

Patry, Jean-Luc

Evaluationsstudien zu Forschungszwecken. Ein Beispiel von 'kritischem Multiplizismus'

Unterrichtswissenschaft 17 (1989) 4, S. 359-374



Quellenangabe/ Reference:

Patry, Jean-Luc: Evaluationsstudien zu Forschungszwecken. Ein Beispiel von 'kritischem Multiplizismus' - In: Unterrichtswissenschaft 17 (1989) 4, S. 359-374 - URN: urn:nbn:de:0111-pedocs-296532 - DOI: 10.25656/01:29653

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-296532>

<https://doi.org/10.25656/01:29653>

in Kooperation mit / in cooperation with:

BELIZ 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
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Digitalisiert

Unterrichtswissenschaft

Zeitschrift für Lernforschung
17. Jahrgang / Heft 4 / 1989

Thema:

Spielforschung und Unterricht

Verantwortlicher Herausgeber:
Wolfgang Einsiedler

- Wolfgang Einsiedler:
Spielforschung und Unterricht: Einführung 290
- Wolfgang Einsiedler:
Zum Verhältnis von Lernen im Spiel und intentionalen
Lehr-Lern-Prozessen 291
- Gerhard Treinies, Wolfgang Einsiedler:
Direkte und indirekte Wirkungen des Spielens im Kindergarten
auf Lernbegleitprozesse/Lernleistungen im 1. Schuljahr 309
- Detlev Leutner, Helmut Schrettenbrunner:
Entdeckendes Lernen in komplexen Realitätsbereichen:
Evaluation des Computer-Simulationsspiels
„Hunger in Nordafrika“ 327
- Detlev Leutner:
Angeleitetes Lernen mit Planspielen:
Lernerfolg in Abhängigkeit von Persönlichkeitseigenschaften
sowie Ausmaß und Zeitpunkt der Anleitung 342

Allgemeiner Teil

- Jean-Luc Patry:
Evaluationsstudien zu Forschungszwecken:
Ein Beispiel von „kritischem Multiplizismus“ 359

Berichte und Mitteilungen 375

Buchbesprechungen 379

Jean-Luc Patry

Evaluationsstudien zu Forschungszwecken: Ein Beispiel von „kritischem Multiplizismus“

Evaluation Studies for Research: An Example of Critical Multiplism

Gemäß den Prinzipien des „kritischen Multiplizismus“ empfiehlt es sich, bei wissenschaftlichen Untersuchungen jeweils von unterschiedlichen Standpunkten aus zu argumentieren. Insbesondere besagt der multiple Operationalismus, daß man wenn möglich jeweils unterschiedliche Operationalisierungen des gleichen Konstruktes in einer Untersuchung verwenden sollte. Es wird gezeigt, wie sich diese Forderung auf die Konzeption der unabhängigen und der abhängigen Variable auswirkt. Dabei wird auf Möglichkeiten hingewiesen, wie man die Konstrukt-Validität der unabhängigen Variable ermitteln kann. Sodann wird gezeigt, daß zwischen Forschungs- und Evaluationsstudien kein grundsätzlicher Unterschied besteht, sondern daß im Prinzip Evaluationsstudien auch zu Forschungszwecken eingesetzt werden können. Dies bedingt allerdings neben dem Einsatz einer wissenschaftlichen Methodologie, daß gewisse Kriterien erfüllt werden, welche im multiplen Operationalismus thematisiert werden. Auch hier steht wiederum die Treatment-Validität im Vordergrund. Falls diese Bedingung erfüllt ist, können Evaluationsstudien gegenüber den üblichen experimentellen Untersuchungen wesentliche Vorteile aufweisen.

According to the criteria of critical multiplism, it is recommended to use differing approaches to answer research questions. Thus, if possible, one should use different operationalizations of the same constructs (multiple operationalism). This has certain consequences for the conception of both the independent as well as the dependent variable. In particular, the construct validity of the independent variable needs to be assessed. In a second part, it is shown that there is no fundamental difference between evaluation and research, so that in principle, evaluation studies can be used to answer research questions. However, this requires a scientific methodology and the fulfillment of certain criteria which are given by multiple operationalism. Again, treatment validity is of central importance. Evaluation studies can have advantages over the traditional experimental studies if these requirements are satisfied.

Im Rahmen der Erprobung, ob gewisse Handlungsprinzipien oder (Lehr-) Methoden praktisch verwendbar sind und zu den erhofften praktischen Resultaten führen, werden immer wieder Evaluationsstudien durchgeführt: Es wird beispielsweise in einem Schulversuch mehr oder weniger systematisch erfaßt, ob sich die dabei geprüfte Methode bewährt; ist dies der Fall, wird man allenfalls diese Methode weiterhin einsetzen (summative Evaluation). Im weiteren soll in der Weiterführung eines früheren Beitrages (Oser & Patry 1986) gezeigt werden, daß und wie summative Evaluationsstudien zu Forschungszwecken verwendet werden können, falls man von einem kritischen Multiplizismus (Cook 1985; Houts, Cook & Shadish 1986) ausgeht und einige damit zusammenhängende Prinzipien berücksichtigt.

1. Der kritische Multiplizismus

Der kritische Multiplizismus ist ein Ansatz, der verwendet wird, um wichtige Kritiken wissenschaftstheoretischer Art an der gängigen Forschungspraxis zu entkräften, dabei aber wichtige methodologische Erkenntnisse weiterhin zu berücksichtigen (*Cook 1985; Houts et al. 1986, p. 53*); es handelt sich um eine Erweiterung des Ansatzes der Triangulation (*Campbell 1966; Campbell & O'Connell 1982; Crano 1981; Fiske 1982*), der darin besteht, zur Beantwortung einer Fragestellung jeweils mehrere Ansätze (auf Methoden- wie auf Theorieebene) zu verwenden. *Houts et al. (1985)* nennen u.a. die folgenden Beispiele für kritischen Multiplizismus:

- Multipler Operationalismus und multiple Methoden (*Campbell & Fiske 1959, vgl. unten*);
- Untersuchung von komplexen multivariaten statt univariaten Modellen (dies bezieht sich auf die zu untersuchenden Konstrukte);
- Prüfung von multiplen rivalisierenden Hypothesen (*Campbell 1969*);
- Verwendung von multiplen, miteinander verbundenen Untersuchungen (Forschungsprogramme, Untersuchungsstrategien mit verschiedenen und verschieden gestalteten empirischen Erhebungen);
- Synthese der Resultate von multiplen, aber unabhängigen Studien (z.B. mittels Meta-Analysen);
- Analyse der gleichen Daten mit unterschiedlichen statistischen Methoden;
- Verwendung unterschiedlicher Theorien, um die gleichen Forschungsfragen hypothetisch zu beantworten und Ergebnisse zu interpretieren (Theorienkonkurrenz).

Im folgenden soll der multiple *Operationalismus* eingehender diskutiert werden ; insbesondere soll analysiert werden, inwiefern er einen Beitrag zur Formulierung von Evaluationsstudien zu Forschungszwecken leisten kann. Dabei werden die anderen multiplen Ansätze als Grundlage für die Arbeiten akzeptiert, aber nicht weiter besprochen.

2. Konstrukte und deren Operationalisationen

Die Besprechung des multiplen Operationalismus setzt gewisse Überlegungen in bezug auf die Rolle und den Stellenwert von Operationalisationen in wissenschaftlichen Untersuchungen voraus. Es geht in diesem Abschnitt darum, ein paar Prinzipien darzustellen, auf die unten zurückgegriffen wird.

Ein Ziel der *Forschungs-Untersuchungen* ist die Generierung oder Überprüfung von „nomologischem Wissen“ (*Perrez & Patry 1982*); dieses steht im Gegensatz zum „technologischen Wissen“, das durch *Evaluationsstudien* angestrebt wird (vgl. dazu unten). In einer Forschungs-Untersuchung soll festgestellt werden, ob bestimmte Aussagen wahr sind. Diese zu prüfenden Aussagen thematisieren in der

Regel Beziehungen zwischen *Konstrukten* (vgl. die Ebene „Konstrukt“ in Abb. 1); typischerweise handelt es sich dabei einerseits um ein Konstrukt der UV (K_{UV} ; Wenn-Komponente in der Aussage) und andererseits um ein Konstrukt der AV (K_{AV} ; Dann-Komponente). Die Untersuchung der Beziehung zwischen den Konstrukten kann jedoch nicht direkt geschehen, sondern diese müssen in irgend einer Art operationalisiert werden (Ebene „Operationalisation“ in der Abbildung).

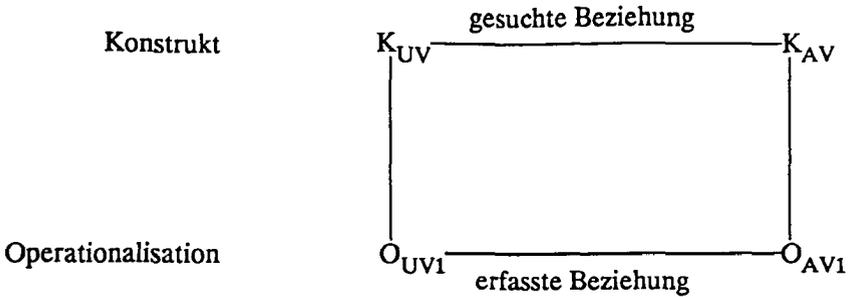


Abbildung 1: Beziehung zwischen Konstrukten und deren Operationalisationen in einfachen experimentellen Designs

Im folgenden wird davon ausgegangen, daß es in bezug auf den Zusammenhang zwischen Konstrukten und deren Operationalisation keinen grundsätzlichen Unterschied zwischen der UV und der AV gibt; in der Literatur werden die nachstehend zu diskutierenden Fragen allerdings fast ausschließlich unter Bezugnahme auf die AV besprochen. Es gibt jedoch keinen Grund, die entsprechenden Überlegungen nicht auch auf die UV anzuwenden; vielmehr spricht der Umstand, daß auch das Treatment als Operationalisation eines Konstruktes zu interpretieren ist, dafür, die sich aus der Diskussion der entsprechenden Probleme bei der AV ergebenden theoretischen Schlußfolgerungen auch auf die UV anzuwenden (vgl. auch *Oser & Patry* 1986).

Die methodische Umsetzung dieser theoretischen Überlegungen kann nicht für beide Variablentypen identisch sein (bei der UV geht es um Treatments, also um Handlungen des Versuchsleiters und um Stimuluskonstellationen, etc., bei der AV um Messungen, d.h. um Handlungen der Versuchspersonen). Im weiteren wird deshalb nur dann zwischen UV und AV unterschieden, wenn es um methodische Fragen (z.B. Wie kann ich Konstruktvalidität ermitteln?) oder um die Beziehung zwischen den beiden Variablen (UV als Antezedens, AV als Konsequens) geht.

Das einfachste Untersuchungsdesign, welches gültige Aussagen über die Beziehungen zwischen diesen Operationalisationen erlaubt (interne Validität), ist das experimentelle Design nach *Campbell & Stanley* (1963), dessen Prinzip in Abb. 1 dargestellt ist; dieses Design wird in den Sozialwissenschaften wohl auch am häufigsten verwendet. Gemäß diesem Design wird eine UV realisiert (darin unterscheiden sich Versuchs- und Kontrollgruppe), und es wird eine AV gemessen. Zeigen sich darin erwartungsgemäße Unterschiede in der AV zwischen Versuchs- und Kontrollgruppe, dann schließt man auf eine „Wenn-Dann-Beziehung“ auf Konstruktebene zwischen diesen Variablen für die untersuchte Population in der untersuchten Situation.

Zu jedem der oben erwähnten Konstrukte kann es jedoch mehrere Operationalisationen geben; dies wird mit dem Index „i“ in der Bezeichnung O_{UVi} für die UV und mit dem Index „j“ in O_{AVj} für die AV angedeutet. In einer empirischen Untersuchung wird dann festgestellt, ob eine Beziehung zwischen *diesen* Operationalisationen O_{UVi} und O_{AVj} besteht.

Im dargestellten einfachsten Design werden für UV und AV je *eine* dieser Operationalisationen gewählt: O_{UV1} und O_{AV1} . Diese Operationalisationen können nicht als repräsentativ für alle möglichen jeweiligen Operationalisationen betrachtet werden; man kann also nicht ohne weiteres von O_{UV1} auf das zugrundeliegende Konstrukt K_{UV} und von O_{AV1} auf K_{AV} schließen. Insbesondere sind dabei zwei Aspekte zu berücksichtigen:

(i) Jedes Konstrukt besteht aus einer Menge von relevanten Komponenten, die in einer theoretisch begründeten Weise zusammenhängen. Ferner weisen Konstrukte eine „über das Beobachtbare hinausreichende ‚Überschußbedeutung‘“ auf (*Herrmann 1973, S. 134*), d.h. sie werden (als AV) „durch Beobachtungen nur teilweise interpretiert“; analog kann man für die der UV zugrundeliegenden Konstrukte sagen, daß sie „Überschußbedeutung“ aufweisen, die über das in einem Treatment realisierbare hinausreicht.

In einer einzelnen Operationalisation eines Konstruktes können höchstens ein paar der entsprechenden Komponenten berücksichtigt werden, während andere Komponenten nicht enthalten sind; *alle* Komponenten eines Konstruktes können nie in einer seiner Operationalisationen vorkommen. Ein Meßinstrument zur Erfassung eines Konstruktes — dessen Operationalisation — wird sich zwangsläufig auf einige wenige der oben genannten Komponenten beschränken und andere nicht berücksichtigen, die ebenfalls als wichtig zu betrachten sind; unterschiedliche Instrumente können aber allenfalls unterschiedliche Komponenten des gleichen Konstruktes erfassen.

(ii) Hingegen kommt in jeder Operationalisation etwas *hinzu*, das für das entsprechende Konstrukt nicht relevant ist. Beispielsweise wird ein Fragebogen eingesetzt; das Beantworten von Fragen eines Fragebogens ist jedoch etwas, das im zugrundeliegenden Konstrukt *nicht* repräsentiert ist: Es setzt u.a. eine Fülle von kognitiven Aktivitäten voraus, die nichts mit dem interessierenden Konstrukt, sondern nur mit den Möglichkeiten, dieses zu erfassen, zu tun haben. Analog kann man für jedes andere Meßinstrument, aber auch für die Operationalisationen der UV argumentieren. Sind die Ergebnisse einer Untersuchung vor allem auf Einflüsse dieser nicht im Konstrukt enthaltenen Aspekte zurückzuführen, spricht man von Artefakten. Man wird sich natürlich bemühen, diese zu minimieren bzw. zu kontrollieren — falls man sie erkennt! —, hingegen lassen sie sich nicht eliminieren, im Gegenteil hat sich gezeigt, daß sie von großer und meist unterschätzter Bedeutung sind (vgl. z.B. *Bungard & Bay 1982; Fiske 1982; Patry 1987a*).

Deshalb stellt sich die Frage, ob es überhaupt möglich ist, Zusammenhänge zwischen Konstrukten aufgrund der Zusammenhänge zwischen deren Operationalisationen zu erschließen. Für das oben angedeutete einfache Design (Abb. 1) lautet die Antwort zweifellos nein. Um eine einigermaßen befriedigende Antwort auf diese Frage zu erhalten, wird es notwendig sein, von der oben dargestellten einfachen experimentellen Untersuchung mit einer UV (O_{UV1}) und einer AV (O_{AV1}) Abschied zu nehmen und stattdessen komplexere, „multiple“ Ansätze zu verwenden.

3. Multiple Konstrukte, multiple Operationalisationen

Der für die weitere Diskussion wichtigste dieser multiplen Ansätze, der multiple Operationalismus, besteht zunächst darin, nicht nur wie in Abb. 1 eine Operationalisation des Konstruktes der UV und eine Operationalisation für die AV zu verwenden, sondern jeweils mehrere Operationalisationen; ferner kann man nicht nur je ein Konstrukt der UV und der AV, sondern mehrere Konstrukte untersuchen (Abb. 2), wobei Hypothesen bezüglich der Gleichheit und der Unterschiedlichkeit der Resultate aufgestellt werden. Die Ergebnisse sollten bei Benützung gleicher Konstrukte (aber allenfalls unterschiedlicher Operationalisationen) gleich sein; dies wird „konvergente Validität“ genannt (k in Abb. 2). Beim Vergleich zwischen unterschiedlichen Konstrukten hingegen sollten sich Unterschiede nachweisen lassen (diskriminante Validität, d in Abb. 2). Der Ansatz (und die Bezeichnung „konvergent“ und „diskriminant“) geht letztlich auf den Vorschlag von *Campbell & Fiske* (1959) für die Validierung der AV zurück.

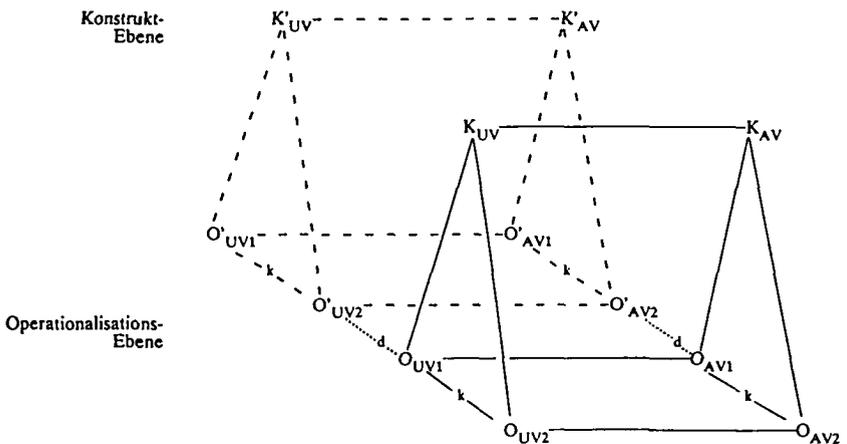


Abbildung 2: Mehrere Konstrukte, mehrere Operationalisationen: Beziehungen zwischen den Konstrukten und deren Operationalisationen
k: konvergente Validitäten; d: diskriminante Validitäten (Auswahl)

Die Bezeichnung „kritisch“ im kritischen Multiplizismus bedeutet hier, daß der Forscher bei der Auswahl der Konstrukte und Operationalisationen kritisch vorgeht. Für die letzteren gelten insbesondere die folgenden Kriterien, die auf die in Abschnitt 2 dargestellte Unterscheidung zwischen der Operationalisationen einiger (nie aller) Komponenten aus dem Konstrukt und dem notwendigen Hinzukommen konstruktirrelevanter Aspekte bezug nimmt:

- Jede Operationalisation sollte eine *plausible Konkretisierung* des thematisierten Konstruktes sein.
- Die Operationalisationen sollten theoretisch relevante Komponenten enthalten, die *nicht in den anderen Operationalisationen sind*.
- Die Operationalisationen sollten sich in bezug auf *theoretisch irrelevante Aspekte unterscheiden*.
- *Kein theoretisch irrelevanter Aspekt* sollte bei *allen* Operationalisationen des gleichen Konstruktes vorhanden sein.

Die multitrait-multimethod-matrix von *Campbell & Fiske* (1959) — den ersten Vertretern eines kritischen Multiplizismus — ist ein Versuch, den multiplen Operationalismus auf die *AV* anzuwenden. Diese Methode soll hier nicht weiter besprochen werden. Der multitrait-multimethod-Ansatz kann aber auch der Ermittlung der konvergenten und diskriminanten Validitäten der Operationalisationen der *UV* (von K_{UV} und von K'_{UV}) zugrundegelegt werden (Treatment-Validität). Die Treatment-Validität wird dann als Konstruktvalidität analog der Konstruktvalidität der *AV* konzipiert; dabei wird das oben angedeutete multitrait-multimethod-Konzept auf das Treatment angewandt, indem relevante Aspekte der Handlung, die als Treatment interpretiert wird, gemessen werden.

Betrachten wir den *einfachsten Fall*, wo nur ein Konstrukt, K_{UV} untersucht werden soll, das in nur einer Art und Weise in Handlung umgesetzt wird (nur eine Operationalisation, O_{UV1}). Multiplizismus besteht dann nur dahingehend, daß mehrere Konstrukte und Methoden zur Analyse der Konstruktvalidität dieses Treatments verwendet werden: O_{UV1} kann nämlich prinzipiell nicht nur eine Operationalisation dieses Konstruktes K_{UV} sein, d.h. Komponenten dieses Konstruktes enthalten (die in der Versuchs-, nicht aber in der Kontrollgruppe vorhanden sind), sondern man kann sich auch ein anderes Konstrukt K'_{UV} vorstellen, von dem wesentliche Komponenten in O_{UV1} enthalten sein könnten. Es gilt nun, folgendes festzustellen:

- Unterscheiden sich Versuchs- und Kontrollgruppe beim Einsatz von O_{UV1} in bezug auf Komponenten, die für K_{UV} wesentlich sind, d.h. hat man das variiert, was man variieren wollte (konvergente Validität)?
- Unterscheiden sich Versuchs- und Kontrollgruppe *nicht* im Hinblick auf Komponenten aus K'_{UV} , die nicht gleichzeitig Komponenten von K_{UV} sind, d.h. wurden im Treatment Aspekt *nicht* variiert, die *nicht* variiert werden sollten (diskriminante Validität)?

Zunächst muß das *Alternativkonstrukt* K'_{UV} bestimmt werden, in bezug auf das theoretisch zwischen Versuchs- und Kontrollgruppe keine Unterschiede bestehen sollen. Sinnvollerweise wird man ein solches Konstrukt wählen, das dem interessierenden Konstrukt K_{UV} genügend fern ist, daß theoretisch in bezug auf die interessierenden Komponenten keine Überschneidungen bestehen. Es ist jedoch auch so nah zu K_{UV} zu wählen, daß die Gefahr besteht, daß das Treatment unerwünschterweise Komponenten dieses Konstrukts enthält. Je näher sich die beiden Konstrukte sind, desto konservativer wird die Überprüfung sein.

Als nächstes werden *Meßinstrumente* zur (validen) Erfassung der interessierenden Komponenten der beiden Konstrukte K_{UV} und K'_{UV} angewandt, und es wird untersucht, ob sich die Versuchs- und die Kontrollgruppe in bezug auf diese Komponenten unterscheiden. Unterscheiden sich die beiden Gruppen in bezug auf Komponenten aus K_{UV} , dann besteht konvergente Validität, unterscheiden sie sich in bezug auf Komponenten aus K'_{UV} , besteht *keine* diskriminante Validität.

Im Sinne des kritischen Multiplizismus empfiehlt es sich, nicht nur ein Treatment durchzuführen, das sich auf das Konstrukt K_{UV} bezieht, sondern zusätzlich noch eines zum Konstrukt K'_{UV} . Es wären somit je eine Versuchsgruppe mit einer entsprechenden Intervention (K_{UV} bzw. K'_{UV}) sowie — nach Möglichkeit — eine Versuchsgruppe mit einer kombinierten Intervention (K_{UV} und K'_{UV}) zusätzlich zur Kontrollgruppe zu realisieren. Da solche Designs in Evaluationsstudien kaum vorkommen, soll dieser Ansatz hier nicht weiter entwickelt werden; eine Analyse der sich dabei ergebenden Prinzipien und Probleme ist in Arbeit (Patry et al. in Vorb.).

Die angedeuteten Prinzipien der Treatment-Validität sind bisher in der Literatur nur an ganz wenigen Stellen (etwa bei *Cook & Campbell* 1979) erwähnt worden, und auch dort nur im Sinne eines Desideratum und nicht dergestalt, daß man sie unmittelbar anwenden könnte. Ein paar Hinweise wurden in *Oser & Patry* (1986) gegeben.

4. Unterschiede zwischen Evaluation und Forschung

Es stellt sich nun die Frage, ob und wie weit Evaluationsstudien zu Forschungszwecken eingesetzt werden können. Um dies diskutieren zu können, müssen zunächst die Unterschiede zwischen „Evaluation“ und (empirischer) „Forschung“ geklärt werden. Die wichtigsten Unterschiede sind in der Tabelle 1 zusammengestellt.

„Evaluation“ kann mit *Wolf* (1987) im Anschluß an Beeby definiert werden als

„the systematic collection and interpretation of evidence, leading, as part of the process, to a judgment of value with a view to action“ (p. 8).

Mittels Evaluationsstudien wird sog. technologisches Wissen untersucht, d.h. es werden Aussagen darüber gemacht, wie (insbesondere mit welchen Handlungen) man bestimmte Ziele erreichen kann und ob Nebeneffekte zu befürchten sind; in der Regel beschränken sich Evaluationsstudien auf die mehr oder weniger systematische

Tabelle 1:

Die wichtigsten Unterschiede zwischen Evaluation und Forschung

	Evaluationsstudien	Forschungsstudien
untersuchtes Wissen	technologisches Wissen Mittel-Ziel-Beziehungen	nomologisches Wissen Wenn-Dann-Aussagen
Untersuchungsziel	Anwendbarkeit	Erkenntnisgewinn
praktisch brauchbare Ergebnisse	ja	möglich
Werturteil	ja: Mittel ist geeignet für das Ziel; keine Werturteile über Ziele	nein: nur deskriptive Aussagen
Handlungsorientierung	Handlungen thematisiert	Konstrukte thematisiert
Theoriebezug	nicht notwendig, aber möglich	notwendig
<u>Design:</u>		
interne Validität	notwendig	notwendig
angestrebte Generalisation	meist auf wenig oder keine anderen Kontexte aber mehr ist möglich	meist auf viele andere Kontexte
ökologische Validität	notwendig	nicht notwendig, aber möglich
<u>abhängige Variable:</u>		
Reliabilität der Messinstrumente	notwendig	notwendig
Konstruktvalidität der Messinstrumente	nicht notwendig, aber möglich	notwendig
<u>unabhängige Variable:</u>		
Reliabilität des Treatments	bezogen auf relevante Merkmale notwendig; selten erhoben	notwendig (Durchführungs-Objektivität); selten erhoben
Heterogenität des Treatments	meist heterogen	Homogenität angestrebt
Konstrukt-Validität des Treatments	nicht notwendig, aber möglich	notwendig, aber selten erhoben

(wissenschaftliche) Gewinnung solchen Wissens, weitergehende Erkenntnisse (wie in der Forschung) werden nicht angestrebt. Die Verwendung des Prinzips des „kritischen Multiplizismus“ drängt sich von daher nicht auf: Es genügt, wenn die Aussagen auf der Ebene „Operationalisation“ (Abbildung 1) gemacht werden, zumal die Operationalisation der unabhängigen Variabel (O_{UVI}) eine Handlung sein kann.

Wie bereits angedeutet, ist Forschung demgegenüber auf *Erkenntnisgewinn* ausgerichtete wissenschaftliche Arbeit. Forschung zielt auf nomologisches Wissen ab, d.h. auf bewährte Aussagen, die u.a. eine gewisse Universalität (Generalisierbarkeit) und Systemzugehörigkeit (Theorienbezug) aufweisen (vgl. *Perrez & Patry* 1982, S. 45 ff.). Dabei ist es durchaus möglich — wenn auch nicht ausgesprochen häufig —, daß Forschungsprojekte praktisch relevante Ergebnisse liefern.

Aus der o.g. Definition von „Evaluation“ ergeben sich einige Unterschiede zum nomologischen (Forschungs-) Ansatz, die sich darauf beziehen, daß Evaluation wert- und handlungsorientiert ist, während Forschung deskriptive Aussagen anstrebt. Hinzu kommt das Problem der Generalisierbarkeit, das bei der Beschreibung der Forschung angedeutet wurde („Universalität“). Diese Unterschiede sind auf den ersten Blick grundsätzlicher Art, ihre Analyse aber ergibt, daß es sich zumindest teilweise eher um unterschiedliche Gewichtungen handelt.

Nach Wolf's Definition sind die Aussagen, die aufgrund von Evaluationsstudien gemacht werden, nicht nur deskriptiv, sondern enthalten — im Gegensatz zu Forschungsergebnissen — auch *Werturteile* bzw. *Normen*: Man soll beispielsweise den Wert einer erzieherischen Unternehmung beurteilen. „Wert“ bezieht sich dabei darauf, ob und wie weit die Intervention geeignet ist, bestimmte Zielzustände zu erreichen. Nun ist aber darauf hinzuweisen, daß es sich bei dieser Wertung um ein bedingtes Urteil handelt. Es ist logisch unmöglich, aus der technologischen, mittels einer Evaluationsstudie festgestellten Aussage „Um z (das Ziel) zu erreichen, kann man x tun“ unbedingte Normen etwa vom Typ „x soll getan werden“ abzuleiten, ohne den naturalistischen Fehlschluß zu begehen, wenn nicht vorher andere Normen (z.B. „Es ist gesollt, z zu erreichen“) gesetzt werden. Deshalb besteht kein Grund, Evaluationsstudien wegen der dort thematisierten Wertungen („values“) von Forschungsstudien abzugrenzen — eine solche Abgrenzung würde sogar implizieren, daß in Evaluationsstudien Wertungen ermittelt werden, was nicht möglich ist. In Evaluationsstudien stehen *Handlungen* im Vordergrund: Es werden primär Informationen darüber gewonnen, was man *unternehmen* kann, um die angedeuteten Werte zu erreichen. Die Erkenntnisse von Forschungsarbeiten sind demgegenüber nicht auf praktische Handlungen, sondern auf Theorien, deren Prüfung und allenfalls Verbesserung ausgerichtet, die in Evaluationsstudien zweitrangig sind. Nun sind aber in vielen Forschungsstudien — vor allem in experimentellen Untersuchungen — die operationalisierten UV (O_{UVj}) ebenfalls Handlungen. Es gibt also vom *Design* her keinen grundsätzlichen Unterschied, der bei Evaluationsstudien Aussagen vom „wissenschaftlichen Typ“ (sog. nomologische Aussagen) verhindern würde, vielmehr sind die Interpretationen allenfalls unterschiedlich.

Eine typische Evaluationsstudie stellt auf der Ebene der Operationalisierung (vgl. Abb. 1) genau die gleiche Frage wie eine Forschungsstudie, der einzige Unterschied besteht

darin, daß in der Forschungsstudie die Handlung (und die AV) mit Konstrukten in Beziehung gesetzt werden, was in Evaluationsstudien nicht notwendigerweise der Fall ist, aber durchaus realisiert werden kann.

Die beabsichtigte *Generalisation* ist ein weiterer zentraler Unterscheidungspunkt. So sagt *Wolf* (1987, p. 12), Evaluation suche nach Wissen, das auf ein bestimmtes Setting beschränkt sei; Generalisierbarkeit ist demgegenüber ein zentrales Ziel wissenschaftlicher Forschung. Dem ist zu entgegnen, daß auch in Evaluationsstudien nach generalisierbaren Erkenntnissen gesucht wird: Letztlich wird die Übertragung auf andere — wenn auch allenfalls ähnliche — Interventionsanwendungen angestrebt. So kritisiert *Müller* (1987, S. 210) rein prozeßorientierte Evaluationsansätze (formative Evaluation) dafür, daß deren Resultate sich nicht auf andere Interventionsbedingungen übertragen lassen; er weist auch darauf hin, daß Vorbehalte gegen die sog. Handlungsforschung gerade in dieser Hinsicht angebracht sind (*Zecha & Lukesch* 1982). Evaluationsuntersuchungen, die nicht u.a. auf Generalisierbarkeit hin angelegt sind, werden somit kritisiert.

Umgekehrt ist die Praxisrelevanz der *Forschung* oft unter anderem auch deshalb gering, weil von diesen Untersuchungen nicht auf die potentielle Anwendungssituation generalisiert werden kann: Es fehlt an ökologischer Validität (*Patry*, in Druck). Gerade diese ökologische Validität aber ist für Evaluationsstudien notwendigerweise gegeben. Zweifellos ist ökologische Validität nicht das einzige Validitäts-Ziel von Forschung; hingegen ist es für viele empirisch-pädagogische Forschungen ein wesentliches Ziel (*Oerter* 1979), und darin sind Evaluationsstudien vielen Forschungsstudien überlegen.

Ein weiterer wichtiger Aspekt bezieht sich nicht auf grundsätzliche, sondern auf faktische Unterschiede zwischen Evaluations- und Forschungsuntersuchungen: In den letzteren wird meist versucht, mit entsprechenden Maßnahmen *homogene* Untersuchungsbedingungen zu realisieren, d.h. die UV soll sich von Versuchsdurchgang zu Versuchsdurchgang möglichst wenig ändern (hohe Durchführungsobjektivität, vgl. *Lienert* 1969, S. 13). Demgegenüber sind die Realisationsbedingungen für das Treatment in Evaluationsstudien meist sehr heterogen, da dem Praktiker nicht bis ins Detail vorgeschrieben werden kann, wie er handeln soll.

Insgesamt kommen wir somit zum Schluß, daß es zwischen „Evaluation“ und „Forschung“ wesentliche Unterschiede gibt (vgl. Tab. 1). Unter bestimmten Bedingungen können aber auch Untersuchungen, die als Evaluationsstudien konzipiert sind, Erkenntnisse von gleichem Wert wie bei „Forschungsuntersuchungen“ ermöglichen (vgl. auch *Patry* 1987b).

5. Evaluationsstudien für die Forschung

Auf dem Hintergrund des kritischen Multiplizismus können diese Bedingungen präzisiert werden. Es geht um die Identifikation einiger

Grundlagen, welche es erlauben, Evaluationsstudien *auch* als Forschungsstudien zu verwenden und damit ein bislang weitgehend ungenutztes und dabei sehr reiches Material der Forschung zugänglich zu machen.

Die nachstehenden Ausführungen sollen dabei nicht als Kritik an der praktizierten Evaluationsforschung verstanden werden, soweit sich diese auf angemessene Art und Weise (vgl. unten) auf die Gewinnung von technologischem Wissen beschränkt.

Zunächst gilt generell, daß nach Möglichkeit die Prinzipien wissenschaftlicher Forschung in Planung und Durchführung der Untersuchung in Forschungen wie in Evaluationen einzuhalten sind.

„Systematisch“ in der Definition von „Evaluation“ nach Wolf (1987; vgl. oben, Abschnitt 4) bedeutet ja nichts anderes, als daß die Untersuchung nach wissenschaftlichen Prinzipien erfolgt; diese Prinzipien brauchen nicht notwendigerweise den strengen Grundsätzen einer Laborforschung zu entsprechen (zu den entsprechenden Problemen vgl. auch Müller 1987), sondern können im Sinne der Feldforschung als ein machbares Optimum gesehen werden: Man hält sich soweit möglich an die strengen Grundsätze, ohne aber die Vorteile der Feldforschung aufzugeben. Zwar mag gelten, daß „(m)eaningful evaluative activity can be carried on that does not follow a research model“ (Wolf 1987, p. 12f.); will man aber gleichzeitig Forschung betreiben, muß eine wissenschaftliche Methodologie verwendet werden.

Insbesondere sind die folgenden in der Literatur häufig diskutierten Kriterien zu erfüllen:

- *Das Design muß intern valide sein, d.h. eine experimentelle, allenfalls quasi-experimentelle Versuchsanordnung (Campbell & Stanley 1963) ist zu realisieren: Es ist in jedem Falle mit Kontrollgruppen zu arbeiten; nach Müller (1987) gilt diese Forderung für Evaluationsstudien generell.*
- *Die Artefaktproblematik ist angemessen zu berücksichtigen (Rosenthal & Rosnow 1969; Bungard & Bay 1982).*
- *Die Messinstrumente müssen valide sein (Lienert 1969).*

Sind diese Bedingungen — auf die oben z.T. eingegangen wurde — erfüllt, kann man zusätzlich extern valide Designs i.S. von *Campbell & Stanley* (1963) konzipieren, welche also auf Generalisierbarkeit hin angelegt sind (vgl. oben); die drei genannten Kriterien haben jedoch Priorität, weil eine Untersuchung nur dann extern valide sein kann, wenn sie auch interne Validität aufweist: Es ist nicht sinnvoll, ungültige Aussagen oder solche, von denen man nicht weiß, ob sie gültig sind, generalisieren zu wollen. Die o.g. Bedingungen sind nicht in allen Evaluationsstudien erfüllt, aber sie sind grundsätzlich erfüllbar.

Hinzu kommen weitere Bereiche, die bei Forschungsuntersuchungen meistens unproblematisch sind — obwohl diesbezüglich auch in der Forschung sehr häufig gesündigt wird —, die aber bei der Verwendung von Evaluationsstudien zu Forschungszwecken besonders zu berücksichtigen sind. Im folgenden seien zwei solcher Bereiche und die damit zusammenhängenden Probleme diskutiert: zum einen die Beziehung der Untersuchung zur Theorie, denn dies ist der wichtigste konzeptuelle Unterschied zwischen Evaluations- und Forschungsstudien, sodann die Heterogenität, welche einen der wichtigsten problemgeladenen Unterschiede in der Realisation der beiden Untersuchungstypen darstellt.

5.1 Beziehung zur Theorie

Für die *Forschung* ist nach den oben genannten Kriterien die Beziehung der Untersuchung (bzw. der verwendeten Variablen) zu einer *Theorie* grundlegend: Es werden theoretische Aussagen angestrebt (Abb. 1: theoretische Ebene). Entsprechend werden in der Forschung die UV wie die AV so gestaltet, daß sie (nach Ansicht des Forschers) jeweils einem *Konstrukt* entsprechen. Die tatsächliche Handlung O_{UVi} (die UV) und die tatsächlich beobachtete AV O_{AVj} sind nur interessant, insofern sie Operationalisationen dieser Konstrukte sind. Dies führt zur Forderung der *Konstrukt-Validität* sowohl der UV als auch der AV (vgl. Abschnitt 3). Im weiteren beschränke ich mich auf die Diskussion der Treatment-Validität, weil die Validität der AV in der Literatur genügend thematisiert ist — auch auf dem Hintergrund des kritischen Multiplizismus (Campbell & Fiske 1959) — und hier nicht noch einmal aufgegriffen zu werden braucht.

In einer Forschungsstudie (mit Ausnahme der Studien mit sog. „natürlichem Treatment“) ist der Zusammenhang zwischen dem Konstrukt K_{UV} und dessen Operationalisation O_{UVj} meist unmittelbar gegeben, weil der Forscher das Treatment nach Bedarf gestalten, also auf das Konstrukt zuschneiden kann. In einer Evaluationsstudie, die zu Forschungszwecken eingesetzt wird, ist das Treatment jedoch aus anderen, pragmatischen Gründen gewählt worden: Es sind diejenigen Handlungen, von denen man am ehesten erwartet, daß sie den entsprechenden Erfolg zeitigen. Ob eine Beziehung zu einem Konstrukt K_{UV} oder zu einer Theorie besteht, ist nebensächlich. Deshalb können forschungsäquivalente Aussagen nur gemacht werden, wenn nachgewiesen wurde, daß die UV eine valide Operationalisation des interessierenden Konstruktes ist: Die Konstrukt-Validität des Treatments ist zu untersuchen.

Es sei darauf hingewiesen, daß es auch in Forschungsstudien sinnvoll ist, diesen Nachweis zu erbringen — und dies nicht nur bei Benützung natürlicher Treatments: oft hat man den Eindruck, in einem Treatment würde etwas ganz anderes variiert, als was zu variieren beansprucht wird (vgl. auch Cook & Campbell 1979, S. 38f.). Die Forderung des Nachweises der Treatment-Validität als Ergänzung zu den oben erwähnten Validitäten des Meßinstrumentes und des Designs (interne und externe Validität) scheint deshalb auch in echten Forschungsstudien sinnvoll zu sein; für den Nachweis der Notwendigkeit in konkreten Untersuchungen vgl. Oser & Patry (1986).

Wie ein solcher Nachweis geschehen kann, wurde oben ausgehend vom kritischen Multiplizismus angedeutet (Abschnitt 3). Bei Zugrundelegung einer Theorie und Sicherstellung der Beziehungen zwischen den verschiedenen untersuchten Variablen und der Theorie (via Prüfung der Konstruktvaliditäten, vor allem auch bei der UV) sowie unter Verwendung optimaler Untersuchungsdesigns (möglichst experimentell, etc.) können somit Evaluationsstudien auch zu Forschungszwecken eingesetzt werden; die wiederholt angedeuteten Vorteile können also ausgenützt werden, um Wissen zu gewinnen.

5.2 Das Problem der Heterogenität

Als Beispiel für eine Evaluationsstudie sei die Untersuchung der Wirkung einer bestimmten Lehrmethode verwendet, die in vielen Klassen mit unterschiedlichen Lehrern erprobt werden soll. Dies führt dazu, daß das Treatment heterogen ist: Die zu prüfende Methode kann nicht von allen Lehrkräften in der genau identischen Art und Weise realisiert werden, vielmehr wird es in der Implementation beispielsweise von Schule zu Schule mehr oder weniger bedeutsame Unterschiede geben.

Diese Heterogenität braucht für die Prüfung der Hypothesen kein Nachteil zu sein. Die Hypothesen im genannten Projekt beziehen sich ja nicht darauf, daß die Schüler durch eine ganz *bestimmte* Art, die Methode praktisch umzusetzen (eine bestimmte Operationalisation O_{UVi}), gefördert werden, sondern darauf, daß eine relativ allgemein (auf theoretischer Ebene) formulierte Methode (K_{UV}), welche auf unterschiedliche Weise operationalisiert werden kann (O_{UV1}, \dots, O_{UVi}) diese Förderung bewirkt.

Im Einklang mit einem der Kriterien zur Auswahl von Konstrukten und deren Operationalisation im kritischen Multiplizismus (vgl. Abschnitt 3) unterscheiden sich somit die Operationalisierungen O_{UVi} in bezug auf theoretisch relevante (fördernde) Aspekte; dies wird durch die konvergente Treatment-Validität erfaßt. Die Intervention in der Kontrollgruppe unterscheidet sich demgegenüber von der Versuchsgruppe in bezug auf theoretisch relevante Aspekte (diskriminante Treatment-Validität). Die Wirksamkeit des Versuchs-, nicht aber des Kontroll-Treatments beweist, daß diese Aspekte nicht nur theoretisch, sondern auch praktisch relevant sind.

Die Variationen in der Realisation der Methode führen allenfalls zu einer Vergrößerung der Fehler-Varianz und damit dazu, daß es schwieriger ist, den Einfluß des Unterrichts nachzuweisen. Gelingt es hingegen, diesen Einfluß für die Versuchs-, aber nicht für die Kontrollgruppe bei gegebener Treatment-Validität zu zeigen, ist diese Erkenntnis wertvoller als bei der Beschränkung auf bestimmte Unterrichtstypen: Die Aussagen, die man machen kann, sind genereller gültig, nämlich für die unterschiedlichsten Implementationen. Ferner können wesentliche plausible rivalisierende Hypothesen (Campbell 1969; vgl. auch oben die Beispiele zum kritischen Multiplizismus: Prüfung multipler rivalisierender Hypothesen), welche zu Artefakten führen könnten, ausgeschlossen werden. Damit erhält die Evaluationsstudie wesentliche Elemente einer Forschungsstudie: Generalisierbarkeit ist untersuchbar (vgl. auch Cronbach et al. 1972), und die Artefaktgefahr ist verringert.

Ich verdanke Don Campbell einige Hinweise in dieser Richtung, die ich an dieser Stelle herzlich danken möchte. Für ein konkretes Beispiel, jedoch ohne Analyse im dargestellten Sinn, vgl. auch Cook & Campbell (1979, S. 61f.).

Ein besonderer Vorteil ist, daß wir es in der oben als Beispiel genannten Evaluationsstudie mit *Klassen* zu tun haben: Innerhalb einer Klasse von durchschnittlich etwa 20 Schülern ist die Heterogenität bezüglich des Treatments weitgehend eliminiert. Die Fehlervarianz wird für die einzelne Klasse also sehr vermindert. Nun wird in der Forschung immer

wieder beklagt, daß zu wenig identische und konzeptuelle Replikationen durchgeführt werden. Man kann aber die Untersuchung für jede einzelne Klasse als konzeptuelle Replikation der Untersuchungen der anderen Klassen interpretieren; in diesem Sinne ist dann die Evaluationsstudie eine Untersuchung mit multiplen identischen und konzeptuellen Replikationen.

In identischen Replikationen werden die Bedingungen der replizierten Untersuchung möglichst identisch nachvollzogen. Unter einer konzeptuellen Replikation versteht man demgegenüber die Durchführung der gleichen Untersuchung unter Verwendung etwas geänderter Randbedingungen; da diese Änderungen von der untersuchten Theorie her irrelevant sind, entsprechen diese Replikationen wiederum den Prinzipien des kritischen Multiplizismus.

Insgesamt erweist es sich somit, daß die Evaluationsstudien zur Beantwortung von bestimmten Forschungsfragen geeignet sind, falls sie in angemessener Weise realisiert werden, vor allem wenn die Treatment-Validität zusätzlich zu den üblichen anderen Validitätstypen erhoben wird.

6. Schlußfolgerungen

Der „kritische Multiplizismus“ (Cook 1985; Houts et al. 1986) ist ein Ansatz, der wichtige Kritiken an der gängigen Forschungspraxis berücksichtigt und ermöglicht, trotzdem (oder erst recht) gültige Aussagen zu machen. Es wird nicht mehr behauptet, aufgrund einer einzigen einfachen Untersuchung grundsätzliche Aussagen machen zu können, sondern es werden systematisch und auf allen Ebenen — methodisch wie auch theoretisch — unterschiedliche Ansätze verwendet; falls die sich daraus ergebenden Schlußfolgerungen konvergieren, können die zugrundeliegenden Aussagen eher als bestätigt angesehen werden, als wenn kein multiplizistischer Ansatz verwendet wurde. Der kritische Multiplizismus ist vor allem in Form des multiplen Operationalismus (vgl. etwa Campbell & Fiske 1959) ein wertvolles Instrument für die Forschung. Dabei wurde insbesondere auf die Bedeutung der Konstruktvalidität des Treatments oder der UV hingewiesen, die in Forschungsuntersuchungen häufig vernachlässigt wird. Bei diesem Validitätstyp geht es um die Beziehung zwischen dem Konstrukt der UV K_{UV} und dessen Operationalisationen $O_{UV1}, \dots, O_{UVi}, \dots$, also um die Frage, ob das Treatment genau diejenigen Komponenten enthält, die durch das zugrundegelegte Konstrukt vorgegeben sind.

Werden einige der Bedingungen des multiplen Operationalismus berücksichtigt, lassen sich, wie gezeigt, Evaluationsstudien gut als Forschungsstudien verwenden. Unter diesen Umständen können Evaluationsstudien sogar gegenüber „reinen“ Forschungsstudien Vorteile aufweisen; genannt wurde insbesondere die Heterogenität in bezug auf Merkmale, die für die zu prüfenden Konstrukte irrelevant sind.

Damit erfüllen die meisten größeren Evaluationsstudien Bedingungen des kritischen Multiplizismus, die in den typischen Forschungsuntersuchungen kaum je berücksichtigt werden können, geht es in diesen doch meist darum, maximale Homogenität der verwendeten Operationalisationen eines Konstruktes zu erreichen, um möglichst scharfe Unterschiede zu erhalten.

Nicht berücksichtigt wurde oben allerdings die Frage, wie eine Theorie gestaltet sein muß, die mit Hilfe solcher Designs untersucht werden kann — denn selbstverständlich sind nicht alle Theorien und Fragestellungen dafür geeignet. Im Einklang mit dem kritischen Multiplizismus ist diesbezüglich allerdings zu sagen, daß sich wahrscheinlich weit mehr Fragestellungen dafür eignen, als man auf den ersten Blick erwarten würde. Umgekehrt besagt der kritische Multiplizismus ebenfalls, daß es sinnvoll ist, Theorien bzw. Hypothesen nicht ausschließlich mittels Evaluationsstudien zu untersuchen, sondern daß es besser ist, sie — in angemessener Formulierung — auch unter kontrollierteren Bedingungen zu prüfen. Erst das Netzwerk von Erkenntnissen wird es erlauben, die Gültigkeit von Hypothesen mit einiger Zuverlässigkeit festzustellen. Die Anwendung des hier diskutierten Konzeptes steht zur Zeit noch aus. Es ist zu hoffen, daß in absehbarer Zeit ein Projekt realisiert werden kann, daß zum einen die Treatmentvalidität oder genereller den kritischen Multiplizismus in der skizzierten Form zu untersuchen erlaubt, andererseits aber auch festzustellen ermöglicht, ob die angedeuteten Bedingungen von „Evaluationsstudien zu Forschungszwecken“ realisierbar und hinreichend sind.

Literatur

- BUNGARD, W./BAY, R.: Feldexperimente in der Sozialpsychologie. In: *Patry, J.-L.* (Hrsg.): *Feldforschung. Methoden und Probleme sozialwissenschaftlicher Forschung unter natürlichen Bedingungen.* Huber, Bern 1982, 183-205.
- CAMPBELL, D.T.: Pattern matching as an essential in distal knowing. In: *Hammond, K.R.* (Hrsg.): *The psychology of Egon Brunswik.* Holt, Rinehart and Winston, New York 1966, 81-106.
- CAMPBELL, D.T.: Prospective: Artifact and control. In: *Rosenthal, R./Rosnow, R.* (Hrsg.): *Artifact in behavior research.* Academic Press, New York 1969, 351-382.
- CAMPBELL, D.T./FISKE, D.W.: Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 1959, 81-105.
- CAMPBELL, D.T./O'CONNELL, E.J.: Methods as diluting trait relationships rather than adding irrelevant systematic variance. In: *Brinberg, D./Kidder, L.H.* (Hrsg.): *Forms of validity in research. New Directions for Methodology of Social and Behavioral Science*, no. 12. Jossey-Bass, San Francisco, 1982, 93-111.
- CAMPBELL, D.T./STANLEY, J.C.: Experimental and quasi-experimental designs for research on teaching. In: *Gage, N.L.* (Hrsg.): *Handbook of research on teaching.* Rand McNally, Chicago 1963, 171-246.
- COOK, T.D.: Postpositivist critical multiplism. In: *Shortland, R.L./Mark, M.M.* (Hrsg.): *Social science and social policy.* Sage, Beverly Hills, Calif. 1985, 21-62.
- COOK, T.D./CAMPBELL, D.T.: Quasi-experimentation. Design and analysis issues for field settings. Rand McNally, Chicago 1979.

- CRANO, W.D.: Triangulation and cross-cultural research. In: *Brewer, M.B./Collins, B.E.* (Hrsg.): *Scientific inquiry and the social science: A volume in honor of Donald T. Campbell.* Jossey-Bass, San Francisco, 1981.
- CRONBACH, L.J./GLESER, G.C./NANDA, H./RAJATRATNAM, N.: *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* Wiley, New York 1972.
- FISKE, D.W.: Convergent-discriminant validation in measurements and research strategies. In: *Brinberg, D./Kidder, L.H.* (Hrsg.): *Forms of validity in research. New Directions for Methodology of Social and Behavioral Science*, no. 12. Jossey-Bass, San Francisco 1982, 77-92.
- HERRMANN, T.: *Persönlichkeitsmerkmale. Bestimmung und Verwendung in der psychologischen Wissenschaft.* Kohlhammer, Stuttgart 1973.
- HOUTS, A.C./COOK, T.D./SHADISH, W.R. jr.: The person-situation debate: A critical multiplist perspective. *Journal of Personality*, 54, 1986, 52-105.
- LIENERT, G.A.: *Testaufbau und Testanalyse.* Beltz, Weinheim 1969 (dritte Auflage).
- MÜLLER, G.F.: Dilemmata psychologischer Evaluationsforschung. *Psychologische Rundschau*, 38, 1987, 204-212.
- OERTER, R.: Welche Realität erfaßt die Unterrichtsforschung? *Unterrichtswissenschaft*, 7, 1979, 24-43.
- OSER, F./PATRY, J.-L.: Interventionsstudien für sozial-kognitive Kompetenz: Beispiele und theoretische Überlegungen zur Transformations-Validität. *Unterrichtswissenschaft*, 14, 1986, 254-268.
- PATRY, J.-L.: *Transssituationale Konsistenz des Verhaltens und Handelns in der Erziehung. Habilitationsschrift, eingereicht bei der Philosophischen Fakultät der Universität Freiburg (Schweiz) 1987a.*
- PATRY, J.-L.: Evaluationsmethodologie zu Forschungszwecken — Ein Beispiel von „Kritischem Multiplizismus“. *Bericht zur Erziehungswissenschaft Nr. 66, Pädagogisches Institut der Universität Freiburg (Schweiz) 1987b.*
- PATRY, J.-L.: *Transssituationale Konsistenz, Generalisierbarkeit und das Laborfeld-Problem. Zeitschrift für empirische Pädagogik, in Druck.*
- PATRY, J.-L., ET AL.: *Der Versuchsplan zur Untersuchung des Lehrereithos II: Ein Beispiel für die Anwendung des „Kritischen Multiplizismus“.* Manuskript, in Vorbereitung.
- PERRES, M./PATRY, J.-L.: *Nomologisches Wissen, technologisches Wissen, Tatsachenwissen — drei Ziele sozialwissenschaftlicher Forschung.* In: *Patry, J.-L.* (Hrsg.): *Feldforschung. Methoden und Probleme sozialwissenschaftlicher Forschung unter natürlichen Bedingungen.* Huber, Bern 1982, 45-66, 1982.
- ROSENTHAL, R./ROSNOW, R. (Hrsg.): *Artifact in behavior research.* Academic Press, New York 1969.
- WOLF, R.M.: The nature of educational evaluation. *International Journal of Educational Research*, 11, 1987, 7-19.
- ZECHA, G./LUKESCH, H.: *Die Methodologie der Aktionsforschung. Analyse, Kritik, Konsequenzen.* In: *Patry, J.-L.* (Hrsg.): *Feldforschung. Methoden und Probleme sozialwissenschaftlicher Forschung unter natürlichen Bedingungen.* Huber, Bern 1982, 367-387.

Anschrift des Autors:

PD Dr. Jean-Luc Patry, Pädagogisches Institut der Universität Freiburg (Schweiz),
Route des Fougères, CH-1700 Freiburg (Schweiz)