), S. 11-26.
- Reichert, J. (2000): Objektive Hermeneutik und hermeneutische Wissenssoziologie In: Flick, U./Kardorff, E. v./Steinke, I. (Hrsg.): *Qualitative Forschung – Ein Handbuch*. Reinbek, S. 514-524.
- Schaeffer, D./Müller-Mundt, G. (Hrsg.) (2002): *Qualitative Gesundheits- und Pflegeforschung*. Bern.
- Seale, C. (1999): *The Quality of Qualitative Research*. London/Thousand Oaks/New Dehli.
- Spencer, L./Ritchie, J./Lewis, J./Dillon, L. (2003): *Quality in qualitative evaluation: A framework for assessing research evidence*. London (www.natcen.ac.uk)
- Steinke, I. (1999): *Kriterien qualitativer Forschung. Ansätze zur Bewertung qualitativ-empirischer Sozialforschung*. Weinheim/München.
- Tashakkori, A./Teddlie, C. (eds.) (2003): *Handbook of Mixed Methods in Social & Behavioral Research*. Thousand Oaks.
- Yardley, L. (2000): Dilemmas in Qualitative Health Research. In: *Psychology and Health*, (15), S. 215-228.
- Terhart, E. (1995): Kontrolle von Interpretationen. In König, E./Zedler, P. (Hrsg.): *Bilanz qualitativer Forschung, Bd. I: Grundlagen qualitativer Forschung*. Weinheim, S. 373-297.