

Wellenreuther, Martin

Willkommen, Mr. Chance. Methodologische Betrachtungen zur Güte empirischer Forschung in der Pädagogik diskutiert vor allem an der neueren Untersuchung über Gewalt von Heitmeyer u.a. (1995)

Zeitschrift für Pädagogik 43 (1997) 2, S. 321-332



Empfohlene Zitierung/ Suggested Citation:

Wellenreuther, Martin: Willkommen, Mr. Chance. Methodologische Betrachtungen zur Güte empirischer Forschung in der Pädagogik diskutiert vor allem an der neueren Untersuchung über Gewalt von Heitmeyer u.a. (1995) - In: Zeitschrift für Pädagogik 43 (1997) 2, S. 321-332 - URN: urn:nbn:de:0111-pedocs-69855

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen

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.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

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.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Digitalisiert

Mitglied der


Leibniz-Gemeinschaft

Zeitschrift für Pädagogik

Jahrgang 43 – Heft 2 – März/April 1997

Essay

- 187 ULRICH HERRMANN
Können wir Kinder verstehen? Rousseau und die Folgen

Thema: Kritik der Werterziehung

- 199 GEORG HANS NEUWEG
Kritische Rationalität und „Werte-Erziehung“
- 219 SÖNKE ABELDT
Das Problem Solidarität. Perspektiven der pädagogischen Ethik und der Kritischen Theorie

Thema: Einheitlichkeit und Differenzierung

- 241 PETER M. ROEDER
Binnendifferenzierung im Urteil von Gesamtschullehrern
- 261 GERHART NEUNER
Das Einheitsprinzip im DDR-Bildungswesen
- 279 ANKE HUSCHNER
Vereinheitlichung und Differenzierung in der Schulentwicklung der SBZ und DDR

Diskussion: Bildungsforschung

- 301 MANFRED LÜDERS
Von Klassen und Schichten zu Lebensstilen und Milieus. Zur Bedeutung der neueren Ungleichheitsforschung für die Bildungssoziologie
- 321 MARTIN WELLENREUTHER
Willkommen Mr. Chance. Methodologische Betrachtungen zur Güte empirischer Forschung in der Pädagogik, diskutiert vor allem an der neueren Untersuchung über Gewalt von Heitmeyer u. a. (1995)

Besprechungen

- 335 BERND-REINER FISCHER
Dietrich Hoffmann/Karl Neumann (Hrsg.): Erziehung und Erziehungswissenschaft in der BRD und der DDR. Band 1: Die Teilung der Pädagogik (1945–1965); Band 2: Divergenzen und Konvergenzen (1965–1989); Band 3: Die Vereinigung der Pädagogiken (1989–1995)
- 342 HANS-WERNER FUCHS
Gert Geißler/Ulrich Wiegmann: Pädagogik und Herrschaft in der DDR. Die parteilichen, geheimdienstlichen und vormilitärischen Erziehungsverhältnisse
- 345 HANS-ULRICH GRUNDER
Fritz-Peter Hager/Daniel Tröhler (Hrsg.): Pestalozzi – wirkungsgeschichtliche Aspekte. Dokumentationsband zum Pestalozzi-Symposium 1996
- 349 BEATE TRÖGER
Johann Heinrich Pestalozzi: Sämtliche Werke und Briefe auf CD-ROM

Dokumentation

- 353 Pädagogische Neuerscheinungen

Willkommen, Mr. Chance

Methodologische Betrachtungen zur Güte empirischer Forschung in der Pädagogik, diskutiert vor allem an der neueren Untersuchung über Gewalt von Heitmeyer u. a. (1995)

Zusammenfassung

Der Nutzen strenger quantitativer Forschung wird – vor allem wegen der unzureichenden Professionalisierung der Methodenausbildung – in der Pädagogik immer noch in Zweifel gezogen. Entsprechend weisen empirisch-pädagogische Forschungen häufig erhebliche Mängel auf. Dies soll im folgenden exemplarisch an der Untersuchung von HEITMEYER u. a. (1995) zum Thema Gewalt gezeigt werden. Kritik wird geübt an (1) der theoretischen Güte, (2) der versuchsplanerischen Güte, (3) der Güte der Operationalisierungen und Messungen sowie (4) an der Güte der statistischen Auswertung. Eine empirische Untersuchung ist nur dann wirklich „gut“, wenn jeder dieser Aspekte hinreichend gut ist. Die Darstellung und Diskussion beschränkt sich auf die Methodologie der Hypothesenprüfung unter besonderer Berücksichtigung der Probleme der Feldforschung.

1. Vorbemerkungen: Die Misere empirisch pädagogischer Forschung als Wunschvorstellung und als reales Problem

In Pädagogik und Sozialwissenschaften bestehen Verständigungsschwierigkeiten, Unsicherheiten bzw. Meinungsverschiedenheiten hinsichtlich der Rolle empirischer Forschung und der dabei relevanten Gütekriterien. Gerade in der deutschen Pädagogik haben qualitative, weiche Verfahren Konjunktur, soweit empirische Methoden überhaupt ernst genommen werden. So behauptet z. B. R. FATKE, daß die realistische Wende (H. ROTH) „mit dem Vormarsch der empirisch orientierten Erziehungswissenschaft ... das Spannungsverhältnis zwischen Theorie und Praxis noch weiter hat auseinandertreten lassen“ (FATKE 1995, S. 675). Und W. LEHMEIER bemerkt (1995, S. 633): „In jüngster Zeit richtet sich die Kritik an der naturwissenschaftlichen Orientierung der Sozialwissenschaften insbesondere gegen die dem quantitativen Paradigma verbundene Forschungsmethodologie. Wohl auch aus Enttäuschung über eine Vielzahl banaler bzw. widersprüchlicher empirischer Ergebnisse wird die (ausschließliche) Angemessenheit quantitativer Forschungsdesigns und -methoden und damit die sie fundierende Erkenntnis- bzw. Wissenschaftstheorie zunehmend in Zweifel gezogen. Kritisiert wird eine Forschungsmethodologie, die ihren Gegenstand hypothesengeleitet zergliedert, den Wirklichkeitsbereich des Gegenstandes standardisiert und die durch Beobachtung und Experiment gewonnene Erfahrung zweckrational verrechnet.“

Daraus wird dann gefolgert, man müsse mehr Fallstudien durchführen bzw. stärker qualitative, weiche Verfahren zum Einsatz bringen, die Forschungsobjekte endlich wieder zu Subjekten machen u. a. m. Dadurch, daß diese „Argumente“ seit Jahrzehnten ständig wiederholt werden, werden sie nicht besser (die

Behauptung von LEHMEIER „In jüngster Zeit ...“ ist schlicht falsch); daß ständig der Niedergang der Pädagogik aufgrund der strengen quantitativen Forschung beschworen wird, zeigt u.U. nur, daß viele Pädagogen Schwierigkeiten haben, den Kenntnisstand in einem Bereich, in dem in den letzten Jahren ein beträchtlicher Erkenntnisfortschritt stattgefunden hat, angemessen aufzuarbeiten und auf pädagogische Fragestellungen anzuwenden.

Damit wird weder behauptet, weiche, qualitative Verfahren seien im Rahmen pädagogischer Forschung unwichtig (vgl. WELLENREUTHER 1976, 1982, S. 20f., S. 114f.), noch wird behauptet, mit der strengen empirischen Forschung stehe alles zum besten. Dies zeigt sich m.E. auch an der Untersuchung von W. HEITMEYER u. a. (1995), die als exemplarisches Beispiel zur Verdeutlichung der Gütestandards dienen soll. Diese Untersuchung ist in der Öffentlichkeit auf ein breites Interesse gestoßen. Zudem gilt HEITMEYER als einer der Experten bei empirischen Forschungen zum Thema Gewalt und Rechtsradikalismus (vgl. z.B. SCHNABEL 1993).

2. Gütekriterien empirischer Untersuchungen

2.1 Die theoretische Güte empirischer Untersuchungen

Um gute empirische Forschung betreiben zu können, muß man sich zunächst einen Überblick über den Erkenntnisstand in dem betreffenden Problembe-
reich verschaffen: Beim Thema Gewalt könnte man sich in neueren Lehrbü-
chern, Zeitschriften, Abstracts informieren, wobei man auch die international
durchgeführte Forschung berücksichtigen sollte.

HEITMEYER u. a. (1995) haben z.B. in ihrer Untersuchung über Gewalt einen
Großteil der einschlägigen Untersuchungen nicht erwähnt, obwohl diese Un-
tersuchungen zur Prüfung von Theorien durchgeführt wurden, die auch allge-
mein auf die Entstehung von Gewalt anwendbar sind (vgl. PATTERSON 1982; PAT-
TERSON/REID/DISHION 1992; SAMPSON/LAUB 1993; LOEBER/STOUTHAMER-LOEBER
1986). Ein Verzicht auf eine umfassende Einbeziehung der fremdsprachigen
Literatur ist m.E. nicht zu rechtfertigen, da ein Großteil der relevanten theo-
retischen und empirischen Literatur nicht in Deutschland, sondern in eng-
lischsprachigen Zeitschriften und Büchern publiziert wird.

Die Darstellung des gegenwärtigen Erkenntnisstands setzt eine zutreffende
Vorstellung davon voraus, was theoretisches Argumentieren und Erklären über-
haupt ausmacht. Gerade in der noch stark geisteswissenschaftlich geprägten
deutschen Pädagogik gibt es hier keinen Konsens. Ich gehe von folgender Vor-
stellung aus: Theorien bestehen aus zusammenhängenden Hypothesen, die em-
pirisch widerlegbar sind und die trotz strenger Tests „überlebt“ haben. Wider-
legbare Aussagen bergen das Risiko in sich, falsch sein zu können. Dafür
erlauben sie aber auch bedingte Prognosen (keine Prophezeihungen), sie infor-
mieren über zu erwartende Wirkungen, die unter bestimmten Bedingungen ein-
treten, sie erweisen sich bei der Lösung praktischer Probleme als nützlich. Wenn
Hypothesen schon aus logischen Gründen nicht widerlegbar sind, dann lohnt
sich auch nicht ihre empirische Überprüfung.

Nach HEITMEYER u. a. (1995) soll die Individualisierung in unserer Gesell-

schaft etwas mit der Entstehung von Gewalt zu tun haben. Dem gestiegenen Individualisierungsdruck entspreche ein steigendes Sicherheitsbedürfnis (HEITMEYER u. a. 1995, S. 40). „Damit ist Verunsicherung strukturell an Individualisierung gebunden“. Und was hat dies alles mit der Entstehung von Gewalt bei Jugendlichen zu tun? Auf Seite 48 kommt HEITMEYER schließlich auf Faktoren, die zu Gewalt führen sollen, in einem Identitätstheoretischen Kontext zu sprechen. Bezug nehmend auf HABERMAS und PIAGET postuliert er: „Insofern ist ein enger Zusammenhang von Autonomie und Gewalt anzunehmen: Je stärker sich gewalttätige Orientierungen zur Selbstdurchsetzung etc. herauschälen, desto weniger ausgeprägt ist die Entwicklung in eine autonomie-orientierte Richtung.“ Im ersten Satz könnte man annehmen, daß Autonomie der Einflußfaktor ist, der gewalttätige Orientierungen beeinflusst. Im nächsten Satz stehen aber in der Satzkomponente mit der unabhängigen Variablen „gewalttätige Orientierungen“, und plötzlich wird von den gewalttätigen Orientierungen auf eine Entwicklung in eine autonomie-orientierte Richtung geschlossen. Aufgrund der langen Ausführungen über die Individualisierung hätte man erwartet, daß das Gewähren großer Autonomieräume unter bestimmten Voraussetzungen, die dann näher erläutert werden, zu gewalttätigen Orientierungen führt.

Unter dem Stichwort „Individualisierungsfallen in Ost und West“ stellt HEITMEYER dann einige Thesen zur Diskussion.

- Für den *Schul- und Ausbildungsbereich* im Osten: „Aus dem stärkeren Wettbewerb um höhere Bildungszertifikate ergeben sich Risiken, die sich in den Bewältigungsmustern in erster Linie als Status- und Leistungsängste äußern“ (S. 53). Diese These ist offensichtlich empirisch nicht widerlegbar: Da es sich um eine Existenzaussage handelt, läßt sie sich ohnehin nur verifizieren, und dies dürfte nicht allzu schwer sein: Wenn man innerhalb eines längeren Zeitraums auch nur ein Risiko feststellen würde, das mit erhöhten Status- oder Leistungsängsten einhergeht – also irgendwann einmal –, wäre die These „verifiziert“. Wenn bei einem „Risiko“ solche Konsequenzen nicht eintreten, dann ist es kein echtes Risiko!
- Über die *Prozesse in der Herkunftsfamilie*: „Aus all dem läßt sich die Annahme ableiten, daß mit dem Funktionswandel der Familie und der Differenzierung ihrer Formen Belastungen entstanden sind, die sich vor allem im Bereich der sozialen Beziehungen und emotionalen Anerkennungen auswirken“ (S. 53) Auch bei dieser These handelt es sich um keine sonderlich informative Aussage: Wenn man irgendwann einmal eine Belastung feststellen kann, die mit der Differenzierung der Familienformen zusammenhängt, ist diese „These“ bestätigt. Wenn eine solche Belastung aber nicht diese Konsequenzen hat, dann ist es auch keine Belastung!

In beiden Thesen wird die Frage der Gewalt nicht angesprochen. Der Argumentationsweise HEITMEYERS ist nur zu entnehmen, daß diese „Desintegrationspotentiale“ irgend etwas mit der zunehmenden Gewalt zu tun haben. Etwas unvermittelt wird dann doch noch ein Zusammenhang zwischen Desintegration und Gewalt hergestellt (vgl. S. 60): „Desintegration ist nun bereits als Kombination von Ausgrenzungs- und Auflösungsprozessen definiert worden ... Der Zusammenhang von solchen Auflösungsprozessen und Gewalt wird allerdings nur dort

angenommen, wo Desintegration als Verlust von Zugehörigkeit, Teilnahmechancen oder Übereinstimmung erfahren und wahrgenommen wird.“

Wann Desintegration als Verlust wahrgenommen wird, wird nicht ausgeführt. So bleibt dem Forscher immer die Möglichkeit, bei widerspenstigen Daten anzunehmen, der Verlust sei als solcher nicht wahrgenommen worden. Nimmt man die These ohne diese Immunisierungsfloskel, dann ist sie offensichtlich falsch: Auch wenn Desintegration als Verlust wahrgenommen wird, wird häufig noch keine Gewalt resultieren, sondern Resignation, Enttäuschung, Rückzug (vgl. WILLEMS 1992). Vermutlich aber ist die These gar nicht so streng gemeint. Vielleicht soll nur behauptet werden, daß bei der subjektiven Wahrnehmung von Desintegration als Verlust von Zugehörigkeit etc. *manchmal* mit Gewalt reagiert wird. Dann allerdings wäre die These wiederum empirisch nicht widerlegbar.

Dieses hier an Beispielen dargelegte Theorieverständnis der Forschungsgruppe um HEITMEYER ist in der neueren „deutschen“ Soziologie und Pädagogik noch häufig anzutreffen. Es entstammt teilweise der sogenannten Kritischen Theorie (HABERMAS, ADORNO), teilweise marxistischen und psychoanalytischen Theorieansätzen (vgl. WELLENREUTHER 1982). Dieses Theorieverständnis erinnert mich an die Rolle von PETER SELLERS in dem Film „Willkommen, Mr. Chance“ (1979, engl. „Being there“). SELLERS spielt in seinem zweitletzten Film einen törichten, durch seine klugen Sprüche auffallenden Gärtner, der seine Informationen ausschließlich aus dem Fernsehen bezieht und der Haus und Garten seines Dienstherrn nie verlassen hat. Durch die Freundschaft mit einem Industriellen und seiner Frau, die er durch seine „Weisheiten“ stark beeindruckt, gewinnt er Einfluß auf die internationale Politik. Trotz des sehr eingeschränkten Erfahrungshorizontes fällt Mr. Chance zu jeder Situation eine kluge Leerformel ein, die sich natürlich immer bewahrheitet. Solche Leerformeln können auf den ersten Blick äußerst überzeugend sein, und in aller Regel verstehen sich die Protagonisten dieses Theorieverständnisses der Leerformeln und Wolken als ausgezeichnete Theoretiker. Allerdings sollte man auch die Nachteile dieser Position nicht übersehen:

- Da die „Theorien“ mit jedem beliebigen Sachverhalt vereinbar sind, gibt es hier statt Erkenntnisfortschritt Wissensstillstand.
- Die Realität spielt als ernstzunehmende Prüfinstanz keine Rolle mehr. Sie wird nur noch dazu verwendet, den einmal vorgefaßten Meinungen zusätzlich eine empirische Weihe zu verleihen.

2.2 Die versuchsplanerische Güte einer empirischen Untersuchung

Der zweite Prüfstein für die Beurteilung einer Untersuchung ist ihre versuchsplanerische Validität, d.h. die Eindeutigkeit des Nachweises, daß die variierte Bedingung eine bestimmte Wirkung erbracht hat. Bestehen begründete Zweifel an der versuchsplanerischen Güte einer Untersuchung, können Ergebnisse alternativ durch die Wirkung verschiedener Einflußfaktoren erklärt werden. Die Untersuchung muß in einem solchen Fall mit einem verbesserten Versuchsplan

erneut durchgeführt werden, um die für das vorliegende Problem relevanten alternativen Erklärungen möglichst auszuschließen.

Das Experiment eröffnet die besten Möglichkeiten, eine solche versuchsplanerische Validität zu realisieren (WELLENREUTHER 1982). Die vielleicht wichtigste Kontrolltechnik im Experiment ist die Randomisierung, d. h. die Zufallsaufteilung der Versuchseinheiten auf verschiedene Gruppen (vgl. CAMPBELL/STANLEY 1963; COOK/CAMPBELL 1979). Randomisierung sollte nicht mit dem Ziehen einer Zufallsstichprobe verwechselt werden. Ziel der Hypothesenprüfung ist ja nicht die möglichst repräsentative Beschreibung einer „Grundgesamtheit“, sondern die Führung des Nachweises, daß die variierte Bedingung eine bestimmte Wirkung erzeugt. Der Bezugspunkt ist nicht eine konkrete, abzählbare, endliche Grundgesamtheit; Hypothesen beziehen sich auf unendliche, nicht verfügbare, „hypothetische“ Grundgesamtheiten, auf die man auch nicht mit Hilfe der besten Statistik schließen kann. Repräsentativität ist somit für hypothesenprüfende Forschung ein irrelevanter Gesichtspunkt (vgl. GADENNE 1976).

Wenn ein Experiment nicht durchführbar ist, dann ist unter den verbleibenden Untersuchungsansätzen eine langfristig angelegte Panel- bzw. Längsschnittuntersuchung, in der z. B. Personen seit ihrer Kindheit mehrfach intensiv in bestimmten zeitlichen Abständen untersucht werden, die beste Möglichkeit, um ein hohes Maß an versuchsplanerischer Validität zu gewährleisten. Und wenn eine solche Paneluntersuchung nicht möglich ist, dann ist eine Querschnittuntersuchung die nächstbeste Wahl, wenn in ihr die Versuchs- und Kontrollgruppen nach theoretischen Gesichtspunkten zusammengestellt wurden (Parallelisierung bzw. Schichtung der Stichprobe nach theoretischen Gesichtspunkten). Positive Beispiele dafür wären die klassische Untersuchung von GLUECK und GLUECK (1950) oder die von SAMPSON und LAUB (1993), die übrigens eine Reanalyse der Daten von GLUECK/GLUECK durchführten, sowie die Untersuchung von QUINTON/RUTTER über „parenting breakdown“ (1988).

In der Untersuchung von HEITMEYER u. a. (1995) erfolgt die Bildung der Stichprobe nicht nach überzeugenden theoretischen Gesichtspunkten, sondern eher nach dem Gesichtspunkt der Repräsentativität. Schichtungsmerkmale waren hier a) Ost- vs. Westdeutschland, b) städtische – ländliche Strukturierung (Metropole, Stadt, Land), c) Geschlecht. Diese Gesichtspunkte mögen zwar dem Gesichtspunkt der Repräsentativität einigermaßen genügen, sie haben aber nur sehr bedingt etwas mit dem theoretisch interessierenden Problem, Gewalt und Gewaltentstehung, zu tun. Man weiß, daß Gewalt überwiegend ein männliches Phänomen ist, eine fünfzigprozentige Berücksichtigung von Frauen ist deshalb überflüssig. Gewalt ist ein großes Problem in bestimmten sozialen Brennpunkten, differiert aber nicht notwendig zwischen Land und Stadt. Entsprechend wäre es sinnvoller gewesen, in solchen sozialen Brennpunkten eine Stichprobe von männlichen Jugendlichen zu ziehen. Noch besser wäre allerdings die Durchführung einer Längsschnittuntersuchung, die sich auf solche soziale Brennpunkte konzentriert und die schon bei der Geburt der Jungen beginnt. Man könnte eine derart ausgewählte Stichprobe dann an verschiedenen Zeitpunkten (z. B. mit einem Jahr, vier Jahren, sieben Jahren, 11 Jahren, 13 Jahren, und 19 Jahren) und mit verschiedenen Meßmethoden (Beobachtung, Fragebogen, Telefoninterview) über verschiedene Meßagenten (Mutter, Vater, Lehrer, Beobachter) untersuchen (vgl. HAAPASALO/TREMBLEY 1994).

Bei der Prüfung theoretischer Aussagen interessiert, ob die zu prüfenden Hypothesen zutreffen, d. h., ob die behaupteten Beziehungen zwischen Einflußfaktoren und den jeweils postulierten Wirkungen bestehen. Dazu ist wichtig, daß nach den theoretisch interessierenden Merkmalen die Stichprobe systematisch so zusammengestellt wird, daß die nach den relevanten Bedingungsfaktoren gebildeten Teilgruppen möglichst gleich groß sind.

Eine Modellrechnung mag dies erläutern: Um Gewalt bei Jugendlichen zu untersuchen, ziehe ich z. B. wie HEITMEYER u. a. (1995) eine Stichprobe von etwa 2000 Jugendlichen. Bei Ziehung der Stichprobe nach dem Zufallsprinzip habe ich etwa 50% männliche und 50% weibliche Jugendliche. Da bei Frauen Gewaltverhalten in Form offener Aggression kaum auftritt, kann von vornherein die Frage der Gewalt nur bei den männlichen Jugendlichen genauer nachgegangen werden. Da solche Probleme eher in Hauptschulen, Realschulen und Gesamtschulen als an Gymnasien auftreten, stehen für eine detailliertere Analyse des Gewaltproblems nur etwa 25% der gezogenen Stichprobe zur Verfügung, also 500 männliche Jugendliche an Haupt-, Real- und Gesamtschulen. Von diesen 500 Jugendlichen zeigen vielleicht 50 Jugendliche ausgeprägtere Tendenzen zur Gewalttätigkeit, zu denen dann mit Hilfe der Parallelisierung eine Vergleichsgruppe gebildet wird. Um zum Problem der Gewalt Analysen durchführen zu können, ist dann diese Gruppe von 50 gewalttätigen Jugendlichen zu klein. Hätte man von vornherein eine Stichprobe von 400 männlichen Jugendlichen aus solchen sozialen Brennpunkten ausgewählt, die nach offiziellen Unterlagen schon gewalttätig geworden sind, und zu diesen Jugendlichen eine Kontrollgruppe von 400 männlichen Jugendlichen gebildet, die nach sozialen und schulischen Merkmalen vergleichbar, bezüglich des entscheidenden Merkmals „Gewalttätigkeit“ aber unvergleichbar (geringe Ausprägung von Gewalttätigkeit) sind, hätte man das gewählte Thema erheblich besser untersuchen können.

2.3 Die Güte von Operationalisierungen und Messungen

Ein weiterer Gesichtspunkt der Güte einer empirischen Untersuchung ist die Güte der Operationalisierungen der Bedingungsfaktoren („Ursachen“, unabhängige Variable) und der Messung der Wirkungen (abhängige Variable). Wenn in einer Untersuchung über Gewalttätigkeit von Jugendlichen der Umfang, die Häufigkeit und die Intensität der Gewalttätigkeit nicht bzw. nur sehr grob gemessen werden, dann ist die Untersuchung von vornherein als weniger gut einzuschätzen als eine Untersuchung, in der mit großer Sorgfalt Meßinstrumente zur Erfassung von Art, Umfang, Häufigkeit und Intensität von Gewalthandlungen erfaßt wurden.

Zunächst muß man sich über den Bedeutungsumfang eines Merkmals klarwerden. „Gewalttätigkeit bedeutet, daß ... Diese Gewalttätigkeit kann in folgenden Situationen mit folgenden Intensitäten oder Modalitäten auftreten.“ Danach bildet man für verschiedene einsetzbare Meßverfahren Items bzw. Kategorien. Eine Verbesserung der Messung kann dadurch erzielt werden, indem man die Messungen verschiedener Meßmethoden kombiniert: Man entwickelt auf diese Weise Konstruktmessungen, die auf Messungen basieren, die über verschiedene Agenten (Jugendliche, Mütter, Lehrer, geschulte Beobachter, Inter-

viewer) erhoben wurden und die verschiedene Meßmethoden berücksichtigen (Fragebogen, Kurzinterview am Telefon, ausführliches Interview, Beobachtung, Rating, offizielle Statistiken und Gerichtsakten). Durch die systematische Kombination verschiedener Messungen, die ja das gleiche Merkmal erfassen sollen, erreicht man eine wesentliche Verbesserung der Meßgüte (vgl. CAMPBELL/FISKE 1959; CRONBACH/MEEHL 1955).

Ein positives Beispiel bezüglich der Entwicklung von Meßinstrumenten liefert auch hier das schon genannte Projekt von PATTERSON, CAPALDI und PATTERSON (1989) haben eigens zu diesem Problem ein Buch veröffentlicht, in dem die verwendeten Meßinstrumente bezüglich relevanter psychometrischer Eigenschaften dargestellt werden.

Auch bezogen auf die Messung der Gewalttätigkeit als entscheidender abhängiger Variable kann die Untersuchung von HEITMEYER u. a. (1995) nicht als positives Beispiel dienen. Verglichen mit der differenzierten Erfassung gewaltbezogener Einstellungen wird die Gewalttätigkeit nur durch eine Frage mit fünf verschiedenen Tätigkeiten erfaßt. Es handelt sich dabei um folgende Frage:

- „Ist es in den letzten 12 Monaten vorgekommen, daß Sie
... Sachen von anderen absichtlich zerstört oder beschädigt haben?
... jemanden absichtlich geschlagen oder geprügelt haben?
... jemanden bedroht haben, damit er/sie tut, was Sie wollen?
... jemandem eine Sache mit Gewalt weggenommen haben?
... irgendwo eingebrochen haben (z. B. in ein Gebäude, ein Auto oder einen Automaten)?“

Es werden weder Umfang, Intensität noch Häufigkeit der Gewalttätigkeit erfaßt. Es macht doch einen Unterschied, ob jemand in Notwehr den Angreifer in die Schranken verweist, ohne ihn zu verletzen, oder ob jemand, ohne angegriffen zu sein, jemanden schlägt. Wenn gewalttätige Handlungen bei einem Jugendlichen an der Tagesordnung sind, dann wird er bei dieser Messung genauso behandelt wie ein Jugendlicher, der im letzten Jahr einmal einen Klaps zurückgegeben hat. Eine differenzierte Erfassung der gewaltaffinen Einstellungen ist dafür kein vollwertiger Ersatz.

Bei der Entstehung von gewalttätigem Verhalten spielt u. a. die Konsequenz des elterlichen Verhaltens eine wichtige Rolle. In der Untersuchung von HEITMEYER u. a. wurden zur Messung der Konsistenz bzw. Inkonsistenz des Erziehungsverhaltens zwei Indikatoren verwendet:

(1) „Bei meiner Erziehung haben sich meine Eltern mal so, mal anders verhalten. Ich wußte eigentlich nie so richtig, wie ich mich verhalten sollte.“

2) „Bei meinen Eltern kann ich mich verhalten, wie ich will, immer ist irgend etwas falsch.“

Erheblich größerer Aufwand bei der Erfassung inkonsistenten Erziehungsverhaltens wurde in der Untersuchung von CONGER u. a. (1995) getrieben. Hier wurde nicht der Jugendliche, sondern die Eltern sowie ein Beobachter befragt. Sie wurden gebeten, auf einer fünfstufigen Skala anzugeben, wie oft die Eltern aufgeben, wenn der Heranwachsende um etwas gebeten wurde, wie oft er einer Bestrafung entgehen kann, wenn eine solche beschlossen wurde, in welchem Umfang die Bestrafung von der Stimmung des Elternteils abhängig sei und wie oft es vorkam, daß ein bestimmtes Fehlverhalten manchmal bestraft und das andere Mal nicht bestraft wurde. Außerdem sollten die Partner die Konsistenz

bzw. Inkonsistenz des Verhaltens des anderen Partners einschätzen. Auch inkonsistente Disziplin der beiden Elternteile wurde zusätzlich von den Beobachtern beurteilt.

Um in der empirischen Forschung weniger widersprüchliche Ergebnisse zu erhalten, scheint es dringend geboten, die vorhandenen Möglichkeiten der Verbesserung der Genauigkeit und Gültigkeit von Meßinstrumenten besser zu nutzen. Dies bedeutet aber auch, daß vor der Durchführung der eigentlichen Untersuchung eine Phase der Entwicklung der Operationalisierungen und Messungen stehen sollte.

2.4 Die statistische Güte einer Untersuchung

Ein letzter Ansatzpunkt für Kritik an empirischen Untersuchungen wäre die geeignete Verwendung statistischer Verfahren. Statistische Verfahren sollen die gesammelten Informationen so auswerten, daß die Fragen, die durch die Untersuchung beantwortet werden sollen, auch möglichst eindeutig beantwortet werden. Dem naiven Betrachter kann die Verwendung komplizierter statistischer Auswertungsverfahren eine Seriosität und Wissenschaftlichkeit einer Untersuchung vorgaukeln, die faktisch keineswegs gegeben ist.

Gerade in der Feldforschung ist eine strenge Prüfung von Hypothesen sehr schwierig. Ein positives Beispiel für die Möglichkeiten der Hypothesenprüfung im Rahmen von Längsschnittuntersuchungen ist die Untersuchung von CHERLIN u. a. (1991). In dieser Untersuchung wird gezeigt, daß die Unterschiede, die zwischen Kindern aus geschiedenen und intakten Familien auftreten, nicht als direkte Folgen der Scheidung zu betrachten sind, sondern weitgehend auf Faktoren zurückgehen, die schon vor der Scheidung, also in sogenannten intakten Familien, vorhanden waren. Man konnte dies dadurch belegen, indem man Kinder, die in einem bestimmten Zeitraum (3. bis 9. März 1958) in Großbritannien geboren worden waren, mehrfach nach der Geburt testete. Damit waren Daten verfügbar, die sich auf die Zeit bezogen, als die Kinder noch in „intakten“ Familien lebten. Wenn man diese Unterschiede, die schon in Zeiten intakten Familienlebens entstanden waren, kontrollierte, verschwanden die Unterschiede zwischen Scheidungs- und Nichtscheidungskindern fast völlig. Solche differenzierten Längsschnittanalysen lenken die Aufmerksamkeit auf die kausalen Prozesse, die hinter der Oberfläche empirischer Ergebnisse stehen (vgl. RUTTER 1994).

Woran kann man erkennen, daß ein Forscher sich auch bei der Verwendung statistischer Verfahren um eine strenge Prüfung der aufgestellten Hypothesen bemüht? Ein wichtiger Anhaltspunkt ist m.E., in welchem Maße der Forscher die gesammelte Information bündelt, um durch möglichst wenige statistische Tests über die aufgestellten Hypothesen entscheiden zu können.

Insbesondere bei Querschnittuntersuchungen ist eine strenge Hypothesenprüfung schwierig. Ein Beispiel dafür ist die Prüfung des „Desintegrations-Verunsicherungs-Gewalt-Theorems“ durch HETMEYER u. a. (1995, S. 141f.). Diese These wird überprüft, indem über viele verschiedene Indikatoren der Zusammenhang zwischen Desintegrationsfaktoren in der Gesellschaft und Verunsicherung geprüft wird mit dem Ergebnis, daß in einigen Fällen ein Zusammen-

hang festgestellt wird, in anderen nicht. Danach wird geprüft, inwiefern Verunsicherung mit Gewalttätigkeit zusammenhängt. Auch hier gibt es mehr oder weniger bestätigende Befunde. Der Schluß scheint somit gerechtfertigt, daß die These insgesamt bestätigt sei. Dieser Schluß ist m. E. jedoch problematisch. Nehmen wir an, der Zusammenhang zwischen einem Desintegrationsfaktor A1 und einem Verunsicherungsindex B1 betrage $r = 0,3$, und der Zusammenhang zwischen dem Verunsicherungsindex B1 und einer bestimmten Form der Gewalttätigkeit C1 betrage ebenfalls $r = 0,3$, dann ist damit keineswegs nachgewiesen, daß die Desintegrationsfaktoren mit der Gewalttätigkeit zusammenhängen. Noch problematischer wird die Argumentation, wenn der Gewalttätigkeit als intervenierender Faktor Einstellungen zur Gewalt noch vorgeschaltet würden (Modell A \rightarrow B \rightarrow C \rightarrow D).

Besonders differenzierte Analysen führen HEITMEYER u. a. mit den sozialen Milieus als moderierender Variablen durch. Diese Analysen machen den kritischen Betrachter etwas ratlos (vgl. S. 242–265). Zu ihrer Methode schrieben die Autoren (S. 242): „Wir vermuten ..., daß spezielle Ursachenzusammenhänge für Gewalt *innerhalb* der einzelnen Milieus sich z. T. deutlicher zeigen oder auch in anderen Fällen unbedeutend werden können. Wir betrachten deshalb die sozialen Milieus als moderierende Variable, die Einfluß auf Ursachenzusammenhänge für Gewalt hat.“

In der Folge wird dann anhand vieler Tabellen demonstriert, daß die Zusammenhänge in den verschiedenen sozialen Milieus sehr unterschiedlich ausgeprägt sind. In Tabelle 9 (S. 243) wird z. B. gezeigt, daß die „Stärke des Zusammenhangs von Beziehungsqualität und Gewalteinstellungen ... in den verschiedenen Milieus im Bereich von $r = 0,03$ bis $r = 0,44$ [schwankt]. Während der Zusammenhang im Konservativen gehobenen Milieu so gering ist, daß nur 0,1% an gemeinsamer Varianz bezüglich beider Faktoren zu finden ist, beträgt diese im traditionellen Arbeitermilieu 19,5%, bei einer Korrelation von 0,44 ...“.

Interessant wird es nun, wenn man sich fragt, auf welche Faktoren diese unterschiedlichen Zusammenhänge zurückzuführen sind. Die Autoren führen aus (S. 244): „Zur Begründung der niedrigen Korrelation im Konservativen gehobenen Milieu ließe sich anführen, daß dieses Milieu milieuvvergleichend die höchste Beziehungsqualität aufweist und somit die Variation auf hohem Niveau stattfindet. Die Einstellungswerte zur Gewalt im Konservativen gehobenen Milieu sind hingegen eher gering ... Die beiden höchsten Korrelationswerte (Traditionelles Arbeitermilieu $r = 0,44$; Kleinbürgerliches Milieu $r = 0,34$) finden sich ferner in Milieus, die sowohl hinsichtlich der Beziehungsqualität als auch des Einstellungswertes zur Gewalt nicht besonders auffällig sind. Unsere Vermutung, daß diese hohen Korrelationen damit zusammenhängen, daß in ihren Milieus der Gemeinschaftssinn noch besonders erhofft wird und aus diesem Grunde ein mangelnder Rückhalt durch die Familie wie auch Freunde sich besonders negativ auf die Gewalteinstellungswerte auswirkt, liegt nahe. Diese Diskrepanz bildet gewissermaßen den Motor für einen starken Zusammenhang von Beziehungsqualität und Gewalteinstellungen.“

Für das Fehlen einer Korrelation werden statistische Gründe ins Feld geführt: Beide Variablen variieren deutlich weniger im Konservativen gehobenen Milieu; auf eine Mitteilung der Standardabweichungen dieser Variablen in den Mi-

lieus wird aber verzichtet. Auch wenn die Beziehungsqualität im Konservativen gehobenen Milieu am höchsten ist, so ist sie im Traditionellen Arbeitermilieu doch nur um 0.19 Punkte niedriger. Falls die Variable auf einer Skala von 1–4 erfaßt wird, ist diese Differenz sicherlich nicht sehr groß. Ob z. B. im Traditionellen Arbeitermilieu „der Gemeinschaftssinn noch besonders erhofft wird“, ist reine Spekulation, die das Fehlen einer überzeugenden Erklärung nur notdürftig überdeckt. Insgesamt bleiben die Ursachen für die vielen berichteten Zusammenhänge dem außenstehenden Betrachter reichlich unklar.

3. Abschließende Bemerkungen

Es ging in dieser Arbeit um eine überblicksartige Darstellung und Diskussion der wichtigsten Gütekriterien empirischer Untersuchungen, wobei möglichst interessante Beispiele zur Verdeutlichung der einzelnen Punkte dargestellt wurden. Eine Einarbeitung in die Anwendung der übergeordneten methodologischen Prinzipien ist für jeden Wissenschaftler erforderlich, weil häufig aufgrund ihrer Unkenntnis nicht nur schlechte empirische Forschung produziert wird, sondern vor allem schlechte Forschung als gute Forschung angesehen wird. Dies trägt dann ganz wesentlich dazu bei, daß Theorien, die z. B. in der Psychologie nicht mehr diskutabel sind, sich in der Pädagogik großer Beliebtheit erfreuen oder daß man seine praktische Arbeit an Theorien orientiert, die durch empirische Forschung längst als überholt anzusehen sind.

Wenn die Realität ein Prüfstein sein soll, dann muß durch entsprechende theoretische Arbeit das, was prüfenswert ist, entwickelt und dargestellt werden: Man muß die zu prüfenden Theorien nach rationalen Kriterien wie Informations- und Wahrheitsgehalt auswählen. Wenn man so tut, als ob mit jeder Untersuchung das Rad neu erfunden wird, muß man sich nicht wundern, wenn kein Erkenntnisfortschritt festgestellt werden kann. Das gleiche gilt, wenn die Ergebnisse einer Untersuchung nicht in zusammengefaßter Form für die Beurteilung des Wahrheitsgehalts der Hypothesen herangezogen werden können.

Die Darstellung legte besonderen Wert auf die theoretische und versuchsplanerische Güte einer Untersuchung; diese beiden Gütestandards werden m. E. in der empirisch-pädagogischen Forschung am ehesten grob vernachlässigt. Eine Untersuchung, die immer wieder als Beispiellieferant herangezogen wurde, war die neue Untersuchung von HEITMEYER u. a. (1995). Dies ist keineswegs die erste Kritik an den empirischen Arbeiten dieser Forschungsgruppe (vgl. HOPF 1989; PFAHL-TRAUGHBER 1993).

Es mag den Leser etwas überraschen, daß diese Arbeit der Gruppe um HEITMEYER unter theoretischen Gesichtspunkten im Hinblick auf die Art des Argumentierens und der Formulierung von Hypothesen am stärksten kritisiert wurde. Gerade in diesem Punkt halten sich viele deutsche Soziologen und Pädagogen z. B. den „amerikanischen Kollegen“ gegenüber für reflektierter, und genau dies ist der Punkt, wo eine Mythenbildung zur Sicherung des Status quo genutzt wird. Solange die einschlägigen amerikanischen Untersuchungen vor dem geistigen Auge als theoretisch unterbelichtet und empiristisch abgetan werden können, braucht man sich nicht mit ihnen auseinanderzusetzen. Noch schlimmer: Man kann dann im Schutze geringer Konkurrenz selbst sein eigenes

empiristisches Tagwerk absolvieren, dies mit Ad-hoc-Interpretationen und klugen Leerformeln garnieren und – weil die gläubige Gemeinde glaubt, relevante Forschung außerhalb des nationalen Horizonts existiere nicht – in Ruhe und unbeirrt seine Forschungen weiterführen.

Willkommen, Mr. Chance.

Literatur

- CAMPBELL, D.T./FISKE, D.W.: Convergent and discriminant validation by the multitrait-multimethod matrix. In: Psychological Bulletin 1959, 56, S. 81–105.
- CAMPBELL, D.T./STANLEY, J.C.: Experimental and quasi-experimental designs for research on teaching. In: N. L. GAGE (Hrsg.): Handbook of research on teaching. Chicago 1963.
- CAPALDI, D.M./PATTERSON, G.R.: Psychometric Properties of fourteen Latent Constructs from the Oregon Youth study. New York 1989.
- CHERLIN, A. J./FÜRSTENBERG, JR., F.F./CHASE-LANSDALE, L. P./KIERNAN, K. E./ROBINS, P.K./MORRISON, D. R./TEITLER, J. O.: Longitudinal Studies of Effects of Divorce on Children in Great Britain and the United States. In: Science, Bd. 252, Nr. 5002, 1991, S. 1386–1389.
- CONGER, R. D./PATTERSON, G. R./GE, X.: It Takes Two to Replicate: A Mediational Model for the Impact of Parent's Stress on Adolescent Adjustment. In: Child development, Bd. 66 (1995), S. 80–97.
- COOK, TH. D./CAMPBELL, T. T.: Quasi-Experimentation. Design & Analysis Issues for Field Settings. Boston 1979.
- CRONBACH, L. J./MEEHL, P. E.: Construct validity in psychological tests. In: Psychol. Bull. 52 (1955), S. 281–302.
- FATKE, R.: Fallstudien in der Pädagogik. Einführung in den Themenschwerpunkt. In: Zeitschrift für Pädagogik 41 (1995), S. 675–680.
- GADENNE, V.: Die Gültigkeit psychologischer Untersuchungen. Stuttgart 1976.
- GLUECK, S./GLUECK, E.: Unraveling Juvenile Delinquency. New York 1950.
- HAAPASALO, J./TREMBLEY, R. E.: Physically aggressive boys from ages 6 to 12: Family background, parenting behavior, and prediction of delinquency. In: Journal of consulting and clinical psychology, 1994, Bd. 62, Nr. 5, S. 1044–1052.
- HEITMEYER, W./COLLMANN, B./CONRADS, J./MATUSCHECK, I./KRAUL, D./KÜHNEL, W./MÖLLER, R./ULBRICH-HERMANN, M.: Gewalt. Schattenseiten der Individualisierung bei Jugendlichen aus unterschiedlichen Milieus. Weinheim 1995.
- HOPF, W.: Rezension von W. HEITMEYER: Rechtsextremistische Orientierungen bei Jugendlichen. Weinheim 1987. In: Deutsche Schule, Nr. 3 (1989), S. 402–404.
- LEHMEIER, W.: Benötigt die handlungswissenschaftlich begründete Pädagogik eine eigene Forschungsmethode? In: Zeitschrift für Pädagogik 41 (1995), S. 631–650.
- LOEBER, R./STOUTHAMER-LOEBER, M.: Family factors as correlates and predictors of juvenile delinquency. In: M. TONRY/N. MORRIS (Hrsg.): Crime and justice. An Annual Review of Research, Bd. 7. Chicago 1986, S. 29–149.
- PATTERSON, G. R.: Coercive family processes. Eugene (Oreg.) 1982.
- PATTERSON, G. R./REID, J. B./DISHION, TH. J.: Antisocial Boys. Eugene (Oreg.) 1992.
- PFAHL-TRAUGHBER, A.: Rechtsextremismus als Folge gesellschaftlicher Modernisierung? Rezension zu HEITMEYER u. a.: Die Bielefelder Rechtsextremismusstudie (1992). In: Politische Vierteljahresschrift 34 (1993), S. 114–118.
- QUINTON, D./RUTTER, M.: Parenting Breakdown. The Making and Breaking of Intergenerational Links. Aldershot 1988.
- RUTTER, M.: Beyond Longitudinal Data: Causes, Consequences, Changes and Continuity. In: Journal of Consulting and Clinical Psychology, Bd. 62, Nr. 5 (1994), S. 928–940.
- SAMPSON, R. J./LAUB, J. H.: Crime in the Making: Pathways and turning points through live. Cambridge, Mass. 1993.
- SCHNABEL, K.-U.: Ausländerfeindlichkeit bei Jugendlichen in Deutschland. Eine Synopse empirischer Befunde in Deutschland. In: Zeitschrift für Pädagogik 39 (1993), S. 798–822.

- WELLENREUTHER, M.: Handlungsforschung als naiver Empirismus. Für ein flexibles Modell theoriegeleiteter Handlungsforschung. In: Zeitschrift für Pädagogik 22 (1976), S. 343–356.
- WELLENREUTHER, M.: Grundkurs: Empirische Forschungsmethoden. Königstein i.Ts. 1982.
- WILLEMS, H.: Fremdenfeindliche Gewalt: Entwicklung, Strukturen, Eskalationsprozesse. In: Gruppendynamik 23 (1992), S. 433–448.

Abstract

The usefulness of strictly quantitative research is – due especially to deficiencies in methodological training – still questioned in educational science. Accordingly, empirical educational research often reveals considerable imperfections. This is illustrated by an analysis of a recent study on violence published by HEITMEYER et al. in 1995. The author criticizes the following aspects: (1) the theoretical quality, (2) the quality of the research design, (3) the quality of the operationalizations and measurements, and (4) the quality of the statistical evaluation. An empirical study is only really “good” if each one of these aspects is treated adequately. The description and discussion is restricted to the methodology of the testing of hypotheses, considering the problems of field research, in particular.

Anschrift des Autors

Dr. Martin Wellenreuther, Universität Lüneburg, Fachbereich Erziehungswissenschaften,
Institut für Pädagogik, Scharnhorststr. 1, 21335 Lüneburg