FORSCHUNGSMETHODEN UND EVALUATION I / MODUL 2

Einführungsskript:

Wie wird aus einer Fragestellung ein Forschungsprojekt?

1. Prüfung,
   - ob die (noch vage) **Forschungsidee allgemein wissenschaftlichen Kriterien genügt** (dazu gehört z.B. präzise Problemformulierung! „Prüfungsangst bei Schulkindern“ ist noch zu vage!) und
   - ob sie **empirisch untersuchbar** ist
   - ob sie **unter ethischen Gesichtspunkten realisierbar** ist (Schutz der Untersuchungsteilnehmer, Informationspflicht, Freiwilligkeit der Teilnahme, Vermeidung von psychischen und physischen Beeinträchtigungen, usw.)

2. Klärung des **Anspruchs** der empirischen Studie **in Abhängigkeit vom Zweck** der Untersuchung

3. **Literaturrecherche ->** gründlich und internetbasiert; sinnvolle Kombination zentraler Keywords (= Schlagworte) -> **Präzisierung der Themenstellung**; wichtig dabei = sorgfältige Dokumentation mit Hilfe eines Literaturverwaltungsprogramms [z.B. EndNote4 von ISI ResearchSoft]

4. **Wahl der Untersuchungsart ->** hängt ab vom Stand der Forschung im jeweiligen Untersuchungsfeld

   ➔ ***explanative Untersuchung***:
   es gibt schon sehr viele Vorkenntnisse, d.h. es kann eine gut begründete Hypothese formuliert werden (z.B. Mädchen haben höhere Prüfungsangst als Buben); hypothesenprüfendes Vorgehen.

   ➔ ***explorative Untersuchung:***
   es gibt noch wenige Vorkenntnisse, d.h. man muss hypothesenerkundend vorgehen (z.B. Gibt es Schultypenunterschiede bezüglich der Prüfungsangst)

   ➔ ***deskriptive Untersuchung:***
   es geht darum, Populationen zu beschreiben (z.B. Erfassung möglichst vieler Facetten von Prüfungsangst bei Schulkindern)

Welche Standards sind bei Forschungs- und Evaluationsprojekten zu beachten?

Es gibt noch keine Kataloge mit Standards, ABER: von den USA ausgehend einen Diskursprozess zur Entwicklung von Standards.

- Die wichtigsten wurden formuliert vom Joint Committee on Standards for Educational Evaluation (**JCSEE, 1994**).
- Für den deutschsprachigen Raum gibt es Standards von der Deutschen Gesellschaft für Evaluation (**DeGEval, 2002**)

Definition von „Standard“:

> “... a principle, mutually agreed to by the people engaged in a professional practise, for example, evaluaton.” (JCSEE)

JSCEE – Standards:

30 Standards in 4 Gruppen

1. ***Nutzen-Standards (Utility)***:

   Evaluationen sollen Bedürfnisse der Öffentlichkeit (d.h. vor allem der Stake-Holder [= von einer Evaluation betroffene Personen oder Gruppen] nach **praktisch verwertbarer Information** befriedigen -> Ergebnisse müssen so aufbereitet werden, dass potentielle Nutznießer auch davon profitieren können.

2. ***Machbarkeits- / Durchführbarkeits-Standards (Feasibility)***:

   Evaluationen sollen **unter natürlichen Bedingungen** (d.h. im Feld!) unter **möglichst ökonomischer Verwendung von Ressourcen, Material, Personal und Zeit** durchführbar sein.

3. ***Standards für ethisches Vorgehen (Propriety):***

   Evaluationen sollen den **Schutz der individuellen Rechte garantieren** (z.B. durch formale Übereinkommen, faire Bewertung, Geheimhaltung von Befunden)

4. ***Genauigkeits-Standards (Accuracy):***

   Evaluationen sollen **methodisch korrekt** durchgeführt werden, damit sie **fundierte und verwertbare Informationen** liefern, die es erlauben, Maßnahmen hinsichtlich ihrer Vor- und Nachteile zu bewerten.

BORTZ: Kapitel 2.1 THEMENSUCHE

Qualität einer empirischen Untersuchung wird auch daran gemessen, ob die Untersuchung zur **Erweiterung des Bestandes an gesichertem Wissen** im jeweiligen Forschungsgebiet beiträgt; ABER: allgemein große Schwierigkeiten beim Finden einer kreativen Untersuchungsidee (Grund = Ausbildung in den Sozialwissenschaften -> 90% des Unterrichts wird für Techniken, wie man Hypothesen prüft, verwendet, anstatt Strategien zur Hypothesenfindung zu vermitteln).

Empirische Arbeiten sind meist sehr zeitaufwendig und arbeitsintensiv –> darum sollte man sich **um die Entwicklung einer Fragestellung bemühen, die einen selber auch interessiert** (wichtig für die Motivation!); Gegenargument: wer sich zu sehr engagiert, fälscht eventuell seine Ergebnisse..., DAHER: sorgfältige Dokumentation der Vorgehensweise und der Ergebnisse, um die Untersuchung genau nachvollziehen bzw. replizieren zu können.

1. ANLEGEN EINER IDEENSAMMLUNG

Möglichst früh im Studium [SUUUPER, dass wir das quasi am Ende lernen!!!!] eine Ideensammlung anlegen -> Einfälle zu Vorlesungen, Gesprächen, etc. mit Datum notieren (erleichtert später die Literaturrecherche bzw. die Eingrenzung des Untersuchungsproblems). Vorteil: **hohes Ego-Involvement** (weil einen diese Themen ja interessiert haben!)

2. REPLIKATION VON UNTERSUCHUNGEN

Replikationen sind unerlässlich zur Festigung und Erweiterung des Kenntnisstandes. Sind vor allem dann **wichtig, wenn Untersuchung zu unerwarteten**, mit dem aktuellen Forschungsstand nur wenig in Einklang zu bringenden **Ergebnissen geführt** haben (durch Bestätigung in der Replikation stärkere Aussagekraft der Ergebnisse!)

Völlig exakte Replikationen sind unmöglich; Gründe:
- veränderte zeitliche Umstände
- andere Versuchspersonen
- andere Versuchsleiter

3. MITARBEIT AN FORSCHUNGSPROJEKTEN

Vorteile: man erhält Einblick in komplexeren Forschungsbereich,
relevante Literatur ist teilweise schon recherchiert,
Vorteile der Teamarbeit,
u.U. finanzielle Unterstützung,
Möglichkeit zur Publikation

Nachteil: eigene Forschungsinteressen könnten zu kurz kommen

4. Weitere Anregungen

Gezielte Themensuche (ABER: Durcharbeiten von Literatur ist zeitaufwendig, aber nicht immer erfolgreich!). Zuerst: Auswahl eines Themenbereich; diesen dann gezielt nach offenen Fragen, Widersprüchlichkeiten, Hypothesen untersuchen.

„Kreative Suchstrategien“:

a) ***intensive Fallstudien***:

gründliche **Beobachtung einzelner Personen** (vgl. Wolfsmensch von Freud); müssen nicht unbedingt herausragend oder irgendwie abnormal sein.

b) ***Introspektion***:

Selbstbeobachtung (z.B. Wieso reagiere ich in bestimmten Bereichen besonders empfindlich auf Kritik [beim Autofahren]? Gibt es Belege dafür, dass auch andere Menschen solche „sensiblen“ Bereiche haben?)

c) ***Sprichwörter***:

enthalten **Erfahrungen vieler Generationen**, warum sollten sie nicht auch für die Gegenwart interessant sein (z.B. „Ein Spatz in der Hand ist besser als eine Taube auf dem Dach“ -> Handlungsregel, bei Wahlentscheidungen lieber die kleinere, gleich verfügbare Belohnung zu wählen als die größere, erst später verfügbare)

d) ***funktionale Analogien***:

Übertragung bekannter Mechanismen auf neuartige Probleme; wird aber erschwert durch „funktionale Fixierungen“, nach denen sich Objekte und Vorgänge nur schwer aus ihrem jeweils funktionalen Kontext herauslösen lassen (z.B. Inokkulations-Theorie von McGuire, 1964 -> Beeinflussung von Personen in verbaler Kommunikation wird reduziert, wenn Person über Beeinflussung vorgewarnt wird / ähnlich wie bei Impfung mit kleiner Menge an Grippeviren, die Abwehreaktion bewirkt und Körper vor Ansteckung mit Grippe schützt)

d) ***Paradoxe Phänomene***:

Wahrnehmungen, die unerklärlich bzw. widersinnig erscheinen (z.B. Warum verursachen schwere Verwundungen in starken Erregungszuständen keine Schmerzen? Wieso kann man trotz totaler Erschöpfung nicht einschlafen?)

e) ***Analyse von Faustregeln***:

beruhen auf jahrelangen Erfahrungen; ihre **Analyse vermittelt oft Einsichten, die eine bessere Nutzung dieser Erfahrungen ermöglicht** (z.B. Warum scheitern so viele Ehen im siebenten Jahr?)

f) ***Veränderungen von Alltagsgewohnheiten***:

Alltagsleben unterliegt einer gesellschaftlichen Normierung, der wir uns meist gar nicht bewusst sind, **erst Veränderungen fallen uns auf** (z.B. Was sind die Ursachen für die Veränderung der Mode, der Freizeitgestaltung, gesellschaftlicher Umgangsformen,...)

g) ***gesellschaftliche Probleme***:

In öffentlicher Diskussion wird oft ein Mangel an Forschungsergebnissen beklagt; in Eigenregie durchgeführte Qualitätsarbeit kann **aktuelle Fragestellung rasch aufgreifen**, während ein größeres Forschungsprojekt oft ein paar Jahre Vorlaufzeit benötigt (z.B. Warum wird eine bestimmte TV-Sendung gar so gern angeschaut?).

h) ***widersprüchliche Theorien***:

Oft widersprechen 2 Theorien einander (scheinbar) -> **eigene Prüfmöglichkeit entwickeln, um das auszuräumen bzw. allgemeineren Ansatz zu entwickeln** (z.B. gewichtetes Durchschnittsmodell bei der Beurteilung von Menschen von Anderson[1], 1967)

[1] Entstanden aus:

- ***Thorstone***: Durchschnittsmodell (Gesamteindruck von einem Menschen entspricht dem Durchschnitt seiner Teilattribute)
- ***Fishbein***: additives Modell (Gesamteindruck ergibt sich aus der Summe der Teilattribute)

Fazit: ***Anderson***: Einzelne Attribute fließen mit unterschiedlichem Gewicht in die Beurteilungmitein

BORTZ: KAPITEL 2.2:

BEWERTUNG VON UNTERSUCHUNGSIDEEN

Anhand der Ideensammlung **überlegen, welche Themen für empirische Untersuchung in engere Wahl kommen**. Wichtig dabei:

➔ Entspricht die Idee allgemeinen ***wissenschaftlichen Kriterien***?
➔ Entspricht die Idee ***untersuchungstechnischen Kriterien***?
➔ Ist sie unter ***ethischen Gesichtspunkten*** empirisch umsetzbar?

1. WISSENSCHAFTLICHE KRITERIEN

a) ***Präzision der Problemformulierung***

Untersuchungsidee = unbrauchbar, wenn:

- **unklar ist, WAS der Untersuchungsgegenstand sein soll** bzw. **wenn er so vielschichtig ist, dass sich daraus sehr viele Fragestellungen ableiten lassen** (z.B. eine Arbeit über „die Leistungsmotivation -> BESSER: Genese von Leistungsmotivation z.B. bei Arbeitslosen)
- sie **unklare, mehrdeutige, schlecht definierte Begriffe** enthält (z.B. Bedeutung der Intelligenz für die Selbstverwirklichung -> zuerst genau erklären, was man unter beidem versteht!)

Begriffe gelten in diesem Stadium als **vorläufig klar formuliert**, wenn sie **kommunikationsfähig** sind.

b) ***Empirische Untersuchbarkeit***

Thema muss empirisch überprüfbar sein. **Ungeeignet** sind daher:

- ***religiöse, metaphysische, philosophische Themen*** (z.B. Leben nach dem Tod, Existenz Gottes, Sinn des Lebens)
- ***Themen, die sich mit unklaren Begriffen befassen*** (z.B. Seele, Gemüt, Charakterstärke, wenn keine Strategie zur Präzisierung verfolgt wird (d.h. Frage nach dem Sinn des Lebens ist schon empirisch untersuchbar, wenn ich z.B. Vorstellungen über den Sinn des Lebens bei Personen in unterschiedlichen Bevölkerungsgruppen, Lebensaltern, Kulturen, etc. untersuche!)
- ***Themen, die einer unangemessenen Arbeitsaufwand erfordern***, wie z.B.
  - Untersuchungen von **ungewöhnlichen Personen** (z.B. psychische Probleme von Zwergwüchsigen)
  - Untersuchungen von **ungewöhnlichen Situationen** (z.B. Ursachen für Panikreaktionen bei Massenveranstaltungen)
  - **sehr zeitaufwendige Untersuchungen** (z.B. Längsschnittuntersuchung zur Entwicklung des logischen Denkens bei Kindern)

c) ***Wissenschaftliche Tragweite***

- **Unbrauchbar sind auch Themen**, die
  - ***keine praktische Bedeutung*** erkennen lassen
  - die ***Grundlagenforschung nicht bereichern*** können

- **abzuraten ist auch von Themen**, die
  - ***schon so intensiv erforscht*** wurden, dass die Untersuchung nur bereits seit langem bekannte Ergebnisse bestätigen würde (z.B. Bestrafungen sind weniger lernfördernd als Belohnungen, Veränderung der Reaktionszeiten unter Alkoholeinfluss)
  - ***triviale Ergebnisse*** erbringen
  - deren ***Ergebnisse aufgrund von „Alttagstheorien" prognostizierbar*** sind.

2. ETHISCHE KRITERIEN

- **Voraussetzung für empirische Untersuchungen** auf dem Gebiet der Humanwissenschaften ist eine **hohe ethische Sensibilität der Untersucher**. Viele Untersuchungsgegenstände (z.B. Gewalt, Liebe, Angst,...) betreffen die **Privatspähre des Menschen** (Schutz durch das Grundgesetz!)

- Unbedingt **vermieden werden muss**:
  - mangelnde Anonymisierung
  - Missbrauch personenbezogener Daten
  - physische und/oder psychische Beeinträchtigung der Probanden

- **Wichtig**: Untersuchungsidee darauf überprüfen, ob **geplante Untersuchung ethisch auch vertretbar** ist!

- **Berufsethische Richtlinien beachten**! Hier wird u.a. geregelt:
  - Umgang mit Menschen und Tieren als Untersuchungsobjekte
  - Umgang mit Mitforschenden z.B. bei Publikationen

- BEISPIEL FÜR EINE ETHISCH UMSTRITTENE STUDIE:

  Untersuchung von Milgram (1963) über Gewissenlosigkeit von Menschen, die sich zum Gehorsam verpflichtet fühlen (wenn als VP getarnter Strohmann in einem Nebenraum etwas falsch machte, bekam er von der wirklichen VP immer stärkere Stromstöße. Bei bestimmter Stromstärke gab es keine Reaktionen mehr vom „Schüler", dennoch befolgten Trainer den Befehl zur weiteren Erhöhung der Stromstärke...)

a) ***Güterabwägung: Wissenschaftlicher Fortschritt oder Menschenwürde***

- Manche Untersuchungen benötigen **Daten, deren Erhebung nur schwer mit der Menschenwürde der Probanden zu vereinbaren ist**, z.B.
  - Untersuchung der Schmerztoleranzschwelle
  - Erzeugung von Depressionen durch experimentell herbeigeführte Hilflosigkeit
  - Reaktion auf angstauslösende Reize
- Können solche Beeinträchtigungen nicht vermieden werden, so sind solche Untersuchungen **nur dann gerechtfertigt, wenn sie Ergebnisse bringen, die anderen Personen zugute kommen** (z.B. Depressive, Phobikern, etc.).
- Wenn **Zweifel an der ethischen Unbedenklichkeit** einer geplanten Untersuchung aufkommen, unbedingt **außenstehende Fachleute und die zu untersuchende Zielgruppe zu Rate ziehen**!

b) ***Persönliche Verantwortung***

- Wer eine Untersuchung durchführt, ist für alle unvorhergesehenen Vorkommnisse zumindest **moralisch verantwortlich**
- Bei ethisch bedenklichen Experimenten, Befragungen, etc. muss **Proband auf mögliche Gefährdung und sein Recht auf Teilnahmeverweigerung hingewiesen** werden
- Bei Gefahr von physischer Beeinträchtigung VOR der Untersuchung **Arzt fragen**!

c) ***Informationspflicht***

- Proband muss **VOR der Untersuchung alle Infos** erhalten, damit er sich entscheiden kann, ob er teilnehmen möchte oder nicht -> „**informed consent**“
- Nach der Untersuchung muss **Proband über seine eigenen Ergebnisse informiert** werden, wenn er das möchte.
- Wenn **Täuschung des Probanden** aus untersuchungstechnischen Gründen **erforderlich** ist -> genaue **Aufklärung über die wahren Zusammenhänge** nach Untersuchung! Wenn Proband darauf Auswertung seiner Daten untersagt, muss man das unterlassen!
- VOR der Untersuchung **überprüfen, ob eine Täuschung nicht doch vermieden werden könnte** (z.B. durch andere Untersuchungsmethode).

d) ***Freiwillige Untersuchungsteilnahme***

- Proband darf **nicht zur Untersuchung gezwungen** werden, hat **jederzeit das Recht abzubrechen** (Probanden, die zur Untersuchung gezwungen werden, verfälschen das Ergebnis!)
- **Untersuchung** muss **so angelegt** werden, dass **freiwillige Teilnahme nicht zum Problem** wird
- Probanden nicht wie „Versuchspersonen“ behandeln!

- **Bezahlung von Probanden**:
  - ethisch bedenklich (finanziell schlechter Gestellte könnten auf diese Bezahlung angewiesen sein)
  - Verfälschung der Untersuchungsergebnisse (Probanden wollen „gute Versuchspersonen“ sein)

  Daher: nur dann, wenn Untersuchung (zeitlich) sehr aufwendig!

e) ***Vermeidung körperlicher oder psychischer Beeinträchtigungen***

3 Arten der Beeinträchtigung nach Lewin:

- ***vermeidbare***:
  - Probanden kommen aus **Mangel an Sorgfalt, Unachtsamkeit**, überflüssiger, nicht unbedingt notweniger Maßnahmen zu **Schaden** (z.B. Bloßstellung, Erschöpfung, Angst, Überforderung,...);
  - **Abhilfe**: sorgfältige Planung der Untersuchung!

- ***unbeabsichtigte***:
  - aufgrund **unvorhersehbarer Zwischenfälle**, z.B. Frage nach dem Berufs des eben verstorbenen Vaters eines Kindes;
  - **Abhilfe**: unverzügliches Eingreifen des VL, persönliches Gespräch, sachliche Aufklärung, etc.

- ***beabsichtigte***:
  - müssen **so wenig wie nur irgend möglich belastend gestaltet** werden, z.B. Untersuchungen von Angst, Schuld- und Schamgefühlen, Verlegenheit,...

f) ***Anonymität der Ergebnisse***

- kann **Anonymität der persönlichen Angaben** nicht gewährleistet werden, muss auf die Untersuchung verzichtet werden

- Proband muss zugesichert werden, dass
  - **Daten nur zu wissenschaftlichen Zwecken** verwendet werden
  - sein **Name nicht registriert** wird (Codierung)

- Bestimmungen des **Datenschutzes** beachten!

FINK Kapitel 2:

EVALUATIONSFRAGEN:

1) ZIELE (GOALS) UND OPERATIONSZIELE (OBJECTIVES)

Evaluationsfragen liefern Daten über die Merkmale und Vorteile eines Programms. In den meisten Evaluationen liegt eines der Hauptinteressen darin, festzustellen, ob die Ziele des Programms erreicht worden sind. Üblicherweise sind die ***ZIELE (GOALS)*** ziemlich **allgemein formuliert** und **langfristig** angelegt.

BEISPIEL: **Typische Programmziele**

Für die Öffentlichkeit:
- Erhöhung der Lebensqualität
- Förderung intellektueller, sozialer und psychologischer Funktionen
- Steigerung der Zufriedenheit mit Schulen

Für die Professionalisten:
- Förderung der Forschungsarbeit auf diesem Gebiet
- Vermehrung des Wissens
- Unterstützung des Zugang zu neuen Technologien und Praktiken
- Verbesserung der Fertigkeiten
- Verbesserung der Ausbildung
- Förderung der Effizienz

Für die Institutionen:
- Verbesserung der Organisation, ihrer Struktur und ihrer Effizienz
- Optimierung ihrer Fähigkeiten, hochqualifizierte höhere Bildung für alle anbieten zu können

Für das System:
- Erweiterung der Kapazität, um hochwertige Ausbildung anbieten zu können
- Unterstützung der Effzienz
- Sicherung der Respektierung der Bedürfnisse aller Bürger

OPERATIONSZIELE (OBJECTIVES) beziehen sich auf den **spezifischen Zweck**, dem ein **bestimmtes Programm** dienen will.

BEISPIEL: **Die Operationsziele eines neuen Kurses**

Der neue 2-semestrige Kurs soll Studenten des 1. und 2. Studienjahres lehren, Evaluationen von Bildungsprogrammen durchzuführen. Eines der primären Ziele ist die Entwicklung eines Handbuchs, das Studenten die Grundprinzipien der Evaluation nahe bringt und das eine kommentierte Bibliographie enthält, damit man sich mehr Informationen verschaffen kann, wenn man sie braucht. Am Ende der beiden Semester sollen die Studenten eine Evaluation planen. Jeder Plann soll Evaluationsfragen, Standards, Untersuchungsdesign, Methoden und Instrumente zur Datenerhebung enthalten.

⇨ *Operationsziele für den Entwickler des Studienplanes:*
- Erstellung eines Handbuches mit einer kommentierten Bibliographie

⇨ *für den Studenten:*
- Vorbereitung eines Evaluationsplanes mit bestimmten Inhalten

---

Operationsziele können von **jedem in die Evaluation Involvierten** angestrebt werden (z.B. von Lehrern, Studenten, Institutionen, Geldgebern, dem System selbst). Die Personen, die von den Ergebnissen Evaluation betroffen sind, werden als „**Stake-Holders**“ bezeichnet.

BEISPIEL: **Evaluationsfragen:**

1. Wurde ein Handbuch zur Programmevaluation entwickelt?
2. Haben die Studenten einen den Anforderungen entsprechenden Evaluationsplan erstellt?

Daraus ergeben sich weitere Fragen:

- Bis wann soll das Handbuch fertig sein?
- Wie wird entschieden, ob es von Nutzen ist?
- Was sind die Kennzeichen eines zufriedenstellenden Evaluationsplans, wer beurteilt das?
- Wie viele Studenten müssen solche Evaluationspläne erstellen?

Diese Fragen müssen in den folgenden Evaluationsaktivitäten geklärt werden. Im nächsten Schritt werden **Standards** erstellt, deren Erfüllung die Zielerreichung und die Effektivität und Effizienz des Programms bestimmen.

Der **Evaluator muss dafür sorgen**, dass

- alle wichtigen Ziele und Operationsziele identifiziert werden,
- die Evaluationsfragen alle wichtigen Ziele abdecken
- alle Fragen aufgrund der vorhandenen Ressourcen auch beantwortet werden können.

2) TEILNEHMER UND EFFEKTIVITÄT

Evaluationsfragen zielen oft darauf ab, die **demografischen und sozialen Merkmale der Teilnehmer** an einem Programm zu **beschreiben** und **effektive Ergebnisse mit spezifischen Teilnehmern zu verbinden**. Ein Evaluator könnte beispielsweise damit beauftragt werden, festzustellen, ob ein Kunst-Programm für alle Kinder effektiv war oder nur für einen Teil (z.B. für Kinder unter 12 Jahren)

BEISPIEL: **Evaluationsfragen, Teilnehmer und Effektivität des Programms**

Nun sollte festgestellt werden, ob das Programm für ALLE Studenten effektiv war oder nur für manche. Folgende Evaluationsfragen wurden dabei gestellt:

1. Wie kann man die Studenten demografisch charakterisieren?
2. Ist das Programm gleich effektive für verschiedene Studenten (z.B. Männer und Frauen)
3. Unterscheiden sich Studenten im ersten Studienjahr von denen im 2. Studienjahr
4. Lernen die Studenten des 1. Studienjahres im 2.Studienjahr dazu?

Evaluationsfragen sollten **mit den vorhandenen Ressourcen zu beantworten** sein. Oft bestimmen praktische Überlegungen die Evaluationsziele.

3) PROGRAMMAKTIVITÄTEN, ORGANISATION UND EFFEKTIVITÄT

Die spezifischen Aktivitäten eines Programms und seine Organisation zu kennen ist oft notwendig, um zu **verstehen, warum ein Programm erfolgreich** ist oder nicht, und um zu erkennen, ob es **auf andere Settings übertragbar** ist.

Typische Fragen zu Programmaktivitäten und Organisation sind:

- Was sind die Schlüssel-Aktivitäten?
- In welchem Ausmaß wurden die Aktivitäten wie ursprünglich geplant auch tatsächlich angewendet?
- Wie gut wurde das Programm administriert?
- Hatte das Programm Einfluss auf andere Programme, Institutionen, Konsumenten?
- Haben sich soziale, politische, finanzielle Umstände geändert und damit die Effektivität des Programms beeinflusst?

BEISPIEL: **Evaluationsfragen, Aktivitäten und Organisation**

Ein neunköpfiges Expertenteam definierte die Lernarten, die einem Kurs für Evaluation angemessen sind. Die Evaluation des Kurses soll 4 Jahre umfassen, damit jeweils 2 Gruppen von Studenten des 1. und des 2. Studienjahres untersucht werden können. Einige der besten Lehrer wurden ausgewählt, um beim Entwurf von Lehrplan und Kursbuch mitzuarbeiten und um einen Kurs zu leiten. Der Evaluator fragte:

- In welchem Ausmaß ist die Auswahl der besten Lehrer verantwortlich für die Qualität des studentischen Lernens und des Handbuchs?
- Beeinflusst der Kurs die nachfolgenden Bildungsaktivitäten der Studenten?
- Hat es während der 4 Jahre irgendwelche Veränderungen hinsichtlich der Unterstützung des Programms durch die Schule oder hinsichtlich der Zahl und Art der beteiligten Institute gegeben?

---

4) WIRTSCHAFTLICHKEIT UND KOSTEN

Programmevaluationen können durchgeführt werden, um Fragen nach den Ressourcen, die eingesetzt werden müssen, um Ergebnisse zu erzielen, zu beantworten

➔ **Ressourcen = Kosten** (z.B. Geld, Personal, Zeit, Ausrüstung, Gebäude, usw.

➔ **Ergebnisse** können sein:

- ***monetär*** (z.B. eingesparte Summe Geldes)

  ⇨ Wenn sich die Evaluationsfragen auf die Relation zwischen Kosten und monetärem Ergebnis beziehen, wird eine solche Evaluation als „**Kosten-Nutzen Analyse**“ (**Cost-Benefit**) bezeichnet.

  BEISPIEL: Inwiefern reduzieren jeweils 100 €, die für eine neues Schuleschwänz-Vorbeugungsprogramm ausgegeben werden, die Kosten für den Polizeischutz lokaler Geschäfte?

- ***substantiell*** (z.B. bessere Lesefähigkeiten, usw.)

  ⇨ Wenn sich die Evaluationsfragen auf die Relation zwischen Kosten und substantiellem Ergebnis beziehen, spricht man von eine „**Kosten-Effektivitäts-Analyse**“ (**Cost-Effectiveness**).

  BEISPIEL: Wie hoch sind die vergleichbaren Kosten des Programms A und eines Programms B gegen Schuleschwänzen für eine Reduktion der Einbrüche?

Fragen nach den Kosten werden aus vielen Gründen **in meist unterschiedlichen Zeitabständen** gestellt. Manchmal ist es **schwierig, Kosten zu definieren und den Nutzen eines Programms zu messen**. Experten sind sich über die am besten dafür geeigneten Methoden uneins, aber auch darüber, ob eine solche Analyse Teil des gesamten Evaluation sein oder als eigene Untersuchung durchgeführt werden sollte.

-> Nachteile von Kostenanalysen im Rahmen einer Gesamtevaluation:

- machen ein meist ohnehin schon komplexes Evaluationsdesign noch komplexer
- es werden Kosten einer Intervention analysiert, von der man ja noch gar nicht weiß, ob sie selber überhaupt effizient ist...

-> **Nachteile von Kostenanalysen als seperater Untersuchung**:

- man braucht dafür eine 2. Studie, die sich von der Evaluationsstudie unterscheidet

---

BEISPIEL: **Evaluationsfragen: Kosten**

Wie sind die Relationen zwischen den Kosten und der Effektivität von 3 Lehrkörper-Modellen.

⇨ **Kosten**: Stundengehälter, Miete, Vergütungen, gesetzliche Gebühren, Versicherung, Ausrüstung, Transport,...

⇨ **Ergebnisse (Effektivität)**: Zufriedenheit von Lehrern, Schülern, Eltern; Fortschritte der Studenten

---

Wichtig: **Kosten und Effektivität / Nutzen müssen definiert werden** (wenn notwendig muss auch der Wert der monetären Kosten angegeben werden).

Manchmal liefern Evaluatoren auf Fragen nach den Programmkosten sogenannte „**sensitive Analysen**“ (d.h. Schätzungen sind unsicher, angebene Maße sind nicht präzise) -> bei Wiederholung zu späterem Zeitpunkt: Möglichkeit zu sehen, ob Effektivität des Programms mit den Kosten zunimmt.

5) **Programm-Umgebung**

Alle **Programme finden in einem institutionellen, sozialen und politischen Rahmen statt** (z.B. Programm A [für Kinder unter 14 Jahren, das ihre Rechenfähigkeiten verbessern soll und das in ländlichen Schulen durchgeführt wird, wird vom Bund finanziert] vs.Programm B [in großstädtischen Schulen und privat finanziert])

Wird das Programm nach mehreren Jahren evaluiert, so können sich mittlerweile soziales und politisches Umfeld verändert haben -> das könnte die Evaluation, aber auch das Programm selbst beeinflussen! **Veränderte Unweltbedingungen** z.B. auch durch Veränderungen in der Programmdurchführung, Entdeckung neuer Technologien, usw.

Für die Evaluation einer Programm-Umgebung braucht man Daten über

- das ***Setting***, in dem das Programm stattfindet,
- die ***Finanzierung*** (durch wen erfolgt sie und wie?)
- die ***Durchführungs-Struktur*** (wer ist verantwortlich für die Ergebnisse? wie effektiv ist die Durchführungs-Struktur? ist das Programm weiterhin effektiv, auch wenn sich die Personen oder Gruppen ändern?)
- den ***politischen Kontext*** (trägt die politische Umgebung zum Erfolg des Programms bei? ist das Programm ausreichend fundiert?)

STANDARDS

1) **Das Aufstellen von Standards**

Programm-Evaluationen zielen darauf ab, überzeugende Beweise dafür zu erbringen, dass ein Programm effektiv ist. Die **Standards** sind die spezifischen Kriterien, mit denen die Effektivität gemessen wird.

BEISPIEL: **Evaluationsfrage**: Haben die Studenten gelernt, Evaluationsfragen zu formulieren?

Standard: 90% der am Programm teilnehmenden Studenten werden lernen, Evaluationsfragen zu formulieren. „Evaluationsfragen formulieren lernen“ heißt: Festlegung von Programmzielen und Nutzen, das Stellen der Frage in verständlicher Sprache (Feststellung der Verständlichkeit durch mindestens 3 Benutzer der Evaluation).

Die **Standards** sind der **Schlüssel zur Glaubwürdigkeit der Evaluation**. Je spezifischer sie sind, desto leichter sind sie zu messen. Um die Spezifität zu erhöhen, müssen **alle möglicherweise doppeldeutigen Begriffe in Evaluationsfragen und Standards definiert** werden. Mehrdeutigkeit entsteht, wenn es keine eindeutigen Definitionen und Sprachlevel gibt.

BEISPIEL: **Evaluationsfrage**: „Hat die „Von-der-Schule-in-die-Arbeit"-Initiative die Studenten für den Arbeitsmarkt besser vorbereitet?" -> „besser für den Arbeitsmarkt vorbereitet" heißt WAS???.
Abhilfe in so einem Fall: In einem Dialog (z.B. mit einem Kollegen) versuchen, die unklaren Begriffe zu klären. Liste all der arbeitsmarktbezogenen Fertigkeiten anlegen, die dokumentieren, dass Student besser vorbereitet ist.

Struktur, Prozess und Ergebnis:

Diese 3 Begriffe werden verwendet bei der Evaluation der Qualität des Gesundheitswesens. Man sollte sie anwenden, wenn man die Standards formuliert

- ***Struktur***: bezieht sich auf die **Organisation und die administrative Umgebung, in der ein Programm stattfindet**, und auf die **Charakteristika**
  - des **Personals und Hilfspersonals** (z.B. Zahl der Lehrer, ihre Erfahrung, demografischer Hintergrund);
  - des **Settings** (z.B. öffentliche oder private Schule, Klasse);
  - der **Organisation** (z.B. Abteilungen, Teams)
- ***Prozess***: bezieht sich auf **das, was getan wird**, beinhaltet die technischen, humanistischen oder wertenden Aspekte von Bildung, Beratung, Dienstleistung, usw.; umfasst auch **Methoden und Materialien**, um Bildung, etc. zu verbessern
- ***Ergebnisse***: = **Resultate oder Konsequenzen** von Schul- oder Sozialsystemen, Settings, Prozesse; beinhalten intellektuelle, soziale, psychologische und moralische **Ziele**.

  ➔ **unmittelbare Ergebnisse**:
  - o direkte, sichtbare Ergebnisse eines Programms,
  - o können schnell und einfach gemessen werden

  ➔ **endgültige Ergebnisse**:
  - o brauchen länger, um sichtbar zu werden;
  - o sind schwieriger zu erfassen und zu messen.

BEISPIEL: **Struktur-, Prozess- und Ergebnis-Standards**
in einer Evaluation zur Frage des Bereitseins für den Arbeitsmarkt

Struktur-Standards:
- o Der gesamte Unterricht erfolgt durch Teams auf 3 verschiedene Arten (Arbeitsstudenten-Programme, Fachunterricht, Arbeit mit Jugendlichen zwischen 15-17 Jahren)
- o Mindestens 4 örtliche Betriebe bieten mindestens 36% der am Programmen teilnehmenden Studenten die Arbeitsmöglichkeiten

Prozess-Standards:
- o Alle Studenten erarbeiten Ziele und Operationsziele für ihren eigenen Kurs
- o Lehrer, Studenten und Arbeitgeber entwickeln gemeinsam Methoden, um die Fortschritte der Studenten zu evaluieren

Ergebnis-Standards / unmittelbare:

- Eine signifikant höhere Zahl von Studenten der VG zeigt bessere Leseergebnisse als Studenten in der KG
- Mindestens 90% der Studenten geben an, mit dem Programm zufrieden zu sein

Ergebnis-Standards / endgültige:

- 5 Jahre nach dem Programm sind 45% der Studenten vollbeschäftigt
- 5 Jahre nach dem Programm sind mindestens 20% ein- bis zweimal befördert worden.

Standards müssen sein:

- zielgerichtet und ***beabsichtigt*** bzw. geplant (willkürliche Standards beeinträchtigen das Programm)
- ***sinnvoll***,
- ***realistisch***,
- ***messbar***.

BEISPIEL für einen **unrealistischen Standard**:

Evaluationsfrage: Unterscheiden sich die Schüler unserer Schule von den Schülern anderer Schulen hinsichtlich ihrer Fähigkeit, ein Stellengesuch zu schreiben?

Standard: Ein statistisch signifikanter Unterschied zugunsten unserer Schule lässt sich feststellen.

⇨ Wenn der Evaluator nicht Zugang zu den Schülern an anderen Schulen hat und sie für die Zeit der Evaluation beobachten kann, ist dieser Standard zwar erwünscht, aber **unrealistisch und kann daher nicht verwendet** werden.

---

2) **Verschiedene Quellen zur Festsetzung von Standards:**

a) Experten,
b) historische Daten,
c) Vergleiche mit anderen Gruppen bzw. über die Zeit hinweg,
d) Normen.

ad a) **Expertengruppen**

Experten können dem Evaluator helfen, Standards zu erstellen. Sie sind **Einzelpersonen oder Repräsentanten von Berufs-, u.a. Gruppen**, und ihre Kenntnisse helfen, die Evaluation genauer oder nützlicher zu machen.

Die Zusammensetzung der Expertengruppen sollte so gestaltet werden, dass **Übereinstimmung zwischen den Experten gefördert** wird.
(BEISPIEL: Wenn ein Evaluator Standards für ein Programm zur Hebung der Qualität des Leseunterrichts erstellen soll, so wird er sich Eltern, Schüler, Lehrplanentwickler, Leseforscher, Bildungsexperten, Psychometriker und Diagnostiker in die Expertengruppe holen.)

Expertengruppen sind **in anderen Fachgebieten sehr üblich** (z.B. im Gesundheitswesen und in der Medizin).

Hauptziel = Herstellung eines Konsens über den Grad der Übereinstimmung bei kontroversiellen Themen und ungelösten Fragen (besonders wichtig, wenn es z.B. bei einem völlig neuartigen Programm noch nicht genügend Daten gibt, die zum Vergleich herangezogen werden können). Oft ist es auch schwierig, einen Konsens zu erzielen, weil dies extensiver Literaturrecherchen zum Thema und komplexer Methoden bedarf.

Expertengruppen haben sich aber als **sehr effizient bei der Festlegung von Standards** erwiesen.

Richtlinien für die Arbeit mit Expertengruppen:

1) Die **Evaluationsfrage muss klar spezifiziert** werden (dabei können die Experten helfen!)

2) **Daten sollten vorgelegt werden**, um die Arbeit der Experten zu unterstützen (z.B. über die Teilnehmer an einem Programm, über die Intervention, über Kosten und Nutzen der Teilnahme, usw.). **Datenquellen**: Literatur, laufende Forschungsarbeiten, finanzielle und statistische Aufzeichnungen, usw.

3) Die **Experten sollten qualifiziert sein**, z.B. weil sie entsprechende Kenntnisse oder Einfluss haben, weil sie die Ergebnisse der Untersuchung verwerten.

4) Der **Prozess sollte sorgfältig geplant und geschickt geleitet** werden (d.h. VOR der Zusammenkunft Tagesordnung festlegen, Anschauungsmaterial mitbringen und gezielt einsetzen, einzelne Mitglieder der Gruppe mit spezifischen Aufgaben betrauen, usw.)

ad b) **Historischen Daten**:

z.B. wenn es ein bestimmtes **Programm schon einmal vor längerer Zeit eingesetzt** und evaluiert wurde -> damals erarbeitete **Standards werden übernommen**. ABER:

- grundsätzlich ist eine ***exakte Wiederholung eines Programms nicht möglich***
- ***Standards von „damals“ können mittlerweile vollkommen ungeeignet geworden sein*** (entweder zu niedrig oder zu hoch)

ad c) **Vergleiche**:

Vergleichende Standards zählen zu den überzeugendsten Standards. **Verglichen werden können:**

- die Leistungen einer Gruppe zu 2 verschiedenen Zeitpunkten
- die Leistungen mehrerer Gruppen zu 1 Zeitpunkt
- die Leistungen mehrerer Gruppen über längere Zeit hinweg

Schwierigkeiten:

- Zusammenstellung möglichst ähnlicher Vergleichsgruppen
- Gruppen zur Mitarbeit gewinnen
- ausreichende Zeitspanne, um Veränderungen auch beobachten zu können, wenn tatsächlich welche existieren

Merke: Wenn in einer Evaluation **Unterschiede zwischen einzelnen Gruppen** gefunden werden, die für das neue Programm sprechen, darf man **nicht automatisch annehmen, dass diese auf die Effektivität des Programms zurückzuführen sind**!

Bevor eine Entscheidung darüber getroffen werden kann, müssen **mindestens 3 Fragen beantwortet** sein:

1) Waren die ***Gruppen von Anfang an vergleichbar***? (Trotz aller Bemühungen könnte eine Gruppe z.B. kooperativer, klüger, etc. sein)

2) Ist die ***Höhe des Unterschieds groß genug, um bedeutsam*** zu sein? (Bei sehr großen Stichproben können sehr geringe Unterschiede zwar statistisch signifikant sein, sie können aber trotzdem z.B. für die Effizienz des Programms nicht von Belang sein)

3) Ist das ***zuletzt gezeigte Verhalten der einzelnen Gruppen bedeutsam im Sinne der Hypothese***? (Wenn man z.B. Verhaltensänderungen feststellt, so muss festgestellt werden, ob diese wirklich auf das Programm zurückzuführen sind. Auch sehr große Verhaltensänderungen können noch immer nicht ausreichend sein, um von einer Effizienz des Programms sprechen zu können)

ad d) **Normen**:

Normen können als Standard herangezogen werden, z.B. wenn man weiß, wie hoch die Lesefähigkeit einer Normgruppe ist, so kann man die Leistung der mit einem neuen Programm behandelten Schüler vorher und nachher daran messen.

Wichtig ist aber, dass man sich **vergewissert, ob diese Normen auch tatsächlich auf die eigene Versuchsgruppe anwendbar** sind (z.B. Normgruppe bestand aus Kindern zwischen 10-12 Jahren, ich aber will Kinder zwischen 13-14 testen)

EVALUATIONSFRAGEN UND STANDARDS

Eine Evaluationsfrage kann mit **einem oder mehreren Standards** in Beziehung gebracht werden

BEISPIEL: **Evaluationsfrage**: Haben die Studenten gelernt, die Literatur verlässlich zusammenzufassen?

Standard 1: 80% der Studenten lernen, die Literatur verlässlich zusammenzufassen

Standard 2: Es gibt einen statistisch bedeutsamen Unterschied in der Fähigkeit, Literatur zusammenzufassen zwischen Schülern der Schule A und Schülern der Schule B. Die Schüler der Schule A haben am neuen Programm teilgenommen, der Unterschied spricht zu ihren Gunsten.

⇨ **ABER**: Um die Verbesserung der Schüler A wirklich dem Programm zuschreiben zu können, braucht man Vergleichsdaten von Schülern, die nicht am Programm teilgenommen haben.

⇨ **Warum braucht man hier Standard 1 überhaupt**, wo man doch ohnehin Standard 2 hat?
Grund: Unterschiede könnten zwar vorhanden, aber nur bei weniger Schülern auftreten; Standard 1 schreibt daher vor, wie viele Prozent das Ziel erreichen müssen, damit Programm effizient ist.

Standards sollen aufgestellt werden, **bevor man das Untersuchungsdesign entwirft und die Auswertung plant.**

Evaluationsfrage und Standards bestimmen:

➔ das Evaluationsdesign
➔ die Art der Datenerhebung
➔ die Art der Auswertung

und zwar deswegen, weil sie die **abhängigen und unabhängigen Variablen** enthalten!

Unabhängige Variablen:

= Prädiktor-Variablen. Sind schon vor Beginn des Programms anwesend bzw. vom Programm nicht abhängig; werden verwendet, um Ergebnisse zu erklären oder vorherzusagen. **In Evaluationen sind das oft**:

- das Programm
- demografische Charakteristika der Teilnehmer (z.B. Geschlecht, Einkommen, Erfahrung, Bildung,...)
- psychologische und kulturelle Merkmale (z.B. geistige und soziale Attribute, Religion, ...)

Abhängige Variablen:

= **Ergebnis-Variablen**. Sind das, was der Evaluator messen will. **In Evaluationen** von Bildungsprogrammen **sind das z.B.**

- Wissen
- Fertigkeiten
- Einstellungen
- Verhalten
- Kosten
- Effizienz

Evaluationsfrage und Standards müssen UV und AV enthalten!

BEISPIEL:

Programmziel: Den Studenten soll das verlässliche Zusammenfassen von Literatur beigebracht werden.

Evaluationsfrage: Haben die Studenten gelernt, Literatur verlässlich zusammenzufassen?

Standard: Ein statistisch signifikanter Unterschied in der Fertigkeit, Literatur verlässlich zusammenzufassen, zugunsten der Studenten, die am Programm teilgenommen haben, ist zu beobachten.

Programmeffekte (UV): Teilnahme vs. Nicht-Teilnahme am Programm

Effekte, gemessen am Ergebnis (AV): verlässliche Zusammenfassung

QSV – Bericht - Formular:

-> QSV = Question (= Evaluationsfrage), Standard, Variablen

dient der Verdeutlichung der Beziehungen zwischen diesen 3 Elementen; sehr nützlich zum Planen einer Evaluation und zur Festlegung der Methoden.

besteht aus einer Tabelle mit 4 Spalten:

1. Spalte: Evaluationsfrage
2. Spalte: Standards
3. Spalte: UV (sie bestimmt das Evaluationsdesign)
4. Spalte: AV

BEISPIEL:

Evaluationsfrage	Standards	UV	AV
Wie effektiv ist Lernprogramm A bei der Modifikation von Wissen, Glauben, Selbstvertrauen und Verhalten hinsichtlich Gesundheitsvorsorge und Krankheitsvermeidung?	Statistische und wissenmäßig bedeutsame Verbesserung der Teilnehmer für mindestens 6 Monate statistische und wissensmäßig signifikante Unterschiede	Teilnahme am Programm in der Gruppe (VG oder KG)	Wissen, Meinungen, Selbstvertrauen, Verhalten