



Hallo,

kurze Anmerkung: Diese Scripte stammen von 1999. Ich kann leider dazu

keine Fragen mehr beantworten! :-)

Euch trotzdem viel Erfolg!

Dorthe

dorthe@luebbert.net

Evaluationsforschung

Geschichte der Evaluationsforschung	1
USA	1
In Deutschland.....	1
Evaluierung als Verfahren systematischen Lernens im Rahmen von „Reformpolitik“.....	2
Evaluierung als „Rückmeldeschleife“ in einem Planungs- und Managementsystem.....	2
Evaluierung als Mittel der Kostenreduzierung	2
Evaluierung als Analysemittel zum Abbau von „Überregelung“ und „Bürokratismus“	2
Evaluierung als Hilfsmittel des Parlaments zur Kräftigung seiner Kontrollfunktion	2
Die Entwicklung der Evaluationsforschung	3
Poblemparadoxe in der Evaluationsforschung in Hinblick auf experimentelle Paradoxe.....	3
Forschungsmethodische Anforderungen	3
Forschungsstrategische Probleme bei komplexen Untersuchungsaufgaben.....	3
Klarheit des Untersuchungsgegenstandes.....	3
Vorab verfügbare Hypothesen über die interne Struktur des Untersuchungsgegenstandes, um Indikatoren, Meßinstrumente, Auswertungsverfahren auswählen und begründen zu können.....	4
Kontrolle über den Projektablauf und über die Randbedingungen durch den Forscher	4
Grundgesamtheit und Auswahl	4
Instrumente der Datenerhebung	4
Lösungsmöglichkeiten der Dilemmas	4
Fazit.....	5
Forschungsdesign anhand eines selbstgewählten Beispiels.....	5

Geschichte der Evaluationsforschung

USA

Die Evaluationsforschung kann auf eine lange *methodologische* Tradition zurückblicken. Erste (kleine) Feldprojekte gab es beispielsweise in der Psychologie, in der Arbeitswissenschaft und Industriosozologie. Eine sehr viel weiterreichende Tradition besitzt die Evaluationsforschung in der staatsphilosophischen und politikwissenschaftlichen Diskussion über Wirkungsweise und Wirksamkeit des Staates. Diese eher staatstheoretisch und verwaltungswissenschaftliche Diskussion ist mit der Entwicklung und Ausdifferenzierung einer empirischen Sozialforschung und ihrer zunehmenden Indienststellung zugunsten einer pragmatischen und praktischen Nutzung der Evaluierung verloren gegangen.

Die ersten sporadischen Beispiele für kritisch-reformerische und instrumentell-rationalisierenden Evaluierungen finden sich im späten 19. und frühen 20. Jahrhundert bei den Experimental Stations und Extension Programs der Land Grant Colleges. In den 50er Jahren des 20. Jahrhunderts verlagerte sich der Schwerpunkt in der Nutzung und Entwicklung auf die Evaluierung von internationalen Entwicklungshilfeprojekten. So war es denn auch die UNESCO, die Ende der 50er Jahre das erste Handbuch zur Entwicklungsforschung herausgab.

Doch erst mit dem Anstieg staatlicher Reformen, Aufgaben und Ausgaben in den sechziger Jahren beginnt der gezielte Einsatz der Evaluierungsforschung für staatliche Aufgaben und Programme. Unter Präsident Johnson sollte ein Sozialreformprogramm die Situation der Armen verbessern, das Bildungssystem verbessern (und damit auch die innenpolitischen Spannungen entschärfen). Evaluierung sollte vor allem ein Lerninstrument in einem insgesamt als lernendes System verstandenen Reformprogramm werden.

Das Scheitern der Reformbemühungen wurde u.a. damit begründet, daß die getroffenen Theorien über die Wirkungsmöglichkeiten des Programms allzu waghalsig und fragwürdig gewesen seien. Um gesichertes Wissen über die Wirksamkeit von Programmen zu erhalten, fing man an, begrenzte, experimentelle Tests durchzuführen. Die experimentelle Vorgehensweise und der wissenschaftliche Test von Politiken sollte kostspielige Irrtümer verhindern helfen und Gewähr für rationale und wirksame Programme erbringen.

In Deutschland

In Deutschland setzt die Evaluationsforschung mit etwa 10 Jahren Verspätung Ende der 60er, Anfang der 70er Jahre ein, ebenfalls im Zuge politischer Reformprogramme. Reformiert werden sollte beispielsweise das Bildungswesen, Städtebau, Gesundheitswesen, Sozialwesen Maßnahmen zur Humanisierung der Arbeitswelt sollten durchgeführt werden um.

Hellstern/Wollmann unterscheiden fünf grundsätzliche Richtungen von Evaluierung als Analyse und **Kontrollverfahren politisch-administrativen Handelns**

1. Evaluierung als Verfahren systematischen Lernens im Rahmen von „Reformpolitik“
2. Evaluierung als „Rückmeldeschleife“ in einem Planungs- und Managementsystem
3. Evaluierung als Mittel der Kostenreduzierung
4. Evaluierung als Analysemittel zum Abbau von „Bürokratismus“
5. Evaluierung als Hilfsmittel des Parlaments zur Kräftigung seiner Kontrollfunktion Evaluierung als Verfahren systematischen Lernens im Rahmen von „Reformpolitik“

Evaluierung als Verfahren systematischen Lernens im Rahmen von „Reformpolitik

Ende der 60er Jahre wurde - vor allem mit unter sozialliberaler Regierung - eine *Politik der inneren Reformen* eingeleitet. Es entwickelte sich eine Bereitschaft, durch „Sozialexperimente“ Erkenntnisse zu gewinnen und dadurch rationale politische Entscheidungen treffen zu können. Beispielsweise evaluierte man die Einführung von bildungspolitischen Experimenten, zum Beispiel den ersten Ganztagschulen. Dabei standen die Forscher im Spannungsfeld zwischen „Aktionsforschung“ und „distanzierter Kontrollforschung“, konnten einerseits experimentell-inovativ mit dem Ziel systematischen Lernens vorgehen, mußten andererseits aber auch Innovationen in einem komplizierten Handlungsfeld entwickeln und durchsetzen.

Im Zuge der ersten bildungspolitischen Evolutionen kam es zum Paradigmenwechsel innerhalb pädagogischer Modellversuche: das traditionelle Paradigma, die erkenntnistheoretische Position des **kritischen Rationalismus**, wurde weitgehend aufgegeben zugunsten eines Paradigmas der **Handlungsforschung**, die nicht nur die empirische Kontrolle fordert, sondern auch Konstruktion, Optimierung und Legitimierung der Modellmaßnahme. Wesentliches methodologisches Charakteristikum ist die Integration von Analyse-, Konstruktions-, Kontroll- und Bewertungsleistungen im Begriff der **handlungsbezogenen Kommunikation**.

Ein weiterer wichtiger Bereich war und ist die Evaluierung von Arbeitsmarkt- und Beschäftigungspolitik, erwähnenswert ist hier das Programm zur „Humanisierung der Arbeitswelt“ (HdA).

Evaluierung als „Rückmeldeschleife“ in einem Planungs- und Managementsystem

Ebenfalls Ende der 60er Jahre wollte man mit einer Regierungs- und Verwaltungsreform ein ressortübergreifendes Informationssystem schaffen zur Koordinierung der Regierungsressorts. Außerdem wollte man so ein mittelfristiges Aufgabenprogramm als Instrument einer ressortübergreifenden Koordinierung aufstellen. Zentraler Bestandteil der Bestrebungen war es, den Informationsfluß über Ablauf und Ergebnisse in den einzelnen politisch-administrativen Handlungsfeldern institutional und prozedural zu sichern. Dazu war es nötig, leistungsfähige „soziale Indikatoren“ für die einzelnen Politikfelder zu entwickeln, was u.a. durch die Entwicklung von leistungsfähiger Software möglich wurde.

Evaluierung als Mittel der Kostenreduzierung

Mitte der 70er Jahre rückten - auch ausgelöst durch die Erdölkrise - die finanziellen Ressourcen und damit Verfahren der Finanz- und Haushaltsplanung in den Vordergrund. Vorausschauende (ex ante) Kosten-Nutzenanalysen, aber auch ex-post-Analysen wurden durchgeführt

Evaluierung als Analysemittel zum Abbau von „Überregelung“ und „Bürokratismus“

Die Diskussion um Bürokratisierung von politisch-administrativen Handlungsabläufen hat u.a. zur Entwicklung anspruchsvoller und leistungsfähiger ex-ante-Techniken (z.B. Kostenvorausschätzungen und Praxistests als Variante von Realsimulation) geführt. Sozialwissenschaftliche Fragestellungen wurden relevant, die die Wirksamkeit und die Wirkungen von gesetzlichen Regelungen und politisch-administrativen Programmen sowie dem Wirkungsprozeß in den Mittelpunkt ihres analytischen und konzeptionellen Interesses rücken, beispielsweise innerhalb der Implementationsforschung, deren Fokus auf den Prozeß der Verwirklichung („Implementation“) von Gesetzen und Handlungsprogrammen gerichtet ist.

Evaluierung als Hilfsmittel des Parlaments zur Kräftigung seiner Kontrollfunktion

Der Bundestag und seine Ausschüsse wollen durch Evaluation empirische Informationen und Einschätzungen von Maßnahmen erhalten. Evaluationsforschung hat beispielsweise den Auftrag, die Wirksamkeit und Wirkung der beschlossenen gesetzlichen Regelungen zu beobachten und innerhalb einer bestimmten Frist zu berichten.

Im Bereich der **Projektauvaluierungen** finden sich die ersten konkreten Ansätze Ende der 60er Jahre bei den Uno-Gremien. Aufgabe dieser Evaluierungseinheiten ist dabei weniger die Evaluierung als vielmehr die Innovierung, Kontrolle und Auswertung von Evaluierungen und Evaluierungsverfahren. Im Mittelpunkt stehen Selbstevaluierungsverfahren, auch unter Aspekten der Finanzierbarkeit.

Die Entwicklung der Evaluationsforschung

Bei der Betrachtung der Entwicklungen der Forschungsvorhaben zeigt sich, daß die Evaluation von den Hochschulen auf die Verwaltungen selber verlagert wird.

Poblemparadoxe in der Evaluationsforschung in Hinblick auf experimentelle Paradoxe

Forschungsmethodische Anforderungen

Die konkreten praktischen Anforderungen an den Evaluationsforscher führen dazu, daß dieser hinsichtlich seiner Methodik mit verschiedenen Problemen umgehen muß. Ausgangspunkt ist, daß der Forscher die folgenden vier Variablenbereiche mit empirischen Daten abbilden muß

1. Programm (Maßnahmen), 2. Programmziele, 3. Programm-Umwelt, 4. beabsichtigte und nicht-beabsichtigte Wirkungen/Effekte (= Kriterien der Zielerreichung)

Als Aufgabe der **Datenerhebung** ergibt sich ein dreifaches Monitoring: Der Forscher muß gleichzeitig die unabhängigen Variablen messen, Umwelt-Ereignisse (exogene Einflüsse) indizieren und messen, sowie die „abhängigen Variablen“ messen, d.h. er muß das Wirkungsfeld mit seinen beabsichtigten und nicht-beabsichtigten Effekten messen.

Die **analytische Aufgabenstellung** ist wesentlich schwerer zu lösen: um die festgestellten Wirkungen der Maßnahme zurechnen zu können, benötigt man die Existenz eines theoriegeleiteten Wirkungsmodells.

Im dritten Aufgabenblock, der **Evaluation/Erfolgskontrolle** müssen die Programmziele und ihre Interdependenzen präzisiert und die festgestellten Wirkungen mit den Zielvorgaben des Programms verglichen werden.

Forschungsstrategische Probleme bei komplexen Untersuchungsaufgaben

Voraussetzungen für die Anwendbarkeit des methodologischen Forschungsprogramms herkömmlicher empirischer Kausalanalysen sind die drei folgenden Punkte:

Erstens muß vor der Entwicklung des Forschungsdesigns **Klarheit über die Untersuchungsziele**, über die Konstanz des Gegenstandes bestehen. Zweitens müssen **vorab verfügbare Hypothesen** (begründete Vermutungen) über die Struktur des Gegenstandes wie über Zusammenhänge und Beziehungen existieren. Drittens muß der Forscher den Projektablauf und seine Randbedingungen kontrollieren können. Im Regelfall, z.B. bei Evaluation eines Pilotprogramms, ist *keine einzige* dieser Voraussetzungen erfüllt.

Klarheit des Untersuchungsgegenstandes

Selbst bei klar definierten Maßnahmenprogrammen (z.B. Wirtschaftsförderung) ist es nicht so, daß zuverlässiges Praxiswissen bereits existiert, das technologisch verwertbar ist. Die Vorstellung der „Steuerung durch Ziele“ wird durch „Steuerung durch Organisation und Verfahren“ abgelöst: wenn keine definitiven Vorstellungen existieren, kann auch kein abgeschlossenes und auf technologischen Ziel-Mittel-Zordnungen beruhendes Maßnahmen-Bündel formuliert werden.

Es ist dem Forscher nicht von vorneherein und eindeutig möglich, die abhängigen Variablen, also die Wirkungsdimensionen festzulegen, auf denen Veränderungen beobachtet werden sollen.

Der Forscher kann sich über die mit seinem Programm verbundenen Ziele *vor* dem Forschungsprogramm keinen Überblick verschaffen, er kann das - wenn überhaupt - nur im Verlauf des Projektes tun.

Die Forschung bewegt sich auf schwankendem Boden, muß sich mit einem als „vorläufig“ definierten Forschungsgegenstand und extrem unpräzisen Untersuchungszielen zu Beginn auseinander setzen.

Die Evaluationsforschung befindet sich in einem Dilemma: entweder hat die Forschung ein gutes, sicheres Design, trägt dann aber nichts zu Veränderungen bei, oder ihr eigenes Design gerät ständig ins Wanken.

Vorab verfügbare Hypothesen über die interne Struktur des Untersuchungsgegenstandes, um Indikatoren, Meßinstrumente, Auswertungsverfahren auswählen und begründen zu können

Auch diese Voraussetzung ist bei üblicher Programmevaluation kaum erfüllt, bei neuen Sachverhalten überhaupt nicht. Da geeignete theoretisch und empirisch abgesicherte Kenntnisse vor Beginn der Untersuchung nicht oder nur in unzureichendem Maße bereit stehen, kann die allgemein empfohlene Sozialforschungsroutine¹ nicht eingehalten werden, jedenfalls nicht in methodologisch vertretbarer Weise.

Ähnliches gilt für die Indikatoren, die in das Forschungsdesign einbezogen werden: in einer längerfristigen Untersuchung aufgrund der Geschichtlichkeit der menschlichen Existenz wird es dazu kommen, daß Indikatoren, die man a priori nicht einbezogen hat, plötzlich Bedeutung gewinnen.

Das benötigte Wissen läßt sich erst im Forschungsvorhaben selbst, im Verlaufe der Projektdurchführung gewinnen. Dabei empfiehlt Elmar Lange, ein Forschungstagebuch zu führen, um Zusammenhänge erfaßbar, interpretierbar und bewertbar zu machen.

Kontrolle über den Projektablauf und über die Randbedingungen durch den Forscher

Die Rolle des Begleitforschers ist beschränkt auf die Rolle eines überwiegend nur reagierenden Beobachters ohne nennenswerte Kompetenzen hinsichtlich der Gestaltung der Rahmenbedingungen - und das in einem Untersuchungsfeld, das sich in ständiger, in seiner Richtung kaum vorhersehbarer Entwicklung befindet. Das Design kann erst während der Untersuchung wirklich konzipiert werden und wird immer modifiziert.

Grundgesamtheit und Auswahl

Probleme entstehen auch hinsichtlich des Objektbereiches. Eine traditionelle Stichprobenauswahl ist i.d.R. nicht möglich, da meist auf Selbstselektion gesetzt wird, die Teilnehmer am Pilotprojekt sich freiwillig melden. Nicht die Konstruktion einer Stichprobe aus der Grundgesamtheit ist Aufgabe, sondern der Schluß von einer sich selbst konstituierenden Stichprobe auf den zugehörigen Objektbereich. Für solche Fälle empfiehlt Kromrey die Methode des theoretical sampling von Glaser/Strauß.

Instrumente der Datenerhebung

Ein weiterer wichtiger Bestandteil jedes Forschungsdesigns sind die einzusetzenden Methoden. Normalerweise legt man nach der Festlegung der Untersuchungsobjekte auch die Erhebungsinstrumente - im voraus - fest und kann sie auch eindeutig begründen, auch über den Grad der Standardisierung kann man Aussagen machen.

Lösungsmöglichkeiten der Dilemmas

Der Forscher befindet sich also in dem Dilemma, daß er die notwendigen theoretischen Kenntnisse über den Untersuchungsgegenstand (noch) nicht hat. Er kann also nicht - ausgehend von der Fragestellung und unter Rückgriff auf empirisch gesicherte Hypothesen - ein problemangemessenes „Modell“ des Untersuchungs-Gegenstandes entwerfen. Es ist ihm nicht möglich, zu den als relevant erkannten Dimensionen des Gegenstandes Indikatoren zu bestimmen und methodisch akzeptabel durch Korrespondenzregeln zu begründen. D. h. er kann zunächst lediglich Oberflächen Erscheinungen beobachten und beschreiben.

Kromrey empfiehlt, da man am Anfang ein möglichst breites Spektrum empirischer Sachverhalte mißt, im Wechsel zwischen Datenerhebung und Interpretation die benötigten Hypothesen über sein Untersuchungsfeld entwickelt.

¹ d.h. Präzisierung der Untersuchungsziele, Modellierung des Untersuchungsgegenstandes (dimensionale und semantische Analyse), Verknüpfung der wesentlichen Dimensionen durch Hypothesen, Auswahl und Begründung von Indikatoren, Entwicklung und Pretest von Meßinstrumenten, Auswahl der Untersuchungseinheiten, Feldarbeit/Datenerhebung, statistische Auswertung, Ergebnis-Interpretation, Erstellung eines Forschungsberichtes

Die traditionelle Methodologie [analytisch-nomologisch orientierter Sozialwissenschaft] stellt hierfür keine akzeptierten Regeln zur Verfügung, wohl aber die *grounded-theory* (Glaser, Strauß) oder die explorativen Strategien (Blumer, Wilson). Diese stellen aber keinen Ersatz der traditionellen Methoden dar, sondern sollen als Ergänzung verstanden werden.

Um die Komplextheit des Untersuchungsgegenstandes zu erfassen, ist eine Kombination von unterschiedlichen Methoden erforderlich. Jede Methode hat Stärken und Schwächen, nicht Wettbewerb sondern Komplementarität von (quantitativen und qualitativen) Methoden sollte im Vordergrund stehen.

Allerdings muß man dazu akzeptieren, daß, wie Wilson es formuliert, nämlich der Gebrauch einer spezifischen Methode nicht mit dem Hinweis auf ein bestimmtes Wissenschaftsparadigma, sondern immer nur von der aktuellen Forschungsaufgabe her begründet werden kann. Jede Methode erfaßt Sachverhalte aus unterschiedlichen Perspektiven, mit unterschiedlicher Breite und Tiefe. Eventuelle Differenzen in den Ergebnissen sind weniger nach den dem Kriterium „Falschheit/Richtigkeit“ zu bewerten, als nach dem Kriterium, aus welcher Perspektive welcher Aspekt des Gegenstands betrachtet wurde. Der Begriff „Triangulation“ versinnbildlicht den Versuch, nach Messungen aus unterschiedlichen Perspektiven, d.h. mit unterschiedlichen Methoden, den eigenen Standpunkt präzise festzustellen.

Stichwort „**partizipative Orientierung**“: aktive, gleichberechtigte Einbeziehung der Befragten in die Forschung notwendig.

Fazit

Je weniger Vorwissen über einen Forschungsgegenstand vorhanden ist, je weniger sich die zu untersuchenden Sachverhalte bereits stabilisiert haben, je komplexer der zu analysierende Gegenstandsbereich ist, desto weniger ist das für die Surveyforschung entwickelte Standardmodell empirischen Vorgehens anwendbar..

In diesen Fällen sollte sich schon das Forschungsdesign an der Varianzmaximierung orientieren: Der Forscher soll möglichst eine weite Vielfalt für alle relevanten Dimensionen des Designs schaffen (Voraussetzung für valide Resultate). D.h. er verwendet verschiedene Auswertungs- und Analyseansätze, nutzt die Vielfalt der in das Forschungsprojekt aktiv eingebundener Personen, verwendet unterschiedliche Instrumente und Datenquellen (Verschiedene Erhebungszeitpunkte, Orte, Situationstypen, Personengruppen)

Forschungsdesign anhand eines selbstgewählten Beispiels

Die Besonderheiten der Evaluationsforschung erfordern, daß der Sozialforscher einige methodische Erfordernisse berücksichtigt, um zu validen und verlässlichen Ergebnissen zu gelangen. Am Beispiel eines fiktiven Programms möchte ich im folgenden die Entwicklung eines adäquaten Forschungsdesigns erläutern.

Es wird geplant, im Englischunterricht an Gymnasien zukünftig computergestützt zu lernen, dazu sollen spezielle, neuentwickelte CD-ROMs eingesetzt werden. An mehreren Gymnasien sollen Modellprojekte durchgeführt werden. Der Evaluationsforscher bekommt den Auftrag, die Wirkung des Medieneinsatzes abzuschätzen.

Zunächst versucht er gemeinsam mit dem Auftraggeber die Ziele so zu konkretisieren, daß die Kriterien zur Messung des Programmerfolges festgelegt werden können. In meinem Beispiel sei dies die Verbesserung der Sprachfähigkeit mit den Indikatoren „Grammatikkenntnis“, „korrekte Aussprache“, „Umfang des Wortschatzes“. Bei anderen Beispielen kann die Zieldefinition schwierig sein, wenn der Bereich komplexer ist und verschiedene Ziele praktisch nicht alle realisiert werden können, obwohl dies wünschenswert wäre.

Nach der Konkretisierung der Ziele sollte sich der Evaluator in die Materie anhand von Literatur einarbeiten und dabei insbesondere prüfen, ob schon vergleichbare Studien vorliegen und wie das Forschungsdesign in diesen Fällen entwickelt und operationalisiert wurde.

Da mein Beispiel die Zielgruppe (alle Schüler an Gymnasien) zunächst nur partiell erfaßt, liegt es nahe, ein Quasi-Experiment mit konstruierten Kontroll- und Experimentalgruppen durchzuführen: moralische Bedenken gegen die Einführung von Kontrollgruppen, die zunächst nicht am PC unterrichtet werden, könnte man äußern, halte ich aber in diesem Fall für vernachlässigenswert, da der „Schaden“ - wenn überhaupt - gering ist, da die Kontrollgruppe ja weiterhin in Englisch unterrichtet wird.

Die Auswahl der Stichprobe ist hier bereits vorgegeben, da die Zielgruppe durch die Schuladministration bereits in Schulklassen geclustert ist. Der Evaluator kann sich sicher, daß er die richtige Zielgruppe erreicht, auch die mit der Selbstselektion verbundenen Probleme kann er vernachlässigen. Allerdings ist die Vergleichbarkeit von Kontroll- und Experimentalgruppe praktisch nicht zu beeinflussen. Im Idealfall wäre für jede Zieleinheit in der Versuchsgruppe (Klasse A) einen „adäquaten“ Partner in der Klasse B („match“), der dem Programm-„Stimulus“ nicht ausgesetzt wird. Diese Parallellisierung läßt sich i.d.R. nicht umsetzen,

Nicht lösbar ist in dem Zusammenhang das Problem, daß die Klassen nach Stadtbezirken mit unterschiedlicher Bevölkerungsstruktur gebildet wurden. Der Forscher hat lediglich die Chance, unter den n in Frage kommenden Klassen vergleichbare Gruppen auszuwählen. Dazu erstellt er eine Liste von Wirkindikatoren, die also wahrscheinlich auf das Evaluationsergebnis Einfluß nehmen.

Indikatoren für die Vergleichbarkeit wären in erster Linie die vorhandenen Englischkenntnisse, d.h. der gegenwärtige Leistungsstand, aber auch die Lernfähigkeit sowie die Kompetenz des Lehrers, der die Klasse unterrichtet. Externe Störfaktoren sind außerschulische Stimuli durch Elternhaus, Freunde (Schüler hat z.B. PC zur Verfügung und hat schon Erfahrung mit Multimedia-Edutainment gesammelt, Schüler bekommt während der Testphase Nachhilfeunterricht).

Bevor der Evaluator irgendwelche Analysen durchführt, muß er im nächsten Schritt zunächst herausfinden, ob die unabhängigen Variablen in der Stichprobe überhaupt realisiert sind, d.h. ob *wirklich* Gruppen am PC unterrichtet wird und andere nicht (Vielleicht blockiert der Lehrer den Modellversuch, die PCs funktionieren nicht etc.).

Um nun die Wirkungen des CD-ROM-Einsatzes zu überprüfen, stehen dem Forscher verschiedene quantitative und qualitative Methoden zur Verfügung, die er möglichst kombiniert einsetzen sollte. Bei dem gegebenen Beispiel ist es ihm durch quantitative Datenerhebung möglich, den Lernfortschritt zu messen. Idealerweise stellt der Evaluator vor Beginn des CD-ROM-Einsatzes anhand eines standardisierten Tests (z.B. Toefl-Test) den Leistungsstand in Kontroll- und Vergleichsgruppe fest, diesen Test wiederholt er nach dem Einsatz der CD-ROMs. Befragungen der Lehrer als Experten sollten Informationen über die generelle Verbesserung von Sprachfähigkeit und Aussprache liefern.