



Pascal Wallisch  
Freie Universität Berlin  
Matrikelnummer 34 0 34 68  
Tel.: 030 / 83 22 98 14

## 1 Einleitung

Der Gegenstand meiner Ausarbeitung ist, wie der Titel bereits erkennen läßt, sehr stark auf die konkrete praktische Durchführung von Evaluationsprogrammen ausgerichtet.

Vor allem im Kontext anderer eher theoretischen Themenkomplexe in der Evaluationsforschung mag dies regelrecht profan erscheinen.

Tatsächlich ist jedoch eine fundierte Erörterung der folgenden Punkte unerlässlich oder zumindest sehr hilfreich, wenn es eben darum geht, eine Evaluation tatsächlich durchzuführen. Es wird im Rahmen dieser Arbeit nicht möglich sein, dieses Thema in einer erschöpfenden Systematik abzuhandeln – dazu ist das Thema zu nahe an der Evaluationspraxis. Es macht Sinn, sich diese Einschränkung und Besonderheit von vorne herein bewußt zu machen – oft wird man sich mit einer reinen Kasuistik begnügen müssen, die den Ansprüchen einer systematischen Theorie nicht genügen kann. Dies tut allerdings der Nützlichkeit und Relevanz dieses Wissens in der Praxis keinen Abbruch. Dabei werden auch bewußt die sogenannten „Dilemmata der Evaluation“ ausgespart, da diese bereits mehrfach in anderen Zusammenhängen besprochen wurden.

Schwerpunkt meiner Arbeit ist die unter anderem die Planungsphase von Evaluationen. Besonders bei größeren Evaluationsprojekten steigt die Notwendigkeit für sorgfältige, langfristige und intensive Planung unverhältnismäßig stark an. Die meisten Fehler, die später gemacht werden und die meisten Probleme die später auftreten können, sind bereits in der Planungsphase angelegt und können präventiv vorweggenommen werden, wenn man die Planung ernst nimmt. Es lohnt sich also, sich dieser Thematik intensiv zu widmen, was im folgenden geschehen wird.

## 2 Durchführung von Evaluationsstudien

Unter Durchführung werden hier die Unterpunkte „**Organisation**“, „**Auswertung**“ und „**Berichtlegung**“ behandelt: Wie gesagt sind – vor allem bei größeren Studien – Schwierigkeiten bei der praktischen Durchführung von Evaluationsstudien geradezu zu erwarten. Unter „**Organisation**“ sollen Ansätze diskutiert werden, die es erlauben, mit den zu erwartenden Schwierigkeiten umzugehen. Bei der Durchführung der **Auswertung** treten sowohl diverse praktische, als auch prinzipielle Probleme auf, die hier besprochen werden. Die **Berichtlegung** ist im Kontext von Evaluationen eine weitere Eigenheit praktischen Arbeitens, da man seine Befunde meist nicht einem Fachpublikum darlegt; vielmehr müssen

die Ergebnisse an – oft fachfremde – Personen wie Auftraggeber oder die Öffentlichkeit kommuniziert werden. Aus dieser Konstellation erwachsen einige Besonderheiten, auf die hier eingegangen werden soll.

## **2.1. Organisation**

Ein Großteil der organisatorischen Fragen sind Fragen der Personalführung. Naturgemäß sind diese Fragen für kleine Betriebe sogar kritischer als für größeren Unternehmen, bzw. treten diese Probleme verschärft zutage. Dies liegt daran, daß in vielen großen Unternehmen eigenständige Personalabteilungen mit entsprechend qualifizierten Fachmännern vorhanden sind, die eine langfristig angelegte Politik der Personalentwicklung- und führung verfolgen. In kleinen Unternehmen fehlen diese meist, was vor allem bei unvorhergesehenen Problemen oder Ausfällen gravierende Auswirkungen haben kann, da eine Kompensation schwierig ist. Da es sich bei viele Evaluationsunternehmen um eher kleine Firmen handelt, ist diese Problematik weit verbreitet.

Auch sonst tun sich kleinere Unternehmen schwerer, da sie sich mit ihrem kleinen Personalstamm auf die „eigentliche Arbeit“ konzentrieren müssen und sich oftmals keinen hinreichenden Verwaltungsapparat leisten können. Dies führt fast zwangsläufig zu einem gewissen Dilletantismus auf diesem Gebiet, was sich in Form ganz profaner Fehler, wie etwa der Vernachlässigung von Lohnnebenkosten oder ähnlichen Fehlkalkulationen, z.B. der Überschätzung der realen Jahresarbeitszeit pro Mitarbeiter niederschlagen kann.

Auf der anderen Seite bieten kleine Evaluations-Institute aber auch einige Chancen, die man bereits in der Organisation der jeweiligen Studie nutzen kann:

Eben der Verzicht auf formale Instrumente der Personalführung, ein Fehlen starrer Hierarchien und eine minimale Bürokratie bedingen oft ein gutes Arbeitsklima, in dem ein guter persönlicher Kontakt zu allen Kollegen möglich ist.

Auch wird den einzelnen Mitarbeitern mitunter eine größere persönliche Verantwortung für das jeweilige Projekt übertragen, was die Identifikation und Motivation des Mitarbeiters erhöht. Ist er diesen, durch die höhere Verantwortung bedingten gesteigerten Herausforderungen gewachsen, kann dies zur verstärkten Entfaltung individueller Kompetenzen führen, was wiederum der Arbeit am Evaluationsprojekt sehr zugute kommt.

Abgesehen von diesen grundlegenden Rahmenbedingungen sind bei der Planung und Durchführung von Evaluationsstudien vor allem 3 Punkte wichtig: **Zeitplanung, Kosten und Qualitätsstandards.**

- **Zeitplanung:** Je größer und aufwendiger das Projekt ist, desto sinnvoller ist die Erstellung eines detaillierten Zeitplanes. Dabei ist auf die Strukturierung des Projektes in Abschnitte und das Setzen von Zwischenzielen wichtig. Die Erreichung dieser Zwischenziele kann dann bei der Durchführung als Maßstab für den Fortschritt des Projektes dienen. Es ist ratsam, hinreichend große Pufferzonen vorzusehen, damit unvorhergesehen auftretende Schwierigkeiten nicht gleich den Erfolg des ganzen Projektes gefährden.
- **Kosten:** Dieser Punkt ist für ein Privatunternehmen eigentlich selbstverständlich.. Allerdings fällt eine vernünftige Kostenplanung erfahrungsgemäß vielen Sozialwissenschaftlern außerordentlich schwer. Bei der Erstellung des Kostenplanes sollten sowohl Erkrankungen von Mitarbeitern als auch inflationsbedingte Preissteigerungen eingeplant werden. Auch hier macht es Sinn, einen wesentlichen Teil des Budgets als Reserve für Notfälle zurückzuhalten. Eine zu riskante Unternehmenspolitik in Kostenfragen stellt eine wesentliche Bedrohung der Existenz des jeweiligen Evaluations-Institutes dar. Wichtig ist hier somit auch das ständige Controlling bei der Durchführung des Projektes, d.h. die wöchentliche Gegenüberstellung von geplanten und tatsächlichen Ausgaben, um die Auswirkungen von Fehlplanungen möglichst schnell auffangen und gegensteuern zu können.
- **Qualitätsstandards:** Natürlich sollten die als „state of the art“ anerkannten Prozeduren der Evaluationsforschung eingesetzt werden, um eine möglichst seriöse Ausführung des Auftrages zu gewährleisten. Dazu gehört unter anderem der Einsatz gängiger Erhebungsverfahren, die Verwendung geprüfter Erhebungsinstrumente, sowie die Erfassung quantitativer Parameter wie etwa Rücklaufzeiten von Fragebögen, Verweigerungsraten von Interviews oder Dropoutraten aller Art. Daneben obliegt es dem Projektleiter und seiner Erfahrung, sich einen Eindruck vom Fortgang des Projektes zu machen, indem er Vergleiche zu früheren Projekten anstellt und gegebenenfalls steuernd eingreift, z.B. durch Interviewerschulungen, Nacherhebungen, etc.

Sonstige organisatorische Probleme fallen in den Personalbereich und können in zwei Hauptkategorien eingeteilt werden:

**a) Die organisatorischen Probleme aus Sicht des Arbeitgebers:**

- Der Arbeitsmarkt für Evaluationsspezialisten ist klein. Dadurch ist die Anwerbung dieser Spezialisten für zeitlich befristete Projekte nur schwer möglich. Dies führt zu einem Mangel an hinreichend qualifiziertem Personal zum Projektbeginn. Die Lösung besteht hier in der rechtzeitigen Einstellung dieser Fachkräfte, bzw. einer Einschulung von Anfängern im Rahmen einer nachhaltigen Personalpolitik, die sich aber wiederum kleinere Institute kaum leisten können.
- Die wenigsten Fachwissenschaftler sind qualifiziert, „Managementaufgaben“ wie die Projektleitung oder den Kontakt zum Auftraggeber verantwortlich wahrzunehmen. Problematischerweise kann ein Mißmanagement in diesen Bereichen die wirtschaftliche Existenz von jungen Instituten gefährden. Unternehmen mit einem erfahreneren Personalstamm tun sich hier leichter.
- Immer wieder kommt es zu kleineren Katastrophen aufgrund der Folgen gruppenspezifischer Prozesse. Diese können zur Kündigung von Schlüsselmitarbeitern oder zu Racheakten, wie der Sabotage des Projektes, z.B. durch Löschung oder Verfälschung von Daten führen, woran die Durchführung des ganzen Projektes scheitern kann. Abhilfe kann auch hier eine sensible, möglichst mitarbeiterorientierte Personalführung und Projektleitung schaffen, die neben den Fachfragen auch die sozialen Aspekte des Arbeitens im Team ernst nimmt und berücksichtigt.

**b) Die organisatorischen Probleme aus Sicht des Arbeitnehmers:**

Oft wird die eigentliche Arbeit durch alle Arten von Störfaktoren beeinflusst, die sich negativ auf die Motivation der Mitarbeiter auswirkt. Dazu gehören:

- Auf die Projektdauer befristete und damit als unsicher empfundene Arbeitsplätze
- Begrenzte Qualifikationschancen aufgrund geringer Beschäftigungsdauer
- Geringe Aufstiegsmöglichkeiten, wenn nur wenige Posten im Unternehmen vorhanden, und diese fest besetzt sind.

Abschließend sind im Rahmen organisatorischer Aspekten von Evaluationsstudien noch die gelegentlich auftretenden „unerwarteten Störfälle“ zu erwähnen.

Auch die sorgfältigste Planung kann nicht völlig vor dem Auftreten dieser Störfälle schützen.

Im wesentlichen kann man sagen, daß es sich hierbei um eine Eigenheit des praktischen Arbeitens handelt, mit der man irgendwie zurecht kommen muß. Als Leitheuristik könnte man sagen: Es macht Sinn, das Unerwartete zu erwarten. Früher oder später werden Fehler und Probleme auftreten – gut wenn der Projektleiter gewappnet ist und diesen mit einer professionellen Einstellung gegenübertritt, anstelle sich z.B. in fruchtlosen Schuldzuweisungen oder Selbstvorwürfen zu verlieren.

Unerwartete Störfälle können aufgrund vielfältigster Ursachen auftreten, z.B.:

- kann sich die Zielsetzung des Projektes ändern, (durch den Auftraggeber)
- der eigentliche Untersuchungsgegenstand kann entfallen, (z.B. die Produktion eines Produkts eingestellt werden)
- unerwartete Reaktionen, d.h. wenn eine Maßnahme nicht akzeptiert wird oder sich die zu evaluierenden Personen einer Evaluation verweigern.
- Auftreten von Artefakten in Folge des Einsatzes ungeeigneter Methoden.
- Wirkliche Katastrophen (z.B. Brände → Datenvernichtung).
- Kündigung und Sabotage (siehe oben).

Gegen einige dieser Ereignisse kann man sich schützen (z.B. durch Datensicherung) oder Versicherungen; das Hauptproblem an diesen Ereignissen ist aber, daß alles mögliche passieren kann, die Einzelwahrscheinlichkeit für das Auftreten eines Ereignisses gering ist, aber die Folgen des tatsächlichen Auftretens eines solchen Ereignisses verheerend sein können. Ein effektiver Schutz gegen alle Eventualitäten zusammen ist entweder zu teuer oder inpraktikabel.

Da man Katastrophen aller Art aber nie ganz ausschließen kann, ist es ratsam, grundsätzlich ein möglichst kulant, auf gegenseitigem Vertrauen und Respekt basierendes Verhältnis zum Auftraggeber zu pflegen. Bringt man diesem für seine Nöte (z.B. plötzliche Modifikationen des Auftrages) Verständnis entgegen, steigt die Chance, bei eigenen Nöten ebenfalls auf Verständnis zu treffen, anstatt mit Konventionalstrafen oder Prozessen bedroht zu werden.

## 2.2. Auswertung

Dieser Abschnitt gliedert sich in **Datenqualität**, **Datenschutz** und **Auswertungsprobleme**.

### 2.2.1. Datenqualität:

In allen Projektphasen, d.h. bei der Datenerhebung, Datenaufbereitung und der Datenauswertung kann es zu Fehlern kommen welche die Datenqualität gefährden. Dieses Problem ist vor allem deshalb kritisch, da es prinzipiell die Gültigkeit der durch die Studie gewonnenen Erkenntnisse und Schlußfolgerungen bedroht. Hierbei handelt es sich tatsächlich um eine Kernfrage, während in anderen Abschnitten teilweise eher Rahmenbedingungen oder grundsätzliche Voraussetzungen einer erfolgreichen Evaluationstätigkeit abgehandelt werden. Kurz: Mit der durch die Datengewinnung erreichten Datenqualität steht und fällt der Erfolg der ganzen Unternehmung – zumindest was den wissenschaftlichen „Nährwert“ der Studie angeht. Dieses Problem wächst mit dem Umfang des Projektes, da der einzelne Projektmitarbeiter dadurch sehr leicht den „Kontakt“ mit den Daten verliert, damit auch eine zuverlässige Einschätzung der Datenqualität. Dies macht die Einführung von Kontrollmechanismen notwendig und sinnvoll.

Potentielle Bedrohungen für die Datenqualität erwachsen in allen Projektphasen in denen Daten im Spiel sind: Vor allem bei der **Datenerhebung**, aber auch bei der **Datenaufbereitung** und der **Datenauswertung**. Die ersten beiden Abschnitte beziehen sich eher auf praktische Probleme, der letztere eher auf prinzipielle.

#### 2.2.1.1. Die Datenqualität bei der Datenerhebung

In diesem Abschnitt konzentriere ich mich vor allem auf die zwei gebräuchlichsten Datenerhebungsmethoden in der Evaluationsforschung: Das Interview und den Fragebogen. Hierbei treten die üblichen Probleme dieser Verfahren auf – sie beschränken sich keineswegs auf die Evaluationsforschung, werden hier aber dennoch nochmals einzeln aufgeführt.

## **Das Interview:**

Generell sollten die Rahmenbedingungen eine ehrliche Datenerhebung begünstigen und die Wahrscheinlichkeit von intentionalen und unintentionalen Verfälschungen minimieren.

Eine Hauptquelle von Fehlern beim Interview ist wohl der Interviewer selbst.

Empirisch gesehen zeigte sich, daß sich folgende Faktoren negativ auf die Zuverlässigkeit der Interviewer und damit auf die Datenqualität auswirkten, da sie die Motivation und Bereitschaft des Interviewers, die Mühe einer ehrlichen Datenaufnahme auf sich zu nehmen senken:

- Schlechte Bezahlung.
- Bezahlung auf „Erfolgsbasis“, d.h. nur für erfolgreich abgeschlossene Interviews. Dies erhöht die Neigung des Interviewers, verstärkt Leute zu befragen, die eine hohe Antwortwahrscheinlichkeit besitzen. Dies kann zu systematischen Verzerrungen führen, evtl. werden Extremgruppen unterrepräsentiert, was die Repräsentativität der Befragung gefährdet. Dies ist problematisch, wenn die Befragung repräsentativ sein sollte (was oft genug der Fall ist).
- Keine Vorgabe von Namen und Adressen. Auch dies macht es wahrscheinlicher, daß der Interviewer sich einfach irgendwelche Leute nimmt, da der Auftraggeber oder das Unternehmen nicht kontrollieren kann auf welche Weise die Stichprobe zustande kam. Dies öffnet der „Bequemlichkeit“ des Interviewers Tür und Tor.
- Zu strikte Quotenvorgaben. Dies kann den Interviewer in große Bedrängnis bringen, wenn er einfach nicht genügend Vertreter der entsprechend geforderten, hochexotischen Randgruppe auffinden kann die zu einem Interview zur Verfügung stehen. Da er meist auf Erfolgsbasis bezahlt wird, steigt wiederum die Versuchung, die Stichprobe im Angesicht der Verzweiflung einfach mit irgendwelchen Leuten aufzufüllen, die gerade greifbar sind – oder gar schlicht und einfach Daten zu erfinden.
- Alle Arten mangelhafter Arbeitsbedingungen, wie z.B. schlechte Adreßvorgaben, schlechte Fragebögen die zu vielen Verweigerungen führen, etc. Diese Rahmenbedingungen sind sehr wichtig, da sie die subjektive Wahrnehmung der „Professionalität“ des Vorhabens des Interviewers beeinflussen und so dessen Motivation modulieren, selbst gute Arbeit zu leisten.
- Identifikation von Interviewern mit Evaluationsergebnissen: Wenn der Interviewer selbst dem Ergebnis und dem Zweck der Studie gegenüber nicht neutral ist, kann es hier zu massiven systematischen Verzerrungen und Fehlern kommen, die man geradezu als Manipulation bezeichnen kann. Die Validität der Studie wird hierdurch sehr stark

gefährdet. Die Glaubwürdigkeit des Institutes auch, wenn bekannt wird, daß dessen Mitglieder nicht unparteiisch und neutral genug sind. Vor allem bei manchen Studien mit politischem Hintergrund oder möglichen politischen Implikationen ist diese Frage hochbrisant.

Möglichkeiten der Reduzierung der oben beschriebenen Gefährdungen:

- Organisatorische Maßnahmen sollten optimale Rahmenbedingungen schaffen, die den Interviewer mit hochqualitativen Arbeitsmaterialien, einem klaren Auftrag und einer angemessenen Bezahlung ins Feld schicken.
- Durch Personalselektionsmaßnahmen kann man gewährleisten, durch entsprechend neutral eingestellte Mitarbeiter zu möglichst unverfälschten Ergebnissen zu kommen.
- Personalschulungsmaßnahmen. Idealerweise vermitteln diese bei den Interviewern zu einem Bewußtsein für die methodischen Probleme die aus unsauber durchgeführten Interviews oder einer unsauber gezogenen Stichprobe erwachsen, was zu einer intrinsischen Motivation der Mitarbeiter, gute Arbeit zu leisten führen kann.
- Verweis auf die Durchführung von Kontrollen sowie anschließend tatsächlich stattfindende Durchführung von Kontrollen, d.h. nachfassende Kontrollen die prüfen, ob angegebene Interviews auch tatsächlich durchgeführt wurden und ob die Durchführung den geforderten Standards entsprach. Diese Maßnahme ist zwar sehr effizient, aber naturgemäß sehr aufwendig und läuft dem Gebot des Datenschutzes zuwider (siehe unten), weshalb dies meist nur stichprobenhaft möglich ist.
- Ordnet man dem jeweiligen Interviewer nur eine kleine Anzahl von Fällen zu (10-15) und setzt viele Interviewer ein, kann man Interviewereffekte reduzieren.
- Setzt man verschiedene Interviewer ein, kann man auch die jeweils erhobenen Substichproben statistisch hinsichtlich Mittelwerten und Varianz miteinander vergleichen und sehen, ob es zu starken Abweichungen im Vergleich zur Gesamtstichprobe kommt. Hier liegt dann der Verdacht der Manipulation oder der unsauberen Arbeitsweise nahe.

### **Die schriftliche Befragung:**

Ähnlich wie beim Interview liegt eines der Hauptbedenken bei der Frage, wie die Daten überhaupt zustande gekommen sind. Während man beim Interview wenigstens noch prinzipiell Kontrolle auf das Zustandekommen der Daten ausüben kann, hält sich dies bei der schriftlichen Befragung noch stärker in Grenzen.

Hier kann man wenig machen, evtl. die Fragebögen mit Lügenitems oder Plausibilitätsitems versehen um sicherzustellen, daß sie aufrichtig und im Vollbesitz der geistigen Kräfte ausgefüllt wurden, bzw. an die Aufrichtigkeit der Befragten appellieren.

Das zweite, eher praktische Hauptproblem der schriftlichen Befragung ist die Rücklaufquote. Diese ist oft unerträglich niedrig, d.h. ein Großteil der Befragten antwortet überhaupt nicht, wobei völlig unklar ist, ob dieses „nicht-antworten“ zufällig ist, d.h. die Repräsentativität der Stichprobe nicht gefährdet oder ob die Wahrscheinlichkeit zu antworten systematisch mit den Ausprägungen der abgefragten Dimensionen korreliert, was einer systematischen Verzerrung gleichkommt. Dies ist in der Psychologie als „non-response“-Problem bekannt.

Ist der Dropout hoch, kann man versuchen die antwortende Population mit bekannten Populationsparametern zu vergleichen (z.B. hinsichtlich Geschlecht, Alter oder sozioökonomischem Status) und evtl. auftretende Diskrepanzen durch Nachfassen oder ähnlichem auszubügeln.

In jedem Falle möchte man den Dropout, d.h. den Anteil der Non-responder möglichst gering halten. Folgende Methoden erwiesen sich empirisch als besonders geeignet, die Rücklaufquote zu erhöhen:

- Man sollte frankierte Rückumschläge beilegen. Dies führt zu höheren Rücklaufquoten als einfach freigestempelte Rückumschläge. Dies ist evtl. darauf zurückzuführen, daß sich die angeschriebenen Personen persönlich verantwortlicher fühlen, die vom Institut offensichtlich eingesetzten Ressourcen nicht zu vergeuden.
- Aufklärung: Ausführliche Instruktionen über Sinn und Zweck der Studie, Erklärung der Auswirkung von Dropouts und Antwortverweigerungen, glaubwürdige Zusicherung der Anonymität durch Erklärung der Anonymisierungsverfahren.
- Möglichst seriöses Anschreiben, in welchem deutlich sichtbare Zeichen der Seriosität enthalten sein müssen, wie etwa der Institutsbriefkopf, etc.
- Ankündigungsschreiben: Es hat sich gezeigt daß die Rücklaufquote stark steigt, wenn die Befragten nicht unaufgefordert gleich mit dem Fragebogen selbst konfrontiert werden, sondern entweder im Vorfeld deren Bereitschaft zur Teilnahme erfragt wird oder zumindest die Befragung vorher angekündigt wird.
- **Nachfassen.** Offenbar sind die meisten Befragten einfach nur nachlässig. Als wirksamste Methode zur Erhöhung der Rücklaufquote erwies sich deshalb das wiederholte Nachfassen durch Erinnerungsschreiben. Dies kann natürlich die Zusicherung der Anonymität diskreditieren, da man zum Nachfassen die Personen identifizieren können muß.

- Ansprechende formale und sprachliche Gestaltung der Bögen, sowie Untersuchungen über aktuelle Themen. Je relevanter und professioneller der Fragebogen von den Befragten wahrgenommen wird und je sinnvoller ihnen die Befragung erscheint, desto eher sind sie bereit, ihre Mühe und Zeit zu opfern, um dem Befragungsersuchen nachzukommen.
- Deutliche Unterscheidung des Briefumschlags von Reklamesendungen. Vor allem in Städten wird die Belastung durch unangeforderte Reklamesendungen langsam unerträglich. Diese wandern zunehmend unbesehen in den Müll. Damit das Befragungsschreiben nicht dem gleichen Schicksal erliegt, sollte es sich dabei leicht erkenntlich nicht um Werbung handeln.
- Auch die Angabe einer Deadline, d.h. eines letzten Rücksendedatums erwies sich einer Erhöhung der Rücklaufquote zuträglich.
- Ist es für das Institut möglich und dem Evaluationsansinnen nicht schädlich, können auch finanzielle Anreize die Motivation der Teilnehmer und damit die Response-Rate erhöhen.

#### **2.2.1.2. Die Datenqualität bei der Datenaufbereitung**

Bei der Datenaufbereitung handelt es sich um den nötigen Schritt der Übertragung (Transkription) der Rohdaten aus den Fragebögen oder der Kodierung der Daten aus den Interviews in eine Form die weiterverarbeitet und ausgewertet werden kann, meist die Eingabe in ein elektronisches Datenverarbeitungssystem.

Bei diesen Kodierungstätigkeiten handelt es sich in der Regel um sehr langweilige und monotone Tätigkeiten, die von Hilfskräften durchgeführt werden. Diese werden oft speziell für diese Tätigkeit beschäftigt und sind daher im allgemeinen schlecht bezahlt und haben keinen Bezug zum Projekt an sich, weshalb sich deren Motivation stark in Grenzen hält. Dadurch sind hohe Fehlerquoten bei der Durchführung dieser Tätigkeit zu erwarten, insbesondere da meist enorme Datenmengen zu verarbeiten sind.

Aus diesem Grund sind Kontrollmechanismen erforderlich, die mögliche Fehler reduzieren. Optimal wäre eine unabhängige Doppelarbeit, d.h. eine doppelte Verarbeitung desselben Datensatzes. Dies wird jedoch aus Kostengründen nur selten praktiziert.

Die üblichsten Methoden zur Fehlerreduzierung sind daher:

- Bezahlung der mit den Arbeiten beauftragten Hilfskräfte nach Zeit anstatt nach Stückzahl verarbeiteter Datensätze, was die Sorgfalt in der Regel erhöht.

- Stichprobenartige Kontrolle der eingegebenen Daten verbunden mit finanziellen Sanktionen für die Hilfskräfte bei entdeckten Fehlern.
- Plausibilitätskontrollen: Manche Konfigurationen von (an sich) unabhängigen Merkmalen sind einfach sehr unwahrscheinlich, wie etwa minderjährige Rentner und ähnliches. Insbesondere bei EDV-gestützten Datenbanken ist eine solche Kontrolle durch entsprechende Software leicht möglich.

In Zukunft wird dieser Schritt durch die verstärkte Nutzung internetgestützter Erhebungsverfahren aber wohl ohnehin zunehmend entfallen, d.h. seine Bedeutung reduziert.

### **2.2.1.3. Die Datenqualität bei der Datenauswertung**

Hier treten im EDV-Zeitalter Fehler nur noch selten auf. Allerdings sind potentiell auftretende Fehler oft gravierend, da sie leicht unbemerkt bleiben und doch verheerende Auswirkungen haben, wie etwa die Verwechslung von Dateien, die Verwechslung von Variablennummern oder die Verwendung von fehlerhaft funktionierenden Auswertungsprogrammen.

Ohne unabhängige Prüfungen der Ergebnisse bleiben diese Fehler fast mit Sicherheit unerkant.

### **2.2.2. Datenschutz**

Ein Nachteil des EDV-Zeitalters ist zweifellos die grundsätzlich prinzipiell vorhandene und dauerhafte Bedrohung der Privatsphäre und der Persönlichkeitsrechte des Individuums durch die elektronische Massenverarbeitung von Daten. Deshalb spielt der Datenschutz zurecht eine gewichtige Rolle bei jeder Art von Unternehmung, die Daten von Personen gewinnt – so auch bei der Evaluation.

Die Wahrung des Persönlichkeitsrechte wird sehr ernst genommen. Im Zweifelsfall werden deshalb (auch bei innerbetrieblichen Evaluationen) Projekte eher gestoppt als genehmigt.

Maßgeblich ist in diesem Zusammenhang in Deutschland das Bundesdatenschutzgesetz, welches den einzelnen vor den Gefahren der Verbreitung von Daten, die sich auf Personen beziehen schützt und die legale Datenverarbeitung personenbezogener Daten regelt.

Glücklicherweise wird bei den meisten Evaluationsprojekten keine personenbezogene Datenauswertung vorgenommen, d.h. die Datenschutzproblematik bereitet für die meisten Projekte keine Probleme.

An sich ist der (an der Evaluation beteiligte) Psychologe rechtlich gesehen jedoch Geheimnisträger, d.h. er darf das Vertrauen der befragten Personen nicht mißbrauchen und sollte sich seiner besonderen Verantwortung bewußt sein, da eine Mißachtung dieses Status für ihn durchaus auch rechtliche Konsequenzen nach sich ziehen könnte. Psychologen sind also an das Berufsgeheimnis gebunden und können sich des Strafbestandes der Geheimnisverletzung schuldig machen.

Ist es jedoch erforderlich, die Daten personenbezogen verarbeiten (aus welchem Grund auch immer), setzt dies laut Bundesdatenschutzgesetz die informierte Einwilligung der Betroffenen voraus.

Informierte Einwilligung ist definiert als: Freiwillig, d.h. ohne Konsequenzen für das Individuum mit voller Aufklärung über den Zweck des Projekts und seine Auftraggeber. Außerdem ist eine Verarbeitung für andere Zwecke ausgeschlossen, die Leiter der Studie sind für Einhaltung der Bestimmungen verantwortlich und können haftbar gemacht werden. Dies führt in der Psychologie im allgemeinen und der Evaluationsforschung im besonderen zu folgenden drei Problemen:

- Oft ist es nicht sinnvoll, die Befragten über den (wahren) Zweck der Untersuchung aufzuklären, da dies entweder die Reaktion der Befragten beeinflusst oder die Verweigerungsraten erhöht.
- Nicht selten ändern sich die Rahmenbedingungen der Untersuchung, d.h. Personen können Einblick in Daten bekommen (z.B. durch Beförderung), zu denen sie keinen Zugang hatten als die Evaluation durchgeführt wurde – mit unvorhersehbaren Konsequenzen für alle Beteiligten – was auch dem Bundesdatenschutzgesetz zuwiderläuft.
- Manchmal entwickelt man noch Hypothesen die zum Zeitpunkt der Datenerhebung noch nicht vorhanden waren. Die Prüfung dieser neuen Hypothesen mit den vorhandenen Daten ist aber wegen des Datenschutzgesetzes nicht möglich, man darf die Daten nicht anderen Zwecken zuführen.

### **2.2.3. Datenauswertung**

Hierbei handelt es sich eher um ein prinzipielles Problem: Man muß sich nämlich für einen bestimmten Auswertungsplan entscheiden. Dieser kann aber kaum alle theoretisch möglichen Auswertungen beinhalten. Schon bei der Erhebung relativ geringer Variablenzahlen

explodiert die Zahl der theoretisch rechenbaren Korrelationen schnell. Läßt man Modelle mit mehreren Variablen als Prädiktor zu, geht die Zahl der berechenbaren multiplen Korrelationen schnell in die Millionen. Dies führt zu enormen Problemen bei der Interpretation der Daten, da einerseits die Übersichtlichkeit völlig verloren geht, andererseits auch bei völlig zufälligen Daten zigtausend signifikante Einzelergebnisse zu erwarten sind.

Man ist also genötigt, sich vorab für eine „theoriegeleitete“ Auswertung zu entscheiden, d.h. vorab selektiv festzulegen, welche Berechnungen durchgeführt werden und welche nicht.

Man hat somit als Projektleiter eine enorme Verantwortung, Variablen auszuwählen die wichtig und relevant erscheinen und gleichzeitig die Perspektive von Auftraggebern und potentiellen Kritikern zu berücksichtigen, da man sonst leicht in den Verdacht kommt, Manipulation zu betreiben, bzw. nur die eigenen Vorurteile und Erwartungen zu bestätigen.

Ideal wäre an sich die Auswahl der auszuwertenden Zusammenhänge aufgrund der bereits in der Literatur berichteten wissenschaftlichen Erkenntnisse. Diese sind jedoch für konkret durchzuführende Evaluationen oft dürftig oder nicht wirklich anwendbar, weshalb sich fast jede Zusammenhangsanalyse plausibel begründen ließe.

Dies bietet allen Interessengruppen denen das Ergebnis der Studie mißfällt (vor allem im politischen Bereich) eine breite Angriffsfläche für Kritik.

Eine praktikable Lösung dieses Problems ist die sogenannte antizipatorische Konsensbildung bei der man bereits im Vorfeld der Studie versucht, mit Vertretern anderer Schulen oder anderer Denkrichtungen einen Konsens über die auszuwertenden Effekten zu erreichen. Bei politisch brisanten Themen ist die Kritik entsprechender Parteien und Verbände auch vorhersehbar, d.h. man sollte die Auswertung durch die Berücksichtigung der Einwände die vermutlich vorgebracht werden wasserdicht machen.

Letztendlich bleibt das Problem jedoch bestehen – jede Art des Auswertungsplanes ist in der Praxis nur eine mögliche Konkretisierung der Menge der möglichen Auswertungspläne.

### **2.3. Berichtlegung**

Die Dokumentation der Evaluationsstudie dient der Kommunikation der Ergebnisse an die Auftraggeber. Diese sollte tatsächlich auch so transparent erfolgen, daß eine klare Kommunikation dieser Ergebnisse gegeben ist. Insbesondere akademisch geschulte Evaluatoren sollten nicht vergessen, daß sie das Abstraktionsniveau, den Einsatz von

Fachbegriffen und den Medieneinsatz dem jeweiligen Publikum anpassen sollten, um Mißverständnissen vorzubeugen.

Eigentlich selbstverständlich, aber dennoch sehr wichtig ist die Tatsache, daß für Auftraggeber und Öffentlichkeit nur die Informationen zugänglich sind, die im Bericht enthalten sind. Der konkrete Ablauf des Projektes bleibt den Auftraggebern weitestgehend verborgen. Deshalb sollte man sicherstellen, daß auch wirklich alle relevanten Informationen im Bericht enthalten sind und nicht als selbstverständlich vorausgesetzt werden.

Formal ist der Bericht aufzubauen wie eine konventionelle Forschungsarbeit auch. Sie wendet sich jedoch an Praktiker und Entscheider. Diesem Umstand sollte durch eine besondere Berücksichtigung der Zielgruppe Rechnung getragen werden. Diese müssen den Bericht verstehen können. Deshalb kann z.B. ein Glossar mit Begriffserläuterungen im Anhang nicht schaden wenn man im Bericht wirklich viele Fachbegriffe verwenden will. Besondere Vorsicht ist im Diskussionsteil geboten. In rein wissenschaftlichen Arbeiten wird dort selten eindeutig Stellung bezogen, da dies als unreflektiert gilt und meist leicht angreifbar ist. Ein Evaluationsbericht sollte jedoch ein möglichst klares Fazit oder eine Handlungsanweisung beinhalten, um den Auftraggeber zufriedenzustellen.

Formaler Aufbau des Berichtes:

### **Danksagungen / Vorwort**

### **Inhaltsverzeichnis**

### **Zusammenfassung (Abstract)**

**Hauptteil:**    **Theorieteil:** Problemdarstellung, bisherige relevante Literatur → Ableitung der Untersuchungshypothesen

**Methodenteil:** Muß Methodik so darstellen, daß Studie replizierbar. Enthält Darstellung von Untersuchungsobjekten (Stichprobe), Material (Erhebungsinstrumente), Geräte, Untersuchungsablauf.

**Ergebnisteil:** Beschreibung des Datenmaterials unter Berücksichtigung des Datenschutzes, Beschreibung der Datenaufbereitung (Kodierung) und verwendeter statistischer Verfahren.

                  Bericht über deskriptive Statistiken und zufallskritische Bewertung der Ergebnisse von Signifikanztests.

**Diskussion:** Kommentierung der Ergebnisse durch den Autor. Deutung und Interpretation der Befunde. Darstellung von Implikationen für die Theorie, Diskussion von Untersuchungsfehlern.

**Ausblick:** Darstellung offener Fragen, ungeprüfter Hypothesen, allgemeine Anregungen über die weitere Entwicklung des jeweiligen Forschungsfeldes.

## **Literaturverzeichnis**

## **Anhang**

## **Personen/Sachregister, Glossar**

Wie oben bereits angedeutet, sollte der Bericht an die entsprechende Zielgruppe angepaßt sein. Dem kann man durch eine **Zielgruppenanalyse** Rechnung tragen, welche als Ergebnis eine bestimmte Komplexität des Satzbaus, den Grad der Verwendung von Fachausdrücken, Konjunktionen, etc. nahelegt. Höchste Priorität liegt darin, daß die Zielgruppe den Bericht verstehen kann.

Diese Lösung ist einfach, wenn es sich beim Adressaten um eine bestimmte, definierte Personengruppe handelt, wie etwa die Auftraggeber. Dann kann man auch den Verwertungszusammenhang und den Sprachstil der Adressaten von vorne herein bei der Berichtlegung berücksichtigen.

Schwieriger wird die Zielgruppenorientierung bei heterogenen oder undefinierten Zielgruppen, wie etwa der „Öffentlichkeit“. Hier sollte man besonders auf die Unmißverständlichkeit seiner Konzepte achten und brisante Punkte klar herausstellen, im Zweifelsfall den Bericht vorab Vertretern der jeweiligen Subpopulationen vor der eigentlichen Veröffentlichung zur Kommentierung vorlegen um darauf noch rechtzeitig reagieren zu können.

Ein weiteres Dilemma ergibt sich im Bereich der sogenannten **Informationsverdichtung**. Diese ist notwendig, da keiner der Auftraggeber die umfangreichen Dokumentationen lesen wird, die bei großen Projekten durchaus Buchstärke erreichen (um wissenschaftlichen Maßstäben zu genügen). Man muß also eine Informationsverdichtung durchführen, um einer Informationsverdichtung durch sachfremde Personen (und damit der Gefahr einer Informationsverzerrung) zuvorzukommen. Dies macht aber wiederum eine Kritisierung der Ergebnisse leichter (und vor allem bei emotionalen Themen wahrscheinlicher), da auf dem verkürzten Raum eine wasserdichte Darstellung der Ergebnisse meist nicht mehr möglich ist.

Hierbei handelt es sich aber um ein grundsätzliches Problem aller Studien, es ist nur im Einzelfall und möglichst im Diskurs zu lösen.

Empfohlen werden jedoch generell folgende Schritte mit entsprechenden Richtwerten:

1. Anfertigung einer wissenschaftlich soliden Dokumentation mit >500 Seiten.
2. Anfertigung eines Berichtes (abgeleitet aus der Dokumentation) mit 80-100 Seiten.
3. Anfertigung einer Zusammenfassung (abgeleitet aus dem Bericht) mit 3-6 Seiten.
4. Anfertigung einer Beschlußvorlage (abgeleitet aus der Zusammenfassung) mit 1 Seite.

Eine Besonderheit der auftraggebundenen Evaluationsforschung ist die obligatorische **mündliche Präsentation** der Studienergebnisse. Hierbei ist die freie Rede und der Einsatz von fortgeschrittenen Präsentationsmedien wie Videobeamern in Verbindung mit Präsentationssoftware wie Powerpoint selbstverständlich. Je nach Anspruch des Publikums sollte der Vortragende möglichst rhetorisch geschult sein und die Übersichtlichkeit sowie die Redundanz der Information am Auditorium orientieren. In keinem Fall sollte man sich – auch bei brisanten Themen – emotional hinreißen lassen. Dies verbietet die kühle Professionalität des akademisch geschulten Wissenschaftlers der in jedem Fall die Form wahren sollte.

### **3 Bewertung von Evaluationsstudien**

Dieser Abschnitt unterteilt sich in die Bewertung von **Einzelevaluationen**, die Evaluation als **gesellschaftliches Grundprinzip** und **Zukunftsperspektiven** der Evaluation.

#### **3.1. Bewertung von Einzelevaluationen**

Die Bewertung einzelner Evaluationsvorhaben läßt sich unter zweierlei Perspektiven subsumieren:

- **Ergebnisabhängige Aspekte:** Hier geschieht die Bewertung unter dem Nutzenaspekt. Dem Nutzen in der Verbesserung bestimmter, konkreter Entscheidungsprozesse auf der Grundlage der Evaluation (und damit auf der Grundlage rationaler Kriterien) im Vergleich zu Entscheidungen ohne Evaluation. Dieser ist allerdings oft sehr schwer zu quantifizieren oder abzuschätzen, da Vergleiche meist fehlen (entweder macht man eine Evaluation oder keine), oder nicht angemessen sind – es fehlt also quasi die Kontrollgruppe. Die

ergebnisabhängige Bewertung ist somit nicht leicht, vor allem wenn man bedenkt, daß die Ergebnisse in der Regel sehr kontextbezogen sind. Leichte Veränderungen in den Rahmenbedingungen können die ganzen Ergebnisse in Frage stellen. Auch die zeitliche Gültigkeit der Ergebnisse ist im allgemeinen sehr stark begrenzt. Darüber hinaus muß der Nutzen der Evaluationsstudie in Bezug auf die Entscheidungsverbesserung mit den Kosten der Evaluationsstudie verrechnet werden. Während die Kosten leicht quantifizierbar sind, ist dies beim Nutzen eben nicht der Fall, was diesen Vergleich schwierig macht. Eine Bestimmung des „Netto“-Nutzens ist evtl. aus grundsätzlichen Inkommensurabilitäten nicht möglich.

- **Ergebnisunabhängige Aspekte:** Unabhängig von den konkreten Ergebnissen ist eine Evaluation mit Kosten verbunden, vor allem an Zeit und Geld. Generell haben Evaluationen einen Effekt auf den Gegenstand der Evaluation. Ob einen eher positiven, oder eher negativen hängt vor allem vom Umfeld ab, z.B. dem „Evaluationsklima“ in einem Betrieb. Die Durchführung einer Evaluation kann Entscheidungsträger sowohl verunsichern, als auch entlasten, sie kann Verhältnisse verbessern oder Ängste auslösen. In jedem Fall sollte man sich bewußt sein, daß die reine Tatsache, daß man eine Evaluation durchführt schon einen Effekt hat. Damit dieser positiv ausfällt (unabhängig vom konkreten Ergebnis der Studie), sollte man nach und nach ein positives, rationales Evaluationsklima kultivieren – sowohl in der Gesellschaft, als auch in Betrieben.

### **3.2. Evaluation als Grundprinzip**

Die Evaluation als Grundprinzip ist nicht wertfrei, sie erfordert vorab eine Zielexplication. Welche Werte dies sind, d.h. welche erstrebenswert sind und welche Rangliste diese Werte haben ergibt sich nicht inhärent aus der Evaluation selbst, sondern muß im gesellschaftlichen Diskurs ermittelt werden.

Generell erlaubt die Evaluation aber die Erreichung dieser konsensuellen Werte auf der Grundlage rationaler Kriterien, erlaubt die Nutzung der Ergebnisse zur Optimierung von Entscheidungsprozessen. Damit wendet sie sich in ihrer Natur an sich gegen verschleiernde, ideologische und irrationale Systeme und leistet dadurch einen prinzipiell aufklärerischen gesellschaftlichen Beitrag.

So gesehen hat die Evaluation als Prinzip (unabhängig von den Ergebnissen konkreter Studien) einen Nutzen für die Gesellschaft, da sie eine Rationalisierung schwieriger, oft emotional behafteter Probleme und Entscheidungsprozesse erlaubt. Dies gestattet den Abbau möglichen Konfliktpotentials, geht allerdings zu Lasten der erwähnten visionären, diffusen und irrationalen Kräfte – Rationalität ist nicht das einzige legitime gesellschaftliche Gestaltungsprinzip.

### **3.3. Zukunftsperspektiven der Evaluationspraxis**

Grob gesagt wird der Bedarf an Evaluationsstudien im Zuge von Qualitätssicherungsmaßnahmen wie der ISO 9000 zukünftig wohl eher noch weiter anwachsen. Obwohl sich die Arbeitsmarktsituation für Psychologen seit den 1970er Jahren stark verschlechterte, nimmt der Bedarf an qualifizierten Evaluatoren weiter zu. Es wird zunehmend deutlich, daß die Evaluation als angemessenes gesellschaftliches Gestaltungsprinzip immer weitere Bereiche der Gesellschaft erschließt. Schlußendlich begrüße ich diese Entwicklung, hoffe das die Evaluation als Gestaltungsprinzip tatsächlich verstärkten Einzug in alle Arten gesellschaftlicher Entscheidungsprozesse und Lebensbereiche nimmt und schließe mit den Worten, mit denen auch Wottawa (1998) im entsprechenden Kapitel schließt:

„Bei Beachtung und Respektierung der Grenzen der Wissenschaft und unter Verwendung von sozialwissenschaftlichen Grundlagen kann Evaluation auf vielen unterschiedlichen Einsatzgebieten einen wertvollen Beitrag leisten. Beachtet man bei der Planung und Durchführung von Evaluationsprojekten die speziellen Ziele und Bewertungskriterien der Zielgruppen und Auftraggeber, wird sich die Evaluation als gesellschaftliches Gestaltungsprinzip bewähren“

## **4 Literatur**

BORTZ, J., & DÖRING, N. (1995). *Forschungsmethoden & Evaluation für Sozialwissenschaftler*. Berlin: Springer.

WOTTAWA, H., & THIERAU, H. (1998). *Lehrbuch Evaluation*. Bern: Hans Huber.