

Heekerens, Hans-Peter

Evaluation von Erziehungsberatung: Forschungsstand und Hinweise zu künftiger Forschung

Praxis der Kinderpsychologie und Kinderpsychiatrie 47 (1998) 8, S. 589-606

urn:nbn:de:0111-opus-9001

Erstveröffentlichung bei:



www.v-r.de

Nutzungsbedingungen

pedocs gewährt 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 dem Gebrauch von pedocs und der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Kontakt:

peDOCS

Deutsches Institut für Internationale Pädagogische Forschung (DIPF)

Informationszentrum (IZ) Bildung

Schloßstr. 29, D-60486 Frankfurt am Main

eMail: pedocs@dipf.de

Internet: www.pedocs.de

Praxis der Kinderpsychologie und Kinderpsychiatrie

Ergebnisse aus Psychoanalyse,
Psychologie und Familientherapie

47. Jahrgang 1998

Herausgeberinnen und Herausgeber

Manfred Cierpka, Heidelberg – Gunther Klosinski, Tübingen –
Ulrike Lehmkuhl, Berlin – Inge Seiffge-Krenke, Mainz –
Friedrich Specht, Göttingen – Annette Streeck-Fischer, Göttingen

Verantwortliche Herausgeberinnen

Ulrike Lehmkuhl, Berlin
Annette Streeck-Fischer, Göttingen

Redakteur

Günter Presting, Göttingen

V&R Verlag Vandenhoeck & Ruprecht in Göttingen

Evaluation von Erziehungsberatung: Forschungsstand und Hinweise zu künftiger Forschung

Hans-Peter Heckerens

Summary

Evaluation of Child Guidance: State of Research and Comments on Future Research

The evaluation of child guidance is placed in the context of quality control with emphasis on the significance of evaluation under field as opposed to laboratory conditions as well as the advantages of experimental or quasi-experimental field research over catamnestic studies. Up until now only three experimental evaluation studies have been carried out in the field of child guidance under natural conditions. These studies are first described in terms of their methods and results and then placed in a larger context on this basis. Proceedings from these studies based on methodological considerations and empirical evidence, comments and recommendations will be made concerning methodologically satisfactory and practically significant future evaluation research. They emphasize in particular the different evaluation strategies as well as various aspects of validity.

Zusammenfassung

Die Evaluation von Erziehungsberatung wird in den Kontext von Qualitätssicherung gestellt, die Bedeutsamkeit von Evaluation unter Feldbedingungen gegenüber der unter Laborbedingungen hervorgehoben und die Vorteile (quasi-)experimenteller Feldforschung gegenüber katamnestic Studien herausgestellt. Die bislang im Feld der Erziehungsberatung durchgeführten drei experimentellen Evaluationsstudien unter natürlichen Bedingungen werden von ihrer Methodik her und nach ihren Ergebnissen dargestellt und nach beiden Aspekten in größere Zusammenhänge gestellt. Ausgehend von diesen Studien werden auf der Basis methodologischer Überlegungen und empirischer Befunde Hinweise und Vorschläge zu einer methodisch befriedigenden und praktisch bedeutsamen zukünftigen Evaluationsforschung unterbreitet. Diese Hinweise und Vorschläge betreffen insbesondere unterschiedliche Evaluationsstrategien und verschiedene Validitätsaspekte.

Prax. Kinderpsychol. Kinderpsychiat. 47: 589-606 (1998), ISSN 0023-7034
© Vandenhoeck & Ruprecht 1998

1 Einleitung

Auch außerhalb des Geltungsbereichs der Krankenversicherung erbrachte Beratung und Psychotherapie, so auch die in Erziehungsberatungsstellen angebotenen Leistungen, werden sich zunehmend mehr der Frage nach ihrer Effektivität und Effizienz ausgesetzt sehen. Das hat teilweise mit bestimmten Neuregelungen bezüglich der Erziehungsberatung im Kinder- und Jugendhilfegesetz zu tun, teilweise mit einem neuen Denken bei den öffentlichen Trägern und Finanziers. Die *Kommunale Gemeinschaftsstelle* hat mit ihrem Bericht Nr. 9/1994 „Outputorientierte Steuerung der Jugendhilfe“ die Geister aufgeschreckt und die Gemüter erhitzt. Die *Kommunale Gemeinschaftsstelle* setzt auf „Qualitätsmanagement“ (Bericht Nr. 6/1995). Selbst wenn sich die wirtschaftlichen und politischen Verhältnisse, in deren Zusammenhang diese Konzeptionen entstanden sind, zu einem Besseren geändert haben, werden diese Themen nicht einfach wieder von der Tagesordnung abgesetzt, weil Evaluation auch im Bereich der beraterischen/therapeutischen Hilfen für Familien unverzichtbar ist (Familie und Beratung, 1993). Dessen ist man sich in Kreisen sowohl der evangelischen wie der katholischen Träger von Erziehungsberatungsstellen offensichtlich bewußt (HARD-LOH 1996; Qualitätssicherung in der Erziehungsberatung, 1996).

Wenn man von „Qualität“ spricht, kann man drei Dimensionen unterscheiden: die technische oder Strukturqualität, die Beziehungs- oder Prozeßqualität und die funktionelle oder Ergebnisqualität. Zwei Aspekte der Ergebnisqualität sind der Grad der Zielerreichung, die Effektivität, einerseits und das Verhältnis von Aufwand und Ergebnis, die Effizienz, andererseits. Diese beiden Aspekte der Ergebnisqualität sind nicht die einzigen; andere, wie etwa die Zufriedenheit von Therapeuten und Klienten oder die Sozialverträglichkeit der beraterisch-therapeutischen Maßnahme kommen hinzu. Aber Effektivität und Effizienz sind unverzichtbare Aspekte bei der Beurteilung von Ergebnisqualität.

2 Unterschiedliche Arten von Effektivität

In einem ersten Versuch (HEEKERENS 1997), etwas beizutragen zur Evaluation der Erziehungsberatung, habe ich einen eher konventionellen Weg beschritten. Ich habe dargestellt, welchen Kenntnissstand uns meta-analytisch verfahrenende Sekundäranalysen von Evaluationsstudien zur Wirksamkeit einzelner Grundformen der kommunikativen Einflußnahme mit psychotherapeutischen Mitteln bei Problemlagen von Kindern und Jugendlichen vermitteln. Als Grundformen wurden dabei unterschieden: Kinder- und Jugendlichen-Psychotherapie, Elterntraining und Familientherapie. Daß damit nur ein bestimmter Ausschnitt von Erziehungsberatung, dessen Größe zudem unklar ist, betrachtet werden konnte, war als ein Schwachpunkt benannt worden. Auf einen anderen wurde ebenfalls hingewiesen: in den zugrundeliegenden Evaluationsstudien – und das gilt für die Psychotherapieevaluation schlechthin – wird ganz überwiegend die Wirksamkeit unter Laborbedingungen und nur zum geringsten Teil die Wirksamkeit unter Feldbedingungen ermittelt.

In der Nachfolge von STARFIELD (1977) wird die erste Art der Wirksamkeit als „efficacy“, die zweite als „effectiveness“ bezeichnet; analoge deutsche Wörter, diese unterschiedlichen Arten von Effektivität zu benennen, fehlen. Neuere Untersuchungen, die sich speziell der Frage gewidmet haben, wieviele Evaluationsstudien eigentlich unter realen Feldbedingungen durchgeführt wurden, geben eine eindeutige Antwort: unser Wissen über die Wirksamkeit von Psychotherapie ist vornehmlich Wissen über deren „efficacy“. Das haben für die Kinder- und Jugendlichen-Psychotherapie die Arbeitsgruppe um WEISZ und WEISS (WEISZ et al. 1992), zuvor hervorgetreten durch eine Meta-Analyse zur Wirksamkeit von Kinder- und Jugendlichen-Psychotherapie (WEISZ et al. 1987), und für die Familientherapie die SHADISH-Gruppe (SHADISH et al. 1995), bekannt durch ihre Meta-Analyse zur Wirksamkeit von Paar- und Familientherapie (SHADISH et al. 1993), gezeigt.

Die Schlußfolgerung ist klar. Wenn wir Näheres über die „effectiveness“ von Erziehungsberatung erfahren wollen, müssen wir uns den Studien zuwenden, die unter normalen Feldbedingungen evaluiert haben. Methodisch betrachtet haben wir dabei für das Feld der Erziehungsberatung zwei verschiedene Arten vor uns: katamnestiche Studien einer- und (quasi-)experimentelle Untersuchungen andererseits.

3 Katamnestiche und (quasi-)experimentelle Studien

Katamnestiche Studien sind bis heute weltweit die vorherrschende Form, die Wirksamkeit von Erziehungsberatung zu evaluieren; in Deutschland gibt es nur diese Art. Mir sind insgesamt zwölf katamnestiche angelegte Studien zur Evaluation der Wirksamkeit von Erziehungsberatung in Deutschland bzw. der früheren Bundesrepublik bekannt; in der Reihenfolge ihres Erscheinens sind das BLECKMANN und BODEMANN (1956), HEGE und BICHLMAIER (1965), BRANDT (1967), BRACKMANN (1977), KEPPLER (1979), SAKOFSKI und KÄMMERER (1986), JÄGER (1987), STRAUS et al. (1988; HÖFER u. STRAUS 1991), ZÜRN et al. (1990), DILLIG (1993), NAUMANN und BECK (1994) sowie LENZ (1994a). Ich habe mich mit der Methodik und den Ergebnissen dieser Studien bereits an anderem Ort ausführlich auseinandergesetzt (HEEKERENS 1997). Hier seien deshalb nur zwei generelle Einwände gegen katamnestiche Studien wiederholt.

Katamnestiche Studien haben, methodisch beurteilt eine zentrale Schwäche: von ihrer Untersuchungsanlage her sind sie wenig gut geeignet, die interne Validität abzusichern. Das Ergebnis einer Evaluationsstudie ist insoweit valide, als ein beobachtetes Kriteriumsverhalten (etwa der Erfolg der Beratung/Therapie oder die Zufriedenheit mit Beratung/Therapie) auf die durchgeführte Beratung/Therapie zurückgeführt werden kann. Gefährdet ist die interne Validität durch alle denkbaren alternativen Erklärungen, die die beobachteten Veränderungen im Kriteriumsverhalten auf andere Einflüsse als auf die Behandlung(en) zurückführen können. Bei den beraterischen/psychotherapeutischen Hilfen für Kinder und Jugendliche hat man dabei neben den allfälligen Ereignissen im normalen Alltag, die Wirkung auf das Kriteriumsverhalten zeigen, verstärkt an Reifungsprozesse zu denken. Zur Kontrolle der genannten zwei Faktoren bieten sich zur Evaluation Kontrollgruppenpläne an, wie sie aus der (quasi-)experimentellen Forschung bekannt sind.

Historisch betrachtet sind (quasi-)experimentelle Versuchspläne die Antwort auf den seit den fünfziger Jahren vorgebrachten Zweifel daran, daß Psychotherapie im allgemeinen und bei Problemlagen von Kindern und Jugendlichen im besonderen überhaupt wirksam sei. Mit der Veröffentlichung von EYSENCKs erster Arbeit über die Wirksamkeit von Psychotherapie im Jahre 1952 geht eine bestimmte Ära der Psychotherapieevaluation zu Ende, die der retrospektiven Globalbeurteilungen des Therapieerfolgs durch Klienten und/oder Therapeuten, und eine andere, die der (quasi-)experimentellen Psychotherapieforschung mit ihren typischen Gruppenvergleichen, beginnt (KAZDIN 1991; KLANN u. HAHLOWEG 1994). Warum auf dem Feld der Erziehungsberatung die frühere Ära noch immer andauert, bedarf einer umfangreicheren Klärung, als sie hier geleistet werden könnte.

4 Katamnese und Experiment im praktischen Vergleich

Katamnestic Studien sind (quasi-)experimentellen darin unterlegen, daß letztere die interne Validität der Evaluationsergebnisse ungleich besser absichern als Katamnesen. Ein zweites kommt hinzu: die bei Katamnesen erzielten Ergebnisse sind – ungeachtet der Unsicherheit, wem oder was wir sie zuzuschreiben haben – weniger gehaltvoll als die von (Quasi-)Experimenten. Das kann man illustrieren an einem Beispiel. Das Beispiel betrifft thematisch nur einen bestimmten Ausschnitt der Tätigkeit von (Integrierten) Erziehungsberatungsstellen (vgl. dazu ABEL 1996; SPECHT 1993), aber es ist dennoch beispielhaft und es gibt kein illustrativeres. Es geht um die zusammenfassende Bewertungen der Wirksamkeit von Paar- und Familientherapie durch eine katamnestic Studie einerseits und eine Sekundäranalyse nach dem meta-analytischen Modell andererseits.

Im Rahmen einer großen Befragung unter 526 Klinischen Mitgliedern der *American Association of Marriage and Family Therapy* aus 15 Bundesstaaten der USA wurden von 492 Klienten(familien) rückblickend Einschätzungen über die Wirksamkeit der Behandlung sowie die Zufriedenheit mit der Behandlung erhoben (DOHERTY u. SIMMONS, 1996). Die Klientenzufriedenheit wurde über acht Einzelindikatoren erfaßt, und es zeigte sich in allen acht Bereichen eine Zufriedenheitsquote von über 90 Prozent. Bei den zehn Indikatoren für Behandlungserfolg schwankt die Zustimmungswahl zwischen 38,9 Prozent („Integration ins Gemeinwesen erhöht“) und 88,8 Prozent („Emotionale Gesundheit verbessert“). Am meisten Zustimmung erfahren die Statements „Emotionale Gesundheit verbessert“ (88,8%), „Therapieziele erreicht“ (83,0%) und „Familienbeziehung verbessert“ (78,2%). Wenn 83 Prozent angeben, das Therapieziel sei erreicht worden, erweckt dies den Eindruck, Paar- und Familientherapie sei eine sehr hoch wirksame Form der Behandlung.

Man kann dieses Ergebnis kontrastieren mit dem, das SHADISH et al. (1993; ergänzend SHADISH et al. 1995) in der jüngsten Meta-Analyse zur Wirksamkeit von Paar- und Familientherapie vorgelegt haben. Die SHADISH-Gruppe hat (unter anderem) die Gesamteffektstärke von Paar- und Familientherapie beim experimentellen Vergleich nur mit nicht-behandelten Kontrollgruppen zum Behandlungsende vorgenommen

und dabei – bei konservativer Schätzung – eine Effektstärke von 0.51 (Basis: 71 Studien) ermittelt.

Was eine Effektstärke von 0.51 bedeutet, kann man sich an zweierlei verdeutlichen. Einem Vorschlag von COHEN (1977) folgend kann man Effektstärken, so sie denn überhaupt signifikant sind, grob folgendermaßen klassifizieren: groß (.80 und mehr), mittel (.50 – 0.79) und gering (0.20 – 0.49); der Wert von 0.51 liegt danach am unteren Rand des Mittelbereichs. Dieses eher ernüchternde Ergebnis bestätigt eine zweite Form der „Übersetzung“ einer Effektstärke von 0.51. Nach ihr steht der mittlere Klient mit Behandlung – beurteilt nach Indikatoren für biopsychosoziales Wohlergehen – besser da als 70 Prozent der Klienten ohne Behandlung. Das ist sehr weit entfernt von der Zielinie, nach der jeder Klient mit Behandlung am Ende besser dastehen müsse als jeder Klient ohne Behandlung.

Paar- und Familientherapie hat also einen recht geringen Wirkungsgrad. Das aber teilt sie mit Psychotherapie überhaupt. Nach der noch immer umfassendsten Meta-Analyse der Wirksamkeit von Psychotherapie von SMITH et al. (1980) liegt die mittlere Effektstärke von Psychotherapie bei 0.85. Das entspricht einer Korrelation von $r=0.32$ zwischen Psychotherapie und Behandlungsergebnis und bedeutet, daß Psychotherapie 10 Prozent der Varianz von Effekten erklärt. Das ist, statistisch beurteilt, recht wenig, praktisch betrachtet aber bedeutsam. Wer etwa würde behaupten, Rauchen habe auf das Risiko, an Lungenkrebs zu erkranken, keinen praktisch bedeutsamen Effekt, wenn er erfährt, daß der Zusammenhang zwischen Rauchen und Lungenkrebs bei nur $r=0.14$, also weniger als halb so hoch wie zwischen Psychotherapie und Behandlungsergebnis, liegt (WITTMANN u. MATT 1986)?

Ergebnisse aus (Quasi-)Experimenten stellen wesentlich nüchternere und realistischere Maße zur Einschätzung der Wirksamkeit von Beratung und Therapie dar als Kattamnese-Resultate. Ferner sind die Ergebnisse von (Quasi-)Experimenten „anschlußfähig“ in dreierlei Hinsicht. Zum einen, dies betrifft den Außenbereich, können die Ergebnisse leicht und gut in Beziehung gesetzt werden zu denen anderer Psychotherapieformen oder alternativer Behandlungen nicht-psychotherapeutischer Art. Zum anderen, hier kommen wir zum Binnenbereich, können diese Ergebnisse nach vielfältigen Gesichtspunkten (z.B. einzelne Methoden, unterschiedliche Problemlagen) differenziert werden. Schließlich kann mit Hilfe von Effektstärken die klinisch relevante Bedeutsamkeit, die sogenannte klinische Relevanz, bestimmt werden (SHADISH et al. 1993; SCHULTE 1993).

5 (Quasi-)Experimentelle Evaluationsstudien

Als Auswahlkriterien wurden für das Feld der Erziehungsberatung dieselben zugrundegelegt, die auch der GRAWE-Gruppe dienen; berücksichtigt wurden nur solche im Feld der Erziehungsberatung durchgeführte Studien,

- in denen eine psychologische Beeinflussung mit therapeutischer Zielsetzung in einer der Methode angemessenen Intensität untersucht und

- in denen mindestens eine Variable (Therapiemethode, Patienten- oder Therapeutenmerkmale, Therapiesetting) experimentell kontrolliert und ein gruppenstatistischer Vergleich entweder mit einer Kontrollbedingung und/oder mindestens einer weiteren Behandlungsbedingung vorgenommen worden war (Grawe et al. 1994, S. 57).

Es gibt nach diesen Selektionskriterien bislang nur drei Evaluationsstudien zur Erziehungsberatung, alle durchgeführt an *Child Guidance Clinics* in den USA (Trankina 1975; Johnson u. Malony 1977; Fisher 1980, 1984). Den drei Arbeiten gemeinsam ist, daß sie Klienten, Therapeuten und Vorgehensweisen betrachten, wie sie sich im jeweiligen Feld vorfinden. Alle drei Untersuchungen sind, methodisch betrachtet, als experimentelle Feldstudien angelegt. In allen drei Studien steht im Mittelpunkt die Frage nach der absoluten bzw. relativen Wirksamkeit eines familientherapeutischen Vorgehens, was wenig verwunderlich ist, mußte sich die Familientherapie als nachgeborene Interventionsform doch (auch und gerade an Erziehungsberatungsstellen) erst legitimieren (Heekerens 1989). Ein Problem ist, daß die Angaben zu dem jeweiligen familientherapeutischen Vorgehen so spärlich sind, daß es keiner theoretischen Grundorientierung (wie etwa strategisch oder strukturell) zugeordnet werden kann (zum Problem vgl. Heekerens 1998).

In dem Feldexperiment von Trankina (1975) wurden 119 Fälle mit familientherapeutisch orientierter Krisenintervention verglichen mit 115 Fällen, die eine traditionelle Behandlung im Sinne klassischer Erziehungsberatung (hierzu ausf. Heekerens 1983, 1989; Heekerens u. Henkelmann-Strickler 1980) erhalten hatten. Von den zahlreichen Einzelergebnissen betrachten wir hier nur diejenigen, die für einen Vorher-nachher-Vergleich der Gruppen relevant sind. „Vorher“ meint: vor Behandlungsbeginn; „nachher“ bedeutet im Falle der familientherapeutisch orientierten Krisenintervention acht Wochen, im Falle der traditionellen Erziehungsberatung neun Monate später. Zu den jeweiligen Zeitpunkten waren die Fälle in der einen wie in der anderen Gruppe (weitgehend) abgeschlossen. Halten wir vorab fest: Die Behandlungsdauer war bei der traditionellen Erziehungsberatung viereinhalbmal so lang wie bei der familientherapeutischen Krisenintervention.

Zur Erfolgseinschätzung wurden drei Kriteriumsvariablen betrachtet: Veränderung im Problemverhalten des Indexpatienten, erfaßt durch Elternangaben, Änderungen in der Wahrnehmung familiärer Zufriedenheit, Anpassung und Kompetenz sowie die Entwicklung der elterlichen Einschätzung des Familienlebens hinsichtlich der Fähigkeit zur Kommunikation, des Grads an Intimität/Bezogenheit und der Zufriedenheit mit der familiären Beziehung. Bei keinem der vorgenommenen Vergleiche waren signifikante Unterschiede zwischen den beiden Behandlungsgruppen zu finden.

Zwei Nebeneffekte außer der schon genannten Zeit- und Arbeitersparnis seien noch erwähnt. Einmal äußern sich die Mütter bei der familientherapeutisch orientierten Krisenintervention über die Arbeit der *Child Guidance Clinic* zufriedener als bei der traditionellen Erziehungsberatung. Zum anderen ist die Ausfallquote bei der familientherapeutisch orientierten Erziehungsberatung geringer als bei der traditionellen Erziehungsberatung; nach acht Wochen betrug sie 28 (Mütter) bzw. 34 (Väter) Prozent gegenüber 40 bzw. 52 Prozent.

In der experimentellen Feldstudie von JOHNSON und MALONY (1977) wurde die Frage verfolgt, ob bei einem familientherapeutischen Vorgehen als spezieller Effekt hervortrete, daß sich die familiäre Kommunikation in Richtung auf größere Gleichverteilung und stärkere Vernetzung hin verändere; dies war als Zeichen für „Normalisierung“ von HALEY postuliert worden. Zur Klärung der Frage wurden jeweils 14 Familien, die sich angemeldet hatten und für die eine Kurzzeitbehandlung angezeigt schien, zufällig einer von zwei Behandlungsbedingungen zugewiesen: einer Familientherapie oder einem alternativen Verfahren aus dem Beratungs- und Therapieangebot der betreffenden Erziehungsberatungsstelle. Beim Vergleich zwischen vier und sechs Wochen nach Behandlungsbeginn zeigte sich die postulierte Veränderung tatsächlich in fünf von neun betrachteten Aspekten.

Inwieweit man dieses Ergebnis als „Erfolg“ ansehen kann, ist aus zwei Gründen fraglich. Zum einen: das genannte Veränderungsmaß ist das einzig erfaßte, und nichts spricht nach empirischen Befunden der einschlägigen Evaluationsforschung dafür, aus besagter Veränderung im Bereich der familiären Kommunikation schon auf eine Verbesserung in anderen Bereichen, etwa der Präsentiersymptomatik des Indexpatienten zu schließen. Zum anderen sind wir uns heute im Vergleich zu damals weitaus weniger sicher, daß man zwischen „normalen/gesunden“ Familien und „abnormalen/kranken“ Familien schon anhand bestimmter Interaktions- und Kommunikationsmuster trennscharf unterscheiden kann.

Das sehr sorgfältig geplante Feldexperiment von FISHER (1980, 1984) war eine Reaktion auf das Problem zu langer Wartezeiten. Eine für die Gesamtklientel repräsentative Stichprobe von Familien mit einem Indexpatienten zwischen 8 und 13 Jahren wurde nach einem Eingangsinterview, das in der Regel binnen 10 Tagen nach der Anmeldung erfolgte, zufällig einer der folgenden vier Bedingungen zugeteilt (in Klammern: Anzahl der Familien, für die komplette Daten vorliegen): 1. einer Behandlungsgruppe (n=9) mit 6 Sitzungen Familientherapie im Zeitraum von 8 Wochen (nach 1–2 diagnostischen Interviews), 2. einer Behandlungsgruppe (n=10) mit 12 Sitzungen Familientherapie im Zeitraum von 16 Wochen (nach 1–2 diagnostischen Interviews), 3. einer Behandlungsgruppe (n=8) mit Familientherapie ohne zeitliche Begrenzung und 4. einer Warteliste-Kontrollgruppe (n=10).

Die vier Gruppen, die sich hinsichtlich soziologischer und sozio-ökonomischer Merkmale nicht signifikant voneinander unterschieden, wurden anhand verschiedener Meßverfahren zu fünf Zeitpunkten miteinander verglichen: beim Aufnahmegespräch, nach der sechsten Sitzung (die Behandlungsgruppe 1 hatte die Behandlung abgeschlossen), am jeweiligen Behandlungsende, ferner sechs Wochen und schließlich ein Jahr nach Behandlungsende. Die wichtigsten Ergebnisse lassen sich in drei Punkten zusammenfassen: (a) alle drei Behandlungsgruppen zeigten Verbesserungen, die auch ein Jahr nach Behandlungsende erhalten blieben. (b) zwischen den einzelnen Behandlungsgruppen ergaben sich keine signifikanten Unterschiede im Behandlungserfolg. (c) im Follow-up nach einem Jahr zeigte sich, daß sieben Familien aus den Behandlungsgruppen zwischenzeitlich weitere Hilfe in Anspruch genommen hatten; auch bei diesem Punkt gab es keine Unterschiede zwischen den Behandlungsgruppen.

Von den sonstigen Ergebnissen sind zwei erwähnenswert. Die Anzahl der tatsächlich durchgeführten Therapieeinheiten lag in der Gruppe ohne zeitliche Beschränkung im Schnitt noch innerhalb des Rahmens von 12 Einheiten (Median: 11,5 Einheiten). Im gegenseitigen Einvernehmen von Familie und Therapeut wurden bis zur sechsten Sitzung ca. 30 Prozent der Behandlungen in der 12-Einheiten-Gruppe und gar 44 Prozent der Behandlungen in der Gruppe ohne zeitliche Limitierung abgeschlossen. Die letztgenannte Zahl steht in Übereinstimmung mit dem Befund, daß an dieser Institution in den letzten drei Jahren 44 Prozent der Fälle nach spätestens sechs Sitzungen abgeschlossen waren.

Wenn man die Ergebnisse der dargestellten experimentellen Evaluationsstudien zusammenfassend betrachtet, so muß man die Ergebnisse der Studie von JOHNSON und MALONY (1977) aus den genannten Gründen außer Betracht lassen. Mit Blick auf die beiden anderen Studien sollte man sinnvollerweise zwischen zwei Arten von Effekten unterscheiden. Die einen könnte man Haupteffekte nennen. Ins Auge gefaßt werden dabei, wie dies in der Psychotherapie(evaluations)forschung üblich ist, solche Kriterien, die den Indexpatienten (oder spezieller noch sein Präsentierproblem) sowie den näheren und weiteren Kontext – von der innerfamiliären Dyade bis zur Beziehung der Familie mit ihrer sozialen Umwelt – in Augenschein nehmen und irgendwelche Maße für (ganz allgemein formuliert) Wohlergehen betrachten. Diesbezüglich kann man zwei Dinge festhalten: im Vergleich mit Nicht-Behandlung erweist sich das familientherapeutische Vorgehen als effektiv (FISCHER 1980) und im Vergleich mit einer alternativen Vorgehensweise im Sinne klassischer Erziehungsberatungsarbeit zeigen sich keine Wirksamkeitsunterschiede (TRANKINA 1975). Beides steht in Übereinstimmung mit und ordnet sich ein in allgemeine Erkenntnisse der Evaluationsforschung (zusammenfassend HEEKERENS 1997b).

Auch zwei Nebeneffekte stehen in Übereinstimmung mit sonstigen Befunden (HEEKERENS 1997b): FISHERS (1980, 1984) Befund, daß auch familientherapeutische Kurzzeitbehandlung effektiv ist, und das Ergebnis aus der Studie von TRANKINA (1975), daß ein familientherapeutisches Vorgehen sich als ökonomisch und effizient erweist.

6 Wie macht man Feldforschung?

Die schon mehrfach erwähnte SHADISH-Gruppe hat zur Feldforschung eine sehr prägnante These aufgestellt: „After all, to test clinic therapy, the only thing one need to do is to go to any real clinical setting, randomly assign patients at intake to treatment or wait-list control, and measure outcome“ (SHADISH et al. 1995, S. 357). Das ist klar, aber nicht ausreichend. Es ist, was die Evaluationsstrategien anbelangt, zu beschränkt; in Frage kommt für die SHADISH-Gruppe offensichtlich nur die Strategie zu prüfen, ob ein bestimmtes Verfahren überhaupt effektiv ist oder nicht. Zum anderen wirft die SHADISH-Gruppe mit dem Vorschlag, nur Experimente mit Kontrollgruppen-Vergleich zuzulassen, einen recht einseitigen und zugleich undifferenzierten Blick auf die Validitätsfrage.

In beiden Fällen bedarf es einer vertieften Auseinandersetzung. Entsprechend geht es im Folgenden zunächst um eine Darstellung unterschiedlicher Evaluationsstrategien

und ihres jeweiligen Nutzens und dann um eine detaillierte Erörterungen verschiedener Validitätsaspekte und ihrer jeweiligen Bedeutungen.

7 Evaluationsstrategien

KAZDIN (1986) hat die unterschiedlichen Strategien, die bei der Evaluation psychotherapeutischer Verfahren eingeschlagen wurden oder prinzipiell verfolgbar sind, in eine Systematik gebracht, anhand derer man sowohl das dargestellte Vorgehen bei den drei Evaluationsstudien zur Erziehungsberatung (FISHER 1980, 1984; JOHNSON u. MALONY 1977; TRANKINA 1975) differenzieren und strukturieren als auch die Perspektive für andere Vorgehensweisen öffnen kann.

Man kann beispielsweise prüfen, ob ein bestimmtes Behandlungspaket überhaupt effektiv ist (*Treatment Package Strategy*) oder wie wirksam es im Vergleich mit einem anderen Vorgehen ist (*Comparative Treatment Strategy*). FISHER (1980) hat die *Treatment Package Strategy* verfolgt, als er die Wirksamkeit von Familientherapie im Vergleich mit einer Wartelistenbedingung überprüfte. JOHNSON und MALONY (1977) sowie TRANKINA (1975) verglichen ein familientherapeutisches Vorgehen mit anderen Behandlungsangeboten, womit sie der *Comparative Treatment Strategy* folgten.

Zwei weitere Strategien bestünden darin zu prüfen, ob eine bestimmtes Verfahren differentiell wirksam ist bei verschiedenen Klienten und unterschiedlichen Therapeuten (*Client and Therapist Variation Strategy*). KAZDINS (1986) Auflistung müßte man an dieser Stelle um einen Punkt erweitern: um die *Setting Variation Strategy*; so wäre etwa die im Bereich der Familientherapie diskutierte Frage, ob die Durchführung einem oder zwei Therapeuten (Ko-Therapie) anvertraut werden sollte, eine der damit zu klärenden Fragen.

Schließlich gibt es noch drei Strategien, denen jeweils eine Variation innerhalb einer Technik zugrunde liegt: ein subtraktives oder additives Vorgehen (*Dismantling/Constructive Treatment Strategy*) sowie eine Parametervariation (*Parametric Treatment Strategy*). Während in den beiden ersten Fällen ein klar identifizierbares Element des Behandlungspaketes weggenommen oder hinzugefügt wird, variiert man im letzten lediglich die Ausprägung eines Elements. Man kann auch sagen, daß der Unterschied einmal ein quantitativer, das andere Mal ein qualitativer ist. Es ist diese letztgenannte Strategie, die FISHER (1980, 1984) verfolgt, wenn er der Frage nachgeht, ob denn die unterschiedliche Dauer von Familientherapie Auswirkungen auf die Wirksamkeit hat. Man könnte eine solche Frage nach möglichen Unterschieden in der „Darreichungsform“ auch unter der genannten *Setting Variation Strategy* einordnen.

8 Validitätsfragen

Kommen wir zum zweiten methodischen Themenkreis. Das Ausmaß, in dem ein bestimmtes methodisches Verfahren zur Prüfung einer Kausalhypothese der Art „Methode x hat die Wirkung y“ geeignet ist, wird Validität genannt. Im Anschluß an CAMP-

BELL (CAMPBELL u. STANLEY 1963) wird traditionellerweise zwischen interner und externer Validität unterschieden. In der Forschungspraxis kommt der Sicherung der internen Validität ungleich mehr Aufmerksamkeit zu als der externen. Die Klage über die mangelnde Übertragbarkeit oder Generalisierbarkeit der Ergebnisse (Aspekt der externen Validität) gehört zu den Standardthemen der Psychotherapieforschung.

Die genannte Zweiteilung in interne und externe Validität ist schon Geschichte. Einmal nämlich hat CAMPBELL später den Begriff der statistischen Validität hinzugefügt, der die Angemessenheit des statistischen (Prüf-)Verfahrens zum Inhalt hat. Zum anderen wurde unter dem Stichwort „Konstruktvalidität“ all das ausgelagert, was die angemessene Repräsentation der in den Hypothesen enthaltenen Konzepte anbelangt. Was die Ausweitung der einzelnen Validitätsarten anbelangt, kann man, dem Sprachgebrauch von HAGER und WESTERMANN (1983) folgend, insgesamt vier einzelne Validitätsarten betrachten: interne Validität, Variablenvalidität, Situations- und Populationsvalidität sowie statistische Validität.

8.1 Interne Validität

Gemessen am Maßstab der Sicherung der internen Validität können Untersuchungen, die eine Kausalhypothese prüfen, in mehr oder minder kontrollierte unterschieden werden. Dabei meint „kontrolliert“, in welchem Maße durch die Anlage der Untersuchung, das (Untersuchungs-)Design, ausgeschlossen ist, daß ein Effekt (auch) durch andere Wirkgrößen als eben die ins Auge gefaßte, etwa eine therapeutische Behandlung, zustande kommt. Lassen wir einmal die für die Evaluation der Erziehungsberatung bis heute faktisch unbedeutenden Einzelfallstudien beiseite, und nehmen die (Quasi-)Experimente, nach Konsens in der Forschungsgemeinschaft die interne Validität am besten absichernd, in den Blick, um an ihnen zu verdeutlichen, was gemeint ist.

Die (Quasi-)Experimente zeichnen sich neben anderen durch drei Merkmale aus: Sie sind prospektiv angelegt, haben mindestens zwei Meßzeitpunkte für einen Vorher-nachher-Vergleich und beinhalten Gruppenvergleiche. Fehlt auch nur eines der drei Merkmale, so kann aus vielerlei Gründen die interne Validität beeinträchtigt sein, bei Einzelgruppenstudien (*Improvement Rates Studies*) mit zwei Meßzeitpunkten ebenso wie bei retrospektiven Analysen/Katamnesen mit einem Meßzeitpunkt selbst dann, wenn ein retrospektiver Gruppenvergleich vorliegt.

Ob sich Experimente und Quasi-Experimente ihrerseits in der Sicherung der internen Validität faktisch unterscheiden, ist eine immer wieder diskutierte Frage. Aus grundsätzlichen theoretischen Erwägungen genießen Experimente (wie bei FISHER 1980, 1984; JOHNSON u. MALONY 1977; TRANKINA 1975) den Vorzug vor Quasi-Experimenten, weil sie durch Zufallszuteilung zu den Gruppen das Risiko von Gruppenunterschieden neben dem, der zur Untersuchung ansteht, minimieren. Ob man bei der Evaluation von Psychotherapie zwischen Experimenten und Quasi-Experimenten differenzieren oder Quasi-Experimente gar ausschließen muß, ist in der Forschungsgemeinschaft umstritten. Der Unterschied trägt nach Auskunft zweier Meta-Analysen nichts bei zur Effektstärke (GIBLIN 1986; GIBLIN et al. 1985; STEINBRUECK et al.

1983), wenig bei einer anderen (HAZELRIGG et al. 1987), viel aber bei der Meta-Analyse der SHADISH-Gruppe (SHADISH et al. 1993; SHADISH et al. 1995).

Verschiedene Studien berichten immer wieder von (teilweise hohen) Ausfallquoten. Wo der Punkt nicht erwähnt wird, besteht Anlaß zum Verdacht, sie seien dort zumindest ebenso hoch gewesen. Die Ausfallquote, die experimentelle Mortalität, beeinträchtigt die interne Validität, wenn sie mit der unabhängigen Variablen zusammenhängt (vgl. HAGER u. WESTERMANN 1983). Bei differentiellen Ausfallquoten (unterschiedliche Ausfallquoten in Experimental- und Kontroll- bzw. Vergleichsgruppe) wird dies im allgemeinen vermutet (BASTINE 1975; CAMPBELL u. STANLEY 1970). In verschiedenen Meta-Analysen wurde gefunden, daß (auch differentielle) Ausfallquoten nur von geringer Bedeutung für die Effektstärke sind (GIBLIN 1986; GIBLIN et al. 1986; SHADISH et al. 1993; SHAPIRO u. SHAPIRO 1982; SMITH et al. 1980). Doch kann daraus nicht geschlossen werden, man müsse dem Thema (differentielle) Ausfallquoten kein Augenmerk schenken. Viel mehr kann das nur heißen, daß man die Sachlage in jedem neuerlichen Einzelfall zu prüfen hat.

8.2 Validität der unabhängigen Variablen

Kommen wir damit zur Variablenvalidität. Die drei betrachteten Untersuchungen zur Wirksamkeit von Erziehungsberatung sind einfaktoriell und multivariat; wir haben jeweils eine unabhängige Variable (mit verschiedenen Ausprägungsgraden) und mehrere abhängige Variablen vor uns. Die unabhängige Variable, das Treatment oder die Behandlungsmethode, hat zwei Ausprägungsgrade bei allen drei Untersuchungen. In der Kontrollgruppen-Untersuchung von FISCHER (1980) sind das „Familientherapie versus keine Familientherapie“ und in den Vergleichsgruppen-Untersuchungen von JOHNSON und MALONY (1977) und TRANKINA (1975) „Familientherapie versus andere Behandlungsform“. Drei Ausprägungsgrade finden wir bei FISCHER (1980, 1984), wenn er prüft, ob drei unterschiedliche Zeitvorgaben für die Familientherapie (6 Sitzungen, 12 Sitzungen oder ohne Vorgabe) unterschiedliche Effekte zeitigen.

Die grundlegende Sicherung der Validität der unabhängigen Variablen besteht darin zu garantieren, daß das intendierte Konstrukt auch wirklich realisiert wird. Ohne große Probleme geht es dabei ab, wenn man auf Zeitvorgaben wie bei FISCHER (1980, 1984) blickt. Größeren Problemen hingegen sieht man sich gegenüber, wenn man fragt, ob denn wirklich dasjenige therapeutische Verfahren geprüft wird, das man zu prüfen vorgibt. Man erfährt aus den Untersuchungen von FISCHER (1980, 1984), JOHNSON und MALONY (1977) und TRANKINA (1975) über die Globalbezeichnung „Familientherapie“ hinaus nicht einmal einen näheren Namen (wie etwa „strategisch“ oder „strukturell“).

Selbst wo dies geschieht, kann man sich nicht sicher sein, daß entsprechendes im therapeutischen Handeln auch realisiert wird. In der Psychotherapie(evaluations)forschung werden deshalb immer Maßnahmen gefordert, die Treatmenttreue garantieren oder sie bezeugen. Im wesentlichen geht es um die Bereitstellung und Befolgung von Behandlungsmanualen, um eine adäquate Überprüfung des tatsächlichen therapeutischen Verhaltens und um Kennzeichnung des Erfahrungs- und Aus-

bildungsstandes der Therapeuten im jeweiligen zur Prüfung anstehenden Verfahren; alles Dinge, die man bei FISHER (1980, 1984), JOHNSON und MALONY (1977) und TRAN-KINA (1975) vermißt – und nicht nur dort (vgl. HEEKERENS 1989). Solche Sicherung der Treatmentreue ist aber dann notwendig, wenn wir Aussagen darüber treffen wollen, was eigentlich wirksam ist. Aus einschlägigen Untersuchungen wissen wir nämlich, daß Therapeuten keineswegs tun, was sie zu tun vorgeben, oder daß sie sich keineswegs so verhalten, wie man es etwa von ihrer benannten theoretischen Orientierung und/oder ihrer bekannten therapeutischen Fort- und Weiterbildung her hätte erwarten dürfen.

Dieses Ergebnis fanden auch DIETZE et al. (im Druck) bei Betrachtung von deutschen Eheberatern im Rahmen des Forschungsprojekts „Beratungsbegleitende Forschung“. Ihr Vorgehen bei dieser Forschung weist zugleich auf eine Möglichkeit hin, wie man bei einer Feldforschung die Frage ganz anders angehen kann, als dies in der feldfernen Forschung üblich ist. Man kann sich nämlich prinzipiell auf den Standpunkt stellen, daß es gar nicht wichtig sei, ob ein ins Auge gefaßtes beraterisch-therapeutisches Vorgehen überhaupt einen wohldefinierten Namen hat, und daß es unter Feldbedingungen auch völlig unrealistisch sei, sich um eine Angelegenheit wie Treatmentreue zu kümmern. Wenn man wissen will, ob Erziehungsberatung wirkt, so die Logik, prüft man einfach, welche Effekte das, was ein Erziehungsberater A tut oder in einer Beratungsstelle X geschieht, hat; katamnestic Studien verfahren oft so.

Der Nachteil einer solchen Einschätzung, der das jeweilige beraterisch-therapeutische Vorgehen als „Black Box“ behandelt, ist ein zweifacher: wir können die Ergebnisse nicht mit anderen – von anderen Erziehungsberatern oder anderen Erziehungsberatungsstellen – vergleichen und wir wissen nicht, was denn am jeweiligen Vorgehen wirksam war. Diese Unsicherheit und Unverbundenheit wird gemildert, wenn in katamnestic Studien nur ein bestimmter Ansatz (wie etwa die Familientherapie) betrachtet wird oder aber in der Auswertung nach wohldefiniert unterschiedlichen Ansätzen getrennt wird.

Noch einen anderen Weg – der bei allen methodischen Bedenken im konkreten Fall erfolgversprechend scheint – schlug die Forschungsgruppe DIETZE et al. (im Druck) ein. Sie erfaßte über ein „Methodeninventar der Beratung“ einzelne Behandlungsmodule wie etwa „Paradoxe Intervention“ oder „Ermutigen zur Eigeninitiative“ und prüfte dann den Zusammenhang der Verwendung dieser Module mit dem Behandlungseffekt. Eine solche korrelationsstatistische Zusammenhangsanalyse ist typisch für eine Prozeß-Ergebnis-Forschung. Die Vorteile dieses Vorgehens liegen klar auf der Hand: die Ergebnisse sind „anschlußfähig“, insbesondere dann, wenn man die einzelnen Behandlungsmodule den von GRAWE (1995) genannten vier therapeutischen Wirkprinzipien zuordnet. Ferner ist ein solches Vorgehen der Praxis, in der einzelne Berater und Therapeuten ihren eigenen Stil entwickelt haben, wesentlich angemessener. Den wesentlichen Nachteil sollte man aber auch nicht verschweigen: Methodiker würden zu Recht darauf verweisen, daß korrelationsstatistisch ermittelte Zusammenhänge nicht ohne weiteres kausal interpretiert werden dürfen, was der Begriff „Wirkfaktoren“ ja beinhaltet. Vielmehr muß dies im Einzelfall immer interpretatorisch glaubhaft gemacht werden.

8.3 Validität der abhängigen Variablen

Die größtmögliche Verletzung der Variablenvalidität bei der oder den abhängigen Variablen bestünde darin, das intendierte Konstrukt nicht zu erfassen. Das intendierte Konstrukt bei Evaluationsstudien heißt allgemein „Verbesserung des biopsychosozialen Wohlergehens“. Welche Effekte bei der Evaluation von Erziehungsberatung von Interesse sind und welche Effektivitätskriterien bei der Bewertung erziehungsberaterischer Maßnahmen relevant sind, bedarf vorgängiger theoretischer Klärung. Erst wenn dies geleistet ist, kann man die Frage beantworten, ob in vorhandenen Evaluationsstudien zur Erziehungsberatung die Validität der abhängigen Variable gewährleistet und wie sie in zukünftigen Studien zu garantieren ist. Dies behält auch Geltung angesichts des meta-analytischen Befundes, daß Charakteristika von Evaluationskriterien keinen bedeutsamen Einfluß auf die Effektstärke haben (SHADISH et al. 1993). Denn ein solches Ergebnis entbindet doch keineswegs automatisch von der geforderten theoretischen Anstrengung.

Recht gering für die Effektstärke ist im einzelnen die Bedeutung einiger immer wieder erwähnter sogenannter kritischer Punkte wie die Zuverlässigkeit oder Reliabilität von Meßinstrumenten und -daten (GIBLIN 1986; GIBLIN et al. 1985; GLASS u. SMITH 1978), der Reaktivität von Effektivitätsmaßen (DUSH et al. 1983; SHADISH et al. 1993; SMITH u. GLASS 1977; SMITH et al. 1980) sowie der Subjektivität von Erfolgsmaßen (GIBLIN 1986; GIBLIN et al. 1985; HAHLWEG u. MARKMAN 1988; SHADISH et al. 1993).

Das freilich darf nicht als Freibrief dafür verwendet werden, die Frage der Qualität der Meßinstrumente als eine zu vernachlässigende Größe anzusehen. Wenn man bei der Evaluation der Erziehungsberatung weiterkommen will in dem Sinne, daß man einerseits im Binnenbereich Evaluationswissen akkumulieren und es mit dem Evaluationswissen außerhalb zu verknüpfen sucht, ist eine bewußte Auswahl und Standardisierung der Meßinstrumente und Evaluationskriterien sinnvoll und notwendig. Eine Forschungsgruppe aus Deutschland, Österreich und der Schweiz hat unter dem Titel „Diagnostik und Evaluation in der Psychotherapie: Empfehlungen zur Standardisierung“ (FYDRICH et al. 1996) einen Standardisierungsvorschlag präsentiert. Dieser Vorschlag bezieht sich vornehmlich auf das Einsatzfeld der Erwachsenenpsychotherapie, könnte aber als Vorbild für ähnliche Bemühungen auf dem Feld der Erziehungsberatung dienen.

8.4 Situations- und Populationsvalidität

Um die Frage der Populationsvalidität mit Exaktheit zu klären, bräuchten wir Angaben über die betrachteten Stichproben einerseits und die Grundgesamtheit andererseits. Über die Grundgesamtheit aller (potentiellen) Nutzer von Erziehungsberatung gibt es aus naheliegenden Gründen keine Angaben. Die Populationsvalidität ist in den bisherigen Untersuchungen (FISHER 1980, 1984; JOHNSON u. MALONY 1977; TRANKINA 1975) gewährleistet und wird in zukünftigen Studien gesichert sein dadurch, daß es zu keiner bewußten oder unbewußten Selektion kommt.

Bei der Abschätzung der Frage, inwieweit in den Studien die Situationsvalidität gewährleistet ist, wird man sich konzentrieren auf die Punkte „Therapeuten“, „Behandlungsort“ und „Behandlungsdauer“. Treffen wir bei diesen drei Punkten nicht auf Abweichungen von dem, was in der täglichen Praxis von Erziehungsberatung vorkommt, kann auch für die Situation Validität angenommen werden.

Zur Situations- oder ökologischen Validität gehört auch die Frage, wann die Messung der abhängigen Variablen erfolgt (HAGER u. WESTERMANN 1983). Vom „Erfolg“ einer Maßnahme kann nur gesprochen werden, wenn er Bestand hat. In der Psychotherapieforschung werden deshalb immer wieder neben Messungen zum Behandlungsende (Posttest-Messungen) solche eine geraume Zeit nach Behandlungsende (Follow-up-Messungen) gefordert, und Studien mit (zusätzlicher) Follow-up-Messung wird in der Regel eine höhere Aussagekraft zugeschrieben als solchen mit (bloßer) Posttest-Messung.

Diese höhere Wertschätzung ist gerechtfertigt, auch wenn meta-analytische Befunde Entwarnung zu signalisieren scheinen. Werden in Meta-Analysen – anders als bei SMITH et al. (1980) – nämlich dieselben, also identische Untersuchungsgruppen zum Post-Test- und Follow-up-Zeitpunkt betrachtet werden, so zeigt sich folgendes: neben einem geringen Anstieg der Effektstärke bei SHOHAM-SALOMON u. ROSENTHAL (1987) und einem ebenfalls geringen Abstieg bei GIBLIN (1986; GIBLIN et al. 1985) dominiert die Gleichheit der ermittelten Effektstärken zu beiden Meßzeitpunkten (ANDREWS u. HARVEY 1981; DUSH et al. 1983; HAHLEWEG u. MARKMAN 1988; LANDMAN u. DAWES 1982; NICHOLSON u. BERMAN 1983; PRIOLEAU et al. 1983; SHADISH et al. 1993; WEISZ et al. 1987).

Wer aus diesen Ergebnissen herausliest, daß sich der Follow-up-Wert durch den Posttest-Wert vorhersagen lasse, muß sich vor Augen halten, daß das allenfalls für eine mittelfristige Perspektive gilt – und im je konkreten Einzelfall gar nicht gelten muß, da Follow-up-Untersuchungen bei (quasi-)experimentellen Studien, wie sie den genannten Meta-Analysen zugrunde liegen, aus praktischen Gründen in der Regel binnen Jahresfrist nach Behandlungsende erfolgen. Es kann aber, wenn man Zeiträume von mehr als einem Jahr nach Behandlungsende ins Auge faßt, bei Psychotherapien zu Verschlechterungen (vgl. zusammenf. SHADISH et al. 1993, S. 1000) oder zum Rückgang einmal erzielter Effekte (CHERNISS u. HERZOG 1996) kommen; es kann aber auch sein, daß sich unter bestimmten Umständen positive Effekte erst recht spät zeigen (HERZOG u. SCHWEITZER 1994). Die Schlußfolgerung ist klar: (Quasi-)Experimente müssen ihren Erhebungszeitraum auf über ein Jahr (wünschenswert bis zu drei Jahren) ausdehnen und neben Posttest- auch Follow-up-Messungen einschließen.

8.5 Statistische Validität

Kommen wir damit abschließend zum Punkt statistische Validität. Gewährleistet sein muß die Angemessenheit sowohl des statistischen Prüfverfahren als auch der gewählten Irrtumswahrscheinlichkeit. Was das zweite anbelangt, so wird man selbst fünf Prozent Irrtumswahrscheinlichkeit als tolerierbar ansehen. Was die Angemessenheit des statistischen Prüfverfahrens betrifft, so wird man im Einzelfall bei der Auswahl eines

statistischen Analyseverfahren prüfen müssen, ob die Daten die dafür erforderlichen Voraussetzungen (Datenniveau, Verteilung) erfüllen.

Einzuhalten ist die oft verletzte Regel, das Verhältnis von Versuchspersonen zu abhängigen Variablen im Verhältnis von 20:1, allenfalls im Verhältnis von 10:1, anzusetzen (SCHMALING u. JACOBSON 1988). Oder es ist Gebrauch zu machen von der Möglichkeit der Bonferroni-Korrektur (KIRK 1968; vgl. KEPPEL 1982), wonach das angenommene Alpha-Niveau dividiert wird durch die Anzahl von Vergleichen.

9 Schlußbemerkungen

Das Fazit ist kurz und klar: Klinische Feldforschung ist nötig und klinische Feldforschung ist möglich – auch und gerade im Feld der Erziehungsberatung. Über die Frage, in welcher Phase der Interventionsforschung (BAUMANN u. REINECKER-HECHT 1991) wir uns im Feld der Erziehungsberatung noch oder schon befinden, wird man noch nachdenken müssen; ebenso darüber, wie man die allfälligen praktischen und technischen Schwierigkeiten überwinden kann.

Literatur

- ABEL, A.H. (1996): Beratung in der Jugendhilfe. *Verhaltenstherapie und psychosoziale Praxis* 28, 49-69.
- ANDREWS, G./HARVEY, R. (1981): Does psychotherapy benefit neurotic patients? *Archives of General Psychiatry* 38, 1203-1208.
- BARKLEY, R.A./GUEVREMONT, D.C./ANASTOPOULOS, A.D./FLETCHER, K.E. (1992): A comparison of three family therapy programs for treating family conflicts in adolescents with attention-deficit hyperactivity disorder. *Journal of Consulting and Clinical Psychology* 60, 450-462.
- BASTINE, R. (1975): Methoden der Psychotherapieforschung. In: SCHRAMML, W.J./BAUMANN, U. (Hg.): *Klinische Psychologie* Bd. 1 (3. Aufl.). Bern: Huber, S. 664-701. (1. Aufl. 1970).
- BAUMANN, U./REINECKER-HECHT, C. (1991): Methodik der klinisch-psychologischen Interventionsforschung. In: PERREZ, M./BAUMANN, U. (Hg.): *Klinische Psychologie* Bd. 2: Intervention. Bern: Huber, S. 64-79.
- BENNUN, I. (1986): Evaluating family therapy: A comparison of the Milan and problem solving approaches. *Journal of Family Therapy* 8, 225-242.
- BLECKMANN, K.H./BODEMANN, H. (1956): Katamnestic Untersuchung bei Patienten aus der Erziehungsberatung. *Praxis der Kinderpsychologie und Kinderpsychiatrie* 5, 11-13.
- BOWERS, T.G./CLUM, G.A. (1988): Relative contribution of specific and nonspecific treatment effects: Meta-analysis of placebo-controlled behavior therapy research. *Psychological Bulletin* 103, 315-323.
- BRACKMANN, S. (1977): Bericht über eine statistische Untersuchung zum Einsatz der Familientherapie und ihrer Auswirkungen im Urteil von Klienten und Therapeuten. In: GERLICHER, K. (Hg.): *Familientherapie in der Erziehungsberatung*. Weinheim: Beltz, S. 107-122.
- BRANDT, G.A. (1967): Probleme und Erfolge der Erziehungsberatung – Ergebnisse einer Erhebung. Weinheim: Beltz.
- BRODY, N. (1990): Behavior therapy vs. placebo: Comment on Bowers and Clum's meta-analysis. *Psychological Bulletin* 107, 106-109.
- BURR, W. (1993). Evaluation der Anwendung lösungsorientierter Kurztherapie in einer kinder- und jugendpsychiatrischen Praxis. *Familiendynamik* 18, 11-21.
- CAMPBELL, D.T./STANLEY, J.C. (1970): Experimentelle und quasiexperimentelle Anordnungen in der Unterrichtsforschung. In: INGENKAMP, K.-H. (Hg.): *Handbuch der Unterrichtsforschung* Bd. 1. Weinheim: Beltz, S. 449-627 (amerikanisches Original: 1963).

- CHERNISS, C./HERZOG, E. (1996): Impact of home-based family therapy on maternal and child outcomes in disadvantaged adolescent mothers. *Family Relations* 45, 72-79.
- DIETZE, F./ENGL, J./HAHLWEG, K./KLANN, N. (im Druck): Zusammenhänge zwischen Methodik und Erfolg von Eheberatung. *System Familie*.
- DILLIG, P. (1993): Katamnese studie an Klienten der Erziehungsberatungsstelle für Landkreis und Stadt Ansbach. In: Eltern-, Jugend- und Familienberatungsstelle für den Landkreis und die Stadt Ansbach (Hg.): Jahresbericht. Ansbach: Eigendruck, S. 15-29.
- DOHERTY, W.J./SIMMONS, D.S. (1996): Clinical practice patterns of marriage and family therapists: A national survey of therapists and their clients. *Journal of Marital and Family Therapy* 22, 9-25.
- DUSH, D.M./HIRT, M.L./SCHROEDER, H. (1983): Self-statement modification with adults: A meta-analysis. *Psychological Bulletin* 94, 408-422.
- EWING, C.P. (1975): Family crisis intervention and traditional child guidance. Unveröff. Diss., Cornell University, Ithaca, USA.
- EYSENCK, H.J. (1952): The effects of psychotherapy. An evaluation. *Journal of Consulting Psychology* 16, 319-324.
- Familie und Beratung (1993): familienorientierte Beratung zwischen Vielfalt und Integration. Stuttgart: Kohlhammer.
- FISHER, S.G. (1980): The use of time limits in brief psychotherapy. *Family Process* 19, 377-392.
- FISHER, S.G. (1984): Time-limited brief therapy with families. *Family Process* 23, 101-106.
- FRIEDMAN, A.S. (1989): Family therapy vs. parent groups: Effects of adolescent drug abusers. *American Journal of Family Therapy* 17, 335-347.
- FYDRICH, T./LAIREITER, A.-R./SAILE, H./ENGBERDING, M. (1996): Diagnostik und Evaluation in der Psychotherapie: Empfehlungen zur Standardisierung. *Zeitschrift für Klinische Psychologie* 25, 161-168.
- GIBLIN, P. (1986): Research and assessment in marriage and family enrichment. *Journal of Psychotherapy and the Family* 2(1), 79-96.
- GIBLIN, P./SPRENKLE, D.H./SHEEHAN, R. (1985): Enrichment outcome research. *Journal of Marital and Family Therapy* 11, 257-271.
- GLASS, G.V./SMITH, M.L. (1978): Reply to Eysenck. *American Psychologist*, 33, 517-519.
- GRAWE, K. (1995): Abschied von den psychotherapeutischen Schulen. *Integrative Therapie* 21, 94-89.
- GRAWE, K./DONATI, R./BERNAUER, F. (1994): Psychotherapie im Wandel – Von der Konfession zur Profession. Göttingen: Hogrefe.
- HAGER, W./WESTERMANN, R. (1983): Planung und Auswertung von Experimenten. In: BREDEKAMP, J./FEGER, H. (Hg.): Hypothesenprüfung (Enzyklopädie der Psychologie, Themenbereich B, Ser. 1, Bd. 5). Göttingen: Hogrefe, S. 24-238.
- HAHLWEG, K./MARKMAN, H.J. (1988): Effectiveness of behavioral marital therapy. *Journal of Consulting and Clinical Psychology* 56, 440-447.
- HAID-LOH, A. (1996): Qualitätsmanagement als Leitungsaufgabe. ezi-Korrespondenz (Evangelisches Zentralinstitut für Familienberatung) H. 14, 4-11.
- HEEKERENS, H.-P. (1983): Institutionelle Aspekte familientherapeutischer Arbeit in der Ambulanz für Kinder und Jugendliche. *Praxis der Kinderpsychologie und Kinderpsychiatrie* 32, 98-105.
- HEEKERENS, H.-P. (1989): Familientherapie und Erziehungsberatung. Heidelberg: Asanger.
- HEEKERENS, H.-P. (1997): Was weiß man von der Wirksamkeit der Erziehungsberatung? *Nachrichten der LAG für Erziehungsberatung in Hessen*, 10-38.
- HEEKERENS, H.-P. (1998): Familientherapie – Zum Stand der Evaluationsforschung. In: MÄRTENS, M. (Hg.): Psychotherapieforschung und die Praxis der Psychotherapie (im Druck).
- HEEKERENS, H.-P./HENKELMANN-STRICKLER, S. (1980): Von der herkömmlichen zur familientherapeutischen Beratung. *Mitteilungen der Evangelischen Landeskirche in Baden* H. 10, 38-42.
- HEGE, M./BICHLMAIER, R. (1965): Katamnestiche Untersuchung über 5jährige Beratungsarbeit an einer öffentlichen Erziehungsberatungsstelle. *Praxis der Kinderpsychologie und Kinderpsychiatrie* 14, 97-111.
- HERZOG, W./SCHWEITZER, J. (1994): Anorexia nervosa – Ergebnisse und Perspektiven in Forschung und Therapie. *Familiendynamik* 19, 182-187.
- HÖFER, R./STRAUS, F. (1991): Familienberatung aus der Sicht ihrer Klienten. Zur Perspektivität der Erfolgsmessung. In: Presting, G. (Hrsg.): Erziehungs- und Familienberatung. Weinheim – München: Juventa, S. 157-198.

- JAGER, G. (1987): Katamnese von Familientherapie: Entwicklung und Erprobung eines Erhebungsinstrumentes an Familienberatungsstellen. Unveröff. Diss., Erziehungswissenschaftliche Hochschule Rheinland-Pfalz, Koblenz.
- JOHNSON, T.M./MALONY, H.N. (1977): Effects of short-term family therapy on patterns of verbal interchange in disturbed families. *Family Therapy* 4, 207-215.
- KATZ, A.J./KRASINSKI, M.D./PHILIP, E./WIESER, C. (1975): Change in interactions as a measure of effectiveness in short term family therapy. *Family Therapy* 2, 31-56.
- KAZDIN, A.E. (1986): Treatment research. In: HERSEN, M./KAZDIN, A.E./BELLACK, A.S. (eds.): *The clinical psychology handbook* (2nd ed.). New York: Pergamon, S. 265-284.
- KAZDIN, A.E. (1991): Effectiveness of psychotherapy with children and adolescents. *Journal of Consulting and Clinical Psychology* 59, 785-798.
- KEPPEL, G. (1982): *Design and analysis*. Englewood Cliffs: Prentice-Hall.
- KEPPLER, E. (1979): Die Effizienz der Erziehungsberatung. Eine empirische Untersuchung an 4 Erziehungsberatungsstellen in Bayern. Unveröff. Diss., Julius-Maximilians-Universität, Würzburg.
- KIRK, R.E. (1968): *Experimental design: Procedures for the behavioral sciences*. Monterey: Brooks/Cole.
- KLANN, N./HAHLWEG, K. (1994): *Beratungsbegleitende Forschung – Evaluation von Vorgehensweisen in der Ehe-, Familien- und Lebensberatung und ihre spezifischen Auswirkungen*. Stuttgart: Kohlhammer.
- Kommunale Gemeinschaftsstelle (1994): *Outputorientierte Steuerung der Jugendhilfe (KGSt-Bericht 4/1995)*. Köln: Eigendruck.
- Kommunale Gemeinschaftsstelle (1995): *Qualitätsmanagement (KGSt-Bericht 6/1995)*. Köln: Eigendruck.
- LAMBERT, M.J./SHAPIRO, D.A./BERGIN, A.E. (1986): The effectiveness of psychotherapy. In: GARFIELD, S.L./BERGIN, A.E. (eds.): *Handbook of psychotherapy and behavior change* (3rd ed.). New York: Wiley, S. 157-211.
- LANDMAN, J.T./DAVES, R.M. (1982): Smith and Glass' conclusions stand up under scrutiny. *American Psychologist* 37, 504-516.
- LASK, B./MATTHEW, D. (1979): Childhood asthma. A controlled trial of family psychotherapy. *Archives of Disease in Childhood* 54, 116-119.
- LENZ, A. (1994a): Die Wirksamkeit von Erziehungsberatung aus der Sicht der Eltern. *Jugendwohl* 75, 303-312.
- LENZ, A. (1994b): Wie bewerten Eltern Erziehungsberatung? In: Bundeskonferenz für Erziehungsberatung (Hrsg.): *Beratung im Blickpunkt*. Fürth: Eigendruck, 32-37.
- LUDWIG, K. (1993): Evaluation systemischer Therapien in einer Kinder- und Jugendpsychiatrie. *System Familie* 6, 21-35.
- Ludwig, K., Schwarz, R. & Kowerk, H. (1984). Systemische Therapie mit Familien von „psychotischen“ Jugendlichen. *Familiendynamik* 9, 108-125.
- MILLER, R.C./BERMAN, J.S. (1983): The efficacy of cognitive behavior therapies. *Psychological Bulletin* 94, 39-53.
- NAUMANN, K./BECK, M. (1994): Effekte von Erziehungsberatung: Eine katamnestiche Studie. In: CREMER, H./HUNDSALZ, A./MENNE, K. (Hg.): *Jahrbuch der Erziehungsberatung Bd.1*. Weinheim: Juventa, S. 253-270.
- NICHOLSON, R.A./BERMAN, J.S. (1983): Is follow-up necessary in evaluating psychotherapy? *Psychological Bulletin* 93, 261-278.
- NICHOLSON, S. (1989): Outcome evaluation of therapeutic effectiveness. *Australian and New Zealand Journal of Family Therapy* 10, 77-83.
- NICHOLSON, S./VIVEKANANDA K. (1985): Client evaluation of therapy in an outpatient adolescent service. *Australian and New Zealand Journal of Family Therapy* 6, 137-143.
- PRIOLEAU, L./MURDOCK, M./BRODY, N. (1983): An analysis of psychotherapy versus placebo studies. *Behavioral and Brain Sciences* 6, 275-310.
- Qualitätssicherung in der Erziehungsberatung (1996): Informationsdienst der Katholischen Bundesarbeitsgemeinschaft für Erziehungsberatung H. 48, 41-44.
- RETZER, A. (1996): Zeit und Psychotherapie. *Familiendynamik* 21, 136-159.
- SAKOFSKI, A./KÄMMERER, A. (1986): Evaluation von Erziehungsberatung. *Zeitschrift für Klinische Psychologie* 15, 321-332.
- SCHMALING, K.B./JACOBSON, N.S. (1988): Social learning family therapy vs. parent training. *Journal of Family Psychology* 1, 292-295.

- SHADISH, W.R./MONTGOMERY, L.M./WILSON, P./WILSON, M.R./BRIGHT, I./OKWUMABUA, T. (1993): The effects of family and marital psychotherapies: A meta-analysis. *Journal of Consulting and Clinical Psychology* 61, 992-1002.
- SHADISH, W.R./RAGSDALE, K./GLASER, R.R./MONTGOMERY, L.M. (1995): The efficacy and effectiveness of marital and family therapy: A perspective from meta-analysis. *Journal of Marital and Family Therapy* 21, 345-360 (deutsch, 1997: Effektivität und Effizienz von Paar- und Familientherapie: Eine metaanalytische Perspektive. *Familiendynamik* 22, 5-33).
- SHAPIRO, D.A./SHAPIRO, D. (1982): Meta-analysis of comparative therapy outcome studies. *Psychological Bulletin* 92, 581-604.
- SHOHAM-SALOMON, V./ROSENTHAL, R. (1987): Paradoxical interventions: A meta-analysis. *Journal of Consulting and Clinical Psychology* 55, 22- 28.
- SIMPSON, L. (1991): The comparative efficacy of Milan family therapy for disturbed children and their families. *Journal of Family Therapy* 13, 267-284.
- SMITH, M.L./GLASS, G.V. (1977): Meta-analysis of psychotherapy outcome studies. *American Psychologist* 32, 542-760.
- SMITH, M.L./GLASS, G.V./MILLER, T.I. (1980): *The benefits of psychotherapy*. Baltimore: John Hopkins University.
- SPECHT, F. (1993): Zu den Regeln des fachlichen Könnens in der psychosozialen Beratung von Kindern und Jugendlichen. *Praxis der Kinderpsychologie und Kinderpsychiatrie* 42, 113-124.
- STARFIELD, B. (1977): Efficacy and effectiveness of primary medical care for children. In: Report of the Harvard Health Project Task Force: Children's medical care needs and treatment Vol. 2. Cambridge, MA: Ballinger, S. 71-76.
- STEINBRUECK, S.M./MAXWELL, S.E./HOWARD, G.S. (1983): A meta-analysis of psychotherapy and drug therapy in the treatment of unipolar depression with adults. *Journal of Consulting and Clinical Psychology* 51, 856-863.
- STRAUS, F./HÖFER, R./GMÜR, W. (1988): *Familie und Beratung; zur Integration professioneller Hilfe in den Familienalltag. Ergebnisse einer qualitativen Befragung von Klienten*. München: Profil-Verlag.
- SZAPOCNIK, J./RIO, A./MURRAY, E./COHEN, R./SCOPETTA, M./RIVAS-VAZQUES, A./HERVIS, O./POSADA, V./KURTINES, W. (1989): Structural family versus psychodynamic child therapy for problematic Hispanic Boys. *Journal of Consulting and Clinical Psychology* 57, 571-578.
- SZYKULA, S.A./MORRIS, S.B./SAYGER, T.V./SUDWEEKS, C. (1987): Child-focused behavior and strategic therapies. *Psychotherapy* 24, 546-551.
- TRANKINA, F.J. (1975): *Aggressive and withdrawn children as related to family perception and outcome of different treatment methods*. Unveröff. Diss., University of Arizona, Tuscon, USA.
- WEISZ, J.R./WEISS, B./ALICKE, M.D./KLOTZ, M.L. (1987): Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology* 55, 542-549.
- WEISZ, J.R./WEISS, B./DONENBERG, G.R. (1992): The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist* 47, 1578-1585.
- WITTMANN, W.W./MATT, G.E. (1986): Meta-Analyse als Integration von Forschungsergebnissen am Beispiel deutschsprachiger Arbeiten zur Effektivität von Psychotherapie. *Psychologische Rundschau* 37, 20-40.
- ZÜRN, H./BOSSSELMANN, R./ARENDET, G./LIEBL-WACHSMUTH, W. (1990): Wie ging's denn weiter? - Ergebnisse und Erkenntnisse aus der Nachbefragung eines Klientenjahrgangs. *Praxis der Kinderpsychologie und Kinderpsychiatrie* 39, 185-190.

Anschrift des Verfassers: Prof. Dr. Dr. Hans-Peter Heekerens, Am Schloßpark 6a, 86949 Windach.