

**Fachbereich Erziehungswissenschaften
der
Philipps-Universität Marburg**

Diplomarbeit

von

Heiko Grunenberg

[g-berg@web.de]

**„Die Qualität qualitativer Forschung. Eine
Metaanalyse erziehungs- und
sozialwissenschaftlicher
Forschungsarbeiten“**

Betreuer: Prof. Dr. Udo Kuckartz
Zweitgutachter: Prof. Dr. Benno Hafeneger
Abgabetermin: 04. September 2001

Inhalt

Verzeichnis der Abbildungen, Grafiken und Tabellen.....	2
0. Einleitung.....	3
1. Allgemeines	5
1.1 Vorgehensweise.....	5
1.2 Einwurf: Metaanalyse und Auswertung	10
1.3 Kennzeichnung qualitativer Sozialforschung	13
2. Die Qualität qualitativer Sozialforschung als Geltungsbegründung mittels Gütekriterien	16
2.1 Die Notwendigkeit von Qualitätskriterien.....	17
2.2 Vorbehalte gegenüber Qualitätskriterien	20
2.3 Der Standpunkt des subtilen Realismus	22
2.4 Die klassischen Gütekriterien	28
2.4.1 Das Konzept der Validität	29
2.4.2 Das Konzept der Reliabilität.....	36
2.4.3 Das Konzept von Objektivität.....	38
2.5 Alternative oder auch methodenangemessene Gütekriterien	41
2.6 Sonstige Kriterien.....	46
3. Die Kriterien	50
3.1 Der Kriteriumskatalog.....	56
4. Auswertung der Zeitschriftenartikel.....	59
4.1 Die Stichprobe	59
4.2 Die vorgefundene Qualität.....	64
4.3 Weitergehende Analysen	78
5. Schlussbetrachtungen und Fazit	83
6. Literatur.....	86
Anhang I: Liste der untersuchten Texte	
Anhang II: Ausgaben der Faktorenanalyse	

Verzeichnis der Abbildungen, Grafiken und Tabellen

	Seite
Abbildungen	
Abbildung 1: Struktur des Forschungsprozesses meta-analytischer Forschung	11
Grafiken	
Grafik 1: Art der Methodenkombination	68
Tabellen	
Tabelle 1: Herkunft der Texte	60
Tabelle 2: Geschlechterbeteiligung an der Forschung	61
Tabelle 3: Forschungsziele	62
Tabelle 4: Themenübersicht der Aufsätze	63
Tabelle 5: Häufigkeit der gewählten Methoden	65
Tabelle 6: Erfüllte Kriterien der Gegenstandsangemessenheit	67
Tabelle 7: Gegenstand von Reflexion	69
Tabelle 8: Übersicht über das Generalisierungsverhalten	72
Tabelle 9: Defizitliste der Dokumentationsbereiche	75
Tabelle 10: Faktorenübersicht	79

0. Einleitung

Es gab Zeiten, da war es eine Art Glaubensfrage sich einem Forschungsparadigma zuzuordnen. Quantitative und qualitative Methoden standen sich samt ihren Verfechterinnen und Verfechtern diametral gegenüber und übten sich in gegenseitiger Herabwürdigung. „Nun sag`, wie hältst du`s mit der Methodologie?“ ist aktuell dagegen nur noch für wenige Personen eine Gretchenfrage von zentraler Bedeutung. Nicht nur, dass eine klare Dichotomisierung kaum mehr aufrecht zu erhalten ist, ferner sind die Vor- und Nachteile der einzelnen Methoden ausgiebig diskutiert und die Anwendungsbereiche abgesteckt worden. Auf dem Gebiet der empirischen Sozialforschung mussten interpretative Verfahren sich den Status weitgehender Akzeptanz in hartem Kampf erobern. Zunehmend ist in den Selbstbeschreibungen von Wissenschaftsbetreibenden ein Interesse für beide Möglichkeiten zu finden.

Dementsprechend haben die heutigen Diskurse in der scientific-community den Streitpunkt der je beanspruchten Legitimität überschritten und bewegen sich nunmehr um Teilaspekte einer dauerhaften nutzbringenden Einrichtung nebeneinander. Es fällt auf, dass die Zusammenschau der Ansätze vordergründig meist zum Nachteil der qualitativen Forschung ausfällt. Insbesondere gilt dies für den Bereich der Sicherstellung von Qualität. Mittels universeller Gütekriterien versuchen quantifizierende Verfahren ihre Standards festzulegen. Die Entwicklung jener Kriterien lässt sich in die Mitte des vergangenen Jahrhunderts datieren, sprich in die Zeit des Aufblühens der empirischen Sozialforschung in den Vereinigten Staaten. Seitdem sind kaum mehr Weiterentwicklungen zu verzeichnen gewesen. In Bezug auf die qualitative Sozialforschung verläuft die Entwicklung ungleichzeitig aber parallel. Entsprechend der Tatsache des erst späteren Erscheinens der Letztgenannten in der „Arena“, sind deren spezifische Diskussionen noch nicht in gleichem Maße ausgetragen worden. Interessanterweise weist die momentane Debatte im Rahmen qualitativer Forschung trotz des Verstreichens einiger Jahrzehnte etliche inhaltliche Übereinstimmungen mit der damaligen auf.

Symptomatisch dafür ist der Disput um Gütekriterien in der qualitativen Forschung, der den folgenden Ausführungen zu Grunde liegt. Eine genauere Betrachtung zeigt

den zeitlichen Rückstand sowie die Einbettung in momentane philosophische bzw. erkenntnistheoretische Entwicklungszustände. Unter dieser Perspektive ist es unangemessen, die erwähnte vordergründige Unterlegenheit der qualitativen Forschung, als dessen eigene Schwäche zu begreifen. Allerdings wirkt mitunter das Selbstverständnis quantitativer Methoden von der eigenen Überlegenheit der „harten“ Daten nach wie vor weiter. Die alte Standpunkt einer strikten Zweiteilung in „harte“ und „weiche“ Verfahren mutet allerdings mittlerweile ewig-gestrig an, zuhauf ist diese Schimäre angegangen worden und längst entkräftet. Nun, wo die grundsätzliche Unterlegenheit der interpretativen Verfahren kaum mehr behauptet wird, gilt es, ihnen den letzten Anschein von Beliebigkeit zu nehmen. Selbstverständlich kann die Umsetzung einer Forschung derartig defizitär sein, dass sie das Etikett der Unwissenschaftlichkeit verdient. Dem Zweck, eine solche Diagnose zu stellen resp. den negativen Abglanz auf die Methode als solche bloßstellen zu können und eine weitgehende Diskreditierung zu vereiteln, gilt die Entwicklung von Gütekriterien.

Wer hier eine Kriterienliste zum abhaken, wie sie lange für die quantifizierenden Verfahren als existent unterstellt wurde, erwartet, wird leider enttäuscht werden. Denn schließlich müsste dafür eine „Wahrheit“ an sich unterstellt werden können, die sich textuell niederschlägt. Wahrheitsvorstellungen spielen in den folgenden Ausführungen eine gewichtige Rolle, jene genannte jedoch nur als überwundenes Stadium.

Um einer Mythenbildung über den Zustand qualitativer Forschung in der Praxis entschieden entgegenzutreten, sei es, dass die momentane Umsetzung vermaledeit wird oder sei es, dass sie sich gegen Kritik immunisiert, werden aktuell durchgeführte qualitative Forschungsarbeiten einer genaueren Betrachtung unterzogen. Dabei wird sich herausstellen, wie es um die Güte bestellt ist. Finden die zahlreichen Bemühungen um die Fortentwicklung einer Fehlertheorie in der Praxis Beachtung oder laufen Theorie und Praxis berührungslos nebeneinander her? Diesen und anderen Überlegungen widmen sich die nun folgenden Ausführungen.

1. Allgemeines

1.1 Vorgehensweise

So wie der Titel dieser Arbeit durch seine Interpunktion in zwei Einheiten geteilt ist, so wird gleichsam die inhaltliche Schwerpunktsetzung eine zweigeteilte sein. Ein zunächst theoretisch gehaltenes Fundament bietet die Grundfläche einer auf diesem fußenden empirischen Analyse.

Zu Beginn wird auf einer theoretischen Ebene erarbeitet werden, welche Merkmale die Qualität qualitativer Sozialforschung ausmachen. Dabei wird eine Auseinandersetzung mit diesbezüglich unterschiedlichen Strömungen und Ansichten innerhalb der aktuellen Debatten im Bereich der qualitativen Sozialforschung gesucht und versucht werden jene auf eine fruchtbare Weise zusammenzuführen, indem gezielt auf die Exploration konkreter Kriterien hingearbeitet wird. Exploration soll an dieser Stelle sowohl verstanden werden als das Herausarbeiten von Qualitätskriterien aus der vorhandenen Literatur, als auch nötigenfalls deren Modifikation oder gar der Generierung gänzlich neuer Elemente.

Das Unterfangen, den aktuellen Forschungsstand im Feld der Qualitätsbestimmung qualitativer Sozialforschung aufzuarbeiten, fördert rasch eine unübersichtliche Vielfalt an Perspektiven und Standpunkten zu Tage. Eine Entwirrung soll möglichst erreicht werden, durch eine letztendliche Statuierung eines Kataloges von Qualitätskriterien, welcher zugleich das Endergebnis und den Abschluss des theoretischen Teils darstellt. Ein solcher kann verstanden werden als eine Art „Checkliste“ (Blaxter 2000:35) und findet Verwendung als Schablone für die im Weiteren vorgenommene empirische Analyse.¹ Den Untersuchungsgegenstand der Analyse bilden Veröffentlichungen, die zunächst einmal zwei notwendige Bedingungen erfüllen müssen: Zum einen müssen sie im engeren Sinne

¹ Eine differenziertere Charakterisierung der Eigenschaften und des Stellenwertes des Kataloges wird im 2. Kapitel vorgenommen, welches sich explizit dessen Generierung zuwendet. Die Festlegung auf jeweilige Charakteristika ist dabei von größtem Folgenreichtum, ist sie doch stets mit Implikationen belegt. So wird beispielsweise ein qualitatives Vorgehen bisweilen als unvereinbar mit starr festgelegten Durchführungsregeln bezeichnet. Dass eine solche Unvereinbarkeit nicht zwangsläufig in einer Sackgasse enden muss und was ein Kriteriumskatalog in seiner Anwendung für Chancen bietet, zeigt eben dieses Kapitel.

empirischen Inhalts sein und zum anderen muss der empirische Zugang in noch zu definierender Weise als qualitativ bezeichnet werden können. Mit anderen Worten wird der erarbeitete Qualitätsrahmen als Maßstab auf jene Art durchgeführter Forschungen angelegt für die er genuin entwickelt wurde. Das eigentlich zu Analysierende, bietet sich jedoch lediglich in einer mittelbaren Form dar, nämlich durch die Veröffentlichung des Projektes als Publikation.² Im Rahmen dieser Arbeit gilt bei der Bestimmung der Grundgesamtheit eine weitere Bedingung, sozusagen neben den obigen beiden notwendigen Bedingungen nun die hinreichende Bedingung. Die Veröffentlichung muss nämlich als Zeitschriftenartikel erschienen sein. Diesem Kriterium liegen keinerlei inhaltliche Überlegungen zu Grunde, vielmehr birgt es praktische bzw. pragmatische Vorteile. Zeitschriftenartikel sind ungleich kürzer als Monographien, sind aber dennoch darauf angewiesen das Essenzielle zu liefern. Sie kommen somit als Beurteilungsobjekt in Frage und bieten zudem die Möglichkeit innerhalb der kontingentierte Zeit eine höhere Zahl an Forschungsprojekten, anhand deren Veröffentlichung, mit in die Evaluation einzubeziehen. Ferner kann das Spektrum verschiedener Richtungen innerhalb der qualitativen Sozialforschung möglichst breit gehalten und das Ziel, eine vergleichende Analyse der zwei im Titel auftauchenden Disziplinen durchzuführen aufrecht erhalten werden. Nicht die Tatsache, wie oft ein qualitatives Vorgehen innerhalb einer Disziplin gewählt wird steht dabei im Blickpunkt, sondern die Güte der Umsetzung. Lassen sich Verschiedenheiten hinsichtlich der methodischen Qualität feststellen? Lassen sich gar Trends eruieren, Stärken und Schwächen aktueller Umsetzung herausarbeiten?

Da an eine Vollerhebung nicht zu denken ist, auf Grund der großen Fülle an Zeitschriften, ist eine Stichprobe und deren Ziehung zu definieren³.

Die Studie ist darauf ausgerichtet ausschließlich explorativ vorzugehen. Eine Prüfung von Hypothesen wird nicht vorgenommen und wäre angesichts des

² Von Ergebnissen eines Forschungsprojektes zu sprechen wäre unangebracht, denn wie gezeigt werden wird, kann in der qualitativen Sozialforschung erst dann etwas bezüglich der Qualität beurteilt werden, wenn die Darstellung eines Projektes in Form einer Veröffentlichung weit über die Darreichung von Ergebnissen hinausgeht. Dies ist nichts ausschließlich einem qualitativen Vorgehen eigenes, vielmehr sind Äquivalente in quantifizierenden Untersuchungen zu finden. Zum Beispiel dann, wenn die methodische Durchführung erläutert oder die zu prüfenden Hypothesen hinsichtlich ihrer von den Forschenden zugeschriebenen Relevanz erläutert werden. Dennoch kann behauptet werden, dass in der qualitativen Sozialforschung der Anteil der Nicht-Ergebnisdarstellung einen weitaus höheren Stellenwert besitzt, wenn eine adäquate Darstellung eines Forschungsprojektes erreicht werden soll.

³ Eine detaillierte Beschreibung befindet sich in Kapitel 4.1.

Forschungsinteresses aus drei Gründen nicht sinnvoll. Erstens müsste eine Verengung des Fokus auf Teilaspekte vorgenommen werden, die einem Falsifikationsversuch unterworfen werden müssten, andere Aspekte blieben ausgeklammert. Dies führt geradewegs zum zweiten Grund, denn warum sollten welche Aspekte ausgewählt werden? Das vorliegende Forschungsfeld liegt noch unter einem Schleier, ist bis dato nicht bearbeitet worden. Wenn aber ein Feld eher unbekannt ist, dann ist eine explorative Analyse in jedem Fall angemessener, als die Prüfung von Hypothesen, zumal diese nicht immanent abgeleitet sein können. Eher entsprängen sie den Köpfen der Forschenden unter der Maßgabe übergestülpter Theorien aus anderen bereits erschlossenen Bereichen. Einer Prädetermination der Erkenntnisse, die unter allen Umständen vermieden bzw. minimiert werden muss, würde auf diesem Wege Vorschub geleistet. Der dritte und letzte Grund ist die rasche Geschwindigkeit mit der sich die qualitativen Methoden und ihre Anwendung entwickeln. Vielmehr als quantitative Methoden befinden diese sich in einem noch entwickelnden Feld. Die Art und Weise der Handhabung qualitativer Methoden hat sich beispielsweise schon innerhalb des relativ kurzen Zeitraumes seit den 1980er Jahren stark gewandelt, in jüngerer Zeit vor allem in die Richtung eines verstärkten Bemühens um kanonisiertes Vorgehen (Flick/ von Kardoff/ Steinke 2000:27). In diesen Bereich fällt auch das Bemühen um „ensuring rigour“ (Seale/ Silverman 1997), also die Sicherstellung von Strenge in der qualitativen Sozialforschung und die Konstituierung von Qualitätskatalogen.⁴ Auch für die mittelfristige Zukunft prognostiziert Flick einen weiteren starken Wandel. (vgl. 1999:23) Ist ein Forschungsfeld aber einem starken Wandel unterworfen, haben Hypothesen eine nur unterdurchschnittliche Lebensdauer. Das heißt, der nach Popper ohnehin nur vorläufige Charakter empirischer Aussagen würde weiter gesteigert, weil ständig eine Falsifizierung von Aussagen erreicht werden könnte. Der Grund liegt im Gegenstand, der praktisch unter der Theorie hinweg mutiert und diese somit zur Beschreibung eines empirischen Relatives nicht mehr greifen kann. Aus diesen Gründen wird in dieser Untersuchung ein exploratives Vorgehen präferiert und deskriptives Niveau kaum überschritten.

⁴ Eine nähere Beschreibung dieser Tendenz befindet sich in Kapitel 2.1

Was ist jedoch das Ziel dieser Analyse und welche Möglichkeiten bietet die Vorgehensweise?

Die Zielsetzung ist keine einseitige und soll in drei Punkten erläutert werden.

1. Zunächst einmal soll der Nebel gelichtet werden, der die Frage nach Qualität in qualitativen Forschungsprojekten umgibt. Es soll eruiert werden, wo die in der scientific-community existierenden Regeln⁵ beachtet werden und an welchen Stellen nicht. Werden also die methodischen Standards beachtet und erfüllt, oder ist der Umgang mit ihnen eher leger? Gerade den Nicht-quantifizierenden Methoden wurde in der Vergangenheit häufiger vorgeworfen, sie würden bisweilen eine methodische Willkür walten lassen. Wie steht es aktuell mit diesem Vorwurf, ist heute eine stärkere Regelgeleitetheit bzw. Standardisierung erreicht? Möglicherweise lassen sich Häufungen finden bezüglich der Stärken und Schwächen, die Hinweise auf die Beschaffenheit des Diskurses innerhalb der Forschungsgemeinschaft geben. Möglicherweise lassen sich interdisziplinäre Unterschiede in der Handhabung der qualitativen Methodik feststellen – in einem solchen Falle wären diese genauestens unter die Lupe zu nehmen.
2. Nicht zuletzt geht es auch darum, eine Momentaufnahme über die momentane Forschungspraxis zu gewinnen. Nicht nur, dass der Standort qualitativer Sozialforschung hinsichtlich der bereits beschriebenen Dimensionen bestimmt werden kann, ferner lassen sich Schlüsse daraus ziehen, in welchem Verhältnis Theorie und Praxis von Forschung zueinander stehen. Sind Vorgabe und Anwendung kongruent, so entspräche es natürlich dem Idealfall; genau dann wäre es den Theoriegenerierenden gelungen, die jeweiligen Konzepte an die Praktizierenden heranzutragen, die Wichtigkeit ihrer Beachtung klar zu machen. Klafft aber eine Lücke zwischen Vorgabe und Anwendung, dann kann dies einerseits darauf beruhen, dass die Anwenderinnen und Anwender einen mangelnden Kenntnisstand aufweisen, weil sie beispielsweise nicht mit der aktuellen Literatur vertraut sind. Andererseits aber auch darauf, dass der Kenntnisstand zwar ausreichend wäre für eine adäquate Methodenanwendung, andere Gründe aber dagegen

⁵ Selbstverständlich sind diese Regeln ständig Gegenstand eines Disputs und so gut wie nie unumstritten. So gut wie immer aber existiert gleichzeitig eine Konvention über das, was als approbiert zu gelten hat. Diese Konvention kann von einem mehr oder weniger großen Anteil Personen getragen werden und ist an dieser Stelle gemeint.

sprechen – zu nennen wären hier die Nichtdurchführbarkeit der Standards oder die Möglichkeit, innovativ zu verfahren. Nichtdurchführbarkeit wird beispielsweise dadurch hervorgerufen, dass qualitative Sozialforschung sich mitunter weitaus umfangreicher gestaltet als ihr quantitatives Pendant. Es ist dann mehr Personal und Zeit nötig – beides lässt häufig die ökonomische Ausstattung nicht zu. Ein Bedingungskreislauf beginnt: Mangelnde ökonomische Ausstattung führt zu mangelhaft eingehaltenen Standards, welche wiederum die Skepsis der Geldgeber gegenüber der Methodik erhöhen und weitere Forschungsanträge in einem kritischen Licht erscheinen lassen. Beides wäre im Rahmen der Veröffentlichung explizit darzulegen, zu begründen und auf seine Berechtigung hin zu prüfen.

3. Ein weiteres Ziel dieser Arbeit, das sozusagen en passant bearbeitet werden kann, ist es, den Stand der Methodendiskussion aus dem anglophonen Raum in den deutschsprachigen Raum zu übertragen⁶. Kelle sieht die deutsche Methodendiskussion von der internationalen Entwicklung abgekoppelt verlaufen, dabei nicht unbedingt schlechter dastehend. (Kelle 2000:588) Der begrenzte Austausch zwischen beiden Bereichen habe aber eine verstärkte Differenzierung qualitativer Forschung zur Folge, zumal sich die Diskussionen um unterschiedliche Themen und Methoden drehen würden (Flick 2001:54) Bezüglich der Debatte um Gütekriterien qualitativer Forschung und speziell um Validität, sieht er die Impulse im Moment und für die nächsten Jahre doch eindeutig aus der Debatte des englischen Sprachraumes kommen. (Kelle 2000:588) Die einzige deutschsprachige Monographie aus diesem Bereich von Steinke (1999) stützt sich zwar zum Teil auf anglophone Quellen, spart aber jüngste Tendenzen aus.⁷ Generell unterscheiden sich die Diskurse in ihrer Ausrichtung. Während der anglophone sehr stark durch den Pragmatismus geprägt wird, ist auf der deutschsprachigen Seite eher eine philosophische und/oder epistemologische Auseinandersetzung zu finden. Folge: Zum Teil zerläuft dann die

⁶ Diese Vorgabe ist im Rahmen dieser Arbeit natürlich vermessen, stellt aber thematisch eine derartig wichtige Aufgabe dar, dass sie zumindest als ideelles Ziel nicht außer Acht gelassen werden darf.

⁷ Diese Tatsache sollte nicht verwundern, denn die Veröffentlichung stellt eine nahezu identische Neuauflage ihrer Dissertation (Steinke 1998) dar, die sie Mitte der 1990-er Jahre begonnen haben muss.

Methodenforschung in winzige philosophische Teilprobleme, deren Berechtigung nicht bestritten werden soll, die aber viel Potenzial binden. Innerhalb des pragmatischen Diskurses kann eine erfrischende Unbefangenheit im Umgang mit Methoden erlebt werden, die allerdings bisweilen ein wenig sorglos anmutet, was ihre Verankerung im philosophischen Fundament angeht. Beide Ausrichtungen sollen hier zum Zwecke der gegenseitigen Befruchtung miteinander verwoben werden, jedoch wird der Standpunkt vertreten, die empirische Sozialforschung letztens als untrennbar von ihrer Praxis anzusehen – auch im Hinblick auf ihre Bewertung.

Wie diese Ziele zu erreichen sind und welche Auswertungen vorgenommen werden, dem widmet sich der folgende Abschnitt. Er geht dabei genauer auf die Methodik der eigenen Untersuchung ein.

1.2 Einwurf: Metaanalyse und Auswertung

Die bisher skizzierte Konstellation, eine Forschung über Forschung durchzuführen entspricht einer Meta-Sichtweise, aus welcher praktisch heliospektiv von oben auf den Gegenstand herabgeblickt werden kann. Der Gegenstand qualitative Sozialforschung bietet sich dem Auge aus einiger Entfernung mit großer Übersichtlichkeit. Bleibt nur noch die Wahl des Betrachtungsinstrumentes, um auch Details wahrnehmen zu können. Diese Metapher steht in der hiesigen Metaanalyse für den zu entwickelnden Bewertungskatalog als erkenntnisleitendes Instrument.

Das Verfahren der Metaanalyse erfreut sich momentan einer steigenden Beliebtheit in den Wissenschaften. In den letzten Jahrzehnten hat eine enorme Ausweitung der Forschungskapazitäten eingesetzt, was unter anderem dazu geführt hat, dass sogenannte Replikations- und Metastudien in den empirisch arbeitenden Wissenschaften vermehrt durchgeführt werden. Nicht nur weil plötzlich Gelder für derartige Studien frei wären, sondern auch weil eine Flut von bereits erhobenem Datenmaterial existiert, das für einen weiteren Umgang zur Verfügung steht. Der Grundgedanke dieser Methode ist die Erstellung eines möglichst objektiven Forschungsüberblicks und einer Forschungsbilanz. (vgl. Drinkmann 1990:2) So weißt Farin darauf hin, dass trotz des scheinbar hierarchischen Aufbaus des

Methodenprinzips (siehe *Abbildung 1*) die Wirkung einer übergeordneten Analyse in zwei Richtungen verläuft: sowohl „bottom up“ – Daten werden zu Metaanalysen integriert, als auch „top-down“ – metaanalytische Resultate dienen der Identifizierung von Forschungslücken und motivieren neue Primärstudien. (1997:163) *Abbildung 1* verdeutlicht das Prinzip der Metaanalyse. Denkbar wäre auch eine Metaanalyse von schon vorhandenen Metaanalysen – eine Analyse 2.Ordnung. Im Falle dieser Arbeit wird die Ebene 1.Ordnung allerdings nicht überschritten.

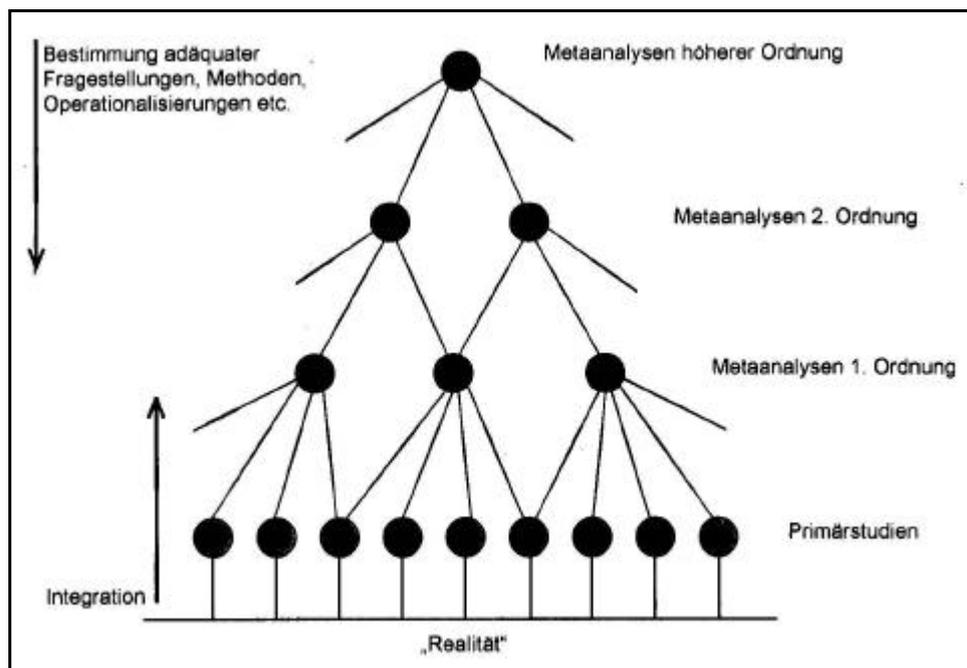


Abbildung 1: Struktur des Forschungsprozesses metaanalytischer Forschung
(Quelle: Farin:163)

Drinkmann definiert folgendermaßen:

„Metaanalyse soll sein: eine an den Kriterien empirischer Forschung orientierte Methode zur quantitativen Integration der Ergebnisse empirischer Untersuchungen sowie zur Analyse der Variabilität dieser Ergebnisse.“
(1990:10)

Insbesondere die Psychologie und die Medizin stellen die Hauptanwendergruppe dar und haben die Entwicklung dieser spezifischen Methodik in diesem Sinne forciert. (vgl. Farin 1997:161f.) Jedoch haben sich parallel mannigfache Ansätze entwickelt, auch qualitative Analysen. Anhand dieser Definition wird deutlich wie die Analyse zumeist eher betrieben wird, nämlich in der Art einer Zusammenfassung

von Ergebnissen. Die vorliegende Arbeit geht also nicht ganz im Drinkmannschen Sinne vor und integriert zwar empirische Ergebnisse dadurch, dass sie deren Hintergründe betrachtet, sie fasst auch gewisse Eigenschaften zusammen, nicht aber die Ergebnisse als solche. Somit liegt sie völlig in dem von der Definition festgelegten Bereich, geringfügige Eigenheiten bleiben ihr aber dennoch vorbehalten.

Welche Auswertungsverfahren werden verwendet? Und womit werden sie durchgeführt?

Der erste Schritt liegt in der Beurteilung mittels des Kriteriumskatalogs. Im strengen Sinne wird die Qualität gemessen, denn messen bedeutet „die Zuordnung erhobener Merkmalsausprägungen zu bestimmten Symbolen oder Codes (meist: Ziffern oder Zahlenwerte).“ (Merten 1995:95) Die Eruiierung der jeweiligen Umsetzung der Qualitätskriterien in eine numerische Codierung, im Sinne von ‚eins gleich gut umgesetzt‘ und ‚fünf gleich weniger gut umgesetzt‘ würde der Definition entsprechen. Das Skalenniveau erreicht in diesem Fall ordinales Niveau. Qualitative Messung kann jedoch besonders bei vorliegender Vorgehensweise nicht den strengsten Anforderungen entsprechen, besonders weil lediglich ein Bewertender die Zuordnung durchführt. Dies ist aber durch die formalen Vorgaben dieser Arbeit nicht zu ändern und soll nicht weiter stören. Die qualitative Messung beginnt mit der Bildung von relevanten Merkmalsdimensionen (vgl. ebd.:98), diese entsprechen den einzelnen Gütekriterien. In einem zweiten Schritt kann mit der Ordnung der empirisch feststellbaren Merkmalszusammenhänge begonnen werden, der Zuordnung eines numerischen Relativs zu einem empirischen.

Bei der Auswertung der Daten, die bei der Bewertung anfallen, werden unterstützend zwei technische Hilfen hinzugezogen. Zum einen das Statistikprogramm *SPSS*⁸ und zum anderen das Textanalysesystem *WinMAX*⁹. *SPSS* kommt dann zum Zuge, wenn es um Verfahren der deskriptiven und inferierenden Statistik geht und die Beurteilung numerisch gestuft vercodiert wurde. Das Programm *WinMAX* dagegen bietet die Möglichkeit inhaltlich differenziertere Beurteilungsschemata anzuwenden, indem ein Kategoriensystem entwickelt wird.

⁸ SPSS – Statistical Products and Service Solutions, Version 10.0 [Vgl. Büh/ Zöfel (2000) und Norusis (1994)].

⁹ WinMax, professionelle Version 2000 [Vgl. Kuckartz 1996 und 1998].

(Kuckartz 1999:3.Kapitel) Ein solcher kann dann durch entsprechende Codierung fein ausdifferenziert werden und zahlreichen Auswertungen unterzogen werden. Es hat während des Umgangs mit WinMax zunächst offenen Charakter (Merten 1995:99) und wird induktiv anhand des Materials ausgebildet. Kuckartz beschreibt Codieren als „die Zuordnung zu Kategorien zu relevanten Textpassagen bzw. die Klassifikation von Textmerkmalen“ (1999:75). In diesem Sinne stellt die Qualität der zu untersuchenden Forschung ein Textmerkmal dar. Es ist nicht nötig, die Texte resp. Publikationen in den Rechner einzulesen, denn die Codierung erfolgt nicht gebunden an Textstellen, sondern ist immer bezogen auf den Gesamttext als Charakteristikum zu verstehen.

Durch eine Zusammenschau sämtlicher Operationen beider Programme ergibt sich schließlich ein interpretierbares Gesamtbild.

Auch angesichts der bereits besprochenen Tatsache, dass es sich um keine konfirmatorische Metaanalyse, sondern um eine explorative handelt, werden die angewandten Verfahren ein höheres Komplexitätsniveau kaum erreichen, deskriptive Analysen stehen im Mittelpunkt.¹⁰

1.3 Kennzeichnung qualitativer Sozialforschung

Zur letztgültigen Festlegung dessen, welchen Bereich die Grundgesamtheit umfasst, welche Veröffentlichungen für die spätere Analyse also in Frage kommen, ist es nötig den Begriff der qualitativen Sozialforschung zu definieren. Wie Steinke allerdings bemerkt, existiert in der Literatur keine feste, allgemeingültige oder verbindliche Bestimmung dessen, was unter qualitativer Forschung zu verstehen ist. (1999:15) Somit kann eine Definition im klassischen Sinne gar nicht geleistet werden, sie ist in diesem Fall eher verwandt mit der Wesensdefinition als einer Unterart der Realdefinitionen. Funktional gesehen wird lediglich eine Festlegung von Eigenschaften als Arbeitsgrundlage benötigt und im Folgenden erarbeitet. Die Begriffsdefinition muss möglichst an der jüngsten Literatur angelehnt sein, denn das Verständnis von dem, was qualitative Sozialforschung eigentlich ist, blieb nicht

¹⁰ Eine genaue Beschreibung befindet sich im 4. Kapitel.

immer gleich. Zumal die zu untersuchende Literatur die allerneueste ist, können auch neueste Definitionen verwendet werden. Das Problem, rückwirkend Qualitätskriterien überzustülpen, die zum Zeitpunkt der jeweiligen Projektdurchführung noch nicht existent waren, stellt sich deswegen nicht.

Folgende Begriffsarbeit soll unter keinen Umständen den Eindruck erwecken, es gäbe eine klare Trennung qualitativer und quantitativer Methoden im Sinne einer Dichotomie. Die Literatur ist sich relativ einig, dass eine klare Zweiteilung nicht mehr aufrecht zu erhalten ist. Von Saldern weist in einem eindringlichen Plädoyer eindrucksvoll die Unhaltbarkeit der Dichotomietheorie nach und bezeichnet „Propagandisten dieser unseligen Dichotomisierung als ‚Verpackungskünstler‘: Eine geheimnisvolle Terminologie verstellt den Blick auf die wahren Verhältnisse“ (1995:340).¹¹ Nicht alle Arbeiten ließen sich der einen oder der anderen Richtung zuordnen (ebd.:331). Entsprechend wird heute bei der Charakterisierung der zahlreichen heterogenen Ansätze, die allesamt unter dem selben Namen firmieren, damit begonnen, gemeinsame Grundannahmen und Kennzeichen festzuhalten, sozusagen den „kleinsten gemeinsamen Nenner“ (von Kardorff 1991:4) zu bestimmen. Eine sehr grundlegende und verbreitete Beschreibung, die hier Verwendung finden soll, ist diejenige von Mayring. Er kennzeichnet die Gemeinsamkeiten, indem er fünf Postulate aufstellt, die erfüllt sein müssen:

1. *Subjektbezogenheit* der Forschung
 2. Betonung der *Deskription* und
 3. der *Interpretation* der Forschungssubjekte
 4. Untersuchung in der *alltäglichen Umgebung*
 5. Auffassung vom *Verallgemeinerungsprozess* als spezifisch begründbar
- (vgl. Mayring 1996:9)

Eine Forschung, die sich an diesen fünf Postulaten orientiert, soll im folgenden qualitative Sozialforschung genannt werden, wenn sie zusätzlich den folgenden

¹¹ Von Saldern spielt an auf das vielverkaufte Handbuch von Lamnek (1988), das eine Dichotomisierung postuliert und große Wirkung seit seinem Erscheinen erzielt hat. Es wird heute als qualitatives Methodenhandbuch der ersten Generation bezeichnet.

Kennzeichen qualitativer Forschungspraxis von Flick, von Kardorff und Steinke entspricht:¹²

1. Gegenstandsangemessenheit der gewählten Methode
2. Orientierung am Alltagsgeschehen und/oder Alltagswissen
3. Kontextualität als Leitgedanke
4. Perspektiven der Beteiligten
5. Reflexivität des Forschers
6. Verstehen als Erkenntnisprinzip
7. Prinzip der Offenheit
8. Fallanalyse als Ausgangspunkt
9. Entdeckung und Theoriebildung als Ziel

(nach Flick/ von Kardorff/ Steinke 2000:24)

An dieser Stelle kann nicht expliziert werden, welche inhaltliche Bedeutung den einzelnen Begriffen zu Grunde liegt, darüber geben die jeweiligen Textstellen Aufschluss, deren Inhalt vorausgesetzt wird. Einzelne Forschungsperspektiven betonen mehr oder auch weniger die verschiedenen Aspekte, stehen diesen aber per definitionem nie diametral gegenüber. Bewusst ist nicht von einem alles entscheidenden Niveau der Informationsreduktion als ausschlaggebendes Faktum die Rede. Zahlen auf der einen Seite und Wörter auf der anderen, sind nicht zwingend Indikator für ein Paradigma, eine Grenzziehung fällt der fließenden Übergänge wegen schwer.

Auf Grund der meist weit auseinander liegenden erkenntnistheoretischen Grundlagen entzog sich manches Mal die qualitative Vorgehensweise im Gegensatz zur quantitativen einer externen Bewertung. Dass dies nicht unbedingt im Wesen der Forschung liegt und wie der Umgang mit Gütekriterien aussieht, soll im Folgenden dargestellt werden.

¹² Es handelt sich hierbei um gegenüber der Quelle leicht modifizierte Kennzeichen qualitativer Forschung. Die ursprünglich allgemein-methodisch gehaltene Kennzeichnung musste auf die an späterer Stelle benötigte Beschreibung einzelner Forschungsarbeiten angepasst werden.

2. Die Qualität qualitativer Sozialforschung als Geltungsbegründung mittels Gütekriterien

Wenn bisher von der Qualität der Forschung die Rede war, dann war das Wesen dieser Qualität noch sehr unklar. Fest steht bisher, dass sie messbar ist anhand der zu entwickelnden Gütekriterien. Jene, den Wert der jeweiligen Publikation bemessenden Kriterien sind zu verstehen als Determinanten ihrer Geltung. Mit anderen Worten ist die Geltung, im Sinne einer allgemeinen Anerkennung, maßgeblich davon abhängig inwieweit die Anforderungen einzuhaltender Qualitätsstandards, repräsentiert durch die Gütekriterien, erfüllt worden sind. Alle Bemühungen, die in die Richtung gehen, die Qualität qualitativer Sozialforschung zu steigern bzw. zu sichern, können somit als Versuch der Geltungsbegründung apostrophiert werden.

Der Begriff der Geltungsbegründung umfasst verschiedene Ansätze der sogenannten Methodisierung des Verstehens als Erkenntnisprinzip. (Flick 1999:259) Dabei werden Kriterien formuliert, die der Beurteilung der verwendeten Verfahren und ihrer Angemessenheit dienen. Die Geltungsbegründung ist folglich zu verstehen als eine Analyse des Forschungsprozesses und nicht ausschließlich der Ergebnisse. Dass diese Problematik nicht undiskutiert geblieben ist, soll an dieser Stelle nicht verschwiegen werden. So wird im Folgenden zunächst die Notwendigkeit der Generierung von Qualitätskriterien erläutert, um dann auf die Kernproblematik einzugehen. Diese stellt sich entsprechend dem kontroversen Diskussionsstand der scientific-community dreigliedrig dar:

- i) Jede Form der Bewertung von qualitativer Forschungspraxis anhand allgemeiner Kriterien oder Standards wird streng zurückgewiesen.
- ii) Bewertungskriterien, Gütekriterien und Standards werden aus den quantitativen Methoden übertragen und gegebenenfalls leicht redefiniert resp. angepasst.
- iii) Es werden eigene, „methodenangemessene Gütekriterien“ (Flick 1987) speziell für die qualitative Sozialforschung formuliert.

Jeder der drei Punkte wird gesondert abgehandelt und die Hauptlinien der gängigen Auseinandersetzungen nachgezeichnet werden. Besonderes Augenmerk gilt den

Implikationen, die aus den im 3. Kapitel vorgenommen erkenntnistheoretischen Betrachtungen resultieren, im Hinblick auf die Auswahl spezieller Kriterien.

2.1 Die Notwendigkeit von Qualitätskriterien

Warum ist es gerade für die qualitative Sozialforschung so wichtig, Gütekriterien zu formulieren?

Die gesamte hier zum Gegenstand gemachte Unternehmung einiger Forscherinnen und Forscher kann summarisch als „ensuring rigour in qualitative Research“ in Anlehnung an eine Aufsatzüberschrift von Seale und Silverman (1997) bezeichnet werden. Die beiden Autoren waren nicht die ersten, die sich mit der Thematik befasst haben, ihre Titelwahl ist jedoch auf ganz besondere Weise bezeichnend.¹³ Nicht nur, dass der anglophone Titel die Fortgeschrittenheit des Diskurses jenes Sprachraumes markiert, er umreißt ferner treffend das angestrebte Unterfangen: Die Sicherstellung von Strenge. Der programmatische Aufsatz mündet in dem Appell, der Qualität qualitativer Forschung höchste Aufmerksamkeit beizumessen, damit deren Inhalte überhaupt erst beurteilbar gemacht werden könnten. (Seale/Silverman 1997:383) Die Problematik von Qualität und Beurteilbarkeit ist allerdings kein Spezifikum von qualitativer Forschung, auch unter quantitativ orientiert Vorgehenden herrscht keineswegs Einstimmigkeit diesbezüglich. (vgl. Hammersley 1990) Die Debatte wird jedoch ungleich leiser geführt und ist weniger heterogen in ihren Extremstandpunkten. Sie zeigt aber, wie ähnlich der Verlauf beider Diskussionen bisher abgelaufen ist, mit der Ausnahme, dass die eine Teildiskussion der anderen um einige Jahrzehnte voraus ist.

Kelle, Kluge und Prein beschreiben, welches vielseitige Misstrauen für gewöhnlich bei Nicht-Beurteilbarkeit von qualitativen Verfahren auftritt. Qualitative Forschungsergebnisse würden als unzulässig betrachtet, weil sie zu sehr der subjektiven Willkür der einzelnen Forschenden unterworfen seien, die Methoden der Analyse

¹³ Seale legte mit diesem Artikel seinen ersten Schritt auf diesem Gebiet zurück, weitere sollten folgen. Thematisch erreicht er noch nicht das ausgesprochen elaborierte Niveau der späteren Arbeiten, sondern setzt sich vorwiegend mit einzelnen Techniken der Standardisierung auseinander, wie Triangulation, Quasi-Statistiken und Computereinsatz. Von speziellen Gütekriterien geschweige denn von einem Kriteriumskatalog ist noch nicht die Rede.

seien zu informell, um verschiedenen Untersuchenden übereinstimmende Urteile und Interpretationen zu erlauben. Ferner seien Schlussfolgerungen in das Belieben der Interpretierenden gestellt (vgl. 1993:5) und die Forschungsergebnisse auf Grund der nicht-geregelten Samplingverfahren nicht generalisierbar. (ebd.:8)

Folge: der qualitativen Sozialforschung wurde bisweilen die „Wissenschaftswürdigkeit“ (Kelle 1994:12) abgesprochen. Abzulesen sei dies daran, „dass qualitative Methoden in den üblichen Hand- und Lehrbüchern in der Regel cursorisch auf nur wenigen Seiten abgehandelt werden“ (ebd.:13) und dabei die mitgeteilten „Urteile in der Regel nicht mehr begründet werden, sondern sich teilweise zu einer Art orthodoxen Lehrmeinung verfestigt haben.“ (Kelle/ Kluge/ Prein 1993:6) Kritik beschränke sich dabei zumeist auf wenige Sätze, in denen Mängel nur noch apodiktisch festgestellt, aber nicht mehr systematisch begründet würden. (ebd.:6f.) Derzeit sind im deutschen Sprachraum „eine Reihe von Aktivitäten zu verzeichnen, die auf eine stärkere Etablierung des Feldes der qualitativen Forschung innerhalb der Soziologie und ähnlich innerhalb der Erziehungswissenschaft hindeuten“. (Flick 2001:59) Kürzlich richtete sich gar in der *Deutschen Gesellschaft für Soziologie* innerhalb der Methodensektion eine Arbeitsgruppe „Methoden der qualitativen Sozialforschung“ ein, unter anderem mit der Begründung, qualitative Methoden besäßen in der Sektion einen niedrigeren Stellenwert als quantitative Methoden und würden nicht ausreichend diskutiert. (Hopf et al. 1998) In der Erziehungswissenschaft lässt sich die Tendenz am verstärkten Literatúrausstoß ablesen. Neben dem disziplinspezifischen Handbuch von Friebertshäuser und Prengel (1997) erscheint unter anderem neuerlich eine Zeitschrift für qualitative Bildungs-, Beratungs- und Sozialforschung (ZBBS).

Der Ausweg aus der bis heute eingenommenen Lage soll aktuell durch die Entwicklung von Gütekriterien aufgezeigt werden. Nicht zufällig kamen wichtige Impulse dazu aus dem Bereich der Medizin, sowie der medizinischen Soziologie. (vgl. Murphy 1998) Gerade in diesem Feld ist es nämlich ungeheuer wichtig, um nicht zu sagen lebenswichtig, die Qualität qualitativer Studien, deren Existenzberechtigung per se dort weniger umstritten war als anderswo, sicherzustellen. (vgl. Blaxter 2000) Der Begriff ‚Lebenswichtig‘ drückt die Problematik dieser Disziplin aus, so sind zum Beispiel in der medizinischen Statistik meist die Signifikanzgrenzen höher als in den Sozialwissenschaften. Äquivalent dazu müssen auch die qualitativen Ergebnisse einer strengsten Prüfung unterzogen werden,

bevor sie möglicherweise in der Praxis auf den Menschen angewandt werden, vielleicht in Form eines neuen Medikamentes. Blaxter bezeichnet die Möglichkeit des Beurteilens als Grundvoraussetzung, um überhaupt erst qualitative Forschungsprojekte veröffentlichen zu können. (ebd.:34) Steinke sieht in Bewertungskriterien ein Mittel, nach außen hin zur Etablierung und Legitimierung qualitativer Forschung, ihrer Vorgehensweisen und Ergebnisse beizutragen. (1999:11) Von der Definition angemessener Kriterien hänge die weitere Etablierung in der Gesamtlandschaft gar ab. (Steinke 2000:319) Die Zuschreibungen von positiven Auswirkungen auf die Wahrnehmung qualitativer Forschung, die in der Literatur zu finden sind, sind vielseitig und Ausdruck der Hoffnung sich von der Stellung der Vergangenheit zu lösen. Seale nennt weitere Gründe für „den Drang zu Bewertungskriterien“ (1999b:467). Ein Nebeneffekt der bereits genannten wäre unzweifelhaft die Durchsetzung gegen Kritik, insbesondere auch gegenüber sogenannten „skeptical audiences“ (ebd.) und der keineswegs zu unterschätzende Bonitätsgewinn gegenüber geldgebenden Institutionen, den „research-funding bodies“. (ebd.) Entsprechend der These „Die Forschungsförderung (re-)produziert das Bild der Forschungspraxis“ (Lamnek 2000:34) stellt es für Forschungsanträge, die Projekte mittels qualitativer Forschungsansätze durchzuführen beabsichtigen, häufig ein schwieriges Procedere dar, Geldmittel zu sichern. Lamnek zeigt den Mechanismus exemplarisch für die *Deutsche Forschungsgemeinschaft* (DFG) auf. So erfolge die Bestimmung der Gutachter in Orientierung am jeweils herrschenden Mainstream (ebd.), mit der Auswirkung, dass qualitativ-empirische Forschung ihrer geringen Wertschätzung wegen benachteiligt werde, allerdings nicht grundsätzlich und nicht immer. (ebd.:36)

Für alle diese Problembereiche kann ein Unternehmen wie „ensuring rigour“ wichtige Fortschritte hinsichtlich der Zukunftsfähigkeit qualitativer Forschung mit sich führen. Das Herzstück dieses Bemühens besteht in der Vielzahl einzelner konkreter Kriterien zur Bestimmung der Qualität. Um letztere geht es auf den kommenden Seiten.

2.2 Vorbehalte gegenüber Qualitätskriterien

Die Problemlösung könnte so einfach sein: zunächst Qualitätskriterien herausarbeiten, die verbindlich sind und im Weiteren diese bei der Bewertung in Anschlag bringen. Leider kann der Weg auf diese einfache Weise nicht beschritten werden. Innerhalb der qualitative-researcher-community herrschte nämlich bislang Unklarheit darüber, ob solche Gütekriterien überhaupt aufstellbar sind oder ob der Gegenstand der Beurteilung einer ebensolchen nicht grundsätzlich entgegensteht. Jene, die sich eher gegen den Einsatz von Kriterien aussprechen, rekurrieren dabei auf hauptsächlich drei epistemologische Standpunkte und einen methodenimmanenten Einwand. Eine radikale Position nimmt innerhalb dieses Spektrums Feyerabend ein.¹⁴ Seine Kritik an der Verwendung von Methoden generell, mündet in der Formel „Anything goes!“ (Feyerabend 1976:45) Darunter versteht er nicht, nach gut Dünken vorzugehen, so wie es den Forschenden gerade in den Sinn kommt. Vielmehr fordert Feyerabend dazu auf, die Zwänge des Herkömmlichen abzulegen und sich jede denkbare Alternative sowie entgegenstehende Bestimmung nutzbar zu machen. Ein derartiges Postulat, rein unter dem Methoden-Gesichtspunkt betrachtet, scheint zunächst kaum mit dem Anspruch universeller Gütekriterien vereinbar zu sein. Bedenken anderer Art kommen aus dem Lager der radikalen Konstruktivisten. Sie behaupten, dass nicht nur das Wissen über die Welt, sondern die Welt selbst kognitiv konstruiert ist. Somit kann objektives Wissen über die Welt nicht produziert werden, jede Beobachtung wird geprägt durch bereits vorhandene Einstellungen oder Theorien. Das bedeutet für die Gütekriterien ein Verlust ihrer Basis, denn eine approximativ objektive Bewertung schließt die konstruktivistische Sichtweise aus, wenn sie auf deren Konstruierbarkeit in Abhängigkeit von vordem vorhandenen Wissensbeständen verweist. Daher könnte ein Qualitätsmaßstab nie eine breitere Legitimation erhalten und bliebe immer im Dunstkreis der Beliebigkeit verhaftet. Für die empirische Sozialforschung bearbeitete diese Thematik am ausführlichsten Smith (1984), der aus obigem Grund

¹⁴ Feyerabends Position wird im Allgemeinen in Anlehnung an seine eigene Verortung als methodischer Anarchismus bezeichnet. Die Wortwahl sollte allerdings in ihrer Bedeutung nicht überstrapaziert werden, denn auch Feyerabend ging in seiner Kritikherleitung methodisch vor. Sein Einwand richtet sich vielmehr gegen den Methodenzwang im Sinne einer ausschließlichen Perpetuierung des bereits vorhandenen.

jede Art von Bewertungskriterien ablehnt. Ausdrücklich betont Smith immer wieder die enge Assoziation von qualitativer Sozialforschung und Relativismus. Im speziellen diese Prämisse wurde häufig bezweifelt, wie gezeigt werden wird.

Eine dritte Ecke, aus welcher Vorbehalte zu hören sind, ist die der postmodernen Philosophie. Kritik von dieser Seite zielt darauf ab, auf die autoritären Tendenzen innerhalb von Forschungsberichten aufmerksam zu machen, um gleichzeitig einem Verständnis von universeller Wahrheit entgegenzutreten. Multiple und gleichsam gültige Welten und Weltansichten, die parallel nebeneinander bestehen können, führen unweigerlich zu einem erkenntnistheoretischen Relativismus (Seale 1999a:17), der sich externen Maßstäben zu entziehen weiß. Lincoln und Denzin bezeichnen den Entwicklungsstand, auf dem sich die qualitative Sozialforschung demnach befindet, als „The Fifth Moment“ (Lincoln/ Denzin 1994:575), in ihrer fünften Phase befindlich. Diese sei gekennzeichnet durch eine Krise der Repräsentation und der Legitimation (ebd.), an die Stelle von Theorien treten Erzählungen. „The social text: telling stories from the field“. (ebd.:582) Klassische Gütekriterien, wie Validität und Reliabilität verlieren ihre Bedeutung in der Dialektik bzw. Kontradiktion von Validität und Authentizität. (ebd.)

Letztlich soll noch ein methodenimmanenter Einwand erwähnt werden, um der Heterogenität der verschiedenen Ansätze innerhalb des qualitativen Paradigmas Rechnung zu tragen. Denn es hat sich, so die Warnung, innerhalb der Forschungsrichtung eine allzu breite Ansammlung verschiedener Programme entfaltet, deren gemeinsame Schnittmenge zunehmend gering geworden ist, so dass gemeinsame Gütekriterien gar nicht mehr formuliert werden können. Der Trend für die nähere Zukunft scheint in einer weiteren Ausdifferenzierung bzw. Hybridisierung zu liegen. (Flick 2001:61) Bei genauem Hinsehen basiert der Einwand lediglich auf dem Niveau der gewählten Definition. Wie Reid und Gough richtig bemerken, ist es unumstritten, dass eine sehr divergierende Vielzahl an Typen, Genres und Formen qualitativer Forschung existiert. Beurteilungskriterien nun beinhalten implizit und explizit eine Ansicht davon, was Forschung ist und wie sie sein sollte. (2000:59f) Die logische Folgerung daraus ist, dass auch umgekehrt, die Definition davon, was unter qualitativer Forschung verstanden wird, die Beschaffenheit von Beurteilungskriterien beeinflusst. Die ex ante vorgenommene Definition des Gegenstandsbereichs legt den Geltungsbereich rahmenartig fest, auf den Kriterien sich formal beziehen können. Folgerichtig muss bei der Kriteriengenerierung stets darauf

geachtet werden, dass die Ergebnisse uneingeschränkt auf alle Definitionselemente anwendbar sind. Im vorliegenden Fall ist es angestrebt worden, alle Forschungsperspektiven mit einzubeziehen, die in den gängigen Lehrbüchern dem qualitativen Paradigma zugerechnet werden. Angesichts der Tatsache, dass die Definitionsarbeit zu Beginn der Arbeit geleistet wurde und nun als verbindliche Arbeitsgrundlage dient, kann jetzt die Arbeit fortgesetzt werden unter der Maßgabe, zu jedem Zeitpunkt mit allen Kriterien alle Definitionselemente umfassen zu müssen. Im Hinblick auf die angeführte Kritik bedeutet das nun eine prinzipielle Unmöglichkeit der Nichtzusammenfassbarkeit der Forschungsperspektiven als Geltungsbereich für Gütekriterien. Jedoch wirkt sich die Bindung an die Definition selbstverständlich desto einschränkender und hemmender aus, je mehr unterschiedliche Forschungsperspektiven per definitionem integriert werden.

Die einschlägigen Einwände gegen universelle Gütekriterien erscheinen in gewisser Weise einleuchtend. In den folgenden Abschnitten allerdings wird die andere Seite der Argumentation zu Wort kommen. Zunächst wird Seales überaus integrative Beschäftigung mit den Einwänden expliziert und daraus ein Standpunkt modelliert, von dem aus sich eine Bearbeitungsbasis für in Frage kommende Beurteilungskriterien anbietet.

2.3 Der Standpunkt des subtilen Realismus

Um seine Position zu differenzieren, setzt Seale sich dezidiert mit philosophischen und erkenntnistheoretischen Diskussionen auseinander und versucht ausdrücklich aus allen Sichtweisen zu lernen. Philosophische Positionen versteht er als Ressource und nicht als ein Hindernis für die Sozialforschung. (Seale 1999a:25) Die Sozialforschung könne sich ein Stück weit freimachen von der philosophischen Bevormundung, die dann auftritt, wenn durch die Forschungspraxis lediglich philosophische Schemen ausgefüllt werden. Zugleich müsse sie den Blick unweigerlich auf jene wertvollen Arenen gerichtet behalten und sich nicht gänzlich von den dort ablaufenden Diskursen lösen. (vgl. Seale 1999b:466) Die Vision für die Qualitätskriterien lautet unter einer solchen Betrachtungsweise dann: „The search for overarching criteria for judging quality ... is held at a distance, and the elusive

nature of quality is preserved.“ (ebd.) Vermieden wird auf diesem Wege ein Steckenbleiben dieser Diskussion auf definitorischem Niveau, die Suche nach übergreifenden Kriterien muss also nicht aufgegeben werden, stellt sich aber gleichzeitig lernbereit jeder Diskussion.

Dementsprechend können auch Feyerabends Gedanken anregend sein – verwahrte sich doch ausgerechnet jener gegen die Anwendung fester Regeln, worunter auch Qualitätskriterien zu subsumieren sind. Seale stimmt mit Feyerabend darin überein, dass eine blinde Regelgeleitetheit die Kreativität blockiert, fügt aber hinzu, dass eine Beschäftigung mit Kriterien nicht zwangsläufig die Kreativität beeinflussen muss. (vgl. 1999a:33) „Der Forschungskontext kann schließlich freier sein als der Bewertungskontext.“ (ebd.) Damit spricht er sich eindeutig gegen den Stellenwert von Kriterien im Sinne von „strict rule following“ (ebd.), dem strikten Befolgen von Regeln aus und möchte ihren Charakter eher als „guiding ideal“ oder „enabling conditions“ (beide ebd.), also als leitendes Ideal bzw. ermöglichende Bedingung verstanden wissen. Diese Charakterisierung ist von großer Bedeutung für den Umgang mit Bewertungskriterien im Rahmen dieser Arbeit und sollte von nun an stets mitgedacht werden. Im Zusammenhang der Auseinandersetzung mit dem methodischen Anarchismus kristallisieren sich weitere Implikationen heraus. Anarchistische Momente finden bei Seale ihren Platz innerhalb eines disziplinierten Forschungskontextes, der gekennzeichnet ist durch nachdrücklich prinzipiengeleitetes, methodisches und systematisches Denken. Qualitativ Forschende sollten in jedem Fall ein geschultes Bewusstsein entwickeln für die jeweiligen Konsequenzen, welche einzelne Entscheidungen während des Forschungsprozesses zeitigen. (vgl. ebd.) Die Grundlage zu einer derartigen Befähigung ist eine methodologische Bewusstheit gegenüber jedweden intelligenten relevanten Diskussionen, wie auch diejenige um Qualität in der qualitativen Forschung eine darstellt.

Angelpunkt radikal-konstruktivistischer, wie auch postmoderner Kritik ist letztendlich eine Vorstellung von universeller Wahrheit (siehe oben). So unterschiedlich die Kritik auch hergeleitet wird, so ähnlich ist sie in ihrer Kernaussage und in ihrer Konsequenz für ein externes Bezugssystem, nämlich vernichtend. Antagonistisch dazu steht das positivistische Weltbild, in welchem Wahrheit als von den Menschen

unabhängig begriffen wird und von universeller Beschaffenheit ist.¹⁵ Ein derartiges Verständnis von Wahrheit kommt selbstverständlich der Berechtigung universeller Gütekriterien entgegen, entsprechend dem Prinzip ‚die wahre Welt kann wahre Gütekriterien bieten, mit deren Hilfe wiederum die Wahrheit gemessen werden kann‘. Die beiden Pole lauten also: an dem einen Ende die Vorstellung von externer Messbarkeit von Güte und starren Vorschriften zu ihrer Messung. Die klassischen Konzepte von Validität, Reliabilität und Objektivität finden hier ihre Heimat. Auf der anderen Seite stehen drohende Beliebigkeit bis hin zu Relativismus. Forschungsergebnisse sind dann kaum mehr durch Zweite einzuordnen. Gütekriterien fehlt die Legitimationsbasis, was häufig in der Sozialforschung Frustration, Hoffnungslosigkeit, Zynismus und oberflächliche theoretische Handhabung zur Folge hat. Allein jedoch aus der Zugrundelegung konstruktivistischer Annahmen sei ein Verzicht auf Kriterien nicht zwingend, so Steinke, denn auch postmodern und konstruktivistisch geprägte Forschende stünden vor dem Problem, andere von dem Wert und der Qualität ihrer Untersuchungen und deren Ergebnis zu überzeugen. (1999:51) Seale wählt einen Mittelweg zwischen den oben skizzierten Polen. Er hält einen fallibilistischen Ansatz in der qualitativen Sozialforschung nicht nur für anwendbar, vielmehr könnte dadurch sogar nihilistischen Tendenzen entgegen gewirkt werden. (1999a:22f) Den häufig dominierenden Naturalismus, der davon ausgeht, mittels eines direkten Zuganges zu den Welten der Beforschten Informationen aus erster Hand sammeln zu können, verweist er ins Reich der Fiktion. Allzu häufig haben ethnographische Feldberichte lediglich literarischen Charakter. An dieser Stelle setzt Seale den Hebel an und stellt dem sein Konzept eines fallibilistischen Realismus gegenüber. (vgl. 1999a:24ff) Der Idealismus, so konstatiert er fast schon polemisch, sehe die Wahrheit ausschließlich bei Gott, sind doch die einzigen Gedankengänge, die jemand kennen kann, die eigenen. Dieses liefere allerdings keine Basis für menschliche Kommunikation, also auch nicht für eine gemeinsame Kommunizierung wissenschaftlicher Aussagen innerhalb einer

¹⁵ An dieser Stelle sollte differenziert werden. Wahrheit wird zwar unterstellt, gleichsam aber auch vernebelt. Der popperianische Wahrheitsbegriff muss die prinzipielle Unerkennbarkeit der Wahrheit betonen, solange er den Falsifikationismus als Erkenntnisprinzip akzeptiert. Demnach ist eine Verifikation von All-Aussagen nicht möglich, Aussagen gelten immer nur als vorläufig bewahrheitet. Folglich entzieht Wahrheit sich der Erschließung. Motor der Wissenschaft ist an ihrer statt die systematische Ausschaltung von Unwahrheit und eine asymptotische Annäherung an das Nicht-Erreichbare. Es soll hier betont werden, was häufig stiefmütterlich behandelt wird: es wird zwar von universeller Wahrheit ausgegangen, nicht aber behauptet, dass jene erreicht werden könnte.

Gemeinschaft von Forschenden. Unter dergestaltigen Prämissen ist der Beliebigkeit bei der Produktion wissenschaftlicher Ergebnisse Tür und Tor geöffnet, ein pragmatischer Umgang mit Forschung wird allerdings außen vor gelassen. Und zwar aus dem Grund, dass Ergebnisse, die beliebig parallel nebenher existieren können und sich doch eventuell widersprechen, eine Strukturierung des Feldes unmöglich machen. Die Sichtweise des Realismus beinhaltet demgegenüber eine Orientierung an Vorgängen im Feld sowie eine mögliche Diskussion darüber in der scientific-community.¹⁶ Diese muss nicht alles hinnehmen, was an Erkenntnissen geliefert wird – Fehlinterpretation und Falschheit sollten identifizierbar sein. Vorsicht ist jedoch geboten, wenn vorschnell abgeurteilt wird, denn die Einschränkung folgt auf dem Fuße, der klassische Begriff von Objektivität ist nach Meinung Seales lang schon nicht mehr haltbar. Gerade die qualitative Forschung habe gezeigt inwieweit die Biographie der Forschenden mit ihrer Forschung verbunden ist. Fazit: Es ist an der Zeit für qualitativ Forschende einen neuen Konsens zu erzeugen, aus einer konstruierenden Forschergemeinde heraus. Exakt diese Forderung führt zum Standpunkt des „subtle realism“, des subtilen Realismus, der diesen kommunikativen Aspekt aufnimmt, wie gezeigt werden soll.

Begründet wurde diese spezielle Sichtweise auf empirische Sozialforschung von Hammersley (1992:Part I), der sich eingehend mit dem Phänomen der Qualität in der Ethnographie auseinandersetzte. Subtiler Realismus wird nicht verstanden als Lösung aller Probleme, sondern als pragmatischer Kompromiss zwischen einigen Extremen. (Seale 1999b:469) Wie oben bereits gezeigt wurde, ist die Grundausrichtung fallibilistisch und geht von der Tatsache aus, dass Forschungsberichte für gewöhnlich einen stark konstruierten Charakter aufweisen. Wie im konstruktivistischen Paradigma auch, wird Wissen begriffen als abhängig und beeinflusst von bereits vorhandenen Denkinhalten, Ideen oder Werten, gleichgültig ob Forschende dies Anmerken oder nicht. Die Wiedergabe innerhalb eines Forschungsberichtes ist demnach eine Funktion der biographischen Erfahrungen der Schreibenden. Nach Hammersley nun ist es das Ziel von Sozialforschung, Realität zu repräsentieren, nicht zu reproduzieren. Dabei werden wie von selbst

¹⁶ Hier wird abermals der theoretische Einfluss Poppers deutlich. Bezüglich des Problems des Letztbegründungszusammenhangs, auch Basissatzproblem genannt, schlägt Popper eine Lösung durch zu kommunizierende Konventionen vor, um diese letztlich nicht mehr empirisch herstellbaren Aussagen absichern zu können.

spezifische Teilaspekte ausgewählt und andere ausgeblendet – multiple sich nicht widersprechende Beschreibungen und Erklärungen des gleichen Phänomens können existieren. (vgl. 1992:51) Aus einer solchen Blickrichtung gesehen, wird Sprache bzw. Text einerseits verstanden als Medium zur Konstruktion neuer Welten und andererseits verweist sie auf eine Realität außerhalb ihrer selbst. Daraus folgert eine Mittelstellung zwischen vergangener Erfahrung und imaginierter neuer Eindrücke. (vgl. Seale 1999b:470)

Trotz und gerade wegen den verschiedenen möglichen Varianten der Realitätsrepräsentation in Forschungsberichten kann ein unterschiedliches Ausmaß an Plausibilität ausgemacht werden, einige Zugänge sind plausibler als andere. Wie kann allerdings Plausibilität beurteilt werden?

Hammersley hält dies für möglich, und Seale folgt ihm in diesem Punkt, indem er eine Beziehung unterstellt zwischen einem Text und der dazugehörigen Realität außerhalb. Diese Beziehung kann zum Beispiel durch die Konsistenz zu anderen bereits existierenden Systemen von Aussagen bewertet werden. Dabei wird keine Beziehung von Sprache und Realität im Sinne einer einfachen Korrespondenztheorie angenommen, sondern ganz einfach ein oppositärer Standpunkt zur reinen konstruktivistischen Sicht eingenommen, die von der Unmöglichkeit ausgehen würde, eine reale Welt erschließen zu können, die separat von der Sprache existiert. Unter diesen Prämissen bedarf es dann nur noch eines Bewertungssubjekts, welches Hammersley in der research community findet. Die Idee einer Gemeinde von Forschenden mit anerkannten Standards zur Beurteilung von Plausibilität, Glaubwürdigkeit und Relevanz von Forschungsberichten nimmt eine zentrale Stellung im subtilen Realismus ein. (Hammersley 1992) Jene community wird idealer Weise gedacht als eine selbstkritische Schar, durchsetzt von gegenseitigem Vertrauen und kritischem Anspruch. Sie sollte ferner fähig sein, Kompromissstrategien anzuwenden, kreative Lösungen zu finden und möglichst reflexiv und selbstbewusst sein. Durch ständige vorsichtige Kritik soll Wahrheit, ähnlich wie bei Popper nur provisorischen Status erlangen, da sie demzufolge lediglich solange Gültigkeit besitzt, bis gute Gründe dafür sprechen, diese zu Gunsten einer widersprechenden Version abzulösen. Realiter zeigt sich demgegenüber eine scientific-community, die in allen Belangen ein Abziehbild der Gesamtgesellschaft darstellt und so ganz und gar nicht dem skizzierten Ideal entspricht. (vgl. Seale 1999a:30) Sie zeichne sich in ihrem Meinungsbildungsprozess durch

Macht- und Statuskämpfe aus, kleinere Meinungsgruppen würden unterdrückt, andere beherrschten dagegen den Diskurs. Angesichts einer solchen Diagnose kann von Subtilität, wie sie Hammersley sich vorstellt, keine Rede sein. Somit stellt der Begriff einen Idealzustand dar, der durch die Diskrepanz zu den existierenden Verhältnissen faktisch zur wissenschaftssoziologischen Analysekategorie wird. Dennoch kann und muss dieser Gedanke hier nicht verfolgt werden, denn fest steht, dass eine community besteht, auf die sich Wissenschaft Betreibende beziehen müssen und die letztlich über verschiedene Mechanismen verfügt, um festzulegen, was qualitativ gute Forschung ausmacht. Hammersley schlägt dementsprechend vor: „der Wahrheitsbegriff sollte umdefiniert werden als ‚Glaube daran, wessen Gültigkeit wir vernünftigerweise folgen‘.“ [Einfache Anführungszeichen im Original, H.G.] (1992:50)

Der subtile Realismus lässt sich deswegen resümierend positionieren zwischen den zwei extremen Spielarten von Realitätsbezug, dem Anti-Realismus, wie ihn manche konstruktivistischen Ansätze vertreten und dem naiven Realismus, wie er bisweilen unter quantitativen Ansätzen zu finden ist. Es wurde gezeigt, dass qualitative Forschung nicht notwendigerweise mit dem Anti-Realismus assoziiert werden muss und dass die Antwort bei Missfallen einer anti-realistischen Position nicht zwangsläufig naiver Realismus lautet.

Dies soweit als Basis. Neben der erkenntnistheoretischen Verortung wurde zugleich der erste Punkt von den drei zu Beginn des zweiten Kapitels genannten Lehrmeinungen bezüglich Beurteilungskriterien abgehandelt. Unter der zuvor entwickelten Perspektive kann deswegen eine fundamentale Skepsis gegenüber Kriterien nicht mehr aufrecht erhalten werden. Die Annahme, Forschungsarbeiten entraten der externen Beurteilbarkeit, muss demzufolge verworfen werden. Da die Bejahung von qualitativen Standards nunmehr hergeleitet ist, stellt sich unmittelbar die Frage, wie diese aussehen sollen. Geklärt werden kann diese Frage, indem im Weiteren die beiden noch ausstehenden Lehrmeinungen eine genaue Betrachtung erfahren. Begonnen wird mit dem Standpunkt, Gütekriterien und Standards lassen sich aus den quantitativen Methoden übertragen. Wie tauglich sind die dort verwendeten Maßstäbe für die qualitative Sozialforschung?

2.4 Die klassischen Gütekriterien

Die in diesem Kapitel diskutierten Gütekriterien werden bisweilen die „klassischen“ genannt. Die Bezeichnung hat ihre Wurzeln in der Entwicklungsgeschichte der empirischen Forschungsmethoden und umfasst eine Anzahl von Ansätzen, welche aus heutiger Sicht gesehen älter sind als andere und deswegen als klassisch gelten. Es handelt sich um die Troika von Validität, Reliabilität¹⁷ und Objektivität, die außerhalb der qualitativen Forschung entwickelt wurde und weitestgehend an naturwissenschaftlichen Idealen orientiert ist. Sie stellen die drei zentralen Gütekriterien dar und stammen hauptsächlich aus der experimentellen klassischen psychologischen Testtheorie und sind nach wie vor bestimmend in der aktuellen Kriteriumsdiskussion in den quantitativen Methoden. Aus diesem Grund ist der Sammelbegriff nicht ganz zutreffend, soll aber nicht verworfen werden, zumal er sich allgemein eingebürgert hat.

Die an der Entwicklung Beteiligten waren sämtlich positivistisch orientiert, der festen Überzeugung, sie würden Methoden ausarbeiten, die eine universale Wahrheit etablieren könnten. Das Forschungsinteresse erreichte seinen Höhepunkt in den 1960er und 70er Jahren, danach vollzog sich die Entwicklung immer langsamer. Auch heute noch werden nicht selten Veröffentlichungen aus diesem Zeitraum als Quelle herangezogen.

Wie in den kommenden Abschnitten noch deutlicher zu sehen ist, stehen die drei zentralen Kriterien in einer nicht voneinander trennbaren Beziehung. „Die drei Gütekriterien ... stehen in einem hierarchischen Verhältnis. Objektivität ist eine notwendige, aber nicht hinreichende Bedingung der Reliabilität. Die Reliabilität ist wiederum eine notwendige aber nicht hinreichende Bedingung der Validität.“ (Diekmann 1995:227) Ziel sei letztlich das Erzielen von Validität. In diesem Sinne seien Reliabilität und damit zugleich Objektivität nur Minimalvoraussetzungen. (ebd.) Im allgemeinen ist es die Regel, „dass die Validität mit der Reliabilität wächst, allerdings nicht im gleichen Maße, sondern wesentlich langsamer.“ (Lienert

¹⁷ Im deutschen Sprachraum wird neben den aus dem Englischen entlehnten Bezeichnungen (dort: Validity und Reliability) zum Teil auch Validität mit ‚Gültigkeit‘ und Reliabilität mit ‚Zuverlässigkeit‘ übersetzt.

1989:294)¹⁸ Und schließlich kann die Validität nicht größer sein als die Reliabilität, wohl aber umgekehrt. (vgl. Schnell/ Hill/ Esser 1992:162) Was jetzt noch verwirrt, wird später deutlicher. In Anlehnung an diese Hierarchisierung werden nun die drei Kriterien von oben herunter abgehandelt und auf ihre Übertragbarkeit auf die qualitative Forschung hin überprüft.

2.4.1 Das Konzept der Validität

Die Auffassung von Validität hat im Laufe der Zeit zahlreiche Veränderungen und Entwicklungen erlebt. Die gebräuchlichste Beschreibung ist, dass Validität angibt, ob die Methode tatsächlich das misst, was sie messen sollte. (Friedrichs 1990:100, Schnell/ Hill/ Esser 1992:162). Swanborn beschreibt, Validität ist „in einem allgemeinen Sinn, dass unsere Untersuchung die empirische Welt auf korrekte Weise beschreibt und erklärt und in einem strikteren Sinn, dass die Untersuchung frei ist von zufälligen und systematischen Fehlern.“ (1996:22) Unter den Begriff summierten sich zunehmend verschiedene Differenzierungen, die sich jeweils Teilgebieten widmeten, so dass sich die *American Psychological Association* (APA) im Jahre 1954 gezwungen sah, der steigenden Verwirrung ein Ende zu bereiten und eine Klassifizierung in drei Kategorien vornahm. Als Oberbegriff für diese drei Formen kursiert die Bezeichnung ‚psychometrische Validitätsformen‘, obwohl sie über die Psychologie hinaus Verbreitung gefunden haben:

1. content validity (Inhaltsvalidität)
2. criterion-related validity (Kriteriumsvalidität)
3. construct validity (Konstruktvalidität)

(vgl. American Psychological Association 1954)

Neben der bisherigen Auffassung von Validität, die sich auf die Qualität der Übersetzung von empirischem Relativ ins numerische Relativ bezieht, gibt es noch einen weiteren Zusammenhang, in welchem von Validität gesprochen wird. Und zwar handelt es sich im Hinblick auf empirische Untersuchungsdesigns um den Grad, inwieweit die Ergebnisse einer Untersuchung einen Anspruch auf Gültigkeit

¹⁸ Lienert weiß auch von Ausnahmen von dieser Faustregel zu berichten, die hier allerdings nur verwirren würden.

besitzen. Diesbezüglich wird unterschieden zwischen interner und externer Validität, wobei versucht wird, das Untersuchungsergebnis hinsichtlich der erreichten approximierten Wahrheit als vermutet wahre Aussage einzuschätzen. (vgl. Cook/ Campbell 1979)

Interne Validität

Interne Validität bezieht sich auf die Gültigkeit von Untersuchungsvariablen innerhalb einer Untersuchungssituation. Eine Untersuchung ist dann intern valide, wenn die Veränderung der abhängigen Variablen eindeutig und ausschließlich auf die zuvor festgesetzten unabhängigen Variablen der Untersuchungsbedingungen zurückzuführen ist. Die Ergebnisse müssen eindeutig interpretierbar sein, die interne Validität sinkt dagegen mit der wachsenden Anzahl plausibler Alternativerklärungen für die Ergebnisse. (vgl. Bortz/ Döring 1995:52) Zentrales Moment ist, ob Kausalinterpretationen nicht nur gemessen oder beobachtet wurden, sondern auch tatsächlich gerechtfertigt sind. Nur die hinzugezogene unabhängige Variable darf Ursache für die abhängige Variable sein. Es muss deshalb geprüft werden, ob nicht Störvariablen Einfluss genommen haben. Zur Sicherung müssen daher die Störvariablen einer Untersuchung kontrolliert werden. (Steinke 1999:160) Interne Validität lässt sich interessanter Weise auch in der quantitativen Tradition nicht letztgültig durch die Anwendung technischer Prozeduren sicherstellen. Es existieren dazu keine allgemeinen und abstrakten Regeln. Sie beruht im wesentlichen auf der Beurteilung der die Validität bedrohenden Gefahren durch die Forschenden. (vgl. Seale 1999a:38) Dies kann geschehen über die Isolation von Bedingungen in der Untersuchungssituation, die Bildung von Kontrollgruppen oder die statistische Kontrolle konfundierender Variablen. Dennoch wird davon ausgegangen, dass selbst in einer Laborsituation oder in einem Experiment eine vollständige Bedingungskontrolle nicht möglich ist, sind doch zahlreiche Störeinflüsse überhaupt nicht bekannt. Steinke weist explizit auf nicht vollständig ausschließbare sozialpsychologische Einflüsse in der Untersuchungssituation hin, die mit den kontrollierten Variablen konfundierten. An dieser Stelle stoße die Experimentalmethodik an ihre Grenzen, die Reflexivität des Subjekts greife in das Untersuchungsgeschehen ein. (1999:161f.)

Was ist davon auf die qualitative Forschung übertragbar, wo liegen die Grenzen?

Das Modell der internen Validität ist nicht eins zu eins auf die qualitative Forschung übertragbar. Problematisch ist die Annahme kausaler Beziehungen, sie beruht auf einer deterministischen Auffassung menschlicher Handlung, die in der qualitativen Forschung als überwunden gilt. (vgl. Seale 1999a:39) Ferner ist die Auffassung der Variablenbehandlung gegensätzlich. Isolierung durch eine Laborsituation auf der einen Seite, Forschung in natürlichen und alltäglichen Kontexten auf der anderen. (vgl. Steinke 1999:164) Dennoch sind zumindest Teilaspekte übertragbar. Dabei ist stets der unterschiedliche Stellenwert von Theoriebildung zu berücksichtigen. Während quantitative Methoden davon ausgehen, ex ante formulierte Hypothesen einer Überprüfung zu unterziehen, ist die qualitative Forschungslogik darauf ausgerichtet, Theorien ex post auf der Basis des empirischen Materials zu bilden. (vgl. Kelle 1994) Um eine vorschnelle Festlegung auf gewisse Erklärungsmuster zu vermeiden, schlagen Kelle, Kluge und Prein drei mögliche Vorgehensweisen vor:¹⁹ Erstens, diskursive Validierungsmethoden, die darauf abzielen, die Gültigkeit von Interpretationen durch Diskussionen in einer Forschungsgruppe zu überprüfen. Darunter fallen eine Handvoll verschiedener Konzepte, denen allen gemein ist, dass durch sie „eine möglichst große Anzahl von Erklärungshypothesen expliziert und gegeneinander gehalten werden“ soll.²⁰ (1993:49) Zweitens, die Überprüfung von Deutungshypothesen am Material. Demnach soll systematisch nach Evidenz und Gegenevidenz im Datenmaterial gesucht werden und gleichsam verhindert werden, dass einmal formulierte Hypothesen im weiteren Verlauf der Auswertung immunisiert bzw. ausschließlich die bestätigenden Instanzen im Datenmaterial registriert werden. (1993:52) Die fallibilistische Perspektive lenkt die Vorgehensweise der qualitativen Forschung dahingehend, dass anstatt das Augenmerk auf die Suche nach konstanten Beziehungen zu legen, eher die Suche nach Falsifizierendem im Mittelpunkt stehen soll. (Seale 1999a:40) In diesem Kontext schlägt Steinke das Konzept der Unterschiedsmethode vor (1999:165), die es erlaubt, den Ausprägungsgrad der unabhängigen Variable als Teilursache des entsprechenden

¹⁹ Die Ansicht, dass theoretische Konzepte aus dem Datenmaterial quasi von selbst emergieren können, wenn die Herangehensweise der Forschenden nur möglichst vorurteilslos ist (vgl. Glaser/ Strauss 1967), wird von Kelle, Kluge und Prein als „naive Fehlinterpretation menschlichen Erkenntnisvermögens“ (1993:45) bezeichnet. Sogenannte ‚Tabula-Rasa‘ Modelle finden folglich auch an dieser Stelle keine Beachtung hinsichtlich ihrer Implikationen für Qualitätskriterien.

²⁰ Die Durchführung dieser und der im weiteren Verlauf vorgeschlagenen Verfahren soll hier nicht detailliert beschrieben werden sondern muss im Einzelfall anhand der entsprechenden Quellen erarbeitet werden. Die Darstellung beschränkt sich lediglich auf die methodisch relevanten Überschneidungspunkte.

Ausprägungsgrades der abhängigen Variable zu betrachten. Realisierbar sei dies in der qualitativen Forschung durch die Auswahl, Analyse und Variation von Fällen, die sich in einem bestimmten Aspekt unterscheiden. (ebd.) Kelle führt das im Rahmen der Grounded Theory entwickelte Prinzip der Kontrastierung von Fällen aus. Es werden Untersuchungseinheiten miteinander verglichen, die „hinsichtlich theoretisch bedeutsamer Merkmale entweder relevante Unterschiede oder große Ähnlichkeiten aufweisen.“ (Kelle 1994:297ff.) Unter der Berücksichtigung des fallibilistischen Standpunktes wird somit klar, wie auf dem Wege der maximalen und minimalen Kontrastierung die interne Validität erhöht werden kann. Drittens und letztens nennen Kelle, Kluge und Prein den Einsatz von Computern in der qualitativen Forschung, mit Hilfe sogenannter CAQDAS („Computer Assisted Qualitative Data Analysis Software“). CAQDAS erleichtern das Handling größerer Datenmengen immens und vermögen es, die Validität der Ergebnisse zu erhöhen, indem sie durch umfangreiche Retrievalfunktionen ein sehr enges Arbeiten am Text zulassen. (vgl. Kelle/ Laurie 1995). Durch den erleichterten Zugriff auf das Datenmaterial kann die Hypothesenbildung enger an die Daten gekoppelt werden. Allerdings, so schränken Kelle und Laurie ein, empfiehlt sich der Einsatz von CAQDAS nicht für jede Vorgehensweise in gleichem Maße wie für Kodierverfahren. Hermeneutische Analysen oder Techniken der Paraphrasierung bieten sich nicht für eine Anwendung an. (1995:24)

Externe Validität

Externe Validität bezieht sich in Anlehnung an Cook und Campbell auf die Verallgemeinerbarkeit von Untersuchungsergebnissen. (1979:39) Sie liegt vor, wenn die Ergebnisse einer Stichprobenuntersuchung auf andere Elemente der Grundgesamtheit, auf andere Situationen der Erhebung und auf andere Erhebungszeitpunkte übertragen werden können, man spricht daher auch von Generalisierbarkeit. (vgl. Bortz/ Döring 1995:52) Es werden Aussagen darüber getroffen, für welche Populationen die Untersuchungsergebnisse gültig bzw. repräsentativ sind. „Die externe Validität sinkt mit wachsender Unnatürlichkeit der Untersuchungsbedingungen bzw. abnehmender Repräsentativität der untersuchten Stichproben.“ (ebd.) Wie den Ausführungen zu entnehmen ist, ist die Verallgemeinerbarkeit sehr stark an die Repräsentativität von Versuchsplänen gekoppelt. Diesbezüglich sind im wesentlichen zwei Teilmomente zu unterscheiden: (vgl.

Steinke 1999:166) zum einen die Repräsentativität der Stichprobe und zum anderen die von Umgebung bzw. Situation. Erstere wird erreicht durch Zufallsauswahlen aus der Grundgesamtheit. Die Zufallsauswahl soll sicherstellen, dass die potenziell störenden oder unbekanntem Merkmale in der Stichprobe annähernd genauso verteilt sind wie in der Grundgesamtheit. Für die Repräsentativität von Umgebung bzw. Situation hat sich der Begriff der ökologischen Validität eingebürgert. Sie ist unter Laborbedingungen kaum gegeben, weil die Situation dort allzu künstlich ist. Diesbezüglich wird von einem Dilemma zwischen interner und externer Validität ausgegangen. Ergebnisse von hoher interner Validität, die unter Bedingungen von hoher Variablenkontrolle entstanden sind, sind nur sehr begrenzt auf Situationen außerhalb des Labors übertragbar und umgekehrt. Beide Validitätsarten sind nicht miteinander vereinbar und stellen ein unauflösbares Problem dar. Für die qualitativen Methoden stellt sich diese Problematik allerdings nicht in gleichem Maße, denn die ökologische Validität ist per definitionem relativ hoch. Das heißt, angesichts der immer in der natürlichen Umgebung der Beforschten durchgeführten Phase der Datengewinnung, kann die Künstlichkeit einer Situation nur gegen Null tendieren. Zu der Thematik des Einflusses der Situation bzw. der Umgebung sind auch alle Formen von Reaktivität zu zählen, wie sie besonders die Sozialpsychologie erforschte. Zu nennen sind der Effekt der sozialen Erwünschtheit, der Hawthorne-Effekt, der Halo-Effekt oder diverse Gruppenleitereffekte, wie der Rosenthal-Effekt. Störende Einflüsse, die durch die Untersuchungssituation überhaupt erst induziert werden, können einzig durch gezielte Variation der Kontexte umgangen werden. Bei Generalisierungen ist es rasch möglich, dass eine derartige Beeinflussung der Ergebnisse sich in den Schlussfolgerungen widerspiegelt als vermeintliches Novum.

Im Bereich der qualitativen Sozialforschung nimmt in der aktuellen Debatte die Frage nach der Generalisierbarkeit von Untersuchungsergebnissen einen breiten Raum ein. Grundsätzlich stellt sie sich dort ebenso wie auf der quantifizierenden Seite, wenn es daran geht, Ergebnisse über die konkrete Untersuchungssituation und die Untersuchungsgruppe hinaus zu verallgemeinern. Durch ihr induktives Vorgehen bekommen Aspekte der Verallgemeinerung gar eine ganz besondere Dimension verliehen, stellt sich doch die entscheidende Frage des Geltungsbereiches der generierten Theorien oder Hypothesen. Anders ausgedrückt ist es von wesentlicher Bedeutung, für welche Population die Hypothese Gültigkeit

beansprucht. Möglicherweise gilt diese lediglich für bestimmte Segmente, aber davon wird noch zu sprechen sein. Generalisierbarkeit lautet gleichwohl nicht immer das Ziel einer qualitativen Forschung und ist kein Wert an sich. (Seale 1999a:107) Sie ist zwar möglich und wünschenswert aber keine Ultima Ratio. (ebd.)

Kelle, Kluge und Prein sehen analog zu den obigen Repräsentativitätsvarianten eine Gefahr für die Verallgemeinerungsfähigkeit in der Auswahl der Untersuchungseinheiten. (vgl. 1993:60ff.) Ihre Auswahl sei bisweilen einseitig und beschränkt, Fragen der Stichprobenziehung würden oftmals salopp beiseite geschoben. Die Autoren weisen mit Nachdruck darauf hin, dass andererseits das Bild von einer Repräsentativität schlechthin zurückzuweisen ist. Repräsentativität kann sich niemals auf alle denkbaren Merkmale beziehen, sondern immer nur auf eine bestimmte Anzahl bedeutsamer Merkmale. „Ist eine Stichprobe theoriebezogen repräsentativ, so bedeutet dies, dass Träger theoretisch relevanter Merkmalskombinationen in der Stichprobe hinreichend vertreten sind ...“. (ebd.:63) Stichprobenverzerrungen, hervorgerufen durch irrelevante Merkmalskombinationen müssen also allgemein vermieden werden. In der quantitativen Methodik wird dies erreicht durch verschiedene Verfahren der Auswahl, wie unter anderem, die Zufallsstichprobe im Falle unbekannter Merkmale oder die Quota-Auswahl im Falle bekannter Merkmale. (vgl. Friedrichs 1990:3.4.1.) Innerhalb eines qualitativen Forschungsdesigns gibt es die Sicherstellung von Verzerrungsfreiheit durch das ‚theoretical sampling‘. Demnach wird die Auswahl der Untersuchungseinheiten theoretisch geleitet und nicht zufällig variiert. Mit Hilfe kontrastierender Fälle wird gezielt variiert bis keine widersprechende Gegenevidenz mehr gefunden wird resp. keine neuen Erkenntnisse mehr gewonnen werden können; genau dann ist eine ‚theoretische Sättigung‘ erreicht. Die Untersuchenden verfügen dabei zu Beginn nicht über einen Stichprobenplan sondern entwickeln diesen erst im Laufe des offenen Verfahrens. Bestehen dagegen schon vor Beginn der Feldphase Orientierungshypothesen über den untersuchten Gegenstand, dann kann durch eine a priori Definition von Auswahlmerkmalen eine bestimmte Merkmalskombination ausgewählt werden für die Sample-Bildung. (ausführlicher: Kelle/ Kluge/ Prein 1993:65) Eine weitere Strategie ist das „Matrjoschka-Sampling“, (Prein/ Kluge/ Kelle 1994:20) das die Auswahl von Untersuchungseinheiten nach

theoretischen Gesichtspunkten aus einer größeren Untersuchung leitet.²¹ Die große Palette an Vorgehensweisen, die in der Literatur ausgearbeitet ist, sollte zum Maßstab der Auswahl gemacht werden, wenn dem keine spezifischen Gründe entgegenstehen. Eine willkürliche Heranziehung von Untersuchungseinheiten ohne Begründung entbehrt jeder Legitimation.

Seale weist auf einen weiteren berücksichtigungswerten Aspekt von Generalisierbarkeit hin. Bei kleinen Fallzahlen nämlich kann die Validität einer Verallgemeinerung durch „thick description“ (Seale 1999a:107), einer dichten Beschreibung erhöht werden. „Thick description“ beschreibt in etwa das, was im Hempel-Oppenheim-Schema als Antecedensbedingungen bezeichnet wird. Diese Rand- oder Kontextbedingungen, die allerdings keine Randstellung einnehmen, müssen bei der Übertragung von einem Bereich auf den anderen von gewisser Ähnlichkeit sein hinsichtlich der relevanten Merkmale. Seale spricht von ‚Sendekontext‘ und ‚Empfängerkontext‘. (1999a:108) Der Erste ist derjenige, in welchem eine Theorie oder Hypothese entstand, das Umfeld der Generierung. Der Zweite ist derjenige Kontext, in welchen diese übertragen bzw. verallgemeinert werden soll. Die Beschreibung beider ist gewiss nur annähernd vollständig möglich, letztlich obliegt die Entscheidung darüber, was alles beschrieben wird, in der theoretischen Operationalisierung begründet. Vor fallibilistischem Hintergrund wird den Lesenden praktisch eine Art Gedankenexperiment dargeboten, welches sie nachvollziehen können oder nicht.

Psychometrische Validitätsformen

Lienert erwähnt eine ganze Reihe weiterer Validitätsformen (1989:11.Kap.), die speziell für die Validierung von Tests entwickelt wurden. Obige, von der APA vorgenommene Dreiteilung bezieht sich genau auf jene Formen. Ihre überaus begrenzte Relevanz für die qualitative Sozialforschung macht eine explizite Definition hier nicht unbedingt erforderlich. Seale subsumiert diese drei unter die Kategorie Validität von Messungen. (1999a:34) Für ihn ist die Bedeutung dieser Kriterien für die qualitative Sozialforschung überaus gering. Probleme der Messung seien zudem auch im quantitativen Ansatz keineswegs überwunden, im qualitativen

²¹ Für weitere Sampling-Strategien und ausführliche Durchführungsdetails siehe Prein/Kluge/Kelle (1994) zum Überblick über Repräsentativität bei kleinen Stichproben, wie sie nahezu stets Wesensmerkmal qualitativer Untersuchungsdesigns ist.

dagegen nicht von Relevanz, wenn Messergebnisse sich durch eine vergleichbare Wiederholung gegenseitig validieren. (1999a:36) Problematisch dabei ist die Annahme eines sich ständig wandelnden Gegenstandes, die starke Fluktuation von Kontextmerkmalen und die Ansicht, die Untersuchungsmethode beeinflusse maßgeblich den Gegenstand. Das letzte Argument ist besonders interessant angesichts der Forderung nach gegenstandsangemessener Forschung in der qualitativen Sozialforschung. Demnach ist die Methode vom Gegenstand abhängig und der Gegenstand wiederum von der Methode. Das dem traditionellen Messbegriff zu Grunde liegende Konzept-Indikatoren-Modell, wonach kurz gesagt, beide Seiten unabhängig voneinander nur auf ihr Pendant zur Beforschung warten, muss bei einem Verständnis der Erkenntnis als methodenabhängig strikt verworfen werden.

2.4.2 Das Konzept der Reliabilität

Das Kriterium der Reliabilität, im deutschen Sprachraum für gewöhnlich mit Zuverlässigkeit übersetzt, wird in der klassischen Testtheorie meist als das nach der Validität wichtigste Gütekriterium erwähnt. Gemeint ist der Grad der Genauigkeit mit dem ein Merkmal gemessen wird, gleichgültig ob der Anspruch besteht, dieses Merkmal auch tatsächlich zu messen, was eine Frage der Validität wäre. (vgl. Lienert 1989:14) Ein Messinstrument ist genau dann vollkommen reliabel, wenn die erzielten Messergebnisse des wahren Wertes fehlerfrei und genau sind. Reliabilität bezieht sich ausschließlich auf die Messung und nicht auf die Interpretation.

Meist wird nach drei Formen der Reliabilität unterschieden (vgl. ebd.:15): Erstens, der Paralleltest-Reliabilität, die durch die Durchführung zweier streng vergleichbarer Tests durch eine Stichprobe von Versuchspersonen bestimmt wird. Anschließend werden die Ergebnisse miteinander korreliert, die Nichtübereinstimmung wird als Messfehler begriffen. Zweitens, der Retest-Reliabilität, die ermittelt wird mittels der Wiederholung ein und des selben Tests durch eine Stichprobe von Versuchspersonen und die abschließende Korrelation miteinander. Und drittens und letztens, die innere Konsistenz eines Testes oder auch als Split-Half-Reliabilität bezeichnet.

Dabei wird der Test in zwei gleiche Hälften geteilt, die zuletzt miteinander korreliert werden.

Konventionell formulierte Reliabilitätskonzepte entpuppen sich in der qualitativen Sozialforschung als „herkulesartiges Problem“, so LeCompte und Goetz. (1982:35). Die Übertragbarkeit ist sehr begrenzt, sie scheitert an den zu Grunde liegenden Kennzeichen der qualitativen Vorgehensweise, die mit denen der klassischen Testtheorie in Bezug auf Fragen der Reliabilität nicht vereinbar sind. Zu nennen sind die „Notwendigkeit von ex-ante Definitionen von Untersuchungsgegenständen, Hypothesen und deren Operationalisierung sowie (...) [die] notwendige Standardisierung der Messinstrumente“. (Steinke 1999:155) Demgegenüber stünden eine induktivistische Orientierung, eine Alltagsbezogenheit und Kontextualität der Untersuchungen, die Zirkularität des Forschungsprozesses und das Prinzip der Gegenstandsentsfaltung. (ebd.) Beide Momente sind nicht miteinander vereinbar, so verwirft Steinke das gesamte Konzept und nimmt keinerlei Übertragungen vor.

Seale verwendet den Begriff der Reliabilität ein wenig anders, nahezu synonym zum Begriff der Replizierbarkeit. In seiner Auffassung von der klassischen Testtheorie liefert ein reliables Instrument im Wiederholungsfall konstante und nahezu identische Ergebnisse. Er adoptiert den Vorschlag der Unterscheidung in interne und externe Reliabilität von LeCompte und Goetz (1982). Demzufolge definiert sich interne Reliabilität als der Grad der Übereinstimmung des Ergebnisses, den verschiedene Forschende erzielen, wenn sie ein zuvor entwickeltes Analysekonzept gleichsam anwenden. Externe Reliabilität bezeichnet dagegen den Grad der Übereinstimmung der Ergebnisse, den voneinander unabhängige Forschende erzielen, wenn sie ein Phänomen mit gleichem oder ähnlichem Setting untersuchen. Dieses Kriterium betrifft die Frage nach der Replizierbarkeit ganzer Studien.

Swanborn und Seale sind sich letztenendes einig und behaupten die prinzipielle Ungleichheit zweier Untersuchungssituationen. Faktisch ist eine Replikation unter identischen Bedingungen unmöglich. (vgl. Swanborn 1996:28) Dies gilt in besonderem Maße für Replikationen, die im Feld durchgeführt werden, ist doch der Gegenstand einem ständigem Wandel unterworfen und präsentiert sich jederzeit anders, als im Moment davor. Hinzu kommen zum Teil unvermeidbare Effekte von Reaktivität, die auf das Ergebnis Einfluss nehmen. Replikationsstudien, die auf einen Feldzugang verzichten und sich auf die Re-Analyse bereits eingeholten Datenmaterials beschränken, sind zumindest vor diesen beiden Bedrohungen

gefeit. Die Annahme, dass gleiches Material zu gleichen Ergebnissen führen kann, wenn nur die Dokumentation aller forschungsrelevanten Randbedingungen ausreichend detailliert ist, soll hier kategorisch verworfen werden. Sie impliziert ungewollt eine Annahme des Tabula-Rasa-Konzeptes (siehe oben) als Startbedingung und degradiert die Forschenden zu Interpretationsautomaten, die nur noch die vorgeschriebenen Wege ausfüllen müssen. Ohne die Tabula-Rasa-Annahme ist allein schon die Ausgangsposition der einzelnen Forschenden ausreichend groß, um gemäß der Auffassung von qualitativer Forschung als Aneinanderreihung von Forschungsentscheidungen, sich am Ende in divergierenden Ergebnissen niederzuschlagen. Latente Vorannahmen pflanzen sich im zirkulären Forschungsprozess nicht nur fort, sondern bedingen zunehmend auseinanderfallende Zwischen- und Endergebnisse.

Jene Aspekte der Dokumentation, die des öfteren der Reliabilität zugeordnet wurden, dort aber nichts vermögen, sollen im folgenden Abschnitt diskutiert werden, weil sie, wie dargestellt, dennoch von unschätzbarem Wert sind.

2.4.3 Das Konzept von Objektivität

Im Rahmen des einstigen untrennbaren Triviums der hier behandelten klassischen Gütekriterien, nimmt das Konzept der Objektivität mittlerweile nur noch eine Randstellung ein. In zahlreichen Publikationen zu Gütekriterien in der qualitativen Forschung taucht die Rubrik sogar überhaupt nicht mehr auf. Möglicherweise liegt dies an der vermeintlichen erkenntnistheoretischen Überkommenheit.

Lienert definiert Objektivität als „den Grad, in dem die Ergebnisse eines Testes unabhängig vom Untersucher sind.“ (1989:13) Steinke verweist auf eine weitere Unterteilung, in ontologische und aperspektivische Objektivität. (1999:132) Erstere thematisiert die Beziehung von Subjekt und Forschungsgegenstand und unterstellt einen korrespondenztheoretischen Wahrheitsbegriff. Dieser geht von der Existenz einer unabhängig vom Subjekt vorhandenen und erkennbaren Realität aus. Allerdings kann die in der Definition enthaltene Trennung von Gegenstand und Subjekt in der qualitativen Methodologie nicht aufrecht erhalten werden, deswegen ist die wahrlich interessantere, die zweite Variante, die aperspektivische Objektivität.

Sie bezieht sich auf die Eliminierung der individuellen Einflüsse von Forschenden und bedeutet in diesem Zusammenhang eine Unabhängigkeit der Erkenntnis von subjektiven Einflüssen des oder der jeweiligen Erkennenden. „Operationalisiert wird diese Nicht-Subjektivität durch das Konzept der Intersubjektivität.“ (ebd.) Demgemäß besteht die Objektivität von wissenschaftlichen Aussagen darin, dass sie intersubjektiv nachvollziehbar sein müssen. Folglich bezieht sich das Kriterium nicht auf eine äußere, repräsentierte Realität, wie die erste Variante der Objektivität, sondern auf eine zwischen unterschiedlichen Forschenden herzustellende intersubjektive Übereinstimmung. „Dieses Objektivitätsverständnis bezieht sich auf die Beziehung zwischen einer Aussage und dem Forschungsprozedere, welches auf interindividuell gleiche Vorgehensweisen und Resultate bzw. die Möglichkeit des Nachvollzugs einer Untersuchung durch andere Forscher ausgerichtet ist.“ (ebd.) Es thematisiert damit die Relationen zwischen interindividuellen Sichtweisen. Intersubjektive Nachvollziehbarkeit kann in diesem Sinne als ein Bewertungskriterium für qualitative Forschung herangezogen werden. Es dient dazu, „Forschung intersubjektivierbar zu machen, d.h. eine (kritische) [alle Klammern im Original, H.G.] Verständigung über eine empirische Studie zwischen Forschern bzw. zwischen Forscher (der eine Studie durchführt) und Leser (der Studie) zu ermöglichen.“ (Steinke 1999:207) Die Nachvollziehbarkeit kann im wesentlichen auf zwei Wegen nicht unerheblich erhöht werden. Der Potenteste ist die umfassende Dokumentation des Forschungsprozesses. Dadurch soll der Weg, wie die Forschenden zu den Ergebnissen gelangt sind und welche Entscheidungen im Laufe der Forschung getroffen und begründet wurden, Transparenz erhalten. Mit der Dokumentation des Vorgehens sei ein Weg gefunden, intersubjektive Nachvollziehbarkeit nicht an formalisierte bzw. standardisierte Verfahren zu knüpfen. Ferner könne mit diesem Kriterium der für jede qualitative Studie einmaligen Dynamik zwischen Gegenstand, Fragestellung und methodischem Konzept Rechnung getragen werden. (ebd.:208) Nebenbei kann gleichermaßen nicht nur aufgezeigt werden, dass Theoriebildung und -prüfung nicht ausschließlich rational erfolgen können und untrennbar mit der Subjektivität der Forschenden verwoben sind, sondern vielmehr kann eine solche Verquickung im Falle guter Dokumentation konkret nachgezeichnet werden.

Was genau muss nun alles dokumentiert werden?

Die Antwort ist genauso kurz, wie unerschöpflich: Alles! Systematisierend soll eine Siebengliederung vorgeschlagen werden: (in Anlehnung an Steinke 1999:6.1.2)

I. Das Vorverständnis, die Erwartungen konstituieren die Wahrnehmung, das Methodenhandling, die Daten und das Gegenstandsverständnis. Es gilt, dieses Vorverständnis zu dokumentieren, um die Forschenden dazu zu zwingen, ihr Vorwissen zu explizieren. Es erlaubt „die Abklärung, inwiefern wirklich Neues erkannt wurde, d.h. nicht nur nach Bestätigung von vorab-Hypothesen gesucht wurde bzw. auch versucht wurde, dieses Vorwissen zu irritieren“. (Steinke 1999:210).

II. Niedergeschrieben werden sollten des Weiteren auch die Erhebungsmethoden und der Erhebungskontext. Nicht nur detailliert, welche Methoden wie angewandt wurden, sondern auch Spezifika aus der Phase der Datengewinnung, wie beispielsweise Details aus der Interviewsituation.

III. Transkriptionsregeln müssen beschrieben werden, um im Nachhinein feststellen zu können, welche Informationen transkribiert worden sind und welche nicht.

IV. Das Datenmaterial sollte zugänglich sein, wenn genügende Anonymität gewährleistet ist. Nicht das gesamte Datenmaterial findet stets im Anhang Platz. Wünschenswert wäre aber alternativ dazu eine Angabe, wo die Daten einzusehen sind. Die Weitergabe kann postalisch oder am Besten per Internet durchgeführt werden. Große Datenmengen sind auf diese Weise schnell und kostengünstig verfügbar.²² Corti sieht in der Möglichkeit der Re-Analyse gar die entscheidende Chance zur Sicherstellung von Validität und Reliabilität, indem durch Einsicht ins Rohmaterial einer Forschungsarbeit eine breitere und tiefere Diskussion stattfinden kann. (2000:[8])

V. Die Auswertungsmethoden müssen dokumentiert werden. Besonders wenn keine kodifizierten Verfahren verwendet werden, ist es nötig, die Prozeduren detailliert zu formulieren. Einzelne Auswertungsschritte und Interpretationen sollten beschrieben sein.

VI. Informationsquellen und ihr Kontext müssen präzise dargestellt sein.

²² In Anlehnung an das Zentralarchiv für empirische Sozialforschung in Köln, das quantitative Daten archiviert und vertreibt, hat sich jüngst ein internationales qualitatives Äquivalent gebildet. Unter dem Namen QUALIDATA (siehe Corti 2001) kann qualitatives Datenmaterial allgemein zugänglich archiviert werden. Diese Initiative ist hinsichtlich der Intersubjektiven Nachvollziehbarkeit und der Transparenz sicherlich ein Zugewinn und ein Meilenstein für die qualitative Sozialforschung.

VII. Im Laufe des Forschungsprozesses anfallende Entscheidungen und Probleme sind zu dokumentieren. Darunter fallen auch angefallene Widersprüche, beispielsweise in der Analyse, die nicht gelöst werden konnten.

Die oben erwähnte Eliminierung der individuellen Einflüsse von Forschenden auf den Forschungsgegenstand, eine Bedingung für die Intersubjektivität, kann zusätzlich auf eine andere Weise angegangen werden und zwar durch Reflexion der Subjektivität. Entsprechend der These, dass Untersuchungsgegenstand und Untersuchende nicht voneinander zu trennen sind, lässt sich folgern, dass die Subjektivität der Untersuchenden ein Teil der Methode ist. Vielmehr noch ist sie unerlässliche Erkenntnisquelle und sollte daher möglichst methodisch reflektiert werden. Später können die Lesenden einer Studie dann beurteilen, wie die Subjektivität die Theoriebildung mitkonstituiert hat.

Expliziert werden kann etwa die persönliche biographische Beziehung zum Untersuchungsthema oder die eigenen Voraussetzungen, wie die kulturelle Herkunft, das Geschlecht, der Beruf uvm. Ferner, welche Beziehung zu den Informantinnen und Informanten besteht, eine vertraute oder eine distanzierte? Vor dem Einstieg ins Feld und in den ersten Feldphasen, sollten die inneren Eindrücke festgehalten und das vorhandene Grundwissen sowie die relevanten Einstellungen bedacht werden. Die Subjektivität wird somit in der qualitativen Sozialforschung nicht als beseitigungswert verstanden, sondern gilt als unentbehrliches Element der Erkenntnisgewinnung. Wirklich fruchtbar macht die Subjektivität erst eine gezielte Reflexion. So kann der sonst als nachteilig empfundene Komplex sich sogar als vorteilig erweisen.

2.5 Alternative oder auch methodenangemessene Gütekriterien

Innerhalb der im Abschnitt 2.) angeführten drei Standpunkte gegenüber Gütekriterien, widmet sich dieser Abschnitt der dritten möglichen Sichtweise. Sie lehnt die Übernahme von Gütekriterien aus der quantitativen Forschung ab und behauptet eine doppelte Gegenstandsangemessenheit. Und zwar sollte nicht nur die Methode dem erforschten Gegenstand angemessen sein, sondern darüber

hinaus sollten auch die Kriterien der Bewertung den eingesetzten Methoden gerecht werden können. (Flick 1987:247) Folglich ergibt sich die Notwendigkeit, eigene, der qualitativen Vorgehensweise angepasste Kriterien der Güte zu entwickeln.

Flicks Postulat der doppelten Gegenstandsangemessenheit könnte als eine Art Monadologisierung der unter der qualitativen Flagge segelnden Verfahren interpretiert werden, denn genau genommen gleicht kein Gegenstand dem anderen, also auch keine Methode der nächsten. Diese Interpretation würde allerdings unterschlagen, dass jede Methode eine gewisse Bandbreite abzudecken vermag, wenn sie in der Detailabstimmung auf den Gegenstand zugerichtet wird. Zur Rettung über den einzelnen Gegenstand hinaus reichender Gütekriterien, muss deren Beurteilungsniveau also auf ein allgemeineres Level gehoben werden. Dieses Level orientiert sich an der Kennzeichnung qualitativer Forschung. „Die Beurteilungskriterien sollten (...) auf einer Einschätzung der allgemeinen Merkmale qualitativer Forschung beruhen“. (ebd.:248) Die zentralen Konzepte werden nun ohne den Anspruch auf Vollständigkeit aufgeführt.

Empirische Verankerung der Theorie und ihrer Bildung

Gewichtiges Merkmal qualitativer Sozialforschung ist die induktive Vorgehensweise. Obwohl die Konzepte von interner und externer Validität Elemente der Theoriebildung berücksichtigen, ist deren Grundorientierung hypothesenprüfend. Qualitative Methoden sind andererseits darauf aus, vom Besonderen auf das Allgemeine zu schließen, eine Theorie zu entwickeln. Kelle (1994) widmet sich in seinem Hauptwerk ausführlich dieser Thematik. Zur Diskussion steht die methodologische Zulässigkeit, die wissenschaftstheoretische Begründbarkeit und, nicht zuletzt, die Gültigkeit der gewonnenen Erkenntnisse. (vgl. Kelle 1994:11) Wie kann Theoriebildung empirisch begründet sein? Was geschieht zwischen der ersten Sichtung des Datenmaterials und der Theorienformulierung?

Der Autor gibt einen weiten Überblick über verschiedene Formen der Theoriebildung bzw. -konstruktion. Abgelehnt werden naive Vorstellungen von Induktivismus – nicht zu verwechseln mit induktivem Vorgehen. Der Induktivismus fällt zwar darunter, doch beinhaltet er besondere Konnotationen, in der Art des bereits beschriebenen Tabula-Rasa-Konzeptes. Die infallibilistische Annahme, Theorie könne unabhängig von menschlicher Erfahrung generiert werden, gilt heute als unhaltbar und ist nicht, wie häufig fälschlicherweise angenommen, unver-

zichtbarer Bestandteil induktiven Vorgehens. Die Alternative liegt einerseits in der Anwendung kodifizierter Verfahren, wie der Grounded Theory, der objektiven Hermeneutik oder der analytischen Induktion, welche standardisierte Vorgehensweisen zur Theorieentwicklung anbieten (vgl. ebd.) und auf bestimmten Pfaden die forschersche Beliebigkeit minimieren. Eine weitere Variante der Theorienbildung bietet der Abduktionsschluss in seinen diversen Spielarten, der jedoch methodologisch nicht unumstritten ist, aber in der Kombination mit anderen Verfahren gerne in Anschlag gebracht wird. Neben den differenziert ausgearbeiteten kodifizierten Verfahren existieren eine Reihe von Heuristiken der Theorienbildung, verschiedener Forschungstraditionen entstammend, in deren Mittelpunkt meist Kodierungsvorgänge stehen. (ebd.:368) Allen gemein ist ein Prozess der Anpassung der entstehenden Theorie an das Datenmaterial durch ständige Kontrastierung an Gegenevidenz im Material. Ständiger fallibilistisch orientierter Theorienwettbewerb unter Einbezug unstimmiger Elemente ist dabei ein wichtiger Qualitätssprung gegenüber sich monoton verteidigenden Verifizierungsversuchen.

Niemand kann erwarten, dass die Forschenden ausgewiesene Experten in der Thematik der Theorienbildung sind. Dennoch muss eine basale Auseinandersetzung mit den Möglichkeiten, Grenzen und Vorgehensweisen der Theorien-generierung der selbigen vorausgegangen sein. Besonders in diesem Punkt darf Gegenstandsangemessenheit nicht in Beliebigkeit ausarten und den Prozess der Gewinnung von Theorie aus dem Datenmaterial zu einem Akt der Willkür werden lassen. Eigene Durchführungsvarianten bedürfen der dezidierten Beschreibung und Erläuterung.

Triangulation

Seit Denzin (1970) den Begriff und die Methode der Triangulation durch eine ausführliche Beschreibung für die Sozialwissenschaften fruchtbar gemacht hat, erlebte die Methode eine breite Rezeption. Der aus der Geodäsie entlehnte Begriff beschreibt eine Situation der Multiperspektivität. Unter Rekurs auf basale Prinzipien der Geometrie erlauben mehrere Bezugspunkte eine genauere Ortsbestimmung. Der Gedanke ist, sich dem Forschungsgegenstand auf mehreren Wegen zu nähern und so ein vielseitiges Bild zu gewinnen. Die im deutschen Sprachraum zum Teil verbreitete Annahme, Triangulation bezeichne den gleichzeitigen Einsatz

qualitativer und quantitativer Methoden innerhalb eines Forschungsprojektes, spart einige andere Variationen aus. Insgesamt zählt Denzin vier Grundtypen: Triangulation der Daten, der Forscher, der Theorien und die methodologische Triangulation. (Denzin 1978:295ff.) Ein fünfter Typ ist die interdisziplinäre Triangulation. (Janesick 1994) Zunächst wurde Triangulation als eine Strategie konzipiert, die die Ergebnisse validieren sollte. Heute ist diese Bedeutung eher in den Hintergrund geraten, an ihrer Statt wird Triangulation eher zu einer Alternative dazu. (vgl. Flick 1992) „Der Fokus hat sich jedoch zunehmend in Richtung der Anreicherung und Vervollständigung der Erkenntnis und der Überschreitung der Erkenntnismöglichkeiten der Einzelmethode verlagert.“ (Flick 1999:250) Besondere Chancen liegen in der Zusammenführung von qualitativen Studien und quantifizierenden Untersuchungen. Als beispielhaft dafür gilt die Marienthal-Studie von Lazarsfeld, Jahoda et al. Erkenntnistheoretisch unhaltbar ist dagegen die Validierungsvermutung. Unter der Prämisse der Methodenabhängigkeit von Ergebnissen, können zwei zusammengeführte Untersuchungen nur unterschiedliche Ergebnisse liefern. Darüber hinaus machen zwei konvergierende Ergebnisse sich nicht gegenseitig valider, schließlich könnten beide mangelhaft sein. Anstatt nach Konvergenzen zu suchen, sollte das Augenmerk auf auftretende Divergenzen gerichtet sein, gerade diese können für die Theorieentwicklung überaus hilfreich sein. (Flick 2000:318)

Die größte Kehrseite triangulativer Forschungsdesigns ist die forschungsökonomische Dimension. Angesichts meist schmaler Budgets der Forschungsprojekte ist an eine redliche Umsetzung triangulativer Forschungsideale kaum zu denken. Zu bedenken bleibt auch noch die Auswirkung der Ungleichzeitigkeit zweier Untersuchungen und die dann entstehende Wandlung des Gegenstandes entsprechend dem Gegenstandsverständnis qualitativer Forschung. Somit präsentiert sich die Konzeption von Triangulation nicht als ein Patentrezept im denzinschen Sinne, sondern als förderliche Vorgehensweise wahlweise „als Validierungsstrategie, als Ansatz der Generalisierung der gefundenen Ergebnisse und als Weg zu zusätzlicher Erkenntnis.“ (ebd.)

Kommunikative Validierungsformen

In der entsprechenden Methodenliteratur wird bisweilen auf eine Möglichkeit hingewiesen die Gültigkeit von Daten oder Ergebnissen der Untersuchung auf

kommunikativem Weg zu erhöhen. Genau genommen unterscheidet sich diese Variante in zwei Unterarten. Die eine und meist einzige, die angeführt wird, bezeichnet eine kommunikative Rückkopplung mit den Beforschten. Das sogenannte member-checking versucht mittels im Nachhinein durchgeführter Erörterung der Äußerungen der jeweiligen Person dem subjektiven Sinn näher auf die Spur zu kommen und konfrontiert darüber hinaus die Teilnehmenden einer Untersuchung mit der Interpretation. (Seale 1999a:62) Fraglich ist jedoch, wie die beforchten Laien die professionalisierten Ausführungen validieren können; dabei ist nicht nur der Professionalisierungsgrad problematisch, sondern auch die tendenzielle Unfähigkeit von der eigenen Person zu transzendieren. Allemal förderlich ist die Vorgehensweise für das Verständnis der Beforschten und deren Äußerungen, Missverständnisse lassen sich eher ausräumen. Wesensmerkmal qualitativer Sozialforschung sollte die Orientierung am Alltagsgeschehen und bzw. oder dem Alltagswissen der Untersuchten sein. Member checking kommt diesem Anspruch sehr entgegen, indem es die jeweilige Person und deren Sichtweise über das obligatorische Maß hinaus in den Mittelpunkt der Forschung rückt.

Die zweite Form, das ‚peer debriefing‘ (vgl. Lincoln/ Guba 1985:308) stellt die Kommunikation in der Forschungsgruppe in den Mittelpunkt. Möglichst nicht an der gleichen Forschung arbeitende Experten sollten die Interpretationen diskutieren und ihre vielseitigen Perspektiven zu einem großen Panoramabild zusammenfügen.

Beide Varianten kommunikativer Validierungstechniken haben durchaus ihre Stärken und Vorteile, stehen und fallen aber mit der Qualität des kommunikativen Austausches. Diskussion als solche ist nicht zwingend von Nutzen, sie benötigt bestimmte Modi, die per Beschreibung im Forschungsbericht keinesfalls vermittelbar sind. Wünschenswert sind beide Techniken, evaluierbar nicht; das Augenmerk dieser Arbeit liegt auf der Dokumentation, nicht auf deren Bewertung.

Umgang mit Zahlen

Lange Zeit besaßen die quantifizierenden Verfahren das Monopol auf die Verwendung von Zahlenmaterial. „Zahlen und Wörter“ lautet der Titel einer Monographie (Erzberger 1998), der ausdrücken soll, was nicht selten als Merkmal der Unterscheidung herangezogen wird. Den einen die harten Fakten, den anderen die soften Erzählungen. Seale gibt einen systematischen Überblick über die vielseitigen Verwendungsmöglichkeiten von Zahlen in der qualitativen Sozial-

forschung. (vgl. 1999a:Kap.9) Die Losung lautet: „Zählen des Zählbaren!“ (ebd.:121) Die Angabe von Zahlen kann der Unterfütterung, Verdeutlichung oder dem Beleg einer Theorie oder einer Generalisierung dienen. „Wider den Anekdotismus!“ – (Seale 1999a:138) eine Art programmatische Parole, die die Funktion von Zahlenmaterial umreißt. Über das einfache Zählen von Häufigkeiten hinaus, werden Möglichkeiten aufgezählt, komplexere Modi anzuwenden, wie unter gegebenen Umständen Chi-Square-Darstellungen oder Clusteranalysen. Die neueren CAQDAS, wie NVivo oder WinMax bieten die Möglichkeit der Kopplung qualitativen und quantitativen Materials, beispielsweise durch die erleichterte Eingabe von Variablen der Textsegmente und deren Import-Export-Funktionen für die gängigen Statistikprogramme. Kuckartz (1995) beschreibt für die qualitative Analyse nutzbare statistische Verfahren, insbesondere für die Bildung von Idealtypen. Eine epistemologisch sinnvolle Integration birgt synergetische Effekte.

Der Umgang mit Zahlen kann aber auch missbräuchlich sein. Sind die notwendigen Anforderungen an die jeweiligen Verfahren nicht gegeben, so sind auch die Verfahren zu verwerfen. Vor vorschnellen Verallgemeinerungen, besonders auch bei meist kleinen Fallzahlen ist ausdrücklich zu warnen.

Zwischen diesen beiden Polen, den Anwendungschancen und den Beschränkungen liegt ein breites Potenzial, das häufig nicht abgerufen wird. Qualitative Forschung sollte sich einen offensiveren aber nicht blinden Umgang mit Zahlen angewöhnen, die Chance gesteigerter Erkenntnis nutzen, ohne zugleich Quantifizierung zum Selbstzweck hochzustilisieren.

2.6 Sonstige Kriterien

Neben den klassischen und den methodenangemessenen Gütekriterien gibt es weitere, die sich nicht direkt einer der beiden Gruppen zuordnen lassen. Diese sind an dieser Stelle unter der Rubrik ‚sonstige Kriterien‘ zusammengefasst. Abgehandelt werden die Frage der Wahl einer angemessenen Methode und forschungsethische Gesichtspunkte.

Indikation

Unter dem Begriff Indikation (Steinke 1999:215) wird in der Medizin die Veranlassung, einen bestimmten Heilungsweg einzuschlagen, verstanden. Im Nachhinein ist am Heilungserfolg überprüfbar, ob der gewählte Weg eher richtig oder eher falsch war. Übertragen auf die qualitative Sozialforschung bedeutet das äquivalent dazu, ob der richtige Weg zur Bearbeitung eines Gegenstandes eingeschlagen wurde. Nicht jede Vorgehensweise ist für jeden Gegenstand geeignet. Der Forschungszweig ist dadurch charakterisiert, dass der jeweilige Untersuchungsgegenstand und die Fragestellung die Auswahl oder Entwicklung der Forschungsmethode bestimmen. Umgekehrt ist natürlich gleichfalls der Gegenstand von der Wahl der Methode beeinflusst, was kein ureigenes Spezifikum der qualitativen Forschung ist. Dennoch geht im Gegensatz zur quantitativen Vorgehensweise, die Wahl des Gegenstandes der Wahl der Methode voraus, Zweitere ist von Ersterem abhängig. Gegenstandsangemessenheit stellt sicher, dass die subjektiven Perspektiven und alltäglichen Handlungsweisen der Untersuchten zur Geltung kommen können und nicht den Zwängen methodischer Strukturen zum Opfer fallen. Die unbedingte Bejahung von Gegenstandsadäquanz beruht aber auch auf der Realisierung des Prinzips der Offenheit, Theorien induktiv oder abduktiv zu generieren bzw. zu entwickeln und nicht nur bereits entwickelte Theorien zu prüfen. Somit ist Gegenstandsangemessenheit „nicht nur ein Kennzeichen qualitativer Forschung, sondern kann (...) gleichzeitig als Kriterium dafür betrachtet werden, inwiefern dieser Anspruch eingelöst wurde.“ (Steinke 1999:215)

Zunächst gilt dieser Anspruch für die Wahl einer qualitativen Vorgehensweise im Allgemeinen. Könnte womöglich ein quantitatives Vorgehen angemessener sein? Aussagen über eine Verteilung von Merkmalen in der Bevölkerung sind zum Beispiel nicht möglich. Sind dahingehende Zweifel ausgeräumt, stellt sich nunmehr die Frage nach der konkreten Wahl der Methode. Sind die Methoden zur Erhebung und Auswertung angemessen? (ebd.) Einzelne Methoden sind in ihrer Verwendbarkeit beschränkt, so sind die Grenzen des narrativen Interviews oder psychoanalytischer Verfahren recht limitiert. Gleiches gilt für eigens entwickelte Methoden. Die Begründung der Bewertung muss im Einzelfall logisch hergeleitet werden, was in der Außensicht auf Forschungen nicht problemlos zu leisten ist. Ferner müssen Methoden von einer Beschaffenheit sein, die Falsifizierbarkeit erlaubt und eine Irritation des Vorwissens ermöglicht. Dazu sei, und dies stellt eine

Art fundamentaler Grundausrichtung dar, eine abduktive Haltung nötig. (ebd.) Angemessen zu sein haben ebenfalls die verwendeten Transkriptionsregeln und die Fallauswahl, also die Samplingstrategie. Stichwortartig sind in diesem Zusammenhang nicht nur das theoretical sampling zu nennen, sondern darüber hinaus auch der Einbezug extremer oder abweichender Fälle oder aber für gewisse Fragestellungen auch die Herstellung eines möglichst homogenen Samples.

Die Beurteilung der Indikation ist sicherlich schwieriger als bei statistischen Verfahren, wo enggefasste, meist mathematische Bedingungen existieren, aber darum noch lange nicht unsinniger. Jedoch benötigt ihre Bewältigung ungleich größere methodische Kenntnis und reduziert sich nicht auf die Abarbeitung von Faustregeln. Letztlich rekurriert sie zum Teil auf epistemologische Abwägungen. Im Ganzen ist das Kriterium der Indikation angesichts der Kennzeichnung der hier vorliegenden Forschungspraktik von zentraler Bedeutung. Unter der Prämisse der Untrennbarkeit von Methode, Gegenstand und Forschungssubjekt kann der Stellenwert, den die gewählte Methode besitzt, nicht hoch genug gehängt werden. Gleichsam ist die Qualität der Forschung auf Gedeih und Verderb verquickt mit der Qualität der Indikation.

Ethik

Die bisherigen Kriterien bezogen sich allesamt auf eine vermeintlich korrekte Durchführung eines Forschungsprozesses. Eine nach Maßgabe wissenschaftlicher Standardisierung adäquate Umsetzung, die es vermochte, den Erkenntnisweg zu verwissenschaftlichen, kann allerdings an ethischen Beeinträchtigungen kranken. Mit anderen Worten, eine richtige Durchführung kann gleichzeitig schlecht sein, wenn ethische Maßstäbe nicht eingehalten werden. Beispiel dafür ist die von Stanley Milgram in den 60er Jahren des 20. Jahrhunderts durchgeführte Versuchsreihe (Milgram 1982), unter der Bezeichnung ‚Milgram-Experiment‘ bekannt. Aus heutiger Sicht eine methodisch glänzende Studie, die aber aus ethischen Gründen schon damals unter massive Kritik geriet und heute nicht mehr durchgeführt werden könnte, so dass neuere Replikationen mit einer entschärften Version des Versuchsaufbaus arbeiten. Unter dieser Perspektive ist die Einhaltung ethischer Standards unabdingbar, wenn eine Forschung höchsten Gütekriterien entsprechen soll. Hopf weist darauf hin, dass neben den normativen Gründen auch klare rechtliche Vorschriften existieren, die die Entwicklung ethischer Richtlinien nötig

gemacht haben. (2000:592) Grundlage für die Bewertung der ethischen Vertretbarkeit bilden die Ethikkodices der je bedeutendsten Fachverbände der beiden untersuchten Disziplinen. Auf Seiten der Erziehungswissenschaft ist dies die ‚Deutsche Gesellschaft für Erziehungswissenschaft‘ (DGfE) und auf Seiten der Soziologie die ‚Deutsche Gesellschaft für Soziologie‘ (DGS), die ihren Kodex zusammen mit dem ‚Berufsverband Deutscher Soziologinnen und Soziologen‘ (BDS) erarbeitet hat. Beide Kodices (Deutsche Gesellschaft für Erziehungswissenschaft 1999, Deutsche Gesellschaft für Soziologie/ Berufsverband Deutscher Soziologinnen und Soziologen 1993) unterscheiden sich nur in Details, die gemeinsame Schnittmenge der relevanten Aspekte liefert die Basis für die Formulierung der Leitfragen. Als solche erscheinen die Bereiche von Forschung, Publikation und der Rechte der Untersuchten. Hervorgehoben werden können zwei Prinzipien: erstens, das Prinzip der informierten Einwilligung und zweitens, das Prinzip der Nicht-Schädigung. (vgl. Hopf 2000) Beide sind besonders in der qualitativen Sozialforschung von Bedeutung, beispielsweise sind verdeckte Undercover-Untersuchungen resp. verdeckte teilnehmende Beobachtung á la Wallraff nicht mehr vorgesehen. Im 3.Kapitel sind die aus den Kodices abgeleiteten Resultate aufgeführt.

3. Die Kriterien

Die im Folgenden vorgeschlagenen Gütekriterien stellen eine zusammenfassende Ableitung aus den im 2. Kapitel beleuchteten Teilaspekten dar. Im einzelnen werden die daraus resultierenden ‚guiding ideals‘ in Form von Leitfragen formuliert, die später in der Analyse der Forschungsberichte abgearbeitet werden²³. Im darauffolgenden Abschnitt (3.1) werden die Kriterien thematisch gebündelt noch einmal übersichtlich aufbereitet. Besondere Beachtung gilt der Tatsache, dass sich die Leitfragen nicht auf die Beurteilung der gesamten Forschung beziehen, sondern lediglich auf die vorliegende Form von Zeitschriftenaufsätzen. Selbstverständlich müssen die Beurteilungskriterien geringfügig anders sein als bei umfassenden Monographien, die ungleich mehr Raum zur Darstellung bieten. Der Zwang zur verkürzten Darstellung nötigt zur Auslassung, enthebt aber nicht von gewissen Pflichten. Gerade in den von den Forschenden gewählten, also auch für wichtig befundenen Darstellungselementen dürfte sich aber eine ebenfalls im Forschungsprozess vorgenommene Akzentuierung widerspiegeln. Gleiches gilt ex negativo für nicht dargestellte Elemente. Summa summarum sollten Zeitschriftenaufsätze eher entgegenkommend beurteilt werden, kommt doch diese spezielle Darstellungsform der qualitativen Forschung alles andere als entgegen. Allgemein wird ihr zuerkannt, extensivere Darstellung zu benötigen, als das bei quantitativen Forschungsergebnissen der Fall ist.

Der hier herausgearbeitete Katalog von Gütekriterien ist bei weitem nicht der Erste. Im anglophonen Sprachraum gehen die Bemühungen schon länger in die Richtung, Güte qualitativer Forschung zu bestimmen. Smith-Sebasto (2000), und vor allem Reid und Gough (2000) geben einen aktuellen Überblick über den Stand der Entwicklung und vorhandene Kataloge. Der Trend der letzten Jahre scheint eine zunehmende Ausdifferenzierung und Autonomisierung gegenüber den quantitativen Methoden zu sein. Die Kriterien werden spezifischer methodologisch zugeschnitten

²³ Bei der Auswahl von Kriterien werden im Zweifelsfall eher bereits in der Literatur vorhandene Vorschläge adaptiert, wenn diese sich für eine Übernahme anbieten, als neue aufzustellen. Dem liegt der Gedanke zu Grunde, dass wenn deren einstmalige Postulierung keinen Widerspruch in der Rezeption ausgelöst hat, sie als gesteigert gültig anzusehen sind. Andererseits kann eine Modifikation oder eine Neuschöpfung ebenso nötig sein. Ausschlaggebend ist allein die inhaltliche Stimmigkeit.

und sind nicht mehr hauptsächlich den quantifizierenden Verfahren entlehnt. Das Selbstvertrauen der einst Gescholtenen ist im Aufschwung begriffen, der stabilisierende Diskurs ausdrücklich erwünscht (vgl. Blaxter 2000:34, Scott 2000) und in Gang gekommen. Die innere Struktur der Kataloge ist sehr vielfältig und zum Teil unübersichtlich, vermutlich tritt in den nächsten Jahren eine Kanonisierung in Abhängigkeit von den je unterstellten erkenntnistheoretischen Grundlagen ein. Der hier vorliegende Katalog reiht sich somit höchstwahrscheinlich ein, in eine noch große, aber sich stets verengende Vielzahl an Vorschlägen.

Das Konzept der **internen Validität** übertragen auf die qualitative Sozialforschung, beinhaltet trotz der zahlreichen Einschränkungen zahlreiche Implikationen, besonders für Aspekte der Theoriegenerierung. Aus den drei Vorschlägen zur Steigerung der internen Validität von Kelle, Kluge und Prein lassen sich folgende Leitfragen ableiten:

GIBT ES EINE ANGEMESSENE DISKUSSION VON EVIDENZ UND GEGENEVIDENZ?

Mit anderen Worten: ist auch abweichendes oder gar widersprechendes Datenmaterial bei der Hypothesengenerierung berücksichtigt worden und wird dieses gar angegeben?

IST BEI DER GENERIERUNG VON HYPOTHESEN DIE VORGEHENSWEISE VON MAXIMALER UND MINIMALER KONTRASTIERUNG ANGEWENDET WORDEN?

Das Verfahren ist nicht nur im strengen Sinne der Grounded Theory zu verstehen, sondern meint auch die gezielte Variation von Bedingungen, wie sie in Quasi-experimentellen Designs zu finden ist. Eine Assoziation von Variablen sollte auf diese Weise systematisch durch Variation untermauert werden.

WURDEN DISKURSIVE VALIDIERUNGSVERFAHREN DURCHGEFÜHRT?

Die Verfahren stehen und fallen mit der Qualität der angewandten Kommunikationsregeln im Diskurs. Für eine Beurteilung wäre deren Angabe unerlässlich, aber kaum angemessen zu leisten. In der Literatur ist diese Art von Verfahren nicht unumstritten, unter anderem weil die Wirkung sehr leicht in eine nicht gewollte Richtung gehen kann. Demzufolge nimmt diese Leitfrage eine nur periphere

Position ein, ihre Aufgabe ist eine dokumentierende, nicht eine bewertende. Evaluiert werden soll, wie verbreitet die Verfahren sind und welche Auswirkungen auf die Ergebnisse dingfest gemacht werden können.

WURDE CAQDA-SOFTWARE BEI DER INTERPRETATION EINGESETZT, WENN DIES SINNVOLL WAR?

Die Empfehlung zum Einsatz von Computern in der qualitativen Sozialforschung bekommt transformiert die Faustregel zugesprochen: Wenn ein Einsatz sinnvoll ist, dann ist der Einsatz besser als kein Einsatz.

Das Konzept der **externen Validität** beinhaltet ebenfalls relevante Implikationen für ein qualitatives Vorgehen. Generalisierbarkeit bzw. Verallgemeinerbarkeit sind da wie dort ubiquitär. Modifiziert für die Methode sind deren Imperative nicht sonderlich verschieden, der Vorgang der Übertragung bleibt identisch. Allein die Voraussetzungen der Übertragungsleistung divergieren und machen eine Neuformulierung nötig. Zur Steigerung der externen Validität werden folgende Leitfragen vorgeschlagen:

KANN, SOLL UND DARF ÜBERHAUPT GENERALISIERT WERDEN?

Diese Frage zielt ab auf den logischen Umgang mit der Generalisierung, nicht auf die Güte der Umsetzung. Ist es im jeweiligen Falle denklogisch überhaupt möglich zu generalisieren; verlangt eine je spezifische Art der Forschung vielleicht gar keine Verallgemeinerung und ist sie dort von Nöten, wo sie unterblieb?

IST DIE AUSWAHL DER UNTERSUCHUNGSEINHEITEN THEORETISCH GERECHTFERTIGT?

Es muss deutlich sein, nach welchen Kriterien die Auswahl erfolgte, möglichst sollte sie mit Hilfe der gängigen Theorien begründet sein. Ferner muss klar sein, auf welche Grundgesamtheit sie sich bezieht.

SIND DER SENDE- UND EMPFÄNGERKONTEXT IM FALLE VON GENERALISIERUNGEN ANGEMESSEN BESCHRIEBEN (,THICK DESCRIPTION')?

Im Hinblick auf relevante Kontextmerkmale muss eine Auslotung vorliegen, die eine Übertragung rechtfertigt. Die Kontextmerkmale müssen vergleichbar sein.

IST DER EINFLUSS DER UMGEBUNG UND SITUATION DER UNTERSUCHUNG BERÜCKSICHTIGT UND EVENTUELL VARIERT WORDEN?

Die Reaktivität der angewandten Verfahren kann auch bei Methoden von hoher ökologischer Validität von Bedeutung sein. Dementsprechend muss der Blick auf mögliche Verzerrungen jener Art wachsam gerichtet sein und gegebenenfalls müssen die maßgeblichen Faktoren variiert werden, um den ihnen eigenen Konsequenzen auf die Spur zu kommen.

Das Konzept von Objektivität wird hier verstanden im Sinne von **Intersubjektivität**, operationalisiert durch intersubjektive Nachvollziehbarkeit. Wie aufgezeigt wurde, ist maßgebliche Anwendung in der qualitativen Forschung anzuraten.

IST DER FORSCHUNGSPROZESS IN EINEM AUSMAß DOKUMENTIERT WORDEN, DASS EINE NACHVOLLZIEHBARKEIT DURCH LESENDE MÖGLICH IST?

INSBESONDERE ZÄHLEN DAZU:

DOKUMENTATION ...

I. DES VORVERSTÄNDNISSES

II. DER ERHEBUNGSMETHODEN UND DES -KONTEXTES

III. DER TRANSKRIPTIONSREGELN

IV. DER DATEN

V. DER AUSWERTUNGSMETHODEN

VI. DER INFORMATIONSQUELLEN

VII. VON ENTSCHEIDUNGEN UND PROBLEMEN

Neben der Beschreibung der je gewählten Richtung, in welche eine Entscheidung fiel, ist jeweils möglichst darzulegen, unter welcher Maßgabe diese gefällt wurde. Insbesondere auch für die Nachvollziehbarkeit von Theoriegenerierung und -prüfung ist Dokumentation unerlässlich. Steinke bezeichnet dieses Kriterium als „Hauptkriterium bzw. als Voraussetzung zur Prüfung anderer Kriterien“. (1999:209)

WURDE DER EIGENE STANDPUNKT DES AUTORS ODER DER AUTORIN KLAR DARGELEGT?

Ist die Perspektive der Forschenden beschrieben, wurden die eigene Rolle und mögliche Einflüsse auf die Untersuchung überprüft? Sehr wichtig ist die Überprüfung der persönlichen Voraussetzungen für eine Untersuchung und die biographische Beziehung zum Forschungsthema.

WURDE DER FORSCHUNGSPROZESS STÄNDIG DURCH GEZIELTE REFLEXIONEN BEGLEITET?

Der gesamte Ablauf der Forschung bis hin zum Abfassen der Publikation sollte von Selbstbeobachtung und Reflexion begleitet sein. Der ständige Zwang, forschungstragende Entscheidungen zu fällen, nötigt zur ständigen Verortung der eigenen Subjektivität. Frühe wie späte Entscheidungen unterliegen den selben Beeinflussungen. Nur eine Einbettung in den eigenen Kontext kann einen Forschungsprozess intersubjektiv nachvollziehbar gestalten.

Methodenangemessene oder alternative Gütekriterien wurden eigens für die qualitative Sozialforschung entwickelt und beziehen sich auf deren Eigenarten und Kennzeichnung. Sie sind nicht übertragbar auf eine quantitative Vorgehensweise und haben bisher innerhalb ihrer Methodologie leider nicht in gleicher Weise eine breite Einigkeit hervorgerufen, wie dies die quantifizierenden Gütekriterien getan haben. Validität und Reliabilität stehen dort eigentlich nicht mehr zur Diskussion.

WIE SYSTEMATISCH IST DIE GENERIERTE THEORIE AUS DEM DATENMATERIAL GEWONNEN WORDEN?

- I. **SIND KODIFIZIERTE VERFAHREN ANGEWENDET WORDEN? UND WENN JA, MIT WELCHER GÜTE?**
- II. **WENN NICHT: WURDE NACH ANDEREN HERRSCHENDEN REGELN DER THEORIEBILDUNG VORGEGANGEN (VGL. KELLE 1994)?**
- III. **ODER WURDE IM FALLE EINER EIGENEN VORGEHENSWEISE DIESER DEZIDIERT ERLÄUTERT?**

Insbesondere die Theoriebildung bedarf einer komplexen Technik, um sich vom Vorwurf der Beliebigkeit freizumachen. Hier sollte strengstens auf Systematisierung geachtet werden. Bedingung ist ein grundsätzlich fallibilistisch orientierter Ansatz, der die Vorannahmen bzw. die eigene Theorie immer potenziell in Frage stellt.

WURDE EINE TRIANGULIERENDE PERSPEKTIVENVIELFALT IN IRGEND EINER IN FRAGE KOMMENDEN WEISE IN DEN FORSCHUNGSPROZESS EINGEFLOCHTEN?

Triangulation ist nicht als notwendige Bedingung für hohe Qualität von Forschungen zu verstehen. Es sind andererseits kaum Situationen denkbar, in denen ein Bezug auf nicht mindestens eine der fünf Formen, samt ihren Sub-Formen sinnvoll wäre.

Folglich ist eine triangulative Sichtweise in nahezu allen Fällen bereichernd aber zugleich schwer durchführbar.

WURDEN QUANTIFIZIERUNGEN ADÄQUAT IN DEN FORSCHUNGSPROZESS EINGEFLOCHTEN?

Adäquat soll bedeuten, dass einerseits die Möglichkeiten ausgeschöpft, aber nicht überansprucht worden sind. Zahlenbenutzung wird nicht zum Selbstzweck, sondern zum probaten Hilfsmittel.

Die unter der Rubrik der **sonstigen Kriterien** aufgeführten Punkte der Indikation und der Forschungsethik sind durch folgende Leitfragen überprüfbar:

IST DIE VORGEHENSWEISE DEM GEGENSTAND ANGEMESSEN? INSBESONDERE:

- I. KÖNNTE EIN QUANTITATIVER ANSATZ DIE THEMATIK BESSER ERREICHEN?**
- II. IST DIE SPEZIELLE METHODENWAHL DEM GEGENSTAND ANGEMESSEN, SEI SIE KODIFIZIERTER NATUR ODER SELBST ENTWICKELT?**
- III. IST DIE SAMPLINGSTRATEGIE DEM GEGENSTAND ANGEMESSEN?**
- IV. IST DIE WAHL DER TRANSKRIPTIONSREGELN DEM GEGENSTAND ANGEMESSEN?**

Wie bereits oben dargestellt, ist die Bedeutung der gewählten Vorgehensweise ebenso hoch, wie schwer zu beurteilen. Im Prinzip gibt es nicht den ultimativen korrekten Weg, sondern lediglich einen angemessenen, einen weniger angemessenen aber dennoch ebenfalls einen offensichtlich überhaupt nicht angemessenen. Externe Bewertbarkeit beruht auf der Verstehensweise der Darstellung. Bestehende Zweifel könnten gegebenenfalls in der Erörterung mit den Forschenden ausgeräumt werden. Somit kann eigentlich die Einordnung nur vorläufig sein, aber im Rahmen dieser Arbeit nicht optimiert werden.

WURDE DIE UNTERSUCHUNG OHNE VERSTÖßE GEGEN ETHISCHE RICHTLINIEN DURCHFÜHRT? INSBESONDERE:

- I. WURDE DIE EINWILLIGUNG ZUR TEILNAHME AN EINER UNTERSUCHUNG VON DEN BEFORSCHTEN EINGEHOLT?**
- II. WURDEN DIE BEFORSCHTEN ÜBER ZIELE UND METHODEN DES FORSCHUNGSVORHABENS INFORMIERT?**

[DIE EINHALTUNG DER PUNKTE I. UND II. KANN AUF GRUND EINER BEFÜRCHTETEN FEHLERWIRKUNG IM ÄUßERSTEN NOTFALL IM NACHHINEIN NACHGEHOLT WERDEN.²⁴]

III. WURDE DIE ANONYMITÄT GEWART?

IV. KÖNNTEN DURCH DIE FORSCHUNG DEN BEFORSCHTEN NACHTEILE ERWACHSEN (ÜBER RISIKEN IST AUFZUKLÄREN)?

V. WURDEN ALLE GELDQUELLEN BENANNT, WENN TEILE DER FORSCHUNG AUS DRITTMITTELN FINANZIERT WURDEN?

Wenn bereits in einer der fünf Fragen Bedenken gegen die ethische Lauterkeit auftritt, kann der Gesamtaspekt Ethik nicht mehr positiv beschieden werden. Es ist jedoch ausgesprochen schwierig, einen Verstoß konkret nachzuweisen, weil ein solcher meist dann nicht auffällt, wenn er verschwiegen wird. Dennoch soll die Messlatte einer ethisch vertretbaren Forschung ihrer Erwünschbarkeit wegen bestehen bleiben.

3.1 Der Kriteriumskatalog

Die oben erarbeiteten Kriterien zur Beurteilung werden in diesem Abschnitt zu einem Katalog zusammengeführt. Sie gliedern sich in vier Bereiche: Methode, Ausführung, Analyse und Präsentation. Obwohl sich der Forschungsablauf qualitativer Sozialforschung nicht eindeutig in Phasen unterteilen lässt, weil die Vorgehensweise zirkulär angelegt ist, Abschnitte sich mitunter eng verzahnen, ist die Vierteilung wegen der dennoch gegebenen deutlichen Funktionstrennung statthaft. Zudem bietet sie eine praktische Übersichtlichkeit, die in den vorhandenen Katalogen bislang vergeblich gesucht wird. Um darüber hinaus einer Verwirrung im späteren Darstellungsverlauf dieser Arbeit vorzubeugen, werden die Kriterien in der Reihenfolge ihrer Auflistung durchnummeriert.

Dementsprechend gestaltet sich der Kriteriumskatalog folgendermaßen:

²⁴ Hopf (2000:600) empfiehlt bei Eintreten eines solchen Falles, sich zuvor an den Ethikrat zu wenden, um das Forschungsdesign durch dessen Einwilligung legitimieren zu lassen.

METHODE

1. Ist die Vorgehensweise dem Gegenstand angemessen? Insbesondere:
 - I. Könnte ein quantitativer Ansatz die Thematik besser erreichen?
 - II. Ist die spezielle Methodenwahl dem Gegenstand angemessen, sei sie kodifizierter Natur oder selbst entwickelt?
 - III. Ist die Samplingstrategie dem Gegenstand angemessen?
 - IV. Ist die Wahl der Transkriptionsregeln dem Gegenstand angemessen?
2. Wurde eine triangulative Perspektivenvielfalt in irgendeiner in Frage kommenden Weise in den Forschungsprozess eingeflochten?
3. Wurden diskursive Validierungsverfahren durchgeführt?

AUSFÜHRUNG

4. Ist die Auswahl der Untersuchungseinheiten theoretisch gerechtfertigt?
5. Wurde der Forschungsprozess ständig durch gezielte Reflexionen begleitet?
6. Ist der Einfluss der Umgebung und Situation der Untersuchung berücksichtigt und eventuell variiert worden?
7. Wurden Quantifizierungen adäquat in den Forschungsprozess eingeflochten?
8. Wurde die Untersuchung ohne Verstöße gegen ethische Richtlinien durchgeführt? Insbesondere:
 - I. Wurde die Einwilligung zur Teilnahme an einer Untersuchung von den Beforschten eingeholt?
 - II. Wurden die Beforschten über Ziele und Methoden des Forschungsvorhabens informiert?
 - III. Wurde die Anonymität gewahrt?
 - IV. Könnten durch die Forschung den Beforschten Nachteile erwachsen (über Risiken ist aufzuklären)?
 - V. Wurden alle Geldquellen benannt, wenn Teile der Forschung aus Drittmitteln finanziert wurden?

ANALYSE

9. Kann, soll und darf überhaupt generalisiert werden?
10. Sind der Sende- und Empfängerkontext im Falle von Generalisierungen angemessen beschrieben („thick description“)?

11. Wie systematisch ist die generierte Theorie aus dem Datenmaterial gewonnen worden?
 - I. Sind kodifizierte Verfahren angewendet worden? Und wenn ja, mit welcher Güte?
 - II. Wenn nicht: wurde nach anderen herrschenden Regeln der Theoriebildung vorgegangen (vgl. Kelle 1994)?
 - III. Oder wurde im Falle einer eigenen Vorgehensweise diese dezidiert erläutert?
12. Ist bei der Generierung von Hypothesen die Vorgehensweise von maximaler und minimaler Kontrastierung angewendet worden?
13. Gibt es eine angemessene Diskussion von Evidenz und Gegenevidenz?
14. Wurde CAQDA-Software bei der Interpretation eingesetzt, wenn dies sinnvoll war?

PRÄSENTATION

15. Ist der Forschungsprozess in einem Ausmaß dokumentiert worden, dass eine Nachvollziehbarkeit durch Lesende möglich ist?
Insbesondere zählen dazu:
Dokumentation ...
 - I. des Vorverständnisses
 - II. der Erhebungsmethoden und des -kontextes
 - III. der Transkriptionsregeln
 - IV. der Daten
 - V. der Auswertungsmethoden
 - VI. der Informationsquellen
 - VII. von Entscheidungen und Problemen
16. Wurde der eigene Standpunkt des Autors oder der Autorin klar dargelegt?

4. Auswertung der Zeitschriftenartikel

Über die Ausführungen des 1. Kapitels hinausgehend, folgt nun zunächst eine genauere Beschreibung der Stichprobe hinsichtlich ihrer Entstehung und ihrer Beschaffenheit (4.1). Danach wird die Umsetzung oder Nicht-Umsetzung der qualitativen Standards überprüft, indem der Verteilung der Merkmale der zentralen Gütekriterien nachgegangen wird (4.2). Eine genauere Analyse auf höherem Komplexitätsniveau schließt sich unmittelbar an, um einige spezielle Ergebnisse genauer zu beleuchten (4.3).

4.1 Die Stichprobe

Die Stichprobe der untersuchten Zeitschriftenaufsätze wurde in ihrem Umfang begrenzt durch den strukturellen Rahmen der vorliegenden Arbeit. Die Anzahl der zu beurteilenden Aufsätze sollte möglichst hoch, gleichzeitig aber noch zu bewerkstelligen sein. Je 30 Fälle sollten allerdings für die Erziehungswissenschaft und die Soziologie nicht unterschritten werden, insgesamt würde die Stichprobe also 60 Fälle umfassen. Diese Anzahl erwies sich vom Arbeitsumfang als machbar und wurde schließlich auch realisiert. Entnommen werden sollten die Aufsätze möglichst renommierten Fachzeitschriften der Disziplinen, die erwiesenermaßen ein gutachterliches Beurteilungsverfahren der Veröffentlichung vorschalten. Damit kann eine jede Aussage der im Rahmen dieser Untersuchung ermittelten Ergebnisse insofern über die Stichprobe hinaus interpretiert werden, als dass das, was die Begutachtenden für veröffentlichenswert befanden, dem jeweils persönlich für erwartbar gehaltenen Standard an Qualität entsprechen dürfte. Somit geht die Vorstellung dieser Personen von guter Forschung implizit mit in die Untersuchung ein und erweitert das Bild von einer Disziplin.

In *Tabelle 1* sind die ausgewählten Zeitschriften und die Anzahl der daraus entnommenen Texte ersichtlich. Der Referenzzeitraum, der dem oben beschriebenen Anspruch der Aktualität entsprechen sollte, wurde auf die Jahrgänge 1998 und folgende festgelegt. Die Entnahmezahlen entsprechen also der Anzahl der

erschiedenen Aufsätze mit qualitativ-empirischem Inhalt im Referenzzeitraum. Für die beiden zentralen Repräsentanten der Disziplin Soziologie, die Kölner Zeitschrift für Soziologie und Sozialpsychologie (KZfSS) und die Zeitschrift für Soziologie (ZfS) wurde der Zeitraum um den Jahrgang 1997 erweitert, um wegen deren hoher Bedeutung für ihr Fach, deren zuvor geringe Fallzahl zu erhöhen. Die Zeitschrift für Erziehungswissenschaft (ZfE) und die Zeitschrift für qualitative Bildungs-, Beratungs- und Sozialforschung (ZBBS) dagegen sind erst jüngeren Alters und konnten erst ab dem Jahrgang 1998 bzw. 2000 berücksichtigt werden.

Zeitschriftentitel	Anzahl der Aufsätze		
	Erziehungswissenschaft	Soziologie	Gesamt
Zeitschrift für Soziologie der Erziehung und Sozialisation (ZSE)	16	6	22
Zeitschrift für Soziologie (ZfS)	-	11	11
Zeitschrift für Biographieforschung und Oral History (BIOS)	2	7	9
Zeitschrift für Erziehungswissenschaft (ZfE)	6	-	6
Zeitschrift für Pädagogik (ZfP)	5	-	5
Kölner Zeitschrift für Soziologie und Sozialpsychologie (KZfSS)	-	4	4
Zeitschrift für qualitative Bildungs-, Beratungs- und Sozialforschung (ZBBS)	1	2	3
Σ	30	30	60

Tabelle 1: Herkunft der Texte

Innerhalb des untersuchten Zeitraumes stammen vier Aufsätze aus dem Jahr 1997, 16 Aufsätze aus 1998, 15 aus 1999. Die meisten, 19 Aufsätze, wurden im Jahre 2000 veröffentlicht und sechs Texte des Jahrganges 2001 konnten berücksichtigt werden.

Die Länge der Texte schwankt zwischen 8 und 41 Seiten, das arithmetische Mittel liegt bei 19,0 Seiten, der Modus liegt bei 18. Es muss dabei bedacht werden, dass die Länge eines Textes mittels dieses Erfassungsmerkmals nur ungenau erfasst

werden kann und wesentlich abhängig von der Formatierung der Seite ist. Eine „je-mehr-desto“-Aussage ist mit Hilfe dieser Variable dennoch möglich.

Die altbekannte Aussage, qualitative Methoden seien eine Sache von Frauen resp. Forscherinnen (thematisch dazu: Sturm 1994), denen das vermeintlich „weiche“ Datenmaterial und die Interpretation sehr entgegenkämen, kann in der Stichprobe nicht bestätigt werden. Wie *Tabelle 2* zu entnehmen ist, sind knapp die Hälfte der Aufsätze von Männern durchgeführt worden, von Frauen dagegen nur gut ein Drittel.

Angaben in %	Anzahl der Aufsätze		
	Erziehungswiss.	Soziologie	Gesamt
Durchgeführt von...			
... Forscherinnen	40	34	37
... Forschern	43	53	48
... Forscherinnen & Forschern	17	13	15
Σ	100	100	100

Tabelle 2: Geschlechterbeteiligung an der Forschung

Einerseits könnte daraus gefolgert werden, dass qualitative Forschung kein überwiegend von Frauen präferierter Ansatz ist und bei Männern in gleicher Weise beliebt ist. Andererseits spricht die im Bereich der öffentlichen Forschung immer noch existierende Männerdominanz dagegen. Folgerichtig müsste dann nämlich ebenfalls der Anteil von Männern im vorliegenden Bereich überproportional hoch sein. Die beiden gegenläufigen Tendenzen können hier jedoch nicht weiter aufgeschlüsselt werden²⁵. Recht auffällig ist dagegen die Ungleichverteilung im Disziplinenvergleich. Frauen sind häufiger in der Erziehungswissenschaft aktiv, Männer eher in der Soziologie – auch diese Aussage wird hier im begründungsleeren Raum verbleiben.

²⁵ Die Autorinnen und Autoren der Stichprobe lassen sich kaum einer exakten Grundgesamtheit zuordnen. Neben institutionsgebundenen Personen treten scheinbar auch Privatiere, sowie in beiden Disziplinen nicht-fachspezifisch ausgebildete Personen und nicht in den deutschen Staatsgrenzen tätige Personen schreibend in Erscheinung. Daher kann auch nicht genau bestimmt werden, ob die Geschlechterverteilung der qualitativ Tätigen, derjenigen in darüber hinausreichenden Bereichen entspricht.

Letztlich ergibt sich innerhalb des festgelegten Auswahlkriteriums, wonach eine Forschungsarbeit dann berücksichtigt werden soll, wenn sie Anteile qualitativer Forschung beinhaltet, noch das Phänomen, dass dieser Anteil sehr unterschiedlich groß ausfallen kann. In vielen Forschungsdesigns werden verschiedenen Verfahren im Sinne der Triangulation miteinander verknüpft. In der Stichprobe kommt dies darin zum Ausdruck, dass auf der einen Seite in 48 Aufsätzen, (80%) ein oder mehrere qualitative Verfahren zum Einsatz kommen, diese also den gesamten Raum einnehmen. Auf der anderen Seite stehen vier Texte (7%), in denen eben solche Verfahren nur eine periphere Stellung inne haben, die von quantifizierendem Vorgehen dominiert sind. In acht Texten (13%) halten sich die Anteile beider Verfahrensmöglichkeiten in etwa die Waage. Trotz dieser vorhandenen Spannweite sind alle Aufsätze völlig egalitär in den Fundus aufgenommen worden.

Welche Forschungsziele werden mit qualitativen Methoden angestrebt?

Die Ziele waren überraschend vielfältig und sind in *Tabelle 3* angeführt.

Forschungsziel	Anzahl absolut
Exploration	37 (62%)
Theorienabarbeitung	19 (32%)
Typologisierung	17 (28%)
Theoriengenerierung	15 (25%)
Hypothesenprüfung	12 (20%)
Methodenentwicklung	4 (7%)
Verstehen	4 (7%)
Illustration	3 (5%)
Σ	111

Tabelle 3: Forschungsziele

Die hohe Anzahl kommt durch das häufige Anstreben mehrerer Ziele, in der Aufzählung sind Mehrfachnennungen enthalten, durchschnittlich 1,85 pro Untersuchung. Die Exploration eines für unbekannt gehaltenen Feldes, dem klassischen Kern qualitativer Forschung, steht ganz oben auf der Liste der Ziele. Allerdings unterscheiden sich die daraus resultierenden Folgerungen. Nur 15 Mal

soll aus dem gewonnenen Datenmaterial eine Theorie erzeugt werden. Eine Theorienaberration, das heißt eine Modifikation einer bereits bestehenden Theorie anhand der neuen Erkenntnisse streben 19 Forschungsdesigns an. Im Endprodukt möchten 17 Forschende eine Typologie gewonnen haben und vier geben sich mit dem Verstehen eines Phänomens zufrieden.

Irritierend ist die in zwölf Fällen angestrebte Hypothesenprüfung. Im Einzelfall sollte dann überlegt werden, ob qualitative Methoden ein solches Ziel nicht besser erreichen könnten. Das Gleiche gilt für die drei illustrativ eingesetzten Forschungen. Zum Zwecke der Illustration bedarf es nicht unbedingt des ausgefeilten qualitativen Instrumentariums resp. sollte kaum das Etikett ‚empirisch‘ angebracht werden.

Interessant sind die Forschungsarbeiten, die sich der Methodenentwicklung verschreiben. Diese ist in allen Fällen nicht der einzige Zweck der Untersuchung, aber die Forschenden verstehen es, nebenher noch gezielte methodenbezogene Beobachtungen einzubinden und diese im Forschungsbericht anschließend zu erwähnen.

Hinsichtlich des Forschungsgegenstandes in den Texten, das heißt, der Thematik, die beforscht wurde, lassen sich nur bedingt Trends ablesen. Das Spektrum der Themen ist sehr weit, auch wenn eine gewisse Häufung wegen der Thematischen Zuschneidung mancher Zeitschriften zu erwarten war. Dennoch ist es ein möglicher Indikator dafür, was derzeit einer induktiven Vorgehensweise bedarf, also eher unbekannt zu sein scheint.

Die häufigsten Themen	Anzahl (abs.)
Gender-Thematiken	7
Schule	7
Kindheit und Jugend	6
Migranten und Migrantinnen	5
Erwachsene n-Kind-Beziehungen	5
Ost-/Westdeutsche Vereinigung	4
Erwerbsarbeit	4
Religion/ Glaube	3
Fernsehwirkung	2

Tabelle 4: Themenübersicht der Aufsätze

Neben den disziplinspezifischen Standardthemen ist die Vielzahl der Aufsätze mit geschlechtsspezifischer Perspektive, sowie der Fokus auf Migrantinnen und Migranten in Deutschland überraschend. Das dritte zu explorierende Feld scheint die deutsche Vereinigung und ihre sozio-psychischen Folgen bzw. die in den Augen der meist westdeutschen Forschenden fremde DDR zu sein.

Kann bei der vorliegenden Stichprobe von Repräsentativität gesprochen werden?

Für einen inferenzstatistischen Schluss von der Stichprobe auf die Grundgesamtheit wäre es von Nöten, die Repräsentativität gewisser relevanter Merkmale der Stichprobe sicherzustellen. In unserem Falle sind die Einheiten der Stichprobe bewusst ausgewählt worden und damit nicht repräsentativ. Aber nicht immer ist die Stichprobenvalidität am besten durch eine Zufallsauswahl zu garantieren. (vgl. Prein/ Kluge/ Kelle 1994:5ff.) Im vorliegenden Falle nämlich ist eine Vielzahl von Bedingungen festgelegt worden, die die Stichprobenelemente erfüllen mussten. Eine Verallgemeinerung kann ohnehin nur auf eine Grundgesamtheit vorgenommen werden, die ebenfalls diese Bedingungen erfüllt. Unter Berücksichtigung des definierten Zeitraumes ist dann aber die Grundgesamtheit gar nicht Mal viel größer als die Stichprobe, die lediglich einige Aufsätze in kleineren oder fachlich hochdifferenzierten Zeitschriften nicht umfasst. Vor dieser Folie steht die Stichprobe praktisch für sich, behauptet aber nicht, universelle Aussagen machen zu können, schon gar nicht über andere Veröffentlichungsweisen, wie beispielsweise Monographien. Wie bereits erläutert (Kap. 1), bestehen für diese völlig andere Voraussetzungen, sie sollten eher von höherer Qualität sein, weil sie schlichtweg mehr Mitteilungsraum einnehmen können.

4.2 Die vorgefundene Qualität

Bevor der Einstieg in die Analyse beginnen kann, erfolgen noch einige Anmerkungen über das Für und Wider subjektiver Bewertungen.

Im Idealfall sollte eine Mehrzahl an geschulten Beurteilenden zur Verfügung stehen, die unabhängig voneinander arbeiten. Eine hohe Intercoder-Reliabilität stünde dann für eine hohe einheitliche Beurteilung. Im vorliegenden Fall musste eine Person

genügen, die einfließende Subjektivität ist entsprechend größer. Das Unterfangen kann aber dennoch einigermaßen gelingen, wenn die Beurteilungskriterien ausreichend deutlich feststehen und in der Rahmung ihrer semantischen Bedeutung Anwendung finden. Die Beurteilungskriterien werden im folgenden jeweils in den zugehörigen Text eingeflochten, damit den Lesenden die Einordnung der Ergebnisse erleichtert wird.

In diesem Abschnitt wird eine homogene Beschaffenheit der Stichprobe unterstellt. Der Korpus qualitative Forschung wird als einheitlich begriffen, um die Standortbestimmung der Forschungsrichtung ganzheitlich darstellen zu können. Dabei treten eventuell trennende Variablen, wie die Disziplinzugehörigkeit, die Länge eines Textes oder gar das Geschlecht bewusst in den Hintergrund. Die folgenden Ergebnisse können interpretiert werden als Zustandsbeschreibung qualitativer Forschung.

Methode

1.

In der Untersuchung wurde zunächst die im Forschungsprozess gewählte Methode erhoben. Nahezu alle denkbaren Varianten sind vertreten (siehe *Tabelle 5*).

Die am häufigsten verwendeten Methoden (mehrfache Methodenverwendung möglich)	Anzahl
Selbst entwickelte Verfahren	18 (30%)
Ethnographische Verfahren	12 (20%)
Grounded Theory	7 (12%)
Diverse hermeneutische Verfahren (außer: Objektive Hermeneutik)	7 (12%)
Biographisch orientierte Verfahren	5 (8%)
Objektive Hermeneutik	5 (8%)
Qualitative Inhaltsanalyse	3 (5%)

Tabelle 5: Häufigkeiten der gewählten Methoden

Immerhin fast ein Drittel der Untersuchungen verwendete eigens entwickelte Verfahren. Das bedeutet, dass die Gutachter und Gutachterinnen der Zeitschriften

diese durchaus zu akzeptieren scheinen und nicht nur approbierten Verfahren ihr Vertrauen schenken.

Die erste Leitfrage untersucht, ob die Wahl einer quantitativen Verfahrensweise das Phänomen hätte besser erfassen können. In 55 Fällen (92%²⁶) konnte dies eindeutig verneint werden, in fünf Fällen (8%) dagegen war der Vorzug einer qualitativen Methode nicht offensichtlich, in nur einem davon war die Anwendung völlig unangebracht.

Etwas kleiner ist die Gruppe der Texte, in denen zweifelsohne eine dem Gegenstand angemessene Methode verwendet wurde (41 Fälle/ 68%). Diese nicht leicht überprüfbare Anforderung konnten 17 Fälle (28%) nur mit Abstrichen erfüllen. Das heißt, die Methode war zwar adäquat, eine andere hätte es aber möglicherweise in einem höheren Maß sein können. In zwei Untersuchungen wurden unangemessene Methoden eingesetzt, die den Gegenstand nicht auf die gewünschte Weise erfassen konnten.

Noch größere Probleme bereitete die Samplingstrategie, nur in 55% der Untersuchungen (33 Mal) war sie nicht zu beanstanden. In 27% (16 Untersuchungen) wies sie kleinere Mängel auf und 12% der Strategien waren stark mangelhaft. In vier Texten wurde sie erst gar nicht erwähnt bzw. war in keiner Weise ersichtlich.

Die Transkriptionsregeln fristen ein vernachlässigtes Dasein. Viele Forschende scheinen diese für nicht besonders wichtig zu halten und würdigen sie keines Wortes. Abgesehen von acht Forschungsdesigns, die ohne jegliche Transkription auskamen und daher ausgeblendet werden können, ist nur in 24 der restlichen Fälle eine Beurteilung durch Betrachtung der direkten oder indirekten Beschreibungen der Autorinnen und Autoren möglich. Zumindest 18 Regelwerke (35%) sind völlig angemessen, die anderen sechs (12%) mit kleineren Abstrichen ebenfalls.

In *Tabelle 6* sind die Beurteilungen summiert dargestellt, innerhalb derer zunächst nichts zu bemängeln war. Dabei ist zu erkennen, dass nur 17% der Untersuchungen dem absoluten Ideal entsprechen. 44% weisen dagegen große und größte Mängel auf.

²⁶ Alle folgenden Prozentangaben sind auf ganze Zahlenwerte gerundet, die Darstellung von Nachkommastellen wird im vorliegenden Zusammenhang als irrelevant eingeschätzt.

Anzahl der Angemessenheitskriterien, die völlig adäquat sind	Fälle
Alle Vier	10 (17%)
Drei	24 (39%)
Zwei	18 (30%)
Eins	7 (12%)
Keine	1 (2%)
Σ	60

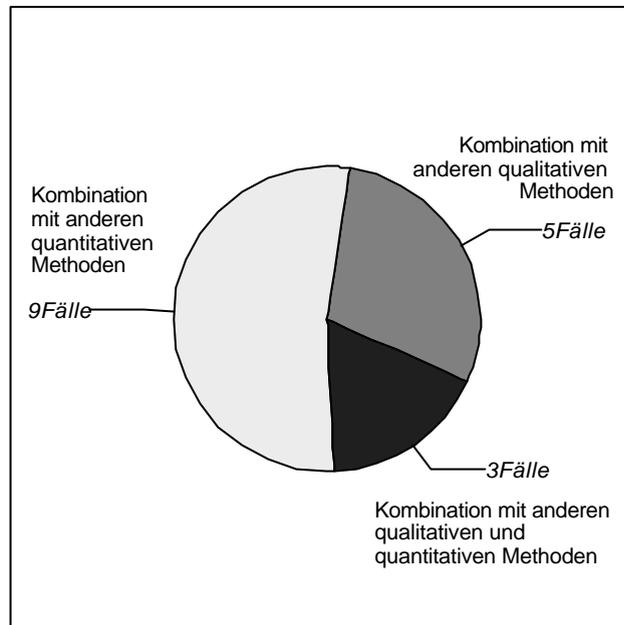
Tabelle 6: Erfüllte Kriterien der Gegenstandsangemessenheit

Die Gegenstandsadäquanz einer Untersuchung insgesamt wurde darüber hinaus vor dem Hintergrund des Aufsatzes noch einmal speziell eingeschätzt. Unter der Berücksichtigung des schwierigen Veröffentlichungsmediums und der Gewichtung auf Methodenwahl und Samplingstrategie, war die Einschätzung recht wohlwollend. 29 Untersuchungen (48%) konnten dabei angemessene Methoden bescheinigt werden, nur in drei Fällen waren grobe Mängel offenbar geworden.

2.

Die zweite Leitfrage bezieht sich auf die Verwendung einer Perspektivenvielfalt im Sinne der Triangulation. 17 Forschungsdesigns (28%) bezogen mehrere Perspektiven ein. Das zeigt, dass von den in der methodischen Literatur häufig beschriebenen Möglichkeiten mittlerweile gerne Gebrauch gemacht wird. Genau zehn Veröffentlichungen beschreiben eine Verwendung von zwei Methoden, sieben sogar von dreien oder noch mehr. Dementsprechend ist der Anteil, den qualitative Verfahren innerhalb der gesamten Forschungsarbeit einnehmen, sehr verschieden. In vier Untersuchungen nehmen sie nur eine Randstellung ein, acht Untersuchungen besitzen in etwa ausgewogene Anteile qualitativer und quantitativer Methoden. Insgesamt kann festgehalten werden, dass in 80% der Untersuchungen (48 Fälle) in der Stichprobe ausschließlich qualitativ geforscht wird.

Zumal alle empirischen Untersuchungen, auch mit einem noch so minimalen Anteil qualitativer Methodenverwendung in die Stichprobe eingegangen sind, kann von einem Status der Hilfs-Methodik für quantifizierend Verfahren kaum die Rede sein.



Grafik 1: Art der Methodenkombination

Am beliebtesten ist die Kombination von qualitativen und quantitativen Methoden, wie sie im Sinne einer verwässerten wissenschaftlich verengten Begriffsauffassung manchmal verstanden wird. Jedes sechste Forschungsdesign verwendet aber auch mehrere qualitative Methoden zur multiperspektivischen Annäherung an ein Phänomen. Kritisch muss angemerkt werden, dass zum Teil höchst aufwendige Verfahren ohne weiteres und scheinbar ohne die geringste Überforderung nebeneinander eingesetzt werden. Helga Kelle (2001:205f.) warnt vor einem allzu leichtfertigen Umgang mit eigentlich schwierig erlernbaren Methoden, für deren richtigen Gebrauch reichlich Übung notwendig sei.

3.

Eine ebenso freudige Diagnose, wie sie für die Triangulation gestellt wurde, kann für die diskursiven Validierungsformen nicht gestellt werden. Nur zwei Mal werden solche erwähnt, auch wenn sie vermutlich, wenn auch unwissentlich, etwas häufiger durchgeführt werden. In beiden Fällen ist dies ein „peer debriefing“ gewesen, also

eine Kommunizierung mit anderen Experten. Die Beforschten wurden in keiner Weise in den Interpretationsprozess miteinbezogen.

Ausführung

4.

Die Auswahl der Untersuchungseinheiten dagegen geschieht recht gewissenhaft. Knapp zwei Drittel der Untersuchungen (37) wählen sie nach theoretisch geleiteten Gesichtspunkten aus. Weitere 21 (35%) machen dies ebenfalls, begehen dabei aber kleine oder mittlere Fehler. Sie berücksichtigen zwar theoretische Abwägungen, doch nicht mit dem untersuchten Phänomenen völlig übereinstimmende. Nur zwei Mal schien die Auswahl der Untersuchungseinheiten völlig willkürlich zu sein und verstieß damit fundamental gegen die Grundregeln der methodischen Theorie.

5.

Die wissenschaftliche Reflexion ist eine der Errungenschaften der interpretativen Verfahren und erfährt zum Teil über ihre Grenzen hinaus eine breite Rezeption. Wenden aber qualitativ Vorgehende selbst gezielt verschiedene Möglichkeiten der Rezeption an?

Die Antwort lautet ja, aber meist nicht gezielt und kontrolliert. Knapp die Hälfte der 29 Veröffentlichungen reflektiert überhaupt nicht und lässt damit diese große Chance der Methode aus.

Reflexionsüberlegungen über ...	Anzahl (abs.)
Grenzen der Untersuchung	22
Theorie	18
Methode	17
Eigenschaften der Untersuchenden	8
Methodisches Handeln im Feld	8
Erkenntnistheoretische Implikationen	4

Tabelle 7: Gegenstand von Reflexion

Nur in vier Untersuchungen (7%) fand sich eine vorbildliche, gezielt-kontrollierte Reflexion von meist mehreren Aspekten zugleich (siehe dazu auch *Tabelle 7*). Die restlichen Forschungsarbeiten gehen mehr oder weniger zufällig oder punktuell auf diverse Schwächen ein.

Die Gegenstände der Reflexion sind zwar vielseitig, aber längst nicht erschöpfend behandelt. Immer noch werden beispielsweise eigene Empfindungen und Gefühle überhaupt nicht erwähnt. Es scheint immer noch die alte Meinung zu bestehen, Gefühle hätten in der Forschung nichts zu suchen.

6.

Der Einfluss der Umgebung und der Situation der Untersuchung ist nur in drei Aufsätzen erwähnt. Die oben erwähnte Tatsache, dass in acht Fällen die Eigenschaften der Untersuchenden reflektiert wurden, führte dann, mit einer Ausnahme, nicht konsequenterweise zur gezielten Variation dieser Merkmale. Ansatzweise und scheinbar planlos wird immerhin in 13 Untersuchungen (23%) variiert, drei non-reaktive Verfahren scheiden für diese Möglichkeit aus.

Im allgemeinen zeigt sich eine sehr starre Handhabung der Erhebung. In nur drei Fällen war ein zirkulär verlaufender Forschungsprozess auszumachen, meist scheinen sich die Forschungsdesigns an der Linearität von quantitativen Untersuchungen zu orientieren. Selbst bei der Vorgehensweise nach der Grounded Theory wurde die eigentlich erforderliche Zirkularität des Forschungsprozesses einfach außer Acht gelassen.

7.

Quantifizierungen scheinen weiter verbreitet, als in der Methodenliteratur vermutet wird. Genau ein Viertel der Veröffentlichungen nutzten diese Möglichkeit. Nur in einem Fall wurde dabei eine ansatzweise nicht haltbare Verteilungsaussage getroffen. Es entsteht der Eindruck, dass noch weitaus mehr Forschungsdesigns von Quantifizierungen Gebrauch machen könnten, als bisher geschehen. Der Umgang mit Häufigkeiten ist im allgemeinen sehr zurückhaltend.

8.

Unter Zugrundelegung der betrachteten Ethik-Kodices ergibt sich ein recht zufriedenstellendes Bild von dessen Einhaltung. Das nahezu einzige Defizit scheint in der

korrekten Informierung der Beforschten über Ziele und Zwecke der Forschung zu liegen. Eine solche ist nur in zwei Texten erwähnt. Es kann allerdings gemutmaßt werden, dass diese vielleicht nicht für erwähnenswert gehalten, aber dennoch zum Teil durchgeführt wurde. Der absolute Großteil (51 Fälle/ 85%) ist, davon abgesehen, ethisch absolut korrekt durchgeführt worden. Die festzustellenden Verstöße sind durchweg geringfügiger Natur. In einem Forschungsbericht sind nicht-anonymisierte Angaben von Lokalitäten zu finden, die den dort aufzufindenden Personen zu Nachteil gereichen könnte. In anderen Fällen sollte eine eingehende Besprechung der Veröffentlichung erfolgen, zum Beispiel im Falle einer psychoanalytischen Analyse einer Einzelperson, vielleicht ist mitunter das Geschriebene nicht immer angenehm und der psychosozialen Stabilität der Beforschten nicht zuträglich, vor allem, wenn es ihrem Selbstkonzept nicht entspricht oder dieses gar in Frage stellt.

In puncto Einwilligung zu einer Forschungsteilnahme ist es bedenkenswert, ob Kinder diese immer gegeben haben, wenn beispielsweise ihre Schulklasse teilnehmend beobachtet wird. Wie Fuhs (1999, 2000) eindringlich appelliert, bedarf es Kindern als Forschungsobjekt gegenüber, einer besonderen Sensibilität und Sensitivität.

Die auftraggebenden Sponsoren wurden in knapp der Hälfte aller Fälle angegeben, bis auf eine Ausnahme waren dies öffentliche Geldquellen. Ob Zuwendungen aus der Wirtschaft tatsächlich derartig selten sind oder diese bisweilen verschwiegen werden, kann hier nicht beurteilt werden.

Es kann insgesamt resümiert werden, dass lediglich in acht durchgeführten Forschungen zwar unter ethischen Gesichtspunkten leichte, aber durchaus diskutabile Problempunkte zu erwähnen sind.

Analyse

9.

Ein sehr gewichtiger Punkt ist die Frage nach der Generalisierbarkeit von Ergebnissen. In der Stichprobe ergibt sich unter anderem durch die Fragestellung, die Samplestrategie und die Festlegung des Geltungsbereichs in genau drei Viertel aller Untersuchungen die Möglichkeit der Generalisierung auf unterschiedliche Reichweiten. Das andere Viertel hegt keinen Wunsch nach Verallgemeinerung auf

über das Sample hinausreichende Zusammenhänge. Nur in einem Fall wurde die Möglichkeit nicht genutzt, wo sie bestanden hätte.

Generalisierung	möglich	nicht möglich	
angestrebt	44	6	50
nicht angestrebt	1	9	10
	45	15	60

Tabelle 8: Übersicht über das Generalisierungsverhalten

Allerdings wird leider in sechs Untersuchungen eine Generalisierung angestrebt, wo sie nicht statthaft ist. Häufig scheinen nicht generalisierte Ergebnisse für wertloser gehalten zu werden, als solche, die in eine Theorie münden. Neun Aufsätze geben sich mit ihrem nicht-verallgemeinernden Status zufrieden.

10.

Dichte Beschreibung des Sende- und Empfängerkontextes scheint als forschungslogisches Muss bislang so gut wie unbekannt zu sein. Von den 50 Aufsätzen, die Generalisierungen vornehmen wollen, beschreiben lediglich vier beide Kontexte mit großer Ausführlichkeit. Weitere 13 (26%) beschreiben recht ungezwungen, aber in einer Weise, die nicht als „dicht“ bezeichnet werden kann. Immerhin 33 Texte (66%) beinhalten gar keine Beschreibung. Innerhalb der zehn Texte, die nicht generalisieren möchten, finden sich sogar nur zwei, die den Kontext beschreiben. Grundsätzlich besteht ein erhebliches Defizit in diesem Bereich.

11.

Eine Einschätzung darüber, wie systematisch eine Theorie aus dem Datenmaterial gewonnen wurde, stellte sich als überaus schwierig heraus. Meist werden kurz die Verfahren angegeben, bestenfalls auch noch erläutert, und anschließend sofort die Ergebnisse präsentiert. Der Prozess der Analyse bleibt vernebelt. Die Ergebnisse wirken dann zwar meist recht plausibel, nur kann keineswegs daraus etwas über die Systematik geschlossen werden.

Insgesamt wurden in 30 empirischen Untersuchungen in der Stichprobe kodifizierte Verfahren angewendet²⁷, die meist außer mit einigen Schlagworten nicht weiter erläutert werden.

Eigene Regeln der Theoriebildung wurden anscheinend in 23 Forschungen angewandt²⁸. ‚Anscheinend‘, weil sie nur zehn Mal genauer erläutert werden. Die restlichen 13 Regeln, sind nur indirekt und ungenau insofern ableitbar, als dass zumindest vermutet werden kann, dass Regeln bestanden haben. In fünf Forschungen scheint dagegen die Theoriebildung völlig ohne Regelanwendungen vorgenommen worden zu sein bzw. ist sie nicht ansatzweise zu erkennen.

Die zugegebenermaßen ziemlich schwierigen Prinzipien empirisch begründeter Theoriebildung (vgl. Kelle 1994) werden in der Forschungspraxis schlichtweg ignoriert – Ausnahmen vorbehalten. Insbesondere wenn selbst entwickelte Theoriebildungsverfahren angewendet werden, könnten sich diese aus Kelles reichhaltigem Fundus bedienen.

Es wird zum besseren Verständnis und zur Beurteilung dringend angeraten, an allen Stellen der Theoriebildung, wie sie hier unter Punkt 11 abgehandelt worden sind, in Zukunft intensivere Arbeit zu leisten. Sicherlich ist hier die absolute Schwachstelle der Praxis zu finden.

12. und 13.

Zwei dieser Möglichkeiten aus der „empirisch begründeten Theoriebildung“, die sich inhaltlich recht nahe stehen, die minimale und maximale Kontrastierung sowie die Diskussion von Evidenz und Gegenevidenz, sollen hier gemeinsam beleuchtet werden.

In reiner, der methodischen Literatur entsprechenden Form waren umgesetzte Prinzipien der Kontrastierung in elf Texten (18%) zu finden. Praktiziert wurde jedoch ausschließlich die maximale Kontrastierung, nicht die Minimale. Evidenz und Gegenevidenz diskutierten gar nur zwei Forschende in ebensolcher Form. Evidenz wurde darüber hinaus häufiger diskutiert, Gegenevidenz nicht. Es ist zu vermuten, dass etwas nicht ins Gefüge passendes immer noch als Fehler empfunden wird, statt darin Vorteile zu entdecken. Ansatzweise benutzen viel mehr Forschungs-

²⁷ Die Differenz zu der Anzahl der oben angegebenen Verfahren ergibt sich durch Methodenkombinationen.

²⁸ Siehe Fußnote 27.

gruppen diese Prinzipien bereits, wenn aber, dann nicht gewollt oder rudimentär. Kontrastierung war in dieser Form 14 Mal (23%) zu finden, eine Evidenzdiskussion 13 Mal (22%). Der Großteil der Forschenden ignoriert diesen Themenkomplex.

14.

Die Ergebnisse zum Computereinsatz in der qualitativen Sozialforschung sind eine der größten Überraschungen dieser Untersuchung. Nur drei Anwendungsfälle sind in der Stichprobe zu finden, darunter ein Videographieprogramm, einmal SPSS zur Chi-Square-Errechnung und ein nicht namentlich genanntes Datenbankprogramm. Von den auf dem Markt vorhandenen großen qualitativen Programmen wird entweder kein Gebrauch gemacht oder der Einsatz aus irgendwelchen Gründen verschwiegen. Es war eher zu erwarten, dass die Klientel, die Zeitschriftenaufsätze verfasst, meist jüngeren Alters und folglich nicht in alten Denkstrukturen verhaftet ist und sich Neuerungen eher abgeneigt gegenüber verhält. Ersteres mag bedingt stimmen, die Folgerung daraus nicht. Ein bewusstes Verschweigen scheint unlogisch, gilt doch der Einsatz von Computern in weitesten Forschungskreisen als eher fortschrittlich, als denn verpönt. Diesem Phänomen nachzugehen wäre eine interessante zukünftige Aufgabe, da vielleicht durch eine Nutzeranalyse andere Ergebnisse ans Licht kommen, als dies zum Teil von den Vertreibern der Computerprogramme angenommen wird. Die vorliegenden Ergebnisse sprechen zumindest dafür, dass der Markt noch sehr lange nicht gesättigt sein dürfte.

Präsentation

15.

Die sieben zu dokumentierenden Bereiche, die der Nachvollziehbarkeit dienen sollen, werden in den Aufsätzen unterschiedlich berücksichtigt.

Am besten dokumentiert ist die jeweilige Informationsquelle, in 57% der Veröffentlichungen (34 Mal) ist diese erschöpfend dargestellt. In wenigen kompakten Sätzen ist dies zum Teil zu leisten, daher auch häufig zu finden. Die Erhebungsmethode wurde noch in einem Viertel aller Aufsätze beschrieben, der Kontext der Erhebung dagegen oftmals ausgelassen.

Bereich der Dokumentation	Anzahl der Fälle, in denen der Bereich nicht berücksichtigt wurde ²⁹
Entscheidungen und Probleme	44 (73%)
Daten	39 (65%)
Transkriptionsregeln	35 (67%)
Vorverständnis	35 (58%)
Auswertung	25 (42%)
Erhebungsmethoden und -kontext	16 (28%)
Informationsquelle	3 (5%)

Tabelle 9: Defizitliste der Dokumentationsbereiche

Leider wurden meistens die beiden eigentlich zusammengehörigen Komplexe, von den Forschenden nicht zusammen beschrieben. Angesichts der eigentlich doch enormen Wertigkeit dieses Bereiches, mutet der Anteil von einem Viertel an adäquaten Dokumentationen sehr gering an.

Die Auswertung und das Vorverständnis dokumentieren nur 12% bzw. 8% angemessen. Ohne erstere sind die Ergebnisse kaum einschätzbar, Zweiteres ist nötig zur Kontrollierbarkeit von Subjektivität.

In fünf Aufsätzen (12%) werden die Transkriptionsregeln beschrieben und zum Teil erläutert. Sicherlich ist dieser Bereich einer der weniger wichtigen, da auch ohne das genaue Wissen darum eine Nachvollziehbarkeit gut möglich sein kann.

Die Dokumentation der Daten, also die freie Einsicht in das Datenmaterial ist nur in fünf Fällen möglich. In einem Fall wird auf eine Internetseite verwiesen, auf der die Daten abrufbar sind, die anderen vier Fälle bedienen sich öffentlich zugänglicher Literatur. In weiteren 22 Texten existiert ein Verweis auf eine umfassendere Veröffentlichung des Forschungsberichtes, meist eine Monographie. Möglicherweise sind dort ausführlichere Dokumentationen aller Bereiche zu finden, möglicherweise aber auch nicht. Einige positive Beispiele zeigen aber, dass es auch auf engem Raum möglich ist, breite Dokumentationen zu entfalten. Deswegen ist die

²⁹ Die Prozentangaben beziehen sich nur auf die jeweils relevanten Fälle. Zum Beispiel bedarf es bei einem auf Literaturvorlagen beruhenden Verfahren keiner Transkriptionsregeln.

Lösung des Problems durch einen Verweis nicht die Ultima Ratio, sondern eher ein fauler Kompromiss.

Zugegebenermaßen machen die Forschenden sich leichter angreifbar, wenn sie ihre Interpretationsgrundlage veröffentlichen, aber eine fundierte und regelgeleitete Interpretation sollte sich jeder Diskussion stellen können. Eine Auseinandersetzung zu vermeiden, indem ihr die Diskussionsgrundlage entzogen wird, kommt einer Kapitulation gleich und steht jedem Bemühen um „ensuring rigour“ in der qualitativen Sozialforschung entgegen.

Zuletzt noch einige Anmerkungen zur Dokumentation von Entscheidungen und Problemen. Nur 10% der Veröffentlichungen greifen diese Themen auf. Nach wie vor wird es als Zeichen der Unzulänglichkeit empfunden, wenn innerhalb des Forschungsprozesses Probleme auftauchen. Entweder ist diese These zutreffend oder 90 von Hundert Forschungen verlaufen völlig problemlos. Im Gegenteil werden Probleme verschleiert oder im Text übergangen, das Gesamtbild muss schließlich stimmen. Eine neue Sicht auf Probleme könnte sich entwickeln, die qualitativen Methoden hier Vorreiter sein.

Summiert man die in den Veröffentlichungen mindestens ansatzweise dokumentierten Bereiche auf, dann zeigt sich einerseits, dass immerhin neun Aufsätze (15%) alle Bereiche, bis auf manchmal einen, ansatzweise dokumentieren. Aber auf der anderen Seite erwähnen aber 27 Veröffentlichungen (45%) die Hälfte der Bereiche oder mehr, mit keinem Wort.

Fazit: Nach dem Urteil des Beurteilenden sind 31 Forschungsprozesse (52%) nur schwer oder gar nicht nachzuvollziehen. Die Schere klafft weit auseinander. Neben bereits teilweise praktizierter Dokumentation herrscht vielerorts noch nicht das Bewusstsein für transparente Forschung. Alles was nicht geschildert wird, kann auch nicht kritisiert werden, denn solange das Ergebnis stimmt, kann der Weg nicht schlecht sein. Dieses Prinzip wird den Vorwurf der Beliebigkeit gegenüber den Methoden weiterhin verstärken.

16.

Die Darstellung der Forschenden ihres eigenen Standpunktes, ihrer biographischen Verstrickung oder ihrer Beziehung zum Fall ist nur zwei Mal angefundenes worden. Als gezielte Explikation von Subjektivität zum Zwecke des Erkenntnisgewinns wird sie, ähnlich wie die gezielte Reflexion, wohl eher als unpassend wahrgenommen.

Vielmehr wird sie hier nicht nur als Chance verstanden, sondern sogar als eine Pflicht. In zehn Veröffentlichungen immerhin werden eigene Standpunkte im Ansatz vertreten, unvollständig und ungezielt. In diesem Punkt besteht noch erhebliches Verbesserungspotenzial.

Sonstiges

Wie sich in den Ausführungen zur Herleitung der Gütekriterien herausgestellt hat, ist eine falsifikatorische Grundhaltung sehr zu begrüßen. Genau ein Drittel in der Stichprobe bauen dieses Prinzip in ihr Handeln ein. Nur fünf Aufsätze (8%) sind deutlich konfirmatorisch ausgerichtet, ihnen liegt vorwiegend an der Bestätigung von impliziten oder tatsächlich vorhandenen Hypothesen. In den restlichen Fällen kann eine Ausrichtung nicht festgestellt werden.

Unter der Rubrik ‚sonstiges‘ soll noch das Sammelsurium an noch nicht aufgeführten Unzulänglichkeiten angeführt werden:

In 16 Fällen war die Samplestrategie unklar, in elf Fällen war überhaupt keine Darstellung der Methode zu finden. Acht Mal war die Fallzahl für eine Verallgemeinerung zu klein (was nicht eine grundsätzliche Unmöglichkeit für kleine Fallzahlen ist), in sieben Aufsätzen ist das Sample für die vorgenommene Generalisierung schlichtweg falsch zusammengestellt worden. Sechs Aufsätze erweckten den Eindruck einer affirmativen Illustration, aber nicht einer induktiv-orientierten Forschung, vier Mal musste ernsthaft bedacht werden, ob die als qualitativ verkaufte Forschung überhaupt eine solche ist, oder ob sie nicht überwiegend quantitative Eigenschaften besitzt. Ebenfalls vier Mal wurde qualitative Forschung innerhalb des Aufsatzes marginalisiert und hätte deshalb ebensogut auch weggelassen werden können, drei Veröffentlichungen beachteten die besondere Reaktivität der Erhebungsmethodik nicht und verzerrten möglicherweise nachhaltig die Ergebnisse. Weitere drei Mal war das Vorverständnis der Forschenden über alle Maße hinaus bestimmend für den Forschungsausgang und einmal bestand überhaupt keine Verbindung zwischen Empirie und deren angeführter theoretischer Anbindung.

Abschließend wurde versucht, die Forschung in ihrer Gesamtheit einzuschätzen. Diese kann nur tendenziösen Charakters bleiben, sorgt aber dennoch an den Rändern für ein wenig Aufschluss. Vier Kategorien sollten zur Orientierung

genügen. Als ‚methodisch hervorragend‘ wurden sieben Veröffentlichungen (12%) eingeschätzt, die zwar auch nie ohne Schwächen waren, aber alles in allem dem, was als ideale Forschung bezeichnet wird, sehr nahe kommen. Weitere 36 Aufsätze (60%) waren ‚methodisch solide‘ durchgeführt, das heißt, mit einem abgesichertem Konzept und zumeist guter Umsetzung, allerdings kleineren Schwächen. Die Beurteilung ‚Methodisch mit Mängeln‘ bedeutet eine Forschung mit einer größeren Anzahl von Unzulänglichkeiten oder grundlegenden methodischen Irrwegen. 14 Mal (23%) wurde diese Kategorie gezählt. Schließlich sind in der Stichprobe drei Texte, die als ‚methodisch schwach‘ bezeichnet werden müssen. Es herrschen darin gröbste methodische Unzulänglichkeiten, wissenschaftliche Standards sind zum Teil nicht eingehalten.

4.3 Weitergehende Analysen

Im folgenden soll über einfache Häufigkeitsverteilungen hinaus, einigen Tendenzen auf die Spur gegangen werden. Ist die Stichprobe wirklich so homogen, wie im letzten Kapitel unterstellt worden ist?

Zunächst wurde mit den in Frage kommenden Variablen eine explorative Faktorenanalyse durchgeführt³⁰. Dadurch konnten sechs Faktoren extrahiert werden (Ladungsmatrix und Kommunalitäten in Anhang 2), die als Dimensionen der Güte verstanden werden können. Nicht alle Variablen konnten sinnvoll miteinbezogen werden, deshalb ist die Dimensionierung nicht erschöpfend aber interpretierbar. Jeder der sechs Faktoren umfasst zwei, drei oder vier Variablen, die hoch auf ihm laden.

Letztlich wurden Faktorenwerte für jeden einzelnen Fall berechnet, die anzeigen, wie stark sich die jeweilige Dimension in einem Fall widerspiegelt. Diese Werte sind

³⁰ Die Durchführung geschah nach der Hauptkomponentenmethode. Die Lösung wurde rotiert nach dem Varimax-Verfahren. Aus den 16 dichotomisierten Variablen konnten auf diese Weise sechs Faktoren extrahiert werden. Diese Zahl entspricht dem Kaiser-Guttman-Kriterium obwohl dieses nicht als maßgeblich bestimmend betrachtet wurde, sondern inhaltliche Konsistenz den Ausschlag geben sollte. Die Faktoren erklären 73,33% der Varianz. In Anhang 2 befinden sich die Ladungsmatrix und die Kommunalitäten.

standardisiert (Mittelwert=0, Standardabweichung=1), so dass Testverfahren, die Intervalskalenniveau voraussetzen, zum Einsatz kommen können.

Tabelle 10 sind die grundlegenden Angaben zu den sechs extrahierten Faktoren zu entnehmen.

Schließlich muss auf das gewählte Signifikanzniveau eingegangen werden. Für angemessen wurde das 5%-Niveau für zweiseitige Signifikanz befunden. Strengere Signifikanzschwellen würden die Ergebnisse wegen der geringen Fallzahl über die Maßen einebnen. Eher sollten sogar knapp darüberliegende Werte als heuristischer Fingerzeig gewertet und nicht unterschlagen werden. Allerdings wird auf deren Status jeweils hingewiesen.

Faktor	Dimension	Variablen-zahl	Reliabilität (Cronbachs α)
1	Materialnähe	4	.75
2	Dokumentation	3	.63
3	Sampling	2	.85
4	Selbstaufmerksamkeit	3	.63
5	Bezug zu Zahlen	2	.70
6	Non-Konfirmatorik	2	.57

Tabelle 10: Faktorenübersicht

Die Disziplin

Auf keine statistische Weise sind irgendwelche Unterschiede zwischen den beiden Disziplinen aufzudecken. Selbst die Faktorenwerte unterscheiden sich in keiner Dimension auch nur annähernd. Leichte Hinweise gibt es lediglich für den Bereich der Ethik, mit deren Prinzipien in der Soziologie nachlässiger umgegangen wird (Sign.: .07 bei Chi-Square) und die ‚thick description‘, die von der Erziehungswissenschaft eher unterlassen wird (Sign.: .042 bei Chi-Square). Auch wenn dies auf den ersten Blick ein signifikantes Ergebnis sein sollte, eine Bonferoni-Korrektur der Signifikanzschwelle würde den Wert deutlich scheitern lassen. Auf der anderen Seite lassen die überaus kleinen Fallzahlen solche Werte als durchaus beachtenswerten Fingerzeig erscheinen.

Das Geschlecht

Abgesehen von der quantitativen Verteilung der Partizipation an qualitativer Forschung, die in Abschnitt 4.1 beschrieben wurde, interessieren vor allem qualitative Unterschiede. Unterscheiden sich Frauen und Männer in der Umsetzung der Gütekriterien? Haben sie eine je eigene Wege der Forschungshandhabung?

Es gab bei der Variable Geschlecht vier Möglichkeiten der Analyse. Entweder drei Gruppen (Männer, Frauen und intergeschlechtliche Forschungsteams) bestehen zu lassen, die intergeschlechtlichen Teams nicht zu berücksichtigen oder zu einem der beiden Geschlechter hin zu dichotomisieren. Die nicht-dichotomisierende Möglichkeit wurde ebenso verworfen³¹, wie die zweite Möglichkeit, die den Verlust von Fällen zur Folge hätte.

Alles in allem fällt die Antwort recht kurz aus: Geschlecht ist kaum eine unterscheidende Variable für die Handhabung der Gütekriterien. Mit einigen wenigen vorsichtigen Ausnahmen³²: Frauen dokumentieren die Erhebungsmethoden und den –kontext angemessener als ihre männlichen Kollegen. Andererseits dokumentieren Männer die Auswertung der Daten ausführlicher als ihre Kolleginnen. Eine logische theoretische Erklärung dafür ist nicht greifbar. Dies könnte durch statistischen Zufall erklärt werden, sollte in folgenden Forschungsarbeiten möglicherweise noch einmal aufgegriffen werden.

Signifikant sind die Unterschiede in zwei Dimensionen der Faktorenanalyse: Frauen weisen eine höhere Materialnähe auf, Männer dagegen zeichnen sich durch eine höhere Selbstaufmerksamkeit aus. Ausgerechnet dieser letzte Punkt verdient inhaltliche Beachtung, wird er doch meist eher als weibliche Eigenschaft verkauft. In naher Zukunft sollte ein grundsätzliches Umdenken stattfinden, was geschlechtsspezifische Attributierungen angeht.

Ein anderer Sachverhalt fällt allerdings noch mit großer Überzufälligkeit auf. Alle neun Fälle von leichten Verstößen gegen die Ethik-Kodices werden von Männern begangen (Sign.: .013 bei Chi-Square). Frauen haben anscheinend mehr Respekt vor den Rechten und Bedürfnissen der Untersuchungssubjekte oder den Verbindlichkeiten der Kodices als Männer.

³¹ Eine Dichotomisierung mindert das Problem der kleinen Besetzung der Felder. Diese führt rasch dazu, dass die Erfordernisse des Chi-Square-Tests oder des T-Tests nicht mehr eingehalten werden.

³² Für die Einordnung der statistischen Absicherung einer Aussage gelten hier, wie auch später, die gleichen Anmerkung, wie für die Disziplin.

Länge des Aufsatzes

An der Länge der Aufsätze sind keine Besonderheiten abzulesen. Die Länge ist allerdings mit der Variable Seitenzahl nur unzureichend operationalisiert, denn die Zeitschriften haben verschiedene Layouts und Formate. Dennoch sollte sich bei zunehmend größerer Datenbasis, die hier nicht behauptet wird, dieses Manko statistisch zunehmend minimieren.

Das überraschende an diesem Ergebnis ist, dass eigentlich große Unterschiede zu erwarten wären. Je länger schließlich ein Text ist, desto mehr Darstellungsraum steht zur Verfügung und detailliertere Schilderungen können angeführt werden. Aber dies scheint trotz des dünnen statistischen Eises nicht der Fall zu sein. Und tatsächlich waren auch kürzere Abfassungen zu finden, die mit umfassenden Darstellungen auf kleinem Raum aufwarteten, längere dagegen, die solche vermissen ließen.

Weder die Dokumentation, noch die Beurteilung insgesamt stehen in irgendeinem Verhältnis zu Seitenlänge der Veröffentlichung.

Aktualität

Ein Gegenstand wie die qualitative Forschung, welcher einem stetigen Wandel unterworfen ist, dürfte auch in der aktuellsten Entwicklung einige Änderungen vollzogen haben. Speziell im Bereich der Gütekriterien sollte die Wandlung zu erkennen sein. Fraglich ist nur, für welchen Zeitraum. Im Vergleich der aktuellen Forschungspraxis mit der aus den 80-er Jahren des 20. Jahrhunderts ist ohne Zweifel ein großer Unterschied festzustellen. Hat sich aber in den letzten fünf Jahren ebenfalls einiges geändert, haben die neuerlichen Debatten auf die Forschung durchgeschlagen?

Um diesen Fragen nachzugehen, wurde die Stichprobe am Median des Erscheinungsjahres dichotomisiert. Somit stehen die Jahrgänge 2000 und 2001 den älteren Jahrgängen, 1997 bis 1999 entgegen.

Es lassen sich auf diese Weise einige positive, aber auch einige negative Trends aufzeigen. Die Gesamtbeurteilung der Veröffentlichungen hat sich dabei nicht verändert.

Verbessert hat sich die Ausführung der Dimension ‚Dokumentation‘ der Faktorenanalyse (Sign.: .006), sowie der Dokumentation von Daten und Auswertung im besonderen (beide: .013). Das heißt, es wird neuerdings angemessener

dokumentiert, jedoch noch längst nicht ausreichend. Reflexion in ihren diversen Facetten wird aktuell eher mehr eingebracht als früher (Sign.: .088) und Quantifizierungen werden in jüngster Zeit häufiger vorgenommen (.012). Vielleicht ist dies ein Anzeichen für den neuen selbstbewussten Umgang mit Zahlenmaterial, gerade auch wenn berücksichtigt wird, dass der Zahlengebrauch recht defensiv ist (siehe Kapitel 4.2), also beispielsweise keine Verteilungsaussagen über eine Population getroffen werden.

Seltener geworden ist dagegen die Ausrichtung an Prinzipien der Falsifikation. Der sechste Faktor der Faktorenanalyse zeigt dies ansatzweise an (Sign: .072). Kann dies dahingehend interpretiert werden, dass der Zwang zur Ergebnislieferung immer größer wird? Ein falsifikativ orientiertes Vorgehen kann dies nicht garantieren. Kein Ergebnis wird nicht immer als Ergebnis akzeptiert, obwohl diese Auffassung absolut der zu Grunde liegenden Wissenschaftstheorie völlig widerspricht.

5. Schlussbetrachtungen und Fazit

Im Laufe dieser Arbeit ist der Bedarf an Maßstäben zur Qualitätsbewertung deutlich geworden. Obwohl die einzelnen Teilbereiche der qualitativen empirischen Sozialforschung alles andere als homogen sind und diese Vielfalt überaus bereichernd ist, gibt es dennoch viele Gemeinsamkeiten. Jede Methode, die dieser Bezeichnung gerecht werden möchte, muss sich von einer Beliebigkeit der Interpretation distanzieren. Auf diesem kleinsten gemeinsamen Nenner wurden zahlreiche Angebote für Gütekriterien herausgearbeitet, die für alle qualitativen Methoden Gültigkeit beanspruchen – Ausnahmen immer vorbehalten, denn mittlerweile haben sich die qualitativen Methoden derartig ausdifferenziert, dass eine fundierte Kenntnis aller Bereiche kaum mehr möglich ist.

Es hat sich als fruchtbar erwiesen, sowohl die Konzepte der quantitativen Methoden auf ihre Übertragbarkeit, als auch methodenangemessene Gütekriterien, die bislang vorgeschlagen wurden, auf ihre Tauglichkeit zu untersuchen. Im ersten Fall konnten wichtige Anleihen genommen werden, im zweiten Fall auf vorhandene Bemühungen direkt oder minimal modifiziert zurückgegriffen werden. Dem Standpunkt der völligen Ablehnung von Gütekriterien konnte nicht gefolgt werden; er konnte lediglich bereichernd in die Diskussion eingeflochten werden.

Die Diskussion um Gütekriterien ist damit längst nicht abgeschlossen, sondern eher erst am Anfang. Diese Arbeit steht vermutlich in einer längeren Reihe noch folgender Arbeiten ähnlicher Art. Die erarbeiteten Ergebnisse stellen somit einen Querschnitt durch aktuelle Theorie und Praxis dar, eine Art Bestandaufnahme des Vorfindlichen. Der Blick auf die Praxis hat erhebliche Defizite offenbart, in allen vier Bereichen war die Streuung der Qualität sehr breit. Die Methode, die Ausführung, die Analyse und die Präsentation der durchgeführten Forschungen war in einigen Fällen von hoher Qualität, in anderen waren größtenteils Handwerksfehler zu finden. Der Löwenanteil der Forschenden präsentierte jedoch solides Forschungsmaterial mit mehr oder weniger sich offenbarenden Schwächen. Im Bereich der Auswahl einer angemessenen Methode wird recht zufriedenstellend gearbeitet, die Umsetzung ist eine andere Sache. Schleierhaft bleibt des Öfteren der Bereich der Analyse. Die Begründung und der Entstehungsweg eines gewonnenen Ergebnisses

bleiben meist im dunkeln und lassen Raum für Spekulationen. Es sollte nicht angenommen werden, dass alle nicht beschriebenen Vorgänge willkürlich zustande gekommen sind, doch zwangsläufig drängt sich der Gedanke auf, wenn Ergebnisse nicht nachvollzogen werden können.

Die Darstellung wird in ähnlicher Weise vernachlässigt und dies, obwohl gezeigt werden konnte, dass sie nicht von der Länge des Aufsatzes abhängt, sondern von dieser unabhängig, in ihrer Qualität variiert.

Ein Großteil von jenen Missständen hätte vermieden werden können, wenn die Auseinandersetzung mit methodologischer und methodischer Literatur ausführlicher ausgefallen wäre. Entsprechend kann Seale nur beigepflichtet werden, wenn er eine Auseinandersetzung der Forschenden mit diesen Grundlagen für äußerst fruchtbar hält, solange diese nicht über alles gesetzt würden und damit die Ausführung einer Forschung zur Erfüllung eines Schemas verkomme. (Seale 1999b:466)

Alle diese Ergebnisse weisen auf bestehende Mängel in der Methodenausbildung hin. Im Gegensatz zur quantitativen Methodenausbildung, die schon lange zum Grundgerüst in den Sozial- und Erziehungswissenschaften gehört, ist die Vermittlung von qualitativen methodischen Fertigkeiten bislang noch eine Randerscheinung an den Hoch- und Fachhochschulen. Es ist naheliegend, dass die in der Vergangenheit vorhandenen Mängel in der Ausbildung für eine heutige defizitäre Praxis mitverantwortlich sind. Nur langsam tritt eine Verbesserung der Situation in der Vermittlung von qualitativem Methodenwissen ein. Die zuständigen Interessengruppen gewinnen nur mühsam gemeinsame Standpunkte, haben aber die Problematik unzweifelhaft erkannt.

Für die Gutachterinnen und Gutachter, die über den Förderungswert einer Forschungsidee befinden müssen, liegen nicht erst seit dieser Arbeit Kriterien vor, die eine grundsätzliche Entscheidung gegen einen qualitativ orientierten Forschungsantrag jeglicher Legitimation entziehen. Vielmehr sollten auch diejenigen, die der Forschungsrichtung wohlwollend gegenüber stehen, im Einzelfall die Methodik sorgfältiger ausloten, um solche Projekte abzulehnen, die die qualitativen Methoden ins Zwielficht bringen.

Insgesamt sollte der deutschsprachige Raum zukünftig noch eingehender die Diskussionsentwicklungen in den anglophonen Ländern verfolgen. Wie dargestellt, hinkt die hiesige Diskussion eher nach, kann aber interessante Aspekte vorweisen,

die im anglophonen Raum untergehen. Eine Angleichung beider Debatten wäre für alle bereichernd. Dazu müssten die dortigen Veröffentlichungen hierzulande stärker rezipiert werden und eigene Beiträge im englischsprachigen Raum veröffentlicht werden – bisher leider noch die Ausnahme.

Darüber hinaus sollten qualitative Methoden in Zukunft verstärkt die von ihnen angeratene Reflexivität in der Methode auf sich selbst beziehen. Eine gezielte Evaluation der Umsetzung von Gütestandards gehört dazu. Eine bewusste Aufsuche von Schwächen sollte nicht mehr als Gefahr empfunden werden für den Fortbestand der Methodik. Das Herausarbeiten von Stärken könnte auf der anderen Seite ebenso als Errungenschaft wahrgenommen werden. Nur so kann sich langfristig eine qualitativ hochwertige und im Vergleich mit anderen Verfahren ohne jede Einschränkung konkurrenzfähige qualitative empirische Sozialforschung etablieren, die ihre Werkzeuge in den Dienst der gesteigerten Erkenntnis stellt.

6. Literatur

- American Psychological Association (1954). Technical Recommendations for Psychological Tests and Diagnostic Techniques. *Psychological Bulletin* 51, Supplement.
- Blaxter, M. (2000). Criteria for Qualitative Research. *Medical Sociology News*, H.2, 34-38.
- Bortz, J.; Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. Springer: Berlin.
- Bühl, A.; Zöfel, P. (2000). *SPSS Version 10. Einführung in die moderne Datenanalyse unter Windows*. Addison-Wesley: München.
- Cook, T.D.; Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Rand-McNally: Chicago.
- Corti, L. (2000). Review note to Clive Seale (1999). The Quality of Qualitative Research [9 paragraphs]. *Forum Qualitative Sozialforschung/ Forum: Qualitative Social Research* [Online Journal], 1(3). Available at: <http://qualitative-research-net/fqs/fqs-eng.htm> (13.03.2001).
- Corti, L. (2001). „QUALIDATA. Qualitative Data Archival Resource Centre.“ <<http://www.essex.ac.uk/qualidata/>> (11.06.2001).
- Denzin, N.K. (1978)². *The Research Act. A theoretical Introduction to Sociological Methods*. New York: McGraw-Hill Book Company (1st Ed.1970).
- Deutsche Gesellschaft für Erziehungswissenschaft [Hg.] (1999). Ethikkodex der Deutschen Gesellschaft für Erziehungswissenschaft. *Mitteilungsheft der Deutschen Gesellschaft für Erziehungswissenschaft*, H.19, 52-57.
- Deutsche Gesellschaft für Soziologie, Berufsverband Deutscher Soziologinnen und Soziologen [Hg.] (1993). Ethik-Kodex der Deutschen Gesellschaft für Soziologie und des Berufsverbandes Deutscher Soziologen. *DGS-Information* H.1, 13-19.
- Diekmann, A. (1995). *Empirische Sozialforschung. Grundlagen, Methoden, Anwendung*. Rowohlt: Reinbek bei Hamburg.

- Drinkmann, A. (1990). Methodenkritische Untersuchungen zur Metaanalyse. Deutscher Studien Verlag: Weinheim.
- Erzberger, Chr. (1998). Zahlen und Wörter. Die Verbindung quantitativer und qualitativer Daten und Methoden im Forschungsprozeß. Deutscher Studien Verlag: Weinheim
- Farin, E. (1997). Metaanalysen: Methodologische Grundlagen und praktische Durchführung. In: Strauß, B.; Bengel, J. [Hg.]: Forschungsmethoden in der medizinischen Psychologie, 161-180. Hogrefe: Göttingen.
- Feyerabend, P. (1976). Wider den Methodenzwang – Skizze einer anarchistischen Erkenntnistheorie. Suhrkamp: Frankfurt/M.
- Flick, U. (1987). Methodenangemessene Gütekriterien in der qualitativ-interpretativen Forschung. In: Bergold, J.B.; Flick, U. [Hg.]: Ein-Sichten. Zugänge zur Sicht des Subjekts mittels qualitativer Forschung. DGVT: Tübingen.
- Flick, U. (1992). Triangulation Revisted – Strategy of or Alternative to Validation of Qualitative Data. *Journal for the Theory of Social Behavior*, 22, 175-197.
- Flick, U. (1999)⁴. Qualitative Forschung. Theorie, Methoden, Anwendung in Psychologie und Sozialwissenschaften. Rowohlt: Reinbek bei Hamburg.
- Flick, U. (2000). Triangulation in der qualitativen Sozialforschung. In: Flick, U.; von Kardorff, E.; Steinke, I. [Hg.]. Qualitative Forschung. Ein Handbuch. Rowohlt: Reinbek bei Hamburg, 309-318.
- Flick, U. (2001). Qualitative Sozialforschung – Stand der Dinge. In: *Soziologie. Forum der Deutschen Gesellschaft für Soziologie*, H.2, 53-66.
- Flick, U.; von Kardorff, E.; Steinke, I. (2000). Was ist qualitative Forschung? Einleitung und Überblick. In: dieselben [Hg.]: Qualitative Forschung. Ein Handbuch. Rowohlt: Reinbek bei Hamburg, 13-29.
- Flick, U.; von Kardorff, E.; Steinke, I. [Hg.] (2000). Qualitative Forschung. Ein Handbuch. Rowohlt: Reinbek bei Hamburg.
- Friedrichs, J. (1990)¹⁴. Methoden empirischer Sozialforschung. Westdeutscher Verlag: Opladen.
- Friebertshäuser, B.; Prengel, A. [Hg.] (1997). Handbuch qualitative Forschungsmethoden in der Erziehungswissenschaft. Juventa: Weinheim.
- Fuhs, B. (1999). Die Generationenproblematik in der Kindheitsforschung. Zur methodischen Relevanz von Erwachsenen-Kind-Verhältnissen. In: Honig, M.-

- S./ Lange, A./ Leu, H.R. [Hg.]: Aus der Perspektive von Kindern? Zur Methodologie der Kindheitsforschung. Juventa: Weinheim, 153-161.
- Fuhs, B. (2000). Qualitative Interviews mit Kindern. Überlegungen zu einer schwierigen Methode. In: Heinzl, F. [Hg.]: Methoden der Kindheitsforschung. Ein Überblick über Forschungszugänge zur kindlichen Perspektive. Juventa: Weinheim, 87-103.
- Hammersley, M. (1990). Reading ethnographic research. Longman: New York.
- Hammersley, M. (1992). What's wrong with ethnography: Methodological explorations. Routledge: London.
- Glaser, B.G.; Strauss, A.L. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine de Gruyter: New York.
- Hopf, Chr. (2000). Forschungsethik und qualitative Forschung. In: Flick, U.; von Kardorff, E.; Steinke, I. [Hg.]. Qualitative Forschung. Ein Handbuch. Rowohlt: Reinbek bei Hamburg, 589-600.
- Hopf, Chr. et al. (1998). Einrichtung der Arbeitsgruppe „Methoden der qualitativen Sozialforschung“. In: *Mitteilungsblatt der Deutschen Gesellschaft für Soziologie*, H.3, 40-42.
- Janesick, V (1994). The dance of qualitative research design: metaphor, methodolatory and meaning. In: Denzin, N; Lincoln, Y.[Ed.]: Handbook of qualitative research. Thousand Oaks: Sage, 209-219.
- Kardorff, E.v. (1991). Qualitative Sozialforschung – Versuch einer Standortbestimmung. In: Flick, U. et al.: Handbuch Qualitative Sozialforschung, 3-8. Psychologie Verlags Union: München.
- Kelle, H. (2001). Ethnographische Methodologie und Probleme der Triangulation. Am Beispiel der Peer Culture Forschung bei Kindern. In: *Zeitschrift für Soziologie der Erziehung und Sozialisation*, H.2, 192-208.
- Kelle, U. (1994). Empirisch begründete Theoriebildung. Zur Logik und Methodologie interpretativer Sozialforschung. Deutscher Studien Verlag: Weinheim.
- Kelle, U. (2000). Literaturbesprechung: Clive Seale: „The Quality of Qualitative Research“, [London: Sage, 1999]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, H.3, 588-590.

- Kelle, U.; Kluge, S.; Prein, G. (1993). Strategien der Geltungssicherung in der qualitativen Sozialforschung. Zur Validitätsproblematik im interpretativen Paradigma. Arbeitspapiere des Sfb 186, Nr.24: Bremen.
- Kelle, U.; Laurie, H. (1995). Computer Use in Qualitative Research and Issues of Validity. In: Kelle, U. [Hg.]: Computer-Aided Qualitative Data Analysis. Theory, Methods and Practice. Sage: London, 19-28.
- Kuckartz, U. (1995). Case-orientated quantification. In: Kelle, U. [Hg.]: Computer-Aided Qualitative Data Analysis. Theory, Methods and Practice. Sage: London, 158-166.
- Kuckartz, U. (1996). MAX für Windows: ein Programm zur Interpretation, Klassifikation und Typenbildung. In: Bos, W.; Tarnai, Chr. [Hg.]: Computer-unterstützte Inhaltsanalyse in den empirischen Sozialwissenschaften. Theorien – Anwendung – Software. Waxmann: Münster.
- Kuckartz, U. (1998). WINMAX professionelle Version. Textanalysesystem für die Sozialwissenschaften – Handbuch. Westdeutscher Verlag: Opladen.
- Kuckartz, U. (1999). Computergestützte Analyse qualitativer Daten. Eine Einführung in Methoden und Arbeitstechniken. Westdeutscher Verlag: Opladen.
- Lamnek, S. (1988). Qualitative Sozialforschung, Bd.1. Psychologie Verlags Union: München.
- Lamnek, S. (2000). Sozialforschung in Theorie und Praxis. Zum Verhältnis von qualitativer und quantitativer Forschung. In: Clemens, W; Strübing, J. [Hg.]. Empirische Sozialforschung und gesellschaftliche Praxis. Bedingungen und Formen angewandter Forschung in den Sozialwissenschaften. Leske+Budrich: Opladen.
- LeCompte, M.; Goetz, J. (1982). Problems of reliability and validity in ethnographic research. *Review of Educational Research*, H.1, 31-60.
- Lienert, G.A. (1989)⁴. Testaufbau und Testanalyse. Psychologie Verlags Union: München.
- Lincoln, Y.S., Guba, E. (1985). Naturalistic Inquiry. Sage: Beverly Hills.
- Lincoln, Y.S.; Denzin, N.K. (1994). The Fifth Moment. In: Denzin, N.K.; Lincoln, Y.S. [Hg.]. Handbook of qualitative research. Sage: Thousand Oaks, 576-586.
- Mayring, P. (1996)³. Einführung in die qualitative Sozialforschung. Eine Anleitung zu qualitativem Denken. Psychologie Verlags Union: Weinheim.

- Merten, K. (1995)². Inhaltsanalyse. Einführung in Theorie, Methode und Praxis. Westdeutscher Verlag: Opladen.
- Milgram, S. (1982). Das Milgram-Experiment. Zur Gehorsamsbereitschaft gegenüber Autoritäten. Rowohlt: Reinbek bei Hamburg.
- Murphy, E. et al. (1998). Qualitative research methods in health technology assessment: a review of the literature. *Health Technology Assessment*, Nr.16.
- Norusis, M.J. (1994). SPSS Advanced Statistics 6.1. SPSS Inc.: Chicago.
- Prein, G.; Kluge, S.; Kelle, U. (1994). Strategien zur Sicherung von Repräsentativität und Stichprobenvalidität bei kleinen Samples. Arbeitspapiere des Sfb 186, Nr.18: Bremen.
- Reid, A.; Gough, S. (2000). Guidelines for Reporting and Evaluating Qualitative Research: what are the alternatives? *Environmental Education Research*, H.1, 59-91.
- Saldern, M.v. (1995). Zum Verhältnis von qualitativen und quantitativen Methoden. In: König, E; Zedler, P. [Hg.]: Bilanz qualitativer Forschung. Bd.1, 331-371. Deutscher Studien Verlag: Weinheim.
- Schnell, R.; Hill, P.B.; Esser, E. (1992)³. Methoden der empirischen Sozialforschung. Oldenbourg: München.
- Scott, W. (2000). Editorial. *Environmental Education Research*, H.1, 5-8.
- Seale, C. (1999a). The Quality of Qualitative Research. Sage: London.
- Seale, C. (1999b). Quality in Qualitative Research. *Qualitative Inquiry*, H.4, 465-478.
- Seale, C.; Silverman, D. (1997). Ensuring rigour in qualitative research. *European Journal of Public Health*, H.7, 379-384.
- Smith, J. (1984). The problem of criteria for judging interpretative inquiry. *Educational Evaluation and Policy Analysis*, H.4, 379-391.
- Smith-Sebasto, N.J. (2000). Potential Guidelines for Conducting and Reporting Environmental Education Research: qualitative methods of inquiry. *Environmental Education Research*, H.1, 9-26.
- Steinke, I. (1998) Kriterien für die Bewertung qualitativer Forschung. Technische Universität Berlin [Diss. Phil.].
- Steinke, I. (1999). Kriterien qualitativer Forschung. Ansätze zur Bewertung qualitativ-empirischer Sozialforschung. Juventa Verlag: Weinheim.

-
- Steinke, I. (2000). Gütekriterien qualitativer Forschung. In: Flick, U.; von Kardorff, E.; Steinke, I. [Hg.]. *Qualitative Forschung. Ein Handbuch*. Rowohlt: Reinbek bei Hamburg, 319-331.
- Sturm, G. (1994). Wie forschen Frauen? Überlegungen zur Entscheidung für qualitatives oder quantitatives Vorgehen. In: Diezinger, A. et al. [Hg.]: *Erfahrung mit Methode. Wege sozialwissenschaftlicher Frauenforschung*. Kore: Freiburg i.Br., 85-104.
- Swanborn, P.G. (1996). A common base for quality control criteria in quantitative and qualitative research. *Quality & Quantity*, H.30, 19-35.

Anhang I: Liste der untersuchten Texte

- Achterberg, S. (2000). Das Kind als Objekt des Begehrens. Die pädophile Ausbeutung der generationalen Hierarchie. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, H.2, 167-180.
- Ayaß, R. (1997). Die kleinen Propheten des „Worts zum Sonntag“. *Zeitschrift für Soziologie*, H.3, 222-235.
- Bauer, K.-O. (1998). Pädagogisches Handlungsrepertoire und professionelles Selbst von Lehrerinnen und Lehrern. *Zeitschrift für Pädagogik*, H.3, 343-359.
- Bender-Szymanski, D. (1999). Wie lernen Lehrer von Migrantenkindern? Eine Prozessanalyse interkulturellen Lernens bei deutschen StudienreferendarInnen in multikulturellen Schulen. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Heft 1, 52-71.
- Blömer, U.; Garz, D. (1998). „Es war ein langsames Getriebenwerden ...“. Biographieanalyse eines nichtjüdischen Emigranten. *BIOS – Zeitschrift für Biographieforschung und Oral History*, H.1, 76-102.
- Böttcher, I.; Weishaupt, H. (1999). Gymnasien in Thüringen und Bayern. Ergebnisse einer vergleichenden Untersuchung. *Zeitschrift für Pädagogik*, H.5, 699-716.
- Böttger, A. (2000). Devianz als Episode – Wege des „Ausstiegs“ aus kriminalisierbarem Handeln. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Beiheft 3, 77-90.
- Bora, A.; Epp, A. (2000). Die imaginäre Einheit der Diskurse. Zur Funktion von „Verfahrensgerechtigkeit“. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, H.1, 1-35.
- Breidenstein, G. (1997). Der Gebrauch der Geschlechterunterscheidung in der Schulklasse. *Zeitschrift für Soziologie*, H.5, 317-336.
- Breitenbach, E.; Kausträter, S. (1998). „Ich finde, man braucht irgendwie eine Freundin“ – Beziehungen zu Gleichaltrigen in der weiblichen Adoleszenz. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Heft 4, 389-402.

- Brinkmann, U.; Seifert, M. (2001). „Face to Interface“: Zum Problem der Vertrauenskonstitution im Internet am Beispiel von elektronischen Auktionen. *Zeitschrift für Soziologie*, H.1, 23-47.
- Bühler-Niederberger, D.; Hungerland, B.; Bader, A. (1999). Minorität und moralische Instanz – der öffentliche Entwurf von Kindern. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Heft 2, 128-150.
- Dallinger, U. (1998). Der Konflikt zwischen familiärer Pflege und Beruf als handlungstheoretisches Problem. *Zeitschrift für Soziologie*, H.2, 94-112.
- Ehret et al. (2000). Stütze – Stigma – Stolperstein? Zu den Auswirkungen strafrechtlicher Intervention auf Erwerbsbiographien von Berufsanfängern. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Beiheft 3, 66-76.
- Faulstich-Wieland, H.; Güting, D.; Ebsen, S. (2001). Einblicke in „Genderism“ im schulischen Verhalten. *Zeitschrift für Pädagogik*, H.1, 67-79.
- Fiege, B.; Dollase, R. (1998). Evaluation kollegialer Beratung in Gruppen von Lehrern und Lehrerinnen. *Zeitschrift für Pädagogik*, H.3, 379-395.
- Frerichs, P. (2000). Klasse und Geschlecht als Kategorien sozialer Ungleichheit. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, H.4, 36-59.
- Granosik, M. (2000). Professionalität und Handlungsschemata der Sozialarbeit am Beispiel Polens. *Zeitschrift für qualitative Bildungs- Beratungs- und Sozialforschung*, H.1, 49-96.
- Heer, H. (1998). „Am Anfang haben wir es aus Überzeugung, später dann aus Pflicht getan“. Kollektive und individuelle Formen der Legitimation. *BIOS – Zeitschrift für Biographieforschung und Oral History*, H.1, 42-68.
- Heintz, B.; Nadai, E. (1998). Geschlecht und Kontext. De-Institutionalisierungsprozesse und geschlechtliche Differenzierung. *Zeitschrift für Soziologie*, H.2, 75-93.
- Herwartz-Emden, L.; Westphal, M. (1999). Frauen und Männer, Mütter und Väter. Empirische Ergebnisse zu Veränderungen der Geschlechterverhältnisse in Einwandererfamilien. *Zeitschrift für Pädagogik*, H.6, 885-902.
- Hof, Chr. (2000). Subjektive Wissenstheorien als Grundlage des Unterrichts. Ergebnisse einer Explorationsstudie. *Zeitschrift für Erziehungswissenschaft*, H.4, 595-607.

- Hopf, Chr. (2001). Gewalt, Biographie, Medien. Qualitative Analysen zur subjektiven Bedeutung filmischer Gewaltdarstellungen. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, H.2, 150-169.
- Hüttermann, J. (2000). Der avancierende Fremde. Zur Genese von Unsicherheits-erfahrungen und Konflikten in einem ethnisch polarisierten und sozialräumlich benachteiligten Stadtteil. *Zeitschrift für Soziologie*, H.4, 275-293.
- Jakob, A.; Mutz, G. (1999). Arbeitslosigkeit in der Erwerbsbiographie. Ergebnisse einer dynamischen Deutungsmusteranalyse. *BIOS – Zeitschrift für Biographieforschung und Oral History*, H.1, 73-91.
- Joerk, Chr. (1999). „Grüne Witwen“ am ostdeutschen Metropolenrand? Lebens- und Handlungsformen von Frauen im suburbanen Raum. *BIOS – Zeitschrift für Biographieforschung und Oral History*, H.2, 236-256.
- Kalthoff, H. (2000). „Wunderbar, richtig“. Zur Praxis mündlichen Bewertens im Unterricht. *Zeitschrift für Erziehungswissenschaft*, H.3, 429-446.
- Kauke, M. (1998). Beobachtungsstudien zu Konstanz und Wandel von prosozialem Verhalten unter Grundschulkindern in Ostberlin. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Beiheft 2, 52-68.
- Kelle, H. (1999). Geschlechterterreitorien. Eine ethnographische Studie über Spiele neun- bis zwölfjähriger Schulkinder. *Zeitschrift für Erziehungswissenschaft*, H.2, 211-228.
- Kirchhöfer, D. (1998). Veränderungen in den alltäglichen Lebensführungen ostdeutscher Kinder – ein qualitativer Längsschnitt *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Beiheft 2, 34-51.
- Klieme, E.; Bos, W. (2000). Mathematikleistung und mathematischer Unterricht in Deutschland und Japan. Triangulation qualitativer und quantitativer Analysen am Beispiel der TIMS-Studie. *Zeitschrift für Erziehungswissenschaft*, H.3, 359-380.
- Koller, Chr. (2000). „Alsacien, Déserteur!“ Die Kriegserfahrungen des Elsässer Bauern Dominik Richert im Spiegel seiner Memoiren. *BIOS – Zeitschrift für Biographieforschung und Oral History*, H.2, 225-239.
- Leiprecht, R. (1999). Von Verunsicherung bis zu „gezielter“ Wahrnehmung. Wie Jugendliche in den Niederlanden mit der Berichterstattung zu asylsuchenden

- Flüchtlingen umgehen. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Heft 4, 421-437.
- Leo, A. (1999). Tabu und Tradition. Der 17.Juni und der 100-Tage-Streik in der Erinnerung der Hennigsdorfer Stahlarbeiter. *BIOS – Zeitschrift für Biographieforschung und Oral History*, H.1, 58-72.
- Matt, E. (1999). Jugend, Männlichkeit und Delinquenz. Junge Männer zwischen Männlichkeitsritualen und Autonomiebestrebungen. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Heft 3, 259-276.
- Mierendorff, J.; Olk, T. (1999). Sozialhilfe und Lebensentwurf. Deutungs- und Bewältigungsmuster von Sozialhilfeempfänger(innen) im ostdeutschen Transformationsprozess. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Beiheft 3, 262-279.
- Moch, M.; Junker, M. (1998). Zur „sozialen Logik“ der Generationenbeziehungen deutscher und amerikanischer Scheidungsväter. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Heft 3, 250-266.
- Morawe, P. (1999). Untersuchungshaft bei der Staatssicherheit der DDR: Realitätsdiffusion infolge psychischer Folter. *BIOS – Zeitschrift für Biographieforschung und Oral History*, H.2, 191-208.
- Nazarkiewicz, K. (1997). Moralisieren über Ethnien. Die Reflexivität der Stereotypenkommunikation. *Zeitschrift für Soziologie*, H.3, 181-201.
- Nolda, S. (1998). Distanzierte Familiaritäten. Zur möglichen Pädagogik von Fernseh-Familienserien. *Zeitschrift für Erziehungswissenschaft*, H.1, 89-102.
- Papastefanou, Chr. (2000). Der Auszug aus dem Elternhaus – ein vernachlässigter Gegenstand der Entwicklungspsychologie. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, H.1, 55-69.
- Peez, G. (1999). Kinderzeichnung im biographischen Rückblick. Erinnerung und Rekonstruktion der frühesten Kritzelerfahrung einer 70jährigen. Eine exemplarische sequentielle Textanalyse. *BIOS – Zeitschrift für Biographieforschung und Oral History*, H.1, 108-115.
- Piechotta, G. (2000). Von der „Schwesterschülerin“ zur Studentin. Bildungs- und Berufserfahrungen und neue Perspektiven in der Pflege. *Zeitschrift für qualitative Bildungs- Beratungs- und Sozialforschung*, H.1, 131-150.

- Popp, U.; Meier, U.; Tilmann, K.-J. (2001). Es gibt auch Täterinnen: Zu einem bisher vernachlässigten Aspekt der schulischen Gewaltdiskussion. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, H.2, 170-191.
- Pries, L. (1998). Das Verhältnis von Ausbildung und Erwerb in vergleichender Perspektive: Sequenz, Übergangsphase oder Parallele? Fallstudienresultate aus einer mexikanischen Großstadt. *Zeitschrift für Soziologie*, H.1, 39-52.
- Reitemeier, U. (2000). Zum interaktiven Umgang mit einbürgerungsrechtlichen Regelungen in der Aussiedlerberatung. Gesprächsanalytische Beobachtungen zu einem authentischen Fall. *Zeitschrift für qualitative Bildungs- Beratungs- und Sozialforschung*, H.2, 253-282.
- Riedel, A. (2001). Doppelter Sozialstatus, späte Adoleszenz und Protest. Algerische Vertragsarbeiter in der DDR. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, H.5, 76-95.
- Rieken, B. (1999). Besuch aus dem Jenseits. Volksglaube im biographischen Kontext. *BIOS – Zeitschrift für Biographieforschung und Oral History*, H.2, 221-235.
- Rieker, P. (2000). Ethnozentrismus im Jugendalter. Ein multiperspektivischer Beitrag zur Sozialisationsforschung. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, H.1, 39-54.
- Schaeper, H.; Kühn, T. (2000). Zur Rationalität familialer Entscheidungsprozesse am Beispiel des Zusammenhangs zwischen Berufsbiographie und Familiengründung. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Beiheft 3, 124-145.
- Schäffer, B. (1999). Stilistische Ein-Findungsprozesse. Zur Rekonstruktion jugendkultureller Stilaneignung im Medium populärer Musik. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Heft 3, 293-310.
- Scheffer, T. (1997). Dolmetschen als Darstellungsproblem. Eine ethnographische Studie zur Rolle der Dolmetscher in Asylanhörungen. *Zeitschrift für Soziologie*, H.2, 159-180.
- Schöll, A. (1998). Verweigerung, stellvertretende Deutung und autonome Gestaltung von Lebenspraxis. Lebensgeschichten von Mitgliedern in religiösen Gruppen und Milieus mit strikten normativen Lebensführungskonzepten. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Heft 4, 357-372.

- Stenger, H. (1998). Soziale und kulturelle Fremdheit. Zur Differenzierung von Fremdheitserfahrungen am Beispiel ostdeutscher Wissenschaftler. *Zeitschrift für Soziologie*, H.1, 18-38.
- Stirn, A. (2000). Überleben und Auseinandersetzung mit dem Holocaust-Trauma in einer Auswahl literarischer Zeugnisse jüdischer Schriftsteller. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, H.4, 720-760.
- Terhart, E. (1999). Zensurengebung und innerschulisches Selektionsklima – die Rolle der Schulleitung. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Heft 3, 277-292.
- Wittel, A. (1998). Gruppenarbeit und Arbeitshabitus. *Zeitschrift für Soziologie*, H.3, 178-192.
- Witzel, A.; Kühn, T. (2000). Orientierungs- und Handlungsmuster beim Übergang in das Erwerbsleben. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Beiheft 3, 9-29.
- Wohlrab-Sahr, M. (1998). Konversion als Re-Sozialisation. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, Heft 4, 373-388.
- Zirfas, J.; Wulf, Chr. (2001). Integration im Ritual. Performative Prozesse und kulturelle Differenzen. *Zeitschrift für Erziehungswissenschaft*, H.2, 191-208.

Anhang II: Ausgaben der Faktorenanalyse

Kommunalitäten

	Anfänglich	Extraktion
Samplestrategie angemessen	1,000	,868
Dokumentation d. Vorverständnisses	1,000	,592
Dokumentation d. Daten	1,000	,732
Dokumentation d. Auswertung	1,000	,757
Dokumentation v. Problemen u. Eigenschaften	1,000	,814
theoret. Auswahl d. Untersuchungseinheiten	1,000	,854
Reflexivität	1,000	,652
Quantifizierung	1,000	,702
"Thick Deskription"	1,000	,696
Kontrastierungen	1,000	,725
Evidenz und Gegenevidenz	1,000	,796
Nachvollziehbarkeit	1,000	,750
Darstellung d. eigenen Standpunktes	1,000	,775
Falsifikation als Grundsatz	1,000	,634
Einfluss des Kontextes berücksichtigt	1,000	,583
Tringulation gewählt	1,000	,803

Extraktionsmethode: Hauptkomponentenanalyse.

Rotierte Komponentenmatrix^a

	Komponente					
	1	2	3	4	5	6
Samplestrategie angemessen			,879			
Dokumentation d. Vorverständnisses	,597	,272		,385		
Dokumentation d. Daten		,535			-,470	-,408
Dokumentation d. Auswertung		,777				
Dokumentation v. Problemen u. Eigenschaften				,844		
theoret. Auswahl d. Untersuchungseinheiten			,897			
Reflexivität		,395		,581	,349	
Quantifizierung		,281			,751	
"Thick Deskription"	,633			,400	,362	
Kontrastierungen			,256			,771
Evidenz und Gegenevidenz	,778		,255			,260
Nachvollziehbarkeit		,799				
Darstellung d. eigenen Standpunktes	,813					
Falsifikation als Grundsatz						,743
Einfluss des Kontextes berücksichtigt				,556		,476
Tringulation gewählt			,273		,843	

Extraktionsmethode: Hauptkomponentenanalyse.

Rotationsmethode: Varimax mit Kaiser-Normalisierung.

a. Die Rotation ist in 10 Iterationen konvergiert.