

Merkena, Hans

## Immunsierung gegen Kritik durch Methodisierung der Kritik

*Benner, Dietrich [Hrsg.]; Borrelli, Michele [Hrsg.]; Heyting, Frieda [Hrsg.]; Winch, Christopher [Hrsg.]: Kritik in der Pädagogik. Weinheim : Beltz 2003, S. 33-53. - (Zeitschrift für Pädagogik, Beiheft; 46)*

urn:nbn:de:0111-opus-39597

in Kooperation mit / in cooperation with:

# BELTZ

<http://www.beltz.de>

### Nutzungsbedingungen / conditions of use

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.  
By using this particular document, you accept the above-stated conditions of use.

### Kontakt / Contact:

**peDOCS**  
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)  
Mitglied der Leibniz-Gemeinschaft  
Informationszentrum (IZ) Bildung  
Schloßstr. 29, D-60486 Frankfurt am Main  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Zeitschrift für Pädagogik · 46. Beiheft

# **Kritik in der Pädagogik**

Versuche über das Kritische  
in Erziehung und Erziehungswissenschaft

Herausgegeben von  
Dietrich Benner, Michele Borrelli, Frieda Heyting  
und Christopher Winch

Beltz Verlag · Weinheim, Basel, Berlin

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder ein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. Auch die Rechte der Wiedergabe durch Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder auf ähnlichem Wege bleiben vorbehalten. Fotokopien für den persönlichen oder sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopie hergestellt werden. Jede im Bereich eines gewerblichen Unternehmens hergestellte oder genutzte Kopie dient gewerblichen Zwecken gem. § 54 (2) UrhG und verpflichtet zur Gebührenzahlung an die VG Wort, Abteilung Wissenschaft, Goethestr. 49, 80336 München, von der die einzelnen Zahlungsmodalitäten zu erfragen sind.

© 2003 Beltz Verlag · Weinheim, Basel, Berlin  
Herstellung: Klaus Kaltenberg  
Satz: Druckhaus »Thomas Müntzer«, Bad Langensalza  
Druck: Druckhaus »Thomas Müntzer«, Bad Langensalza  
Printed in Germany  
ISSN 0514-2717

Bestell-Nr. 41147

# Inhaltsverzeichnis

Vorwort der Herausgeber.....	7
<i>Christopher Winch</i>	
Die Entwicklung kritischer Rationalität als pragmatische Aufgabe der Erziehung ..	13
<i>Hans Merkens</i>	
Immunsierung gegen Kritik durch Methodisierung der Kritik .....	33
<i>Helmut Heid</i>	
Domestizierung von Kritik. Legitimationsprobleme des Kritischen im Kontext pädagogisch bedeutsamen Denkens und Handelns .....	54
<i>Frieda Heyting</i>	
Die Relativierung von Kritik. Zur erziehungsphilosophischen Kraft antifundamentistischer Kritik .....	75
<i>Dietrich Benner</i>	
Kritik und Negativität. Ein Versuch zur Pluralisierung von Kritik in Erziehung, Pädagogik und Erziehungswissenschaft .....	96
<i>Jörg Ruhloff</i>	
Problematisierung von Kritik in der Pädagogik .....	111
<i>Jan Masschelein</i>	
Trivialisierung von Kritik. Kritische Erziehungswissenschaft weiterdenken .....	124
<i>Michele Borrelli</i>	
Utopisierung von Kritik. Pädagogik im Spannungsverhältnis von utopischem Begriff und kontingenter Faktizität .....	142

Hans Merkens

# Immunisierung gegen Kritik durch Methodisierung der Kritik

## 1. Einleitende Bemerkungen: Vom Ungenügen der rohen Erfahrung als Grundlage der Erziehungswissenschaft

Versuche, Wissenschaften auf eine Erfahrungsgrundlage zu stellen, stehen vor der Schwierigkeit, dass die Erfahrung – bei genauer Betrachtung – keine zuverlässige Basis bieten kann, weil sie selbst auf einem Irrtum beruhen<sup>1</sup> oder allzu schnell nahe legen kann, die Möglichkeiten des Handelns auf eine Variante zu beschränken, die keineswegs optimal sein muss, indem die Bestätigung dieser Variante durch Erfahrung als Anlass genommen wird, andere mögliche Varianten nicht mehr zu erproben.<sup>2</sup> Blankenburg (1975) hat zusätzlich darauf verwiesen, dass nicht einmal der Sinn des Wortes Erfahrung einheitlich sei; er wandle sich vielmehr, „je nachdem ob wir Erfahrung gegen Erfindung, Phantasie, Interpretation, Deutung, gegen Theorie, deduktives Vorgehen oder (als ‚bloße‘ Erfahrung) gegen die Einsicht in die Notwendigkeit einer Sache abgrenzen“.

In der Erziehungswissenschaft ist bereits von Herbart (1806/1964) Erfahrung, die auf der ungeprüften und nicht weiter reflektierten Gewohnheit aufbaut, nachdrücklich kritisiert worden.<sup>3</sup> Das Sich-Berufen auf die rohe Erfahrung unterliegt seitdem dem Verdikt, dass sich – zumindest auf Erfahrung allein – keine Regeln für die rechte Form der Praxis begründen lassen. Nicht nur die Zufälligkeiten des Augenblicks beim Sammeln von Erfahrung bilden in dieser Hinsicht ein Hindernis, viel gefährlicher ist die Tatsache, dass nicht mehr alle Möglichkeiten durchgespielt werden, sondern nur die eine, zufällig aufgetretene, als Lösung für ein Problem akzeptiert und als hinreichend betrachtet wird. Bei diesem Vorgehen würde der Alltag der bisherigen Praxis als hinreichende Begründung für die Gestaltung der zukünftigen Praxis hingenommen. Deshalb ist der Versuch, Regeln für das praktische Handeln allein über die Beobachtung eben des Praktizierten zu gewinnen, zumindest für die pädagogische Situation, fragwürdig. Die Erfahrung bedarf offensichtlich neben der Tatsache des Machbaren einer weiteren Be-

- 1 Optische Täuschungen sowie die Luftspiegelung „Fata Morgana“ können aus dem Bereich von Erfahrungen, die den Naturwissenschaften zuzuordnen sind, als Beleg für diese Aussage herangezogen werden.
- 2 Das ist vor allem bei angewandten Wissenschaften und der Anwendung wissenschaftlicher Erkenntnisse der Fall, wenn das Finden einer Lösung als hinreichend angesehen wird und weitere Experimente nicht mehr unternommen werden.
- 3 Später hat Stegmüller (1965) zwei Wirklichkeitsbegriffe unterschieden, einen, den er im Anschluss an Carnap (1966a) als sinnlos metaphysisch bezeichnet, und einen weiteren, der als sinnvoll empirisch gekennzeichnet wird. Die folgenden Überlegungen beziehen sich ausschließlich auf diesen zweiten Begriff und dessen Verwendung in den so genannten empirischen Wissenschaften.

gründung, die ich vorläufig als kritische Haltung gegenüber dem alltäglich Gewohnten bezeichne. Experiment (Trapp 1913) und Spekulation (Herbart 1806/1964) sind traditionell als kritischer Habitus in diesem Sinne entwickelt worden. Nur diejenigen Erfahrungen sollten einbezogen werden, die einer systematischen Kontrolle (Experiment) oder einer theoretischen Einordnung (Spekulation) genügten. Beide Vorgehensweisen sollten dazu führen, neue Formen der Praxis anzuregen bzw. zu erproben. Auf der Basis von Spekulation sollten Praxen des Erziehens und Unterrichtens entworfen und damit neue Erfahrungen ermöglicht werden,<sup>4</sup> auf der Basis von Experimenten neue Praxen getestet werden, neue Erfahrungen in der Praxis also durch systematisches Testen im Vorfeld gewertet werden. In beiden Fällen ist aber noch nicht ausgeschlossen, dass das auf diese Weise Erfahrene auf einem Irrtum beruht und sich als nicht angemessen bzw. haltbar erweist. Eine weitere Gefahr besteht darin, dass es sich bei dem, was man als Erfahrung aus der Praxis von Erziehung und Unterricht ansieht, um Einbildungen handeln könnte. Die Fragilität des Erfahrens in sozialen Situationen hat Laing (1969, S. 26f.) anschaulich beschrieben. Mit diesem Hinweis ist eines der Probleme angesprochen, welches diejenigen, welche Erfahrung als Grundlage wählen, immer wieder beschäftigt hat: Wie lassen sich Erfahrungen vom Traum, aber auch von Halluzinationen unterscheiden (Stegmüller 1965, S. 28)? Sollen Träume oder Halluzinationen von Erfahrungen unterschieden werden können, dann bedarf es entsprechender Prüfungen: Das, was als Basis dienen soll, muss aus dieser Sicht einer kritischen Überprüfung standhalten. Das ist nicht nur in der Pädagogik, sondern auch in anderen Disziplinen und nicht zuletzt in der Wissenschaftstheorie eines der wichtigsten Grundthemen. Damit ist eine der wesentlichen Voraussetzungen für den Rückgriff auf Erfahrung in der Wissenschaft benannt: Das, was erfahren wird, bedarf der kritischen Überprüfung, wenn es in die Wissenschaft z.B. als Bestätigung einer Erkenntnis übernommen werden soll. Vor diesem Hintergrund hat sich in den Wissenschaften die Vorgehensweise des Methodisierens der Kritik ausgebildet, die im Folgenden näher untersucht werden wird. Diese Methodisierung der Kritik war erforderlich, weil nur die Wissenschaftlichkeit einer auf Erfahrung aufbauenden Erkenntnis gesichert werden konnte, wie sie für die Naturwissenschaften, aber auch die empirischen Sozialwissenschaften, typisch ist.

Die bisher benannten Schwierigkeiten beim Bauen auf Erfahrung vergrößern sich nämlich noch, wenn sich Tatsachen dem direkten Zugriff des Erfahrens verschließen (Blankenburg 1975), wie Levitt/March (1996) für das Organisationslernen am Beispiel des Lernens als Interpretation von Erfahrungen dargestellt haben.<sup>5</sup> Merkens (1974) hat für das Beobachten auf die Rekursivität des Prozesses beim Gewinnen von Beobachtungsaussagen hingewiesen, wenn beim Sammeln neuer Erfahrungen auf biografisch

4 Die Spekulation bedurfte der Bewährung aus der Erfahrung (vgl. Ruhloff in diesem Band).

5 In diesem Zusammenhang ist von besonderem Interesse, dass man in der Wissenschaft ähnlich wie im Alltag, wenn man ein Merkmal betrachten kann, das man für konstitutiv für eine Sache hält, davon spricht, dass man die Sache beobachtet habe, z.B. das Innere der Sonne, obwohl man nur ein Merkmal beobachten konnte, das man dieser Sache zuordnet, Neutrinos (Schickore 1999, S. 275); vgl. auch Bac 2000, 40 zum gleichen Thema: Verdichtung von Messungen durch Interpretation zu für beobachtet Gehaltenem.

früher gemachte Erfahrungen zurückgegriffen wird. Diese vorangehenden Erfahrungen unterliegen bei ihrem Gewinnen den gleichen Restriktionen wie sie für das Gewinnen der Neuen gelten: Das, was beim Sammeln von Erfahrungen als Filter wirkt, ist selbst nur begrenzt durch die Aufmerksamkeit beeinflussbar. Wenn es beeinflusst wird, gibt es einen interpretativen Anteil, der aus früheren Erfahrungen entstanden ist und im Akt des Erfahrens bzw. Beobachtens kaum kontrollierbar ist.<sup>6</sup> Beeinträchtigungen dieser Art verstärken sich, wenn ein Vorgang wie der Prozess der Erziehung beobachtet werden soll, der selbst nicht direkt beobachtbar ist. Vielmehr kann in diesem Fall zumindest im Alltag erst aus den Ergebnissen des Prozesses darauf geschlossen werden, dass Erziehung stattgefunden hat, wie bereits Tausch/Tausch (1970, S. 39ff.) beschrieben haben.

In diesen Varianten wird ersichtlich, dass Erfahrungen roh nicht verfügbar sind. Über die Erfahrung wird nicht der unmittelbare Zugang zu Ereignissen oder Tatsachen vermittelt. Vielmehr bedarf es beim Gewinnen von Erfahrungen des Durchgangs durch zwei Filter: Erstens müssen die Sinne so intensiv affiziert werden, dass sich Empfindungen ausbilden; hierbei handelt es sich um einen physiologischen Filter. Daran anschließend können zweitens Wahrnehmungen Gestalt annehmen, indem frühere Erfahrungen aktiviert werden, wie Gehlen (1962) anschaulich beschrieben hat. Erst danach entwickelt sich die neue Erfahrung, die auf dieser Wahrnehmung aufbaut (Merkens 1972). Speziell im Fall des Erziehens wird interpretierend durch den Vergleich mit früheren Erfahrungen oder Berichten Dritter geschlossen, dass erzogen worden ist (Tausch/Tausch 1970). Bei der Erfahrung handelt es sich also um einen komplexen Prozess, bei dem das, was man alltagssprachlich für Erfahrung hält, mit interpretativen Anteilen verschränkt ist, bei deren Genese frühere Erfahrungen eine wichtige Rolle gespielt haben.

Beim Gewinnen wissenschaftlicher Erfahrungen gibt es einen ähnlichen Ablauf, weil neben der Messung die Interpretation des Gemessenen erforderlich ist, damit es sich um eine Erfahrung handeln kann (Schickore 1999, S. 283). Für diese Interpretation gibt es Regeln, die vom Gewinnen der Aussagen bis hin zu deren Bearbeitung bzw. zum Umgang mit ihnen reichen, wie beispielsweise Prim/Tilmann (1973) am Beispiel der Sozialwissenschaften dargestellt haben, zu denen sie auch die Erziehungswissenschaft rechnen. Opp (1970, S. 12) hat entsprechend die sozialwissenschaftliche Methodologie dahingehend charakterisiert, sie versuche, „die Arbeit des Sozialwissenschaftlers einer Kritik zu unterziehen und Vorschläge für eine verbesserte sozialwissenschaftliche Praxis zu machen.“ Die Regeln der Durchführung sozialwissenschaftlicher Untersuchungen sollen dabei einer Kritik unterzogen werden. Dazu bedarf es bestimmter Kriterien, die Hülst (1973, S. 14) mit Prüfbarkeit, Präzision und Informationsgehalt bezeichnet hat.

Das, was eingangs als rohe Erfahrung gekennzeichnet wurde, ist demnach eher eine Art der Erfahrung, der es an Reflexion mangelt. Man könnte die folgenden Überlegungen auch unter das Thema stellen, wie die für wissenschaftlich verwertbare Erfahrung erforderlichen reflexiven Prozesse gestaltet und welchen Methoden der Prüfung sie un-

6 Fernsehsendungen wie „Achtung, Kamera“ haben aus Situationen, die so entstehen, ihre Beispiele gewonnen.

terworfen werden. Diese Methoden kann man als den Versuch ansehen, die zufälligen, nicht kontrollierbaren Anteile der Erfahrung zu minimieren. Die Prüfung erfolgt also kritisch gegenüber dem, was man erfahren hat. Dazu dient die Methodisierung der Kritik, die sich gegen die vorschnelle Übernahme einzelner Erfahrung richtet, die vielmehr als Instanz zur Überprüfung verstanden werden muss. Die These ist dabei, dass eine nachprüfbar Kritik der jeweiligen Erfahrungen ohne Methodisierung nicht gelingen kann. Erfahrungen, die der Kontrolle der methodisierten Kritik nicht unterworfen worden sind, werden im Folgenden als vorkritisch gekennzeichnet.<sup>7</sup>

## 2. Wiederholbarkeit und Konsens als Programm

Wenn Wissenschaftlichkeit von Aussagen angestrebt wird, dann ist ein möglicher Aspekt, dass sie über den Zeitpunkt hinaus, zu dem sie gewonnen werden, als gültig angesehen werden können. Diesem Aspekt kommt in den empirischen Wissenschaften eine besondere Bedeutung zu. Die Aussagen sollen in dem Sinne angemessen sein, dass die Ereignisse, die sie erfassen bzw. wiedergeben, als vorhanden bzw. bestätigt angesehen werden können. Das kann naturgemäß nur so lange sein, wie nichts Gegenteiliges bekannt geworden ist. Singuläre, vom Zufall generierte oder nicht nachprüfbar, Erfahrungen sollen ausgeschlossen werden. Vielmehr soll aus der Beobachtung einmaliger Eigenschaften eines Dings, dass z.B. ein Körper zu einem bestimmten Zeitpunkt rotes Licht reflektiert, auf eine dauernde Eigenschaft geschlossen werden können, nämlich, dass der Körper rot ist (Carnap 1966b, S. 7). Häufig soll die Erfahrung zusätzlich in dem Sinne genutzt werden, dass sie unbedenklich auf ähnliche Situationen übertragen werden kann. Außerdem sollen sie nicht nur vom Urheber, sondern auch von anderen Personen als zutreffend angesehen werden (Prim/Tilmann 1973, S. 14).

Um diesen Anforderungen zu genügen, müssen zur Gewinnung und Sicherung von Erfahrungen bestimmte Verfahren bzw. Methoden eingesetzt werden, die als unstrittig angesehen werden. Mit dem Ziel der Präzisierung eines entsprechenden Katalogs von Anforderungen lassen sich unterschiedliche Probleme identifizieren, die gelöst sein müssen, wenn das Bauen auf Erfahrung erfolgreich gestaltet werden soll. Einerseits muss gesichert werden, dass die Zuordnung der empirischen Daten zu den Aussagen über sie in einer Weise erfolgt, die keinerlei Zweifel zulässt (Carnap 1961). Das lässt sich am Beispiel des Beobachtens beschreiben (Merkens 1974). Eine Anforderung, die auf der Seite der Ereignisse daraus resultiert, lautet, dass die Ereignisse präzise bestimmt und Verwechslungen ausgeschlossen werden müssen. Das gelingt eher, wenn es sich um Elementarereignisse handelt; dies sind Ereignisse, die kleinste Einheiten bilden. Ähnlich wie in der Molekularchemie setzen sich größere Ereignisse auch in anderen Disziplinen aus mehreren Elementarereignissen zusammen. Elementaraussagen können dann Elementarereignissen zugeordnet werden. Dieses Problem wird später noch unter dem

7 Diese Verwendung des Begriffs ‚vorkritisch‘ unterscheidet sich von der bei Benner (2003) in diesem Band, der die Urteilsformen behandelt.



Thema Interpretation aufgegriffen werden. Es sei aber schon hier darauf verwiesen, dass in den Sozialwissenschaften Versuche, solche Elementarereignisse oder Basishandlungen gemäß dem Vorbild von Naturwissenschaften wie z.B. Chemie oder Physik zu bestimmen, gescheitert sind (vgl. Danter 1977; Baier 1977).

Außerdem muss sichergestellt werden, dass die den Daten korrespondierenden Ereignisse systematisch und nicht zufällig auftreten. Mit dieser Zielrichtung sind in den empirischen Wissenschaften entsprechende Vorgehensweisen entwickelt worden. Wenn die gleiche Erfahrung an verschiedenen Orten auf dieselbe Weise gewonnen werden kann – z.B. im Experiment – bzw. wenn verschiedene Betrachter über das Geschehen zum gleichen Ergebnis kommen – z.B. beim Beobachten –, oder man mit verschiedenen Methoden im Feld ähnliche Ergebnisse erhält – man also trianguliert (Flick 2000) –, dann wird die entsprechende Erfahrung als gesichert angesehen. Das kann man als kritischen Vorbehalt gegenüber der Einzelerfahrung ansehen. Dieser wird durch Formen der Reproduzierbarkeit ausgeräumt. Das ist eine methodische Vorschrift. Die kritische Einstellung wird über ein methodisches Vorgehen – Wiederholung – realisiert. Wiederholbarkeit ist eine Anforderung, die auf den ersten Blick einfach zu erfüllen scheint, die aber dennoch in der Praxis große Probleme bereitet, weil man ebenso davon ausgehen kann, dass Elementarereignisse prinzipiell dann nicht wiederholt werden können, wenn Menschen beteiligt sind. Das bedeutet im Kern, dass nur ähnliche Elementarereignisse stattfinden können, wie Benner/Ramseger (1981, S. 8) für Erziehungsverhältnisse erläutert haben: „Entgegen einer weit verbreiteten Annahme ist pädagogische Praxis grundsätzlich nicht wiederholbar. Kein Lehrer kann eine gelungene pädagogische Situation oder einen erfolgreichen Prozess einfach reproduzieren.“ Dies muss nicht im Widerspruch zu quantitativen empirischen Forschungen stehen, wie noch gezeigt werden wird. Nur wird bei diesen die jeweilige individuelle Abweichung als zufällig vom jeweiligen „wahren“ Wert angesehen.

Nun ist Reproduzierbarkeit nicht einmal bei den Experimenten in der Physik vollständig möglich, d.h. sie kann nur unter Vernachlässigung von Merkmalen gesichert werden, die für marginal gehalten oder erklärt werden. Damit ist eine Besonderheit der Erfahrung benannt, die in Widerspruch zu der Bedeutung zu stehen scheint, die ihr innerhalb der empirischen Wissenschaft, das gilt bei quantitativen wie bei qualitativen Methoden, zugebilligt wird: die prinzipielle Singularität der einzelnen Erfahrung, die ausschließt, dass sie in toto wiederholt werden kann. Es bedarf bestimmter Absicherungen, damit trotz dieses Merkmals dennoch auf Erfahrung gebaut werden kann. Es muss geklärt werden, wieweit die Replizierbarkeit von relevanten oder für relevant gehaltenen Merkmalen einer Erfahrung gesichert werden kann, wenn das Gesamt aller Merkmale nicht repliziert werden kann. Voraussetzungen dieser Art lassen sich in der Physik erfüllen, weil es Theorien gibt, die es gestatten, das Auftreten von Ereignissen zu prognostizieren bzw. zu erklären. Auf diese Weise lässt sich die Identität unterschiedlicher Einzelereignisse in Relation zu dieser Theorie herleiten bzw. genau genommen interpretieren.

Die Wiederholbarkeit erfordert offensichtlich in jedem Fall ein gewisses Maß von Abstraktion vom singulären Ereignis. Das kann man als Form einer kritischen Position gegenüber dem einzelnen Ereignis ansehen: Nicht das Ereignis an sich, sondern nur das

für typisch Gehaltene des Ereignisses muss repliziert werden können. Man verhält sich gegenüber dem Ereignis als Ganzem kritisch, indem man bestimmte Eigenschaften oder Merkmale als relevant erklärt. Sie werden als konstitutiv für das Ereignis angenommen. Erfahrungswissenschaftliche Aussagen sind daher in Bezug auf das Geschehen, auf das sie sich beziehen, reduktionistisch, weil sie kritisch sind. Das wird durch die Notwendigkeit der Reproduzierbarkeit erzwungen. In erfahrungswissenschaftlichen Aussagen werden jeweils nur Aspekte der Situation abgebildet bzw. erfasst, auf die sie sich beziehen oder beziehen lassen. Das folgt aus der Annahme, dass man die Erfahrungen, auf denen Theorien basieren bzw. durch die sie bestätigt werden, replizierbar sein müssen. Das ist ein Aspekt, der nicht ohne Folgen für die empirische Forschung in der Erziehungswissenschaft ist. Als Ausweg bietet es sich an, den institutionellen Rahmen pädagogischen Handelns – Familie, Schule, Weiterbildungseinrichtungen etc. – und typische Konfigurationen wie z.B. das Lehr-Lern-Verhältnis als Ausgangspunkt für empirische Untersuchungen zu wählen, also auf die abstrakte Begrifflichkeit zu setzen (vgl. dazu Ruhloff in diesem Band).

Erst auf der Basis von Theorien kann bestimmt werden, welcher Wert der Erfahrung beigemessen werden kann (Métraux/Graumann 1975). Allerdings ist damit noch nicht geklärt, wie mit Erfahrungen verfahren werden kann, die auf der Grundlage von Theorien zu Stande kommen. Es könnte im Sinne der eingeschränkten Wahrnehmung geschehen, dass man auf der Basis der Wiedererkennung selektiert, die Reproduzierbarkeit also nicht im Sinne einer erneuten Bestätigung, sondern nur als Wiederauftreten genommen werden kann und auf diese Weise Marginales zu Relevantem umstilisiert wird. Deshalb fügt Popper (1966) im Kritischen Rationalismus der Notwendigkeit einer Erfahrungsbasis das Erfordernis einer vorangehenden Theorie hinzu. Es sollen also nur solche Erfahrungen einbezogen werden, die auf der Basis der Theorie erwartet werden können. Andere Erfahrungen bleiben, auch wenn sie widersprechen, solange als Sätze über singuläre Ereignisse aus dem Corpus der als wissenschaftlich akzeptierten Basissätze ausgeschlossen, bis eine theoretische Erklärung für ihr Auftreten gefunden worden ist. Diese Form der empirischen Forschung setzt also voraus, dass Theorien entwickelt worden sind, die an der Empirie scheitern oder durch sie verstärkt werden können, indem sich entweder auf der Basis der Theorie erwartete Erfahrungen nicht einstellen oder doch in der vorhergesagten Weise angetroffen werden. Es werden methodische Regeln des Vorgehens entwickelt, mit deren Hilfe sich feststellen lassen soll, ob die Zuordnung von Ereignissen zu Aussagen in der Art geschieht, dass die Angemessenheit der Aussagen in Bezug auf Vorgänge in der Umwelt als wahrscheinlich angesehen werden kann (Verstärkung). Diese Prüfung geschieht kritisch. Die kritische Einstellung gegenüber den Basissätzen ist eine Haltung, die als aufwendig erachtet wird.

Um diese kritische Einstellung auch gegenüber anderen Wissenschaftlern zu dokumentieren, bedarf es methodischer Regeln, aus denen sich die erforderlichen einzelnen Schritte herleiten lassen. Aus diesem Grunde steht am Beginn aller Überlegungen zur Kontrolle der Verhältnisse von Aussagen und Ereignissen die Methodisierung der Kritik.

In der Erziehungswissenschaft haben Ansätze dieser Art in der quantitativen bzw. qualitativen Forschung wenig Beachtung gefunden, obwohl Poppers (1966) Hinweis auf

die Notwendigkeit einer vorangehenden Theorie für die Praxis der erziehungswissenschaftlichen Forschung eine wichtige Vorbedingung enthält: Das Sammeln von Aussagen über Praxis (z.B. qualitativ) durch Befragung von Experten bzw. das Sammeln von Daten in der Praxis (z.B. quantitativ) durch Befragung von Schuljugendlichen setzt schon eine Vorstellung von dem Pädagogischen in der Praxis voraus. Das heißt, die Sammlung von Daten oder Aussagen erfolgt nicht im Sinne der ‚grounded theory‘ (Glaser/Strauss 1967), sondern unter dem Vorbehalt der Angemessenheit der eigenen Annahmen über das Pädagogische in der Praxis. Diese Annahmen werden in der Regel nicht expliziert, sie müssten aber bekannt sein, wenn das Zutreffen der berichteten Erfahrungen von Dritten beurteilt werden soll. In einer vorkritischen Haltung wird das Authentische des eigenen Erlebens für eine hinreichende Erfahrungsbasis gehalten, ohne dass die „Dignität der Praxis bzw. Erfahrung“ kritisch geprüft wird. Arbeiten dieses Typs finden sich vor allem in der pädagogischen Migrationsforschung, wenn das eigene Erleben der Forscher mit der wissenschaftlichen Kriterien genügenden Erfahrung der Praxis verwechselt wird.<sup>8</sup>

Es lässt sich also in der Erziehungswissenschaft eine Form der Erfahrung beobachten, die häufig den Charakter von Praxisberichten aufweist und bei der die Nähe zur Praxis als wesentliches Merkmal behauptet wird.<sup>9</sup> Diese Arbeiten lassen sich in der Systematik der hier vorgenommenen Betrachtung als „vorkritisch“ klassifizieren. In dieser Variante des Bauens auf Erfahrung ersetzt nämlich das Authentische das Kritische. Dabei wird interessanterweise angenommen, dass dieses für authentisch Gehaltene dem Kritischen überlegen sei. Richtig ist an dieser Annahme, dass die durch die Forderung der Reproduzierbarkeit erforderlichen Reduktionen nicht erfüllt werden müssen, man also ganzheitlicher vorgehen kann. Bedenklich ist, dass der interpretative Gehalt der Erfahrung nicht überwunden, ja nicht einmal thematisch werden kann, sondern, soweit er methodisch nicht zu kontrollieren versucht wird, unkritisch übernommen wird. Wird er aber aufzuarbeiten versucht, wie das bei Bohnsack/Nentwig-Gesemann/Nohl (2001) für das Beispiel der dokumentarischen Methode geschieht, dann treten im Verlauf des Forschungsprozesses ebenfalls Reduktionen auf, die dem methodischen Vorgehen geschuldet sind. Bereits Kockelmanns (1975, S. 45) hat das für eine Bedingung gehalten, die mit Abstraktion, Idealisierung und Formalisierung notwendigerweise verbunden ist.

Eine andere Variante findet sich bei Brandom (1998), der darauf verweist, dass alle Äquivalenzannahmen beim Gebrauch von Begriffen normative Annahmen eben über die Äquivalenz des Begriffs zu der Sache implizieren, d.h. es handelt sich bei Begriffen nicht um eine genaue Abbildung, sondern um eine Zuordnung, die nach Regeln erfolgt,

8 Zu diesem Genre gehören viele Erfahrungs- und Projektberichte aus diesem Arbeitsfeld, in denen in der Regel keine kritische Überprüfung der eigenen Erfahrungsbasis angegeben wird, sondern eher empathische Haltungen deutlich werden (vgl. z.B. Bauer 2000). Eine Variante wird bei Schad (2000) ersichtlich, die die ihren Aussagen zu Grunde liegende Untersuchung methodenkritisch nicht prüft.

9 Merkens (1986) hat am Beispiel der kritischen Reanalyse eines eigenen Projektberichtes verdeutlicht, wie leicht sich Vorurteile in die konkrete Falldarstellung einschleichen und den Duktus der Beschreibung bestimmen, ohne dass sie hinreichend geprüft worden sind.

wenn Nachprüfbarkeit wichtig ist. „Die Bildung eines Begriffes besteht in der Aufstellung eines Gesetzes über die Verwendung eines Zeichens (z.B. eines Wortes) bei der Darstellung von Sachverhalten“ (Carnap 1966b, S. 4). Wie Jäger (2001) nachgewiesen hat, resultiert daraus das Problem, dass es Normen mit universalem Anspruch geben müsste, wenn man das Prinzip der Allgemeingültigkeit von Aussagensystemen nicht zugunsten eines relativierenden Kulturalismus aufgeben will.<sup>10</sup> Dieser Frage soll nicht weiter nachgegangen werden. Sie würde zur Klärung einer eigenen Analyse bedürfen. Jedoch ist im hier diskutierten Zusammenhang wichtig, dass innerhalb der vorgestellten Auffassungen von empirischer Wissenschaft ein Klärungsbedarf besteht, der über das hinausreicht, was im Folgenden unter dem Thema Methodisierung der Kritik behandelt werden wird. Festzuhalten bleibt an dieser Stelle erstens, dass in der erfahrungsbasierten Berichterstattung, die in der Erziehungswissenschaft häufig anzutreffen ist, ein vorkritischer Typ vorherrscht, der die Adäquatheit der berichteten Erfahrungen in Bezug auf die Ereignisse oder Tatsachen, die Gegenstand des Berichts sind, nicht hinterfragt. Dieser Typ von Erfahrung weist den Vorteil der Unmittelbarkeit auf, der neben dem Preis der Unkontrollierbarkeit durch andere auch dadurch erkaufte wird, dass die Erfahrung in ihrer Gänze nicht wiederholbar ist. Scheint sie wiederholt zu werden, so wird das über ungeprüfte Interpretationsanteile im Akt des Erfahrens bewirkt. Zweitens ist festzuhalten, dass das Verhältnis von Theorie und Erfahrung, wie es im kritischen Rationalismus vorausgesetzt wird, in der Erziehungswissenschaft in aller Regel nicht einzuhalten ist, weil es an entsprechenden Theorien mangelt, die erst fundiert das notwendige Maß an Konsens sicherstellen, damit auf der Basis eines theoretisch begründeten Reduktionismus Wiederholbarkeit der Erfahrung möglich wird. Regeln, die hierzu entwickelt worden sind, werden im nächsten Abschnitt vorgestellt.

### **3. Aussagen korrespondieren Ereignissen nicht automatisch**

Für die Zuordnung von Aussagen zu Ereignissen bedarf es bestimmter Verfahren (Carnap 1966b). Bei komplexeren Ereignissen werden hierzu in den Naturwissenschaften und in den quantitativen Sozialwissenschaften Instrumente entwickelt. Eine klassische Variante des Vorgehens ist der Operationalismus (Bridgman 1970). Vor allem in den Naturwissenschaften ist dem Problem des Zuordnens von Ereignissen zu Aussagen große Aufmerksamkeit geschenkt worden (Carnap 1961). In Sozialwissenschaften sowie Erziehungswissenschaft ist in der Praxis der quantitativen Forschung weitgehend dem Vorbild des Operationalismus gefolgt worden. Die bei PISA verwendeten Instrumente zur Feststellung der Schulleistungen sind z.B. insbesondere in dieser Tradition entwickelt worden: Die Schulleistungen in Mathematik und Physik wurden über die entspre-

<sup>10</sup> Daraus wird ersichtlich, dass sich das Problem bei quantitativen Methoden und in den Naturwissenschaften noch radikaler stellt als in dem hier diskutierten Zusammenhang, bei dem sich eher eine moderate Variante ergibt, weil man letzten Endes auch eine kulturelle Haltung mit dem methodischen Anspruch vereinbaren kann.

chenden Items operationalisiert. Kriterien bzw. Normen wurden über diese Tests definiert bzw. entwickelt (Deutsches PISA-Konsortium 2001). Das Beispiel demonstriert, dass der Operationalismus in der empirisch orientierten erziehungswissenschaftlichen Forschung praktiziert wird.

In der naturwissenschaftlich orientierten Wissenschaftstheorie muss in diesem Zusammenhang ein weiterer Ansatz erwähnt werden: Die Diskussion über das Gewinnen und die angemessene Formulierung von Protokoll- (Carnap (1933), Neurath (1933)) und Basissätzen (Popper 1966). Durch die Formulierung bzw. Zuordnung der Sätze sollte die mögliche Kritik an einer fehlerhaften auf Erfahrung aufbauenden Basis sowohl vorweggenommen als auch ihr begegnet werden. Dabei kann als eine Grundüberlegung angesehen werden, dass die Sätze, welche die Verbindung zur Erfahrung herstellen, formal so formuliert werden, dass sie einer kritischen Prüfung standhalten können: Sie dürfen beispielsweise keine widersprüchlichen Prognosen bzw. Erwartungen auslösen oder unterstützen, sondern sollen im Fall des kritischen Rationalismus, um eines der Beispiele fortzusetzen, genau beschreibbare Ereignisse prognostizieren (Popper 1966). Der Interpretationsspielraum, welcher bei Übersetzungen immer gegeben ist, sollte auf die Messproblematik der Physik reduziert werden. Die Kritik der Methoden konnte auf diese Weise in eine Messkritik transformiert werden. Das ist deshalb ein interessanter Vorschlag, weil versucht worden ist, ein inhaltliches Zuordnungsproblem in ein weitgehend statistisches Messproblem zu transponieren. Im kritischen Rationalismus wurde vorausgesetzt, dass Aussagen, in denen das Auftreten von Ereignissen vorhergesagt oder erklärt wurde, nur auf der Basis von Theorien hergeleitet werden sollten. Dabei sollten einschränkende Bedingungen hinsichtlich Raum und Zeit bei der Theorie entfallen. In der Erziehungswissenschaft ist es allenfalls möglich, auf bestimmte Räume (Institutionen) und Zeitpunkte (Alter, Klassenstufe z.B.) eingegrenzte Aussagen theoriegeleitet zu formulieren (Krapp 1979). Das impliziert im Prinzip eine kulturrelativistische Position. Auf dieses Problem wird im Zusammenhang mit der Validität nochmals eingegangen.

Die empirischen Sozialwissenschaften haben, am Vorbild des in den Naturwissenschaften praktizierten Operationalismus orientiert, einen abgekürzten Weg gewählt, indem die jeweiligen Probanden aufgefordert werden, selbst eine Zahl oder ein Zahlenäquivalent anzukreuzen, um die Intensität von Zustimmung oder Ablehnung der Aussagen in einzelnen Items zu vermitteln. In dieser Tradition werden den Untersuchten die Übersetzungsleistungen abgefordert, wenn sie ihre Einstellung auf einer Likert-Skala markieren. Sie sollen die Intensität ihrer Zustimmung bzw. Verneinung zu Aussagen in Zahlen- oder auf Vergleichswerte (größere/kleinere) transformieren. Die auf diese Weise gewonnenen Zahlenwerte werden dann in die weitere statistische Auswertung übernommen. Daraus resultieren Probleme, die später noch genauer dargestellt werden. Die Annahme, dass die Zuordnung einer Zahl zu einer Aussage nur deshalb zutreffend ist, weil sie vom Untersuchten selbst vorgenommen wird, erweist sich nämlich bei genauem Hinsehen als nicht hinreichend, um sicher gehen zu können, dass zutreffend angekreuzt worden ist. Vielmehr erfordert das umfangreiche weitere Überprüfungen. Das gilt für die quantitativen empirischen Sozialwissenschaften ganz allgemein.

Bei den qualitativen Methoden scheint zumindest dieser Teil des diskutierten Problems aus der Perspektive der Forschenden einfacher zu liegen, weil die Forschenden selbst es in der Regel sind, welche die entsprechenden Zuordnungen vornehmen, z.B. in der Ethnographie. Die Annahme der Authentizität wird mit dem eigenen Erleben begründet. Gerade in pädagogischen Kontexten bietet sich ein solches Vorgehen an, weil die Tätigkeit in der Praxis einen solchen Zugang erfordert; allerdings hat bereits Flitner (1963) mit seinem Verweis auf die Notwendigkeit der Reflexion in der Situation die Grenzen eines solchen Vorgehens sichtbar gemacht, das eher dem Besonderen als dem Allgemeinen zugeordnet ist (Benner 1978). Geht es aber um das Allgemeine, dann sind andere Vorgehensweisen erforderlich. Dabei erweisen sich schnell die Grenzen der Annahme, das Authentische reiche hin, wenn bei Methoden wie der Objektiven Hermeneutik der Diskurs der Interpreten theoretisch für nicht beendbar erklärt wird. Rekonstruktive Verfahren erfordern ebenfalls eine anschließende Bearbeitung, auch wenn sie aus der Forschungspraxis emergiert sind (Bohnsack 2000). Vielmehr wird dann das Besondere in den Kontext des Allgemeinen gestellt und daraus zu verstehen versucht. Sobald die Korrespondenz von Aussagen zu Ereignissen bzw. das Wissen um Zusammenhänge von Interpretationen mit Ereignissen sichergestellt werden sollen, bedarf es demnach kritischer Überprüfungen. Diese müssen, damit sie von anderen Beteiligten am wissenschaftlichen Diskurs akzeptiert werden, in ihrem Ablauf überprüfbar sein. Das erfordert ein methodisches Instrumentarium. Darin liegt begründet, dass die notwendige Kritik im und am Prozess des Gewinnens von Erfahrungen ohne Methodisierung nicht in dem erforderlichen Umfang erfolgen kann.

#### **4. Validitätsproblem**

Die Angemessenheit der Zuordnung von Ereignissen zu Aussagen geschieht nicht selbstverständlich, wie die bisherigen Überlegungen belegen. Wenn eine Handlung im Kontext eines Erziehungsprozesses als Lob bewertet wird, bedarf es beispielsweise einer entsprechenden Interpretation der Beteiligten oder eines Fremdbeobachters. Dabei können die Urteile differieren. Deshalb muss die Validität der Aussagen geprüft werden. Das geschieht, indem sie der Kritik unterworfen werden. Allgemeiner ausgedrückt, muss kontrolliert werden, ob das gemessen wird, was gemessen werden soll (Bortz/Döring 1995, S. 185), bzw. ob ein Indikator tatsächlich den Sachverhalt anzeigt, „der mit dem definierten Begriff bezeichnet worden ist (Kromrey 1998, S. 170). Binet hatte die Absicht, Intelligenz unabhängig von Schulleistung zu messen und musste doch als Außenkriterium für die Validität der Messung die Schulleistung wählen. Ebenso soll mit dem Schulleistungstest die Schulleistung „objektiver“ festgestellt werden als mit der traditionellen Methode des Zensierens. Wiederum wird die Schulleistung als Außenkriterium gewählt, um die Validität zu überprüfen. Obwohl diese Vorgehensweise widersprüchlich erscheinen mag, weist sie doch eine gewisse Rationalität auf: Im Durchschnitt wird nicht angezweifelt, dass die Schulleistung auf traditionellem Wege angemessen festgestellt wird. Es wird nur zusätzlich geprüft, ob sich dabei im Einzelfall sys-

tematisch Fehler ergeben, die der Korrektur bedürfen. Deshalb kann folgerichtig die Güte eines Tests, seine Validität, durch den Vergleich mit Schulleistungen festgestellt werden.

In der Forschungspraxis und der dazu gehörenden Literatur über Forschungsmethoden werden drei verschiedene Herangehensweisen benannt, wenn es darum geht, die Validität von Instrumenten bzw. Indikatoren festzustellen: Es wird zwischen Inhalts-, Kriteriums- und Konstruktvalidität unterschieden (Bortz/Döring 1995, S. 185ff.; Diekmann 1995, S. 224). Bei der Inhaltsvalidität eines Tests wird beispielsweise überprüft, ob die Auswahl der Items den jeweiligen zu messenden Inhalt, Intelligenz z.B., angemessen repräsentiert. Diese Formulierung lässt erkennen, dass es in diesem Fall sehr schwer fallen dürfte, Inhaltsvalidität tatsächlich zu kontrollieren, weil zunächst die Menge der Items bestimmt werden müsste, die den Inhalt – Intelligenz z.B. – wiedergibt und dann aus diesen Items eine Stichprobe gezogen werden müsste, von der, wenn es sich um eine Zufallsstichprobe handelt, angenommen werden könnte, dass die in ihr enthaltenen Items tatsächlich den Inhalt – Intelligenz – repräsentieren. Wie Diekmann (1995, S. 225) zutreffend beschreibt, lässt sich dieser Forderung in aller Regel nicht genügen, weil schon die Menge der Items nicht produziert werden kann. Bortz/Döring (1995, S. 185) ergänzen, dass es sich bei der Inhaltsvalidität nicht um ein Testgütekriterium, sondern um eine Zielvorgabe handle. Dennoch kommt der Inhaltsvalidität im Kontext der Unterrichts- und Schulforschung eine große Bedeutung zu. So kann für Schultests geprüft werden, ob in ihnen die in Lehrplänen oder Rahmenrichtlinien vorgegebenen Lernziele abgebildet sind. Bei der PISA-Studie ist der Kontrolle der inhaltlichen Validität große Aufmerksamkeit geschenkt worden (vgl. z.B. Artelt u.a. 2001, S. 78ff.). Es wird festgelegt, was in diesem Fall unter Lesekompetenz verstanden wird, und anschließend werden entsprechende Items formuliert. Dabei zeigt sich, ohne dass das näher diskutiert wird, ein Dilemma international vergleichender Studien. Wer definiert die für relevant gehaltenen bzw. erklärten Inhalte? Eine substantielle Veränderung bei den Inhalten würde wahrscheinlich substantielle Veränderungen bei den Ergebnissen nach sich ziehen. Das führt sogar dazu, dass eine nationale Variante angeschlossen wird. In einer konkreten Untersuchung bleibt im Allgemeinen der Ausweg, die Interpretationsgemeinschaft der Forscher als Gremium mit der Auswahl der Inhalte zu beauftragen. Das bedeutet eine erhebliche Einschränkung gegenüber dem eigentlichen Anspruch. Gerade in der empirischen Unterrichtsforschung kommt der inhaltlichen Validität eine große Bedeutung zu, wie die Hinweise auf PISA belegen.

Auch bei der Kriteriumsvalidität ergeben sich Schwierigkeiten, wenn sie überprüft werden soll. Sie wird praktisch über einen Korrelationskoeffizienten mit einem Außenkriterium ermittelt. Dabei kann man in der Regel verschiedene Kriterien heranziehen, was zu unterschiedlichen Werten und damit auch zu einem nicht genau bestimmbar Resultat führt (Diekmann 1995, S. 224). Neben der Schulleistung kann man bei Schulleistungstests ebenso den Berufserfolg als Kriterium einbeziehen, um nur ein Beispiel zu nennen. Erst bei der Konstruktvalidität befindet man sich auf den ersten Blick auf gesicherterem Boden. Sie kann kontrolliert werden, indem mit dem gemessenen Ergebnis, Differenzen überprüft werden, die man als gesichert annehmen kann, wenn z.B. der

gemessene Grad der Einsamkeit dem Selbstwertgefühl insoweit korrespondiert, dass Menschen, die als einsamer klassifiziert werden, auch einen niedrigeren Selbstwert aufweisen (Bortz/Döring 1995, S. 186). Mit der Multitrait-Multimethod-Methode (Campbell/Fiske 1959) ist hier eine Vorgehensweise gefunden worden, die mehrere Konstrukte mit mehreren Methoden kombiniert und auf diese Weise die Bestimmung der Konstruktvalidität verbessern soll (Bortz/Döring 1995, S. 187). Hier ergeben sich bestimmte Ähnlichkeiten zur später noch vorzustellenden Triangulation. Diese Vorgehensweise ist in erziehungswissenschaftlichen Untersuchungen bisher wenig praktiziert worden.

Allen vorgestellten Überlegungen und Verfahren zur Validität ist die Grundüberzeugung gemeinsam, dass die nach bestem Wissen und Gewissen erfolgte Zuordnung von Aussagen zu Ereignissen allein als nicht ausreichend betrachtet wird, um hinreichende Sicherheit über die Wahrheit der auf diese Weise gewonnenen Aussagen zu erlangen. Indem Kritik via Kontrolle und Überprüfung methodisiert wird, soll der Wahrheitswert der Aussage nachgewiesen werden. Allerdings gibt es, wie die Beispiele belegen, wenige gesicherte Verfahren der Validitätskontrolle. Das hängt mit dem Reduktionismus beim Gewinn der Aussage zusammen, weil auch Validitätskriterien reduktionistisch gewonnen werden müssen. Sie unterliegen ähnlichen Restriktionen wie die Aussagen selbst. Wenn aber sowohl das gemessene Ergebnis als auch ein Konstrukt, mit dessen Hilfe seine Validität überprüft werden soll, nach den gleichen Vorgehensweisen gewonnen worden sind, z.B. indem operationalisiert worden ist, dann unterliegen beide den gleichen Restriktionen in Bezug auf das Zustandekommen des Ergebnisses. Daraus resultiert mit Notwendigkeit ein großer Erklärungsbedarf, obwohl scheinbar alle erforderlichen Maßnahmen zur Absicherung getroffen worden sind.

Dieses Manko resultiert aus einer weiteren Besonderheit, die bisher nicht hinreichend dargestellt worden ist. Die Beispiele Intelligenz- und Schulleistungstest lassen sie als eine der auf Erfahrung basierenden Aussagen erkennen, die ebenso in der Physik gilt: Es handelt sich um Bedingungsansagen. Wie Carnap (1966b, S. 8) erläutert hat, behaupten solche Aussagen mehr, als überhaupt beobachtet worden ist. So wird von der Anzahl gelöster Aufgaben eines bestimmten Typs auf die Schulleistung in einem bestimmten Fach zurückgeschlossen und die Beantwortung einzelner Fragen soll Aufschluss über die Intelligenz des Antwortenden geben. Annahmen diesen Typs machen die Kontrolle der Validität zwingend erforderlich und zeigen ebenso die Grenzen des kritischen Habitus auf.

Jenseits dieser Probleme ist versucht worden, das Validitätsproblem zu lösen, indem zwischen externer und interner Validität unterschieden worden ist (Campbell/Stanley 1963). Unabhängig von der aus wissenschaftstheoretischer Sicht hieran geübten Kritik (Gadonne 1985) ist von Interesse, dass es wiederum methodische Überlegungen gewesen sind, die eine kritische Kontrolle, vor allem der internen Validität, ermöglichen sollten. Es sollten Kriterien formuliert werden, die es gestatteten, die Menge der Aussagen dahingehend zu überprüfen, ob sie trennscharf in Bezug auf diese Kriterien ausfielen, beispielsweise Messungen des Zustandes der Ängstlichkeit in einer standardisierten Situation in Bezug auf das Geschlecht der Untersuchten. Das ist ein Modus der Absicherung von Aussagen, der das Vorwissen – Menschen weiblichen Geschlechts sind im Durchschnitt ängst-



licher als Menschen männlichen Geschlechts – als Kriterium für die Gültigkeit der Messung einführt. Gerade die interne Validität fügt keine neuen Erkenntnisse in Bezug auf das Gemessene hinzu, sondern trägt allenfalls zur Stabilisierung von Vorwissen und bisher Bekanntem bei. Ähnlich verhält es sich mit der externen Validität, die letzten Endes als Prüfstein dient, wieweit das Gemessene im Vergleich mit schon bekannten Außenkriterien eine Differenz bestätigt, die bereits vorher bekannt gewesen ist.

Bei allen Gegenargumenten und Mängelrügen im Detail bleibt festzuhalten, dass mit der Kontrolle der Validität die Methodisierung der Kritik einen Standard erreicht hat, der empirische Untersuchungen erst möglich macht: Das, was mit erfahrungswissenschaftlichen Methoden gewonnen wird, wird kritisch auf seinen empirischen Gehalt überprüft, indem gefragt wird, ob das gemessen worden ist, was gemessen werden sollte. Auf diese Weise wird es möglich, dass die Erfahrung eine verlässliche Basis für Aussagen bleiben kann.

Im Rahmen empirischer Untersuchungen in der Erziehungswissenschaft ist die interne Validität häufig überprüft worden, ohne dass das unter diesem Thema geschehen ist, wenn in Studien zum Rechtsextremismus der Jugendlichen beispielsweise Differenzen beim Geschlecht oder der Schulform als Beleg für die Angemessenheit der Stichproben herangezogen worden sind.<sup>11</sup> Demgegenüber sind Kontrollen der externen Validität sehr selten. Wie bereits dargestellt worden ist, gibt es allein in der Unterrichtsforschung vermehrt die Tendenz, die inhaltliche Validität zu kontrollieren. Bezüglich der Validitätskontrolle ist demnach festzuhalten, dass es bisher nicht zum normalen Repertoire empirischer Untersuchungen in der Erziehungswissenschaft zählt, die Validität zu kontrollieren. Auch in diesem Fall zeichnet sich eine vorkritische Haltung ab.

Der kritische Impetus bezieht sich auf die Gewinnung von Aussagen. Er lässt sich fortsetzen in Richtung der Prüfung von Theorien. Bevor hierzu Erläuterungen folgen, wird aber die kritische Prüfung von gewonnenen Aussagen über Erfahrungen noch einen Schritt weiter vorangetrieben.

## 5. Die Methodisierung des kritischen Habitus<sup>12</sup>

Wie die bisherige Darstellung bereits verdeutlicht hat, ist eine der wesentlichen Hilfsannahmen, dass zwischen Aussagen, die sich auf das gleiche Ereignis oder eine Ereignisfolge beziehen, Konsistenz vorausgesetzt wird. Daraus folgt, dass es ein Monotonieprinzip gibt: Reaktionen auf verschiedene vorgegebene Reize sollen auch zu verschiedenen Zeitpunkten in sehr ähnlicher Weise, ideal sogar gleich, erfolgen. Das ist eine weitere reduktionistische Annahme, weil die mögliche Vielfalt oder Abweichung als Zufallsvarianz betrachtet wird. Von Interesse ist nur der als gemeinsam erachtete Kern, das Systematische: Diese Annahme ist angesichts der vielfältigen Möglichkeiten in pädagogi-

11 Das ist eine so häufig geübte Praxis, dass es nicht angemessen ist, einzelne Untersuchungen anzuführen.

12 Vgl. hierzu Heid (2003) in diesem Band.

schen Kontexten zu reagieren, so selbstverständlich nicht. Getreu dem naturwissenschaftlichen Vorbild wird aber auch in den Sozialwissenschaften angenommen, dass sich Reaktionen auf der Basis von Regelmäßigkeiten heraus bilden. Das hat Folgen für die Konzeption eines weiteren Gütekriteriums empirischer Sozialforschung, der Reliabilität.

### 5.1 *Reliabilität als Kriterium*

Bei der Erhebung empirischer Daten besteht in der Praxis der Sozialforschung eine Methode darin, eine Reihe von einander sehr ähnlichen Ereignissen zu präsentieren, wobei erwartet wird, dass dies konsistente Reaktionen bei denen hervorruft, denen diese Ereignisse nacheinander vorgelegt werden. Dabei ist es das Ziel, zufällige falsche Reaktionen, welche es geben kann, in ihrer Wirkung zu minimieren. Dieses Ziel lässt sich erreichen, wenn eine möglichst große Anzahl von sehr ähnlichen Ereignissen präsentiert werden kann. Die Überprüfung, ob es gelungen ist, Ereignisse vorzugeben, die die erwarteten ähnlichen Reaktionen hervorrufen, geschieht anschließend mit der Kontrolle der Reliabilität. „Die Reliabilität eines Messinstrumentes ist ein Maß für die Reproduzierbarkeit von Messergebnissen“ (Diekmann 1995, S. 217). Es kann nicht überraschen, dass angesichts der einfachen Überprüfbarkeit mit statistischen Methoden der Forschungspraxis der empirischen Sozialwissenschaften der Reliabilität große Aufmerksamkeit gewidmet wird. Das wird noch deutlicher, wenn man sich die Praxis der Reliabilitätsmessung in die Betrachtung einbezieht. Reliabilitätskoeffizienten sind entweder ein Maß für die Konsistenz der Reaktionen über die Zeit (Stabilität der Reaktionen mit einem Zeitintervall dazwischen) oder zu einem bestimmten Zeitpunkt auf ähnliche Vorgaben (split-half-Methode oder Cronbachs Alpha). Insbesondere die zweite Variante der Reliabilitätskontrolle dominiert die Praxis der empirischen Sozialforschung. Sie ist aus vielen Gründen auch das für erziehungswissenschaftliche Untersuchungen vorzuziehende Verfahren. So ist im pädagogischen Feld die Stabilität des Ergebnisses bei der Wiederholung von Messung, im Unterricht z.B., nicht erwünscht, vielmehr wird über die Zeit ein Lernfortschritt erwartet. Dieser kann auch dadurch entstehen, dass sich bei der zweiten Bearbeitung der Items Lösungen, die beim ersten Mal noch nicht gefunden wurden, aus dem Kontext ergeben. Vor allem ist in pädagogischen Kontexten nicht erwünscht, dass Lernfortschritte bei allen Jugendlichen gleich groß sind. Dann würden Lernausgangsleistungen die Leistungen am Ende einer Schulstufe bedingen, wie das für Hamburg der Fall zu sein scheint (Lehmann u.a. 2002). Das würde die gezielte Förderung von Jugendlichen mit Defiziten ausschließen. Ähnlich ist bei PISA die große Abhängigkeit des Schulerfolgs vom Sozialstatus der Eltern (in Deutschland) beispielsweise einer der Hauptkritikpunkte (Baumert/Schümer 2001, S. 390). Ergebnisse dieser Art besagen, dass Zugewinne beim Lernen durch Voraussetzungen bedingt sind, die selbst nicht mehr im Einflussbereich der Schule liegen. Das kann aber dem Selbstverständnis der Schule nicht entsprechen und muss durch eine Änderung des pädagogischen Konzeptes der Schule in seiner Wirkung korrigiert werden. Aus diesem Grund kommen als Maße der Reliabilität nur die interne Konsistenz und die split-half-Methode in Betracht.

Demnach lässt sich die Reliabilität entweder über die interne Konsistenz, das ist z.B. im Prinzip die Kontrolle in wie weit alle Items in der gleichen Weise zum Gesamtwert eines Index beitragen, oder über die Wiederholung der Messung und die Korrelation beider Messungen feststellen. In beiden Fällen hängt der Erfolg im Prinzip davon ab, das zu messende Ereignis durch eine Mehrzahl von Items zu erfassen, die untereinander ein hohes Maß an Ähnlichkeit aufweisen. Eine hohe Reliabilität bestätigt im Prinzip, dass dieses Vorhaben gelungen ist. Es ist also keineswegs gesichert, dass etwas Relevantes gemessen wird; es wird nur überprüft, ob das Messinstrument zuverlässig funktioniert. Dem Ergebnis der Messung wird demnach einerseits mit Misstrauen gegenübergetreten; wenn dieses Misstrauen sich als unbegründet erweist, weil der Koeffizient für die Reliabilität zufrieden stellend ausfällt, verwandelt sich dieses Misstrauen allerdings in die Annahme einer zutreffenden Messung: Die methodische Kritik wird instrumentell genutzt. Das ist ein Merkmal bei quantitativen Methoden. Wenn sich die Kontrolle von Messergebnissen auf die Überprüfung der Reliabilität beschränkt und anschließend unter Verweis auf die Reliabilitätskoeffizienten das Zutreffen der Ergebnisse behauptet wird, dann zeichnet sich ab, wie durch die Methodisierung der Kritik Immunisierung betrieben wird, denn bewiesen ist nur, dass etwas zuverlässig gemessen worden ist. Das kann nur als Voraussetzung dafür betrachtet werden, dass die Validität hinreichend gesichert ist.

Eine Bilanz empirischer erziehungswissenschaftlicher Untersuchungen lässt allerdings erkennen, dass nicht einmal diesem Kriterium der Reliabilität eine hinreichende Aufmerksamkeit geschenkt wird. Häufig werden Indizes mit nur zwei oder drei Items verwendet, die dann allein schon aus statistischen Gründen sehr niedrige Reliabilitäten erzielen, d.h. der kritische Habitus wird nicht in dem für erforderlich gehaltenen Maß eingehalten. Das führt keineswegs zu qualitativ besseren Daten, sondern diese Datenqualität ist in vielen erziehungswissenschaftlichen empirischen Untersuchungen unzureichend. Damit wird eine Mindestanforderung an quantitative Methoden häufig nicht erfüllt. Wiederum zeigt sich, dass in vielen Fällen ein vorkritischer Stand der erziehungswissenschaftlichen Forschung zu konstatieren ist. Das gilt nicht für die PISA-Studie, die vielmehr als beispielhaft für den kritischen Status angesehen werden kann.

## 5.2 *Wiederholbarkeit als Kriterium*

Bei qualitativen Methoden werden Verfahren der Triangulation im Sinne der Reliabilitätskontrolle eingesetzt; sie wird aber auch als Validierungsstrategie bezeichnet (Flick 2000, S. 309ff.). Das kann zunächst als ein Hinweis dafür genommen werden, dass vom Anspruch her – im Unterschied zu den quantitativen Methoden – der Validität größere Beachtung geschenkt wird als der Reliabilität. Methoden der Triangulation stellen den Versuch dar, auf andere Weise wieder die Haltbarkeit von Aussagen durch einen geeigneten Umgang mit den Datenquellen zu sichern. Beispielsweise wird die Wiederholbarkeit der Datengewinnung zu kontrollieren versucht, indem mehrere Personen im Feld die gleichen Ereignisse beobachten bzw. beschreiben sollen. Eine dabei erreichte Ähn-

lichkeit der Ergebnisse wird im Sinne der Wiederholbarkeit interpretiert. Gerade dieser Typ der Triangulation stellt eine Vorgehensweise dar, die im quantitativen Paradigma mit Reliabilitätskontrolle bezeichnet wird. Wenn es jedoch bei der Triangulation darum geht, verschiedene Methoden auf den gleichen Gegenstand anzuwenden und die Ergebnisse prüfend gegeneinander auszuspielen, um die Validität von Feldforschungen zu maximieren, handelt es sich um eine Überprüfung der Validität (Denzin 1978, S. 304). Hier ist von Interesse, dass in beiden Fällen auch bei qualitativen Methoden Kritik in der Form des Misstrauens gegenüber einem Resultat methodisiert wird. Erweist sich das Resultat als wiederholbar, wird das als Beleg für dessen Qualität genommen. Vom Prinzip her lassen sich also Parallelen zwischen den Ansprüchen qualitativer und quantitativer Methoden bei der Methodisierung der Kritik betrachten. Allerdings sind die methodischen Feinheiten der Kritik bei den quantitativen Methoden weiter voran getrieben. Wenn allerdings Kritik methodisiert wird, dann liegt der Akzent bei den qualitativen Methoden in den theoretischen Diskursen eher bei der Kontrolle der Validität, in der Praxis wird aber ebenfalls häufig in Richtung Reliabilität Kritik methodisiert, wenn beispielsweise mehrere Beobachter den gleichen Gegenstand untersuchen.

Beide Vorgehensweisen werden in qualitativen erziehungswissenschaftlichen Untersuchungen selten befolgt. Hier wird sich allzu häufig auf das Prinzip der Authentizität verlassen. Insgesamt spielt der kritische Habitus bei der Anwendung qualitativer Methoden bei der Datengewinnung eine noch geringere Rolle als bei den quantitativen Methoden. Bei den qualitativen Methoden wird dieser kritische Habitus vor allem bei der Integration der Ergebnisse praktiziert (vgl. z.B. Bohnsack 2000). Das hängt auch damit zusammen, dass dem Datum als Datum weniger Aufmerksamkeit geschenkt, sondern angestrebt wird, das Datum aus dem Kontext heraus zu verstehen. Am Beispiel der pädagogischen Feldforschung hat Zinnecker (2000) zusätzlich argumentiert, dass dem kritischen szientistischen Habitus ein an humanistische Lebenswelt orientierter gegenübersteht, der in der Erziehungswissenschaft über Anhänger verfügt. Einen guten Überblick bietet das Handbuch von Friebertshäuser/Prengel (1997). Dabei zeigt sich ähnlich wie bei den quantitativen Methoden eine große Offenheit in Richtung andere sozialwissenschaftliche Disziplinen, wenn es um die Übernahme von Methoden geht. Gleichzeitig wird deutlich, dass sich zahlreiche Anwendungsfelder herausgebildet haben. Das Stichwortverzeichnis weist dann allerdings aus, dass Begriffen wie Kritik und Reliabilität fast gar keine und dem der Validität oder Validierung nur geringe Aufmerksamkeit geschenkt wird. Das kann als Indikator dafür genommen werden, dass der vorkritische Status auch in diesem Bereich noch vorherrscht.

### 5.3 *Nachvollziehbarkeit als Kriterium*

Das eigentliche Kriterium, welches bei allen betrachteten Verfahren leitend ist, kann man mit Nachvollziehbarkeit kennzeichnen. Die methodische Form der Kritik soll letzten Endes Nachvollziehbarkeit sichern. Damit ergibt sich eine Differenz zu der wissenschaftstheoretischen Position des kritischen Rationalismus. An die Stelle der methodi-

sierten kritischen Überprüfung von Aussagen tritt die methodisierte kritische Position bei der Datengewinnung bzw. Interpretation.

Mithilfe der kritischen Überprüfung der Datengewinnung bzw. der Nachvollziehbarkeit der Interpretation soll die Verlässlichkeit der Aussagen gesichert werden. Damit nimmt der Ansatz der empirischen Vorgehensweise eine interessante Wende: Durch Kritik als Habitus kann die Frage der Zusammenhänge von Ereignis – Aussage – Interpretation methodisch zwar bearbeitet werden, der ursprünglich inhaltliche und konkrete Zusammenhang wird dabei aber in einen formalen transformiert. Das unterscheidet den kritischen vom vorkritischen Zustand. Mit dem Betonen der Überprüfung der Reliabilität zeichnet sich eine Hinwendung zur Immunisierung durch Methodenkritik ab.

## 6. Die Überprüfung von Theorien

Theorien werden in den empirisch orientierten Sozialwissenschaften als Aussagensysteme betrachtet, die durch Erfahrung bestätigt oder widerlegt werden können.<sup>13</sup> Für beide Varianten sind umfangreiche metatheoretische Überlegungen vorgelegt worden, deren gemeinsamen Kern man dahin zusammenfassen kann, dass ein methodisches Vorgehen als Leitlinie gewählt wird (Wolters 2000, S. 211). Ist eine Theorie durch Erfahrung bestätigt, dann sollen beispielsweise im kritischen Rationalismus anschließend Prüfungen an der Erfahrung vorgenommen werden, die für ein Nichtzutreffen bzw. Scheitern wahrscheinlicher sind, d.h. es wird ein kritischer Habitus eingeführt, für den sich auch noch Regeln des Vorgehens angeben lassen, die dahin zusammengefasst werden können, dass die Klasse der Falsifikationsmöglichkeiten erweitert und nicht verkleinert werden soll, wenn eine Theorie geprüft wird.

Eine Kernfrage ist gewesen, in welcher Weise man Protokoll- bzw. Basissätze mit allgemeinen theoretischen Aussagen in eindeutiger, nachprüfbarer Weise in Beziehung setzen kann. Die Auseinandersetzungen über die angemessene Vorgehensweise – Induktion versus Deduktion – müssen hier nicht weiter betrachtet werden.<sup>14</sup> Wichtig ist ausschließlich, dass es sich um Vorgehensweisen handeln soll, bei denen mithilfe methodischer Regelungen – die formale Logik dient dazu als Instrument – die Verknüpfung von Aussagen unterschiedlicher Reichweite und Allgemeinheit – Gesetzesaussagen mit Aussagen über einzelne Ereignisse – gewährleistet werden soll. Diese Verknüpfungen werden mithilfe von allgemein anerkannten Regeln durchgeführt, die selbst wiederum nicht Gegenstand empirischer Überprüfungen sind, unabhängig von dem betreffenden Fall entwickelt worden sind und deren angemessener Gebrauch ohne Mühe kontrolliert werden kann.

Damit wiederholt sich im Prinzip ein Muster, welches nun bereits mehrfach beschrieben worden ist: Indem man sich an bestimmte anerkannte methodische Regeln

13 Die positiven Folgen dieser Vorgehensweise in Bezug auf eine bessere Praxis des pädagogischen Handelns hat Winch (in diesem Band) dargelegt.

14 Eine interessante neuere Arbeit ist die Würdigung von Wolters (2000) für Hempel anlässlich dessen Todes.

hält, entwickelt man eine Vorgehensweise, die gegen Kritik resistent ist. Die schon bekannte Immunisierungsstrategie wird auch in diesem Fall angewendet: Die Verknüpfungen erfolgen kritisch geprüft und kontrolliert nach den Regeln der Kunst der Logik. Kritik wird formalisiert, und der kritische Habitus ist in das Verfahren inkludiert.<sup>15</sup> Eine erneute Überprüfung der Aussagen findet nicht statt. Wenn sie aber praktiziert wird, dann geschieht diese mithilfe mechanistischer Kriterien, wie bereits an den Beispielen Protokoll- und Basissätze erläutert worden ist. Der Aspekt der Methodisierung beginnt den der Kritik zu dominieren. Das ist eine Immunisierungsstrategie (vgl. Heid in diesem Band). Gegenüber diesem anspruchsvollen Programm, welches Krumm (1983) eindrucksvoll beschrieben hat, weist die empirische Erziehungswissenschaft deutliche Rückstände auf. Zwar ist vermutet worden, der kritische Rationalismus sei durch Brezinka (1971) in die Erziehungswissenschaft eingeführt worden, aber abgesehen davon, dass Brezinka (1997) diese Zuschreibung ablehnt und der wissenschaftstheoretische Ansatz häufig erwähnt worden ist, hat sich die empirische Erziehungswissenschaft selbst nur selten diesen Bemühungen angeschlossen. Sie ist im Wesentlichen auch in diesem Bereich vorkritisch geblieben.

Dem steht als Position der kritische Habitus des Wissenschaftlers gegenüber, der sich bemüht, Aussagensysteme – Theorien bzw. Hypothesen – zu falsifizieren, indem er nach der Bestätigung für Basissätze sucht, die in Widerspruch zu diesen Aussagensystemen stehen. Wissenschaftliche Theorien und Hypothesen können in dieser Tradition als Aussagensysteme angesehen werden, die allen Bemühungen der Widerlegungen standgehalten haben. Diese Bemühungen um Widerlegung bilden das Grundelement der Kritik. Die Kritik ist nicht mehr eine Haltung von außen gegenüber dem für wissenschaftlich sicher Gehaltenen, sie ist selbst ein Bestandteil des wissenschaftlichen Prozesses, genau genommen dessen Motor (Popper 1966). Das ist eine Haltung, die in den empirischen Sozialwissenschaften allgemein und der Erziehungswissenschaft speziell selten praktiziert wird, die aber verdeutlicht, welchen Wert methodisierte Kritik für die Wissenschaft haben kann.

## **7. Durch Kritik zur Wahrhaftigkeit der Modellannahmen**

Reifikation der Realität wird für kritisch geprüfte Modelle angenommen. Weil die Kritik nach den Regeln der Kunst erfolgte, kann das Modell als wahr oder bestätigt angenommen werden (je nach wissenschaftstheoretischer Position). Die Kritik verändert damit ihren Stellenwert im Wissenschaftsprozess. Ihre Funktion liegt nicht mehr im reflexiven Überprüfen des für richtig Angenommenen, sondern erschöpft sich in der Kontrolle des methodischen Vorgehens. Wird dieses als den Regeln der Kunst entsprechend darstellbar, dann resultiert daraus die Annahme, eine adäquate Aussage über Vorgänge in der

15 Ein wichtiges Zusatzargument findet sich bei Benner (in diesem Band), der auf die Negativität der kritischen und den aus der notwendigen Voraussetzung der Allgemeinheit der Negation folgenden Aporien verweist.

Realität gefunden zu haben. Die der Wissenschaft eigene Form des Diskutierens, die sich im methodischen Vorgehen noch identifizieren ließ, geht verloren. Das angemessene methodische Vorgehen wird im Sinne einer Immunisierung gegen Kritik argumentativ verwendet. Die Methode, der Weg, wird zum Symbol der Wissenschaftlichkeit. Das ist zwar gegenüber dem vorkritischen Status, den die empirische Erziehungswissenschaft in ihrer alltäglichen Praxis noch immer einnimmt, ein Fortschritt, genügt aber keineswegs den Erwartungen, die mit dem Programm verbunden werden.

## 8. Schlussfolgerung

Bei den empirischen Wissenschaften wird, wenn sie den vorkritischen positivistischen Status überwinden, der kritische Habitus immunisierend gegen Kritik eingesetzt. „Nur wenn ich weiß, in welcher Hinsicht etwas ‚nicht in Ordnung‘ sein könnte, weiß ich auch, welchen Test ich vorzunehmen habe, um den Verdacht zu entkräften. Wenn ich hingegen ohne derartige Andeutung einfach mit der Frage überfallen werde, ist dies ein wirkliches ‚x‘, so stehe ich ratlos da und weiß nicht, was ich tun und sagen soll“ (Stegmüller 1965, S. 29). In dieser Sicht lassen sich nochmals die Vorteile der Methodisierung der Kritik erkennen. Gleichzeitig wird aber auch die Tendenz zur Immunisierung deutlich, wenn unter Hinweis auf das Wissen um den Test, das Instrumentell-Handwerkliche des Vorgehens in den Vordergrund geschoben wird.

## Literatur

- Artelt, C./Stanat, P./Schneider, W./Schiefele, U. (2001): Lesekompetenz: Testkonzeption und Ergebnisse. In: Deutsches PISA-Konsortium (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen. Leske + Budrich, S. 69-137.
- Bac, M. (2000): Structure versus Process: Mach, Hertz, and the Normative Process of Science. In: Zeitschrift für Allgemeine Wissenschaftstheorie 31, S. 39-56.
- Baier, A. (1977): Auf der Suche nach Basishandlungen. In: Meggle, G. (Hrsg.): Analytische Handlungstheorie, Band 1: Handlungsbeschreibung. Frankfurt a.M.: Suhrkamp, S. 137-168.
- Bauer, M. (2000): Bildungsarbeit zu religiösen und kirchlichen Themen mit jugendlichen Ausiedlern. In: Interkulturelles Forum für interkulturelle Kommunikation, Erziehung und Beratung. H. 3/4. PH Freiburg: Forschungsstelle Migration und Integration, S. 173-180.
- Baumert, J./Schümer, G. (2001): Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In: Deutsches PISA-Konsortium: PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich, S. 323-407.
- Benner, D. (1978): Hauptströmungen der Erziehungswissenschaft. Eine Systematik traditioneller und moderner Theorien. München: List.
- Benner, D./Ramseger, J. (1981): Wenn die Schule sich öffnet. Erfahrungen aus dem Grundschulprojekt Gievenbeck. Juventa Materialien, B 50. München: Juventa.
- Blankenburg, W. (1975): Was heißt ‚Erfahren‘? In: Métraux, A./Graumann, C.F. (Hrsg.): Versuche über Erfahrung. Bern: Huber, S. 9-20.
- Bohnsack, R. (2000): Rekonstruktive Sozialforschung. Einführung in die Methodologie und Praxis qualitativer Forschung. Opladen: Leske + Budrich.

- Bohnsack, R./Nentwig-Gesemann, I./Nohl, A.-M. (2001): Die dokumentarische Methode und ihre Forschungspraxis. Grundlagen qualitativer Sozialforschung. Opladen: Leske + Budrich.
- Bortz, J./Döring, N. (1995): Forschungsmethoden und Evaluation. Berlin: Springer.
- Brandom, R.B. (1998): Making it Explicit: Reasoning, Representing, and Discursive Commitment. Harvard University Press.
- Brezinka, W. (1971): Von der Pädagogik zur Erziehungswissenschaft. Weinheim: Beltz.
- Brezinka, W. (1997): Rückblick auf 50 Jahre erlebte Pädagogik. In: Uhl, S. (Hrsg.): Wolfgang Brezinka – Fünfzig Jahre erlebte Pädagogik. Rückblick, Lebensdaten, Publikationen. München: Reinhardt, S. 11-29.
- Bridgman, P.W. (1970): Die Natur einiger unserer physikalischen Begriffe. In: Krüger, L. (Hrsg.): Erkenntnisprobleme der Naturwissenschaften. Texte zur Einführung in die Philosophie der Wissenschaft. Neue Wissenschaftliche Bibliothek, Bd. 38. Köln: Kiepenheuer & Witsch, S. 57-70.
- Campbell, D.T./Fiske, D. (1959): Convergent and discriminant validation by the multitrait-multimethod matrix. In: Psychological Bulletin 56, S. 81-105.
- Campbell, D.T./Stanley, J.C. (1963): Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally.
- Carnap, R. (1933): Über Protokollsätze. In: Erkenntnis, Bd. 3, S. 215-228.
- Carnap, R. (1959): Induktive Logik und Wahrscheinlichkeit. Heidelberg: Springer.
- Carnap, R. (1961): Der logische Aufbau der Welt. Hamburg: Felix Meiner Verlag.
- Carnap, R. (1966a): Scheinprobleme in der Philosophie. Das Fremdpsychische und der Realismusstreit. Frankfurt a.M.: Suhrkamp.
- Carnap, R. (1966b): Physikalische Begriffsbildung. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Danter, A.C. (1977): Basis-Handlungen. In: Meggle, G. (Hrsg.): Analytische Handlungstheorie, Bd. 1: Handlungsbeschreibung. Frankfurt a.M.: Suhrkamp, S. 69-88.
- Denzin, N.K. (1978): The Research Act. A Theoretical Introduction to Sociological Methods. New York: McGraw Hill.
- Deutsches PISA-Konsortium (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich.
- Diekmann, A. (1995): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Hamburg: Rowohlt.
- Flick, U. (2000): Triangulation in der qualitativen Forschung. In: Flick, U./Kardorff, E.v./Steinke, I. (Hrsg.): Qualitative Forschung. Ein Handbuch. Reinbeck: Rowohlt, S. 309-318.
- Flitner, W. (1963): Das Selbstverständnis der Erziehungswissenschaft in der Gegenwart. Heidelberg: Quelle & Meyer.
- Friebertshäuser, B./Prenzel, A. (Hrsg.) (1997): Handbuch qualitative Forschungsmethoden in der Erziehungswissenschaft. München: Juventa.
- Gadamer, H.-G. (1985): Theoretische Begriffe und die Prüfbarkeit von Theorien. In: Zeitschrift für allgemeine Wissenschaftstheorie 16, S. 19-24.
- Gehlen, A. (1962): Der Mensch. Seine Natur und seine Stellung in der Welt. Frankfurt a.M.: Athenäum.
- Glaser, B./Strauss, A. (1967): The discovery of grounded theory. Chicago: Aldine.
- Herbart, J.F. (1964): Die ersten Vorlesungen über Pädagogik. In: Herbart, J.F.: Kleinere Pädagogische Schriften, hrsg. v. Asmus, W. Düsseldorf: Küpper, S. 121-139.
- Hülst, D. (1973): Regelmäßigkeit und Erfahrung. Zum Problem sozialwissenschaftlicher Theoriekonstruktion. In: Hülst, D./Tjaden, K. H./Steinhauer-Tjaden, M. (Hrsg.): Methodenfragen der Gesellschaftsanalyse. Frankfurt a.M.: Fischer Athenäum Taschenbuch Verlag.
- Jäger, M. (2001): ‚Was ist Begründung?‘ In: Zeitschrift für Allgemeine Wissenschaftstheorie 32, S. 167-192.



- Kockelmanns, J.J. (1975): Empirische, phänomenologische und hermeneutische Psychologie. In: Métraux, A./Graumann, C.F. (Hrsg.): *Versuche über Erfahrung*. Bern: Huber.
- Krapp, A. (1979): *Prognose und Entscheidung*. Weinheim: Beltz.
- Kromrey, H. (1998): *Empirische Sozialforschung. Modelle und Methoden der Datenerhebung und Datenauswertung*. Opladen: Leske+Budrich.
- Krumm, V. (1983): Kritisch-rationale Erziehungswissenschaft. In: D. Lenzen, K. Mollenhauer (Hrsg.): *Theorie und Grundbegriffe der Erziehung und Bildung. Enzyklopädie Erziehungswissenschaft, Bd.1*. Stuttgart: Klett-Cotta, S. 139-154.
- Laing, R.N. (1969): *Phänomenologie der Erfahrung*. Frankfurt a.M.: Suhrkamp.
- Lehmann, R.H./Peek, R./Gänsfuß, R./Husfeldt, V. (2002): LAU 9. Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe 9. Ergebnisse einer Längsschnittuntersuchung in Hamburg. Freie und Hansestadt Hamburg, Behörde für Bildung und Sport, Amt für Schule.
- Levitt, B./March, J.G. (1996): Organizational Learning. In: Cohen, M.D./Sproull, L.S. (Hrsg.): *Organizational Learning*. Thousand Oaks (Sage), S. 516-540.
- Merkens, H. (1972): Probleme und Schwierigkeiten bei der Beobachtung als einer empirischen Methode. In: *Programmiertes Lernen* 9, S. 75-82.
- Merkens, H. (1974): Überlegungen zu einer Theorie der Beobachtung. In: *Unterrichtswissenschaft* 2, S. 14-20.
- Merkens, H. (1986): Vorwissen und Hypothesenbildung beim Prozess des Beobachtens – Überlegungen zu den Grenzen der Beobachtung bei der Arbeitsmigrantenforschung. In: Hoffmeyer-Zlutnik, J.H.P. (Hrsg.): *Qualitative Methoden der Datenerhebung in der Arbeitsmigrantenforschung*. Mannheim: Quorum Verlag Berlin, S. 78-108.
- Métraux, A./Graumann, C.F. (Hrsg.) (1975): *Versuche über Erfahrung*. Bern: Huber.
- Neurath, O. (1933): Protokollsätze. In: *Erkenntnis*, Bd. 3, S. 204-214.
- Opp, K.-D. (1970): *Methodologie der Sozialwissenschaften. Einführung in Probleme ihrer Theoriebildung*. Reinbeck: Rowohlt, S. 339-341.
- Popper, K.R. (1966): *Logik der Forschung*. Tübingen: J.C.B. Mohr.
- Prim, R./Tilman, H. (1973): *Grundlagen einer kritisch-rationalen Sozialwissenschaft*. Heidelberg: Quelle & Meyer.
- Schad, U. (2000): Rechte Jugendliche im breiten Strom – über die unterschiedlichen Dimensionen von Rassismus und Rechtsextremismus. In: *Interkulturelles Forum für interkulturelle Kommunikation, Erziehung und Beratung*, H. 3/4. PH Freiburg: Forschungsstelle Migration und Integration, S. 64-79.
- Schickore, J. (1999): Sehen, Sichtbarkeit und empirische Forschung. In: *Zeitschrift für Allgemeine Wissenschaftstheorie* 30, S. 273-287.
- Stegmüller, W. (1965): *Glauben, Wissen und Erkennen. Das Universalienproblem*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Strauss, A./Corbin, J. (1990): *Basics of qualitative Research. Grounded theory. Procedures and techniques*. Newbury Park (Sage).
- Tausch, R./Tausch, A.-M. (1970): *Erziehungspsychologie*. Göttingen: Hogrefe.
- Trapp, E.C. (1913): *Versuch einer Pädagogik*. Hrsg. v. T. Fritsch. Leipzig: Quelle und Meyer.
- Wolters, G. (2000): Die pragmatische Vollendung des logischen Empirismus. In memoriam Carl Gustav Hempel. In: *Zeitschrift für Allgemeine Wissenschaftstheorie* 31, S. 205-242.
- Zinnecker, J. (2000): Pädagogische Ethnographie. In: *Zeitschrift für Erziehungswissenschaft* 3, S. 381-400.

*Anschrift des Autors:*

Prof. Dr. Hans Merkens, Freie Universität Berlin, FB 12: Erziehungswissenschaften (WE 1), Fabeckstr. 13, 14195 Berlin.