



Datenerhebung innerhalb der Programmevaluation

Moira Atria

mailto: moira.atria@univie.ac.at

Fokus: quantitative und qualitative Erhebungsmethoden innerhalb der Programmevaluation

Workshopelemente

TeilnehmerInnen: Fragen, Themen, Anmerkungen

Datenerhebung in der Evaluationsausbildung nach DeGEval

Datenerhebung im Kontext der DeGEval-Evaluationsstandards

Zur Bedeutung der Datenerhebung bei verschiedenen Autoren

Überblick über gängige Datenerhebungsverfahren

Diskussion

Beispiel aus der Programmevaluation *Aggressionsprävention*

Diskussion

Arbeit an einem Beispiel aus TeilnehmerInnenkreis

Zusammenfassung

**„Ausgangspunkt“:
Datenerhebung innerhalb
der
Evaluationsausbildung
nach den Empfehlungen
der DeGEval, 2004**

Aus Empfehlungen für die Aus- und Weiterbildung in der Evaluation (DeGEval, 2004)

4 Kompetenzfelder:

- ✓ Theorie und Geschichte der Evaluation
- ✓ **Methodenkompetenz**
- ✓ Organisation und Feldkenntnisse
- ✓ Sozial- und Selbstkompetenz
- ✓ Praxis der Evaluation

Aus Empfehlungen für die Aus- und Weiterbildung in der Evaluation (DeGEval, 2004)

4 Kompetenzfelder:

- ✓ Theorie und Geschichte der Evaluation
- ✓ **Methodenkompetenz**
 - ✓ Grundzüge empirischer Sozialforschung und Untersuchungsdesign
 - ✓ **Datenerhebung**
 - ✓ Statistische Kenntnisse
 - ✓ Datenverarbeitung
 - ✓ Projektorganisation

Aus Empfehlungen für die Aus- und Weiterbildung in der Evaluation (DeGEval, 2004)

4 Kompetenzfelder:

✓ Theorie und Geschichte der Evaluation

✓ **Methodenkompetenz**

✓ Grundzüge empirischer Sozialforschung und Untersuchungsdesign

✓ **Datenerhebung**

- Grundlagen von Erhebungsformen (hier: Erhebungsmethoden genannt)
- Entwicklung von Erhebungsinstrumenten

Schritte einer Programmevaluation (vgl. DeGEval, 2002).

Entscheidung über die Durchführung einer Evaluation

Definition des Evaluationsproblems

Planung der Evaluation

Informationsgewinnung

Informationsauswertung

Berichterstattung zur Evaluation

Budgetierung der Evaluation

Evaluationsvertrag

Steuerung der Evaluation

Personelle Ausstattung der Evaluation

Datenerhebung im Kontext der Evaluationsstandards

N Nützlichkeit

D Durchführbarkeit

F Fairness

G Genauigkeit

Datenerhebung im Kontext der Evaluationsstandards (vgl. DeGEval, 2002).

Nützlichkeit (Auswahl):

- ✓ N1 Identifizierung der Beteiligten und Betroffenen
- ✓ N2 Klärung der Evaluationszwecke
- ✓ N4 Auswahl und Umfang der Informationen

Durchführbarkeit (Auswahl):

- ✓ D1 Angemessene Verfahren
- ✓ D3 Effizienz von Evaluation

Fairness (Auswahl):

- ✓ F2 Schutz individueller Rechte
- ✓ F3 Vollständige und faire Überprüfung

Durchführung (Auswahl):

- ✓ G5 Valide und reliable Informationen
- ✓ G7 Analyse qualitativer und quantitativer Informationen

Wesentliche Anmerkungen zur Datenerhebung innerhalb der Programmevaluation bei verschiedenen Autoren

Bortz & Döring (2003)

Hager & Hasselhorn (2000)

Kirkpatrick (1998)

Datenerhebungsarten bei Bortz & Döring (2003) – Grundsätzliches

„Die Methoden der empirischen Datenerhebung haben die Funktion, Ausschnitte der Realität, die in einer Untersuchung interessieren, möglichst genau abzubilden. Im Vordergrund bei den sog. Quantitativen Methoden steht die Frage, wie die zu erhebenden Merkmale **operationalisiert** werden sollen [...]. In der Regel wird eine empirische Untersuchung nicht nur mit einer dieser Erhebungsarten auskommen.“

„Qualitative Methoden eignen sich nicht nur zu Explorationszwecken, sondern werden im Forschungsalltag oft mit quantitativen Methoden kombiniert [...]. Ein erstes Unterscheidungsmerkmal [...] ist die Art des verwendeten Datenmaterials“:
Verbalisierung vs. Quantifizierung

Quantitative Methoden der Datenerhebung nach Bortz & Döring, 2003

- ✓ **Zählen:** z.B. Indexbildung; Quantitative Inhaltsanalyse
- ✓ **Urteilen:** z.B. Ratingskalen; Vergleichsmethoden

Besonderheit: Kategorienerstellung

- ✓ **Testen:** Testtheorie
- ✓ **Befragen:** mündlich oder schriftlich
- ✓ **Beobachten:** systematische Beobachtung
- ✓ **Physiologische Messung:** Grenzwerte

Gütekriterien quantitativer Datenerhebung:
Objektivität, Reliabilität, Validität

Wesentliche Aspekte der Datenerhebung bei Hager & Hasselhorn (2000)

Unterscheidung zwischen

- ✓ **Wirksamkeit und Wirkmodell**
- ✓ **Programmgebundene und interventionsunabhängige Wirkungen**
- ✓ **Performanz und Kompetenz**
- ✓ **Fokus auf Transferleistungen**

Ebenen der Evaluation - das Modell von Kirkpatrick (1998)

- 1. Evaluation der Reaktionen**
- 2. Evaluation des Lernens**
- 3. Evaluation des Transfers**
- 4. Evaluation auf „Ergebnisebene“**

„Ultimative“ Evaluationsebene
Bewertung der Wirkungen von Programmen, Maßnahmen etc. an ihrem Einfluss auf die „Ausgangssituation“

Planung der Datenerhebung beinhaltet die Klärung von 4 Fragen

Was will ich untersuchen?

Zumeist: Programmziel, andere programmbezogene Merkmale (z.B. „Bausteine“; Wirkmodell)

Welche Merkmale weist der
Untersuchungsgegenstand auf?

z.B. Indikatoren für das Programmziel

Wie kann ich es untersuchen?

Welche Ressourcen stehen mir zur Verfügung?

Messinstrumente bzw. Leitlinien für Dokumentation

Zur Klärung der grundlegenden Fragen:

Was will ich untersuchen?

- ✓ *Wissen? Kompetenz? Einstellungen? Meinungen?
konkrete Handlungen? implizite Inhalte, Merkmale?
dynamische Prozesse?*

Wie kann ich es untersuchen?

- ✓ Fragebogen? Tests? Interviews? Schriftliche Befragungen? Text-Analyse? (Inhaltsanalyse) Arbeitsproben? Szenario? Vignetten? Falldarstellungen? Verhaltensbeobachtung?

Gütekriterien: Objektivität, Reliabilität, Validität

Diskussion
Anmerkungen
Fragen

Beispiel

Datenerhebung innerhalb des Wiener Sozialen Kompetenztrainings (WiSK)

Hintergrund: zur Problematik der
Evaluation schulbasierter
Aggressionsprävention

WiSK ist ein schulklassenbasiertes Programm für Risikojugendliche (14-19J.).

Findet während Unterrichtszeit durch externe(n)
Trainerin (Trainer) statt

Konzipierte Einheiten bei adaptivem Vorgehen

Genaue Dokumentation als integraler Bestandteil

„Individuelles“ Moment und „gruppenbezogenes“
Moment (Klasse)

Untersuchung zur Programmentwicklung: Wiener BMS: 2
WiSK-Schulklassen; 6 Vergleichsklassen (vgl. Atria &
Spiel, in Druck) – Vergleich TG-KG: uneindeutige
Ergebnisse; Vergleich Trainingsklassen –
Vergleichsklassen

Wiener Soziales Kompetenztraining

Grundgedanke der Intervention

ad Wirkmodell (vgl. Hager & Hasselhorn, 2000)



Leitgedanke 1 „Soziale Inkompetenz als Ausdruck eines eingeschränkten Handlungsrepertoires“

Leitgedanke 2 „Es gibt keine Unbeteiligten – wir sind alle involviert“

Wiener Soziales Kompetenztraining WiSK - Programm

Impulse 6 Einheiten	Reflexion 1 Einheit	Aktionen 6 Einheiten
<p><i>Was sich so alles abspielt ...</i></p> <p>Kommunikation & Konfliktmanagement anhand von Beispielsituationen bzw. aktuellen</p>	<p><i>Was ich bisher gelernt habe ...</i></p>	<p><i>Was wir sonst noch wollen ...</i></p> <p>Gruppe übernimmt Verantwortung Kriterien!</p>

ad Lerntransfer (vgl. Hager & Hasselhorn, 2000; Kirkpatrick, 1998)

Gruppenn

Einzel/Kleingruppe

Schulklasse

WiSK-Datenerhebung im Kontext der Evaluationsstandards (Auswahl N).

**N1 Identifizierung der Beteiligten und Betroffenen:
Zielgruppen und deren soziales Umfeld definieren!**

SchülerInnen und LehrerInnen

**N2 Klärung der Evaluationszwecke: Priorität setzen!
(Programmziele: Steigerung der sozialen Kompetenz,
Senkung der Aggression bei SchülerInnen);**

**Evaluationsszweck: v.a. Konzeptoptimierung (auch:
Effekte und Anregungen für die wissenschaftliche
Diskussion)**

**N4 Auswahl und Umfang der Informationen: Unabdingbares
von Interessantem unterscheiden! Ressourcen im Sinne
der Prioritäten einsetzen!**

**Selbst- und Fremdeinschätzung - Fragebögen; Vignetten;
Verhaltensbeobachtung**

WiSK-Datenerhebung im Kontext der Evaluationsstandards (Auswahl D).

D1 Angemessene Verfahren: Belastungen in Grenzen halten!
Evaluation innerhalb des Unterrichts; Instrumente der LehrerInnen nur sehr kurz; beziehen sich auf gesamte Klasse

D3 Effizienz von Evaluation: Aufwand in Grenzen halten!

SchülerInnen

FB (je Untersuchungszeitpunkt: ca. 25 Minuten)

Vignetten (ca. 10 Minuten)

VB-Video (während des normalen Unterrichts)

LehrerInnen

Alle 2 Wochen 1 Blatt bezogen auf Interventionsklasse und Parallelklasse

TrainierIn

Systematische Trainingsdokumentation

Evalteam: VB-Auswertung; Vignetten-Auswertung (vgl. DeGEval, 2002)

WiSK-Datenerhebung im Kontext der Evaluationsstandards (Auswahl F, G).

F2 Schutz individueller Rechte: Anonymisierte Daten verwendet

G5 Valide und reliable Informationen: Gütekriterien der Instrumente überprüft

Beispiel zu F3 in Ergebnisauswahl

F3 Vollständige und faire Überprüfung: vgl. Ergebnisauswahl z.B. Schulstufeneffekte und Ansteigen der Aggression über das Schuljahr!

Untersuchungsdesign

Jun00	Sep00		Jun01	Okt01
				9 (4 KI)
	2 Wiener Handelsschulen			10 (4 KI)
	9 (4 KI)	1 Trainingsklasse	9 (4KI)	11 (4 KI)
	10 (4KI)	1 Trainingsklasse	10 (4KI)	
9 (4KI)	11 (4KI)		11 (4 KI)	
10 (4KI)				

**Erhebungen sowohl bei SchülerInnen und LehrerInnen als auch bei angehenden TrainerInnen
genaue Dokumentation**

WiSK-Datenerhebung: Quantitative und qualitative Methoden

SchülerInnen

FB (je Untersuchungszeitpunkt: ca. 25 Minuten)

Vignetten (ca. 10 Minuten)

VB-Video (während des normalen Unterrichts)

LehrerInnen

Alle 2 Wochen 1 Blatt bezogen auf Interventionsklasse und Parallelklasse

TrainierIn

Systematische Trainingsdokumentation

Evalteam:

VB-Auswertung

Vignetten-Auswertung

SchülerInnen: „Demokratie-FB“ (nach Eder, 1998) und „Bully/Victim-FB“ (nach Olweus, 1989)

„Demokratie“ – Skalen

Gleichwertigkeit (8 Items, $\alpha = 0,82$);

Information (7 Items, $\alpha = 0,76$);

Partizipation (7 Items, $\alpha = 0,72$);

Vielfalt (4 Items, $\alpha = 0,79$)

„Bully/Victim“ – Skalen

Viktimisierung – Schüler (8 Items, $\alpha = 0,80$)

Aggression – Schüler (8 Items, $\alpha = 0,85$)

Viktimisierung – Lehrer (3 Items, $\alpha = 0,86$)

Aggression – Lehrer (4 Items, $\alpha = 0,84$)

**Hintergrund: Erfassung der Programmziele „Partizipation“ und
„Senkung der Aggression“**

SchülerInnen: Vignette

Eine Schülerin aus deiner Klasse veranstaltet nächsten Samstag eine Party und hat dich dazu eingeladen. Heute nach der Schule kommen Karin und ein anderer Schüler auf dich zu. Karin sagt: „Tut mir leid, du kannst doch nicht zu meiner Party kommen“. Paul lächelt und sagt: „Das ist aber schade...“

Wie reagierst du auf diese Situation?

Warum glaubst du haben die beiden das zu dir gesagt?

Wie könnte man sonst noch reagieren?

Quelle: Lösel & Averbek; Hintergrund: Erfassung „Erweiterung des Verhaltensrepertoires“ Wirkmodell

SchülerInnen: Verhaltensbeobachtung, vgl. Faßnacht, 1995

Videomaterial: 4 Schulklassen, 2 Videokameras, 4 Unterrichtsstunden prä und 4 Unterrichtsstunden post

Analyse 1: Kategorien- und Kodiersystem

- systematische Auswahl von Filmsequenzen (3 mal 2 Minuten pro Stunde, Beginn, Mitte, Ende) = 192 Minuten
- Erstellung eines **Kategorien- und Kodiersystems (VB1)**
 - Einschulung (Studierende)
 - Randomisierte Vorgabe, Kodierung
 - BeurteilerInnenübereinstimmung (Kappa-Maße)

Hintergrund: Erfassung „Kompetenz-Performanz“

Evaluationsteam: Kategorien- und Kodiersystem (VB 1)

Erfüllung des Geforderten	0.754*	0.675*	0.660*	0.634*	0.623*	0.712*
Aktive Mitarbeit	0.86*	0.886*	0.851*	0.828*	0.794*	0.867*
Positive Reaktion auf direkte Aufforderung	0.532*	0.857*	0.705*	0.614*	0.373*	0.666*
Unaufmerksamkeit	0.547*	0.492*	0.446*	0.404*	0.393*	0.489*
Störung des Unterrichts	0.737*	0.692*	0.701*	0.676*	0.675*	0.704*
Physisch aggressives Verhalten – Austeiler	0.909*	0.909*	0.769*	1.000*	0.857*	0.857*
Physisch aggressives Verhalten – Einstecker	x	x	x	0.000	-0.001	0.500*
Verbal aggressives Verhalten	y	y	y	y	y	y
Hilfsbereitschaft	1.000*	0.857*	0.750*	0.857*	0.750*	0.889*
Gesamt	0.702*	0.643*	0.621*	0.595*	0.581*	0.657*

SchülerInnen: Verhaltensbeobachtung, vgl. Faßnacht, 1995

Videomaterial: 4 Schulklassen, 2 Videokameras, 4 Unterrichtsstunden prä und 4 Unterrichtsstunden post

Analyse 2: Ratingsystem

- Nochmalige Reduktion des Filmmaterials (3 mal 4 Minuten vor und nach Intervention pro Klasse = 24 Min pro Klasse) = 96 Minuten
- Erstellung eines **Ratingsystems (VB2)**
 - Einschulung (angehende WiSK-TrainerInnen)
 - Randomisierte Vorgabe Rating
 - BeurteilerInnenübereinstimmung (Kappa-Maße)

Hintergrund: Erfassung „Kompetenz-Performanz“

Rating: Verhaltensbeobachtung, vgl. Faßnacht, 1995

Videomaterial: 4 Schulklassen, 2 Videokameras, 4 Unterrichtsstunden prä und 4 Unterrichtsstunden post

Analyse 2: Ratingsystem

- Nochmalige Reduktion des Filmmaterials (3 mal 4 Minuten vor und nach Intervention pro Klasse = 24 Min pro Klasse) = 96 Minuten
- Erstellung eines **Ratingsystems (VB2)**
 - Einschulung (angehende WiSK-TrainerInnen)
 - Randomisierte Vorgabe Rating
 - BeurteilerInnenübereinstimmung (Kappa-Maße)

Hintergrund: Erfassung „Kompetenz-Performanz“

Angehende TrainerInnen: Ratingsystem (VB 2)

Ruhe während des Unterrichtes	1—2—3—4—5—6—7
Mitarbeit am Unterricht	1—2—3—4—5—6—7
Aufmerksamkeit	1—2—3—4—5—6—7
Freude am Lernen	1—2—3—4—5—6—7
Physische Aggression unter den SchülerInnen	1—2—3—4—5—6—7
Verbale Aggression unter den SchülerInnen	1—2—3—4—5—6—7
Hilfsbereitschaft unter den SchülerInnen	1—2—3—4—5—6—7
Zusammenhalt, Klassengemeinschaft	1—2—3—4—5—6—7

LehrerInnen: Ratingsystem (Auswahl)

Einschätzung alle 2 Wochen, gesamte Klasse

Ruhe während des Unterrichtes	1—2—3—4—5—6—7
Mitarbeit am Unterricht	1—2—3—4—5—6—7
Aufmerksamkeit	1—2—3—4—5—6—7
Freude am Lernen	1—2—3—4—5—6—7
Physische Aggression unter den SchülerInnen	1—2—3—4—5—6—7
Verbale Aggression unter den SchülerInnen	1—2—3—4—5—6—7
Hilfsbereitschaft unter den SchülerInnen	1—2—3—4—5—6—7
Zusammenhalt, Klassengemeinschaft	1—2—3—4—5—6—7

Trainingsdokumentation: nach jeder Einheit

EINHEIT:

TrainerIn:

Klasse (plus Anzahl der anwesenden SchülerInnen):

Gab es zum Standardablauf eine geplante Ergänzung (mit Vorlage vergleichen)?

Was wurde tatsächlich gemacht (Stichworte)?

Was lief gut? (plus konkretes Beispiel)?

Was lief weniger gut? Warum? (Hat etwas gefehlt?)

Worauf sollte ich in der nächsten Einheit besonders achten?

Diskussion
Fragen
Anmerkungen