FORSCHUNGSMETHODEN UND EVALUATION I / MODUL 1

EINFÜHRUNGSSKRIPT:

Es gibt keine ideale Forschungsmethode -> jede Fragestellung erfordert ein für sie spezifisches Vorgehen; Vor- und Nachteile diskutieren und begründete Entscheidung treffen.

Was versteht man unter wissenschaftlicher Forschung?

Bortz & Döring (2002): „Empirische Forschung sucht nach Erkenntnissen durch systematische Auswertung von Erfahrungen."

Das geschieht durch **Aufstellen und Prüfen von wissenschaftlichen Hypothesen**.

3 Kriterien einer wissenschaftlichen Hypothese:

1) allgemeingültige, über den Einzelfall hinausgehende Behauptung (**All-Satz**)
2) in Form eines **Konditionalsatzes** transferierbar (wenn-dann; je-desto)
3) durch Erfahrungsdaten widerlegbar (**Falsifizierbarkeit**)

BEISPIEL: „Mit der Höhe des Alkoholspiegels sinkt die Reaktionsfähigkeit."
- über den Einzellfall hinausgehende Behauptung
- kann als Konditionalsatz formuliert werden
- ist falsifizierbar.

⇨ ist eine *wissenschaftliche Hypothese*

„Ich glaube, Kinder sollten weniger fernsehen."
- Einzelmeinung
- kann nicht als Konditionalsatz formuliert werden
- ist nicht falsifizierbar

⇨ ist *keine wissenschaftliche Hypothese*

In Psychologie (bzw. in allen Sozialwissenschaften) ist es **nicht möglich**,
- wissenschaftliche **Hypothese durch ihr widersprechende Einzelfälle prinzipiell zu widerlegen**
- wissenschaftliche Hypothese **vollständig zu verifizieren** (dafür müssten ALLE möglichen Fälle überprüft werden....)

daher: empirische Prüfung der Hypothese nicht an **allen möglichen Fällen (= Population)**, sondern nur an einem **Ausschnitt (= Stichprobe)**

Fazit: Wissenschaftliche Hypothesen = **Wahrscheinlichkeitsaussagen**, zu deren Prüfung von der Scientific Community begründete Prüfkriterien festgelegt wurden (z.B. statistische Signifikanz)

Unterschied zwischen Grundlagenforschung und Angewandter Forschung:

Grundlagenforschung: fragt nicht nach Nutzen oder Anwendungsmöglichkeiten der Forschungsergebnisse (z.B. hirnphysiologische Korrelate von systematischen Unterschieden im schlussfolgernden Denken sollen identifiziert werden)

Angewandte Forschung (zu ihr gehört auch die Evaluatinsforschung): Ziel = Nutzbarmachung von Erkenntnissen der Grundlagenforschung und ihre Umsetzung in effiziente Handlungsanleitungen.

Was versteht man unter wissenschaftlicher Evaluation?

Definition von Spiel:

Wissenschaftliche Evaluation untersucht **wissenschaftsgestützt** unter Berücksichtigung der **geltenden Qualitätsstandards** die
➔ ***Effektivität*** (= Ausmaß der Zielerreichung) und
➔ ***Effizienz*** (= ökonomische Ressourcennutzung; Verhältnis von Kosten und Nutzen)
von gesetzten Maßnahmen.

Qualitätsmanagement:

- hat **mit Evaluation identische Zielsetzungen**, und zwar: Vorgang soll auf Basis empirisch feststellbarer Sachverhalte bewertet und optimaler gestaltet werden
- **unterschiedliche Ursprünge**: Evaluation kommt aus Sozialwissenschaft; Qualitätsmanagement aus Wirtschaft.

Wissenschaftliche Evaluation = Teilgebiet der angewandten Forschung
Ziele werden von Forschern / Evaluatoren teils selbst gesetzt, teils von externen Auftraggebern vorgegeben (hier spielt Effizienz = Kosten-Nutzen-Realtion zentrale Rolle!).

Fließende Übergänge zwischen Grundlagenforschung, angewandter Forschung und Evaluationsforschung (alle drei verwenden dieselben empirischen Forschungsmethoden)

Evaluationskonzepte (= Evaluationsmodelle):

werden unterschieden in Abhängigkeit von Zielsetzung und Realisierungsmöglichkeiten.

- ***formative Evaluation***: dient der Begleitung einer Maßnahme; Ziel = laufende Optimierung
- ***summative Evaluation***: bewertet Qualität und Einfluss bereits abgeschlossener Maßnahmen

Bei umfangreichen Maßnahmen muss man auch durchführen:

- ***Baseline-Erhebung***: dient der Beschreibung der Ausgangslage
- ***prospektive Evaluation***: dient der Bewertung der Programmkonzeption
- ***Evaluation der Programmeffizienz***

Außerdem unterscheidet man:

- ***Selbstevaluation***: Mitglieder einer Institution bewerten sich selbst bzw. ihre Maßnahme
- ***Fremdevaluation***: unabhängige Dritte (z.B. außenstehende Forscher) nehmen die Evaluation vor
- ***intern konzipierte Evaluation***: Institution selbst setzt sich die Evaluationsziele und überprüft, ob sie erreicht wurden
- ***extern konzipierte Evaluation***: übergeordnete Institution gibt die Evaluation in Auftrag.

JCSEE – Definitionen:

vom Joint Committee on Standards for Educational Evaluation

Evaluation: Systematische Untersuchung des Wertes oder Nutzens einer Maßnahme (z.B. Unterrichts-, Trainingprogamm, Projekt, Materialien).
Metaevaluation = Evaluation einer Evaluation

Programm: kontinuierlich durchgeführte Bildungsmaßnahme (z.B. Leseprogramm für alle Schüler eines Bezirks, Trainingsprogramm für Arbeitslose, usw.)

Projekt: zeitlich begrenzte Bildungsmaßnahme (z.B. dreitägiges Workshop für Unterrichtsstrategien, zweijährige Testentwicklung, usw.). Institutionalisierte Projekte werden zu Programmen.

Materialien: inhaltlich auf die Maßnahme bezogene Bildungsunterlagen (z.B. Bücher, Lernprogramme, Filme, Tonbänder, usw.)

Evaluationsstandard: Prinzip, dem sich Evaluatoren freiwillig unterwerfen, das bei Einhaltung zur Verbesserung von Qualität und Fairness der Evaluation beiträgt.

Evaluator: Person, die eine Evaluation durchführt

Information: Numerische und verbale Präsentation (z.B. Tatsachen, Beschreibungsberichte, Grafiken, Karten, Ansichten, Statistiken, mündliche Reports), die hilft, bestimmte Aspekt zu illustrieren, Fragen zu beantworten und das Wissen und das Verständnis für ein Programm oder ein anderes Evaluationsobjekt zu erhöhen.

Klient: Einzelner, Gruppe oder Organisation, bei der die Evaluation durchgeführt wird.

Stakeholder: Einzelne oder Gruppen, die in die Progamm-Evaluation involviert bzw. von ihr betroffen sind.

FINK Kapitel 1: Programm Evaluation

Was ist Programm Evaluation?

- = sorgfältige **Untersuchung der Charakteristika und des Nutzens eines Programms**.
- Soll **Informationen über die Effektivität des Projekts** liefern, um **Ergebnisse, Effizienz und Qualität optimieren** zu können.
- Evaluation **analysiert Aufbau, Aktivitäten und Ablauf** eines Programms und untersucht **politische und gesellschaftliche Rahmenbedingungen**.
- Evaluation kann eingesetzt werden, um die **Zielerreichung zu überprüfen** und um festzustellen, ob das Ergebnis den Auswand rechtfertigt **(Kosten-Nutzen-Relation)**

Evaluationsmethoden

Eingesetzt werden **dieselben Methoden wie in anderen Sozialwissenschaften**, um reliable und valide Beweise zu sammeln:

- Formulierung von **Fragestellungen**,
- Aufstellen von **Standards zur Überprüfung der Zielerreichung**,
- Erstellung eines **Evaluationsdesigns**,
- **Datenerhebung**, **Datenauswertung** und **Interpretation**,
- **Bericht über die Ergebnisse**.

Wichtig ist **sorgfältige Wahl und Anwendung der Methoden**.

Programm oder Intervention

- = **systematische Bestrebung, vorher geplantes Ziel zu erreichen** (z.B. Wissenszuwachs, Veränderung von Verhalten und Einstellungen, Übung).
- Programme können **relativ klein** sein (z.B. Elternschule in einer einzigen Schule), aber auch **relativ umfangreich** (z.B. Head Start).
- Sie können in **verschiedensten geografischen und politischen Umgebungen** stattfinden und sich, was Zweck, Aufbau, Organisation und Teilnehmer betrifft, stark voneinander unterscheiden.

Nutzen eines Programms

- Ihn festzustellen, ist eines der **Hauptziele der Evaluation**.
- Ein **nützliches Programm** hat wertvolle Ziele, entspricht den Effektivitätskriterien, bringt Vorteile für die Teilnehmer, informiert die Teilnehmer vollständig über mögliche Risken bei der Teilnahme und richtet absichtlich keinen Schaden an.

- **Evaluation unterscheidet sich von anderen Forschungsmethoden** in Bildung, Psychologie und den Sozialwissenschaften dadurch, dass überzeugende Beweise für die Nützlichkeit eines Programms identifiziert, der Nutzen des Programms gemessen und die nutzbringendsten Aspekte des Programms beschrieben werden.

Programmziele und Aktivitäten

- **Programmziele = geplant**, z.B. Optimierung der Effizienz einer Ausbildung. Eine der Hauptfunktionen der Evaluation = Daten zu liefern -> Evaluator kann helfen beim Festlegen von Programmzielen bzw. beim Herausarbeiten von Prioritäten (z.B. bei Auswahl von Lehrbüchern gemeinsames Vorgehen zwischen Direktion, Lehrern, Eltern, Schülern)

- Evaluation beantwortet auch die Frage, welche Aktivitäten bei einem Programm gesetzt werden sollen; bietet Einsicht in Anwendung und Durchführung des Programms -> **Prozess-Evaluation** (= Evaluation der Programm-Aktivitäten; wichtig, um Einsichten zu gewinnen, zum Formulieren von Hypothesen für spätere Testungen, zur Feststellung der Erreichung bzw. Nichterreichung der Programmziele).

  BEISPIEL: Evaluation zweier Programme zur Förderung des Anlegens von Sicherheitsgurten bei jugendlichen Autofahrern -> mögliche Fragen des Evaluators: Haben alle betroffenen Jugendlichen an dem Programm teilgenommen? Haben alle Lehrer die selbe Unterrichtsmethode verwendet? Waren manche Lehrer einfach besser als andere? War die Teilnahme an beiden Programmen vergleichbar bzw. wurde sie vergleichbar forciert?

Outcome eines Programms

- Ziel vieler Programme im Bereich Bildung, Psychologie und Sozialwissenschaft = Wissen, Verständnis, funktionierende Persönlichkeit in einer demokratischen Gesellschaft -> ist aber sehr schwer zu messen... (keine übereinstimmenden Definitionen; zu wenig Zeit, um das wirklich beobachten zu können).

- Daher: **Überprüfung solcher Zielerreichungen, die einfacher zu erfassen sind**, wie z.B. Wissenszuwachs, erwünschte Veränderungen in Werthaltungen, Einstellungen, Fertigkeiten und Verhaltensweisen. Kann das Programm das bewirken, so ist auch die Erreichung fernerer Ziele eher wahrscheinlich.

Impact eines Programms

- = **Ausmaß der Effekte, Andauer der Ergebnisse, Ausmaß der Beeinflussung** in anderen Umgebungen und bei anderen Personen.

BEISPIEL: Evaluation zweier Programme zur Verbesserung der Lesefähigkeit.

- Programm A -> Leseleistung wurde verbessert; Effekte hielten mindestens 3 Jahre an; auch in anderen Schulen feststellbare Verbesserungen

- Programm B -> ebenfalls Verbesserung der Leseleistung, Effektdauer mindestens 3 Jahre, ABER: bei weniger Personen waren Verbesserungen feststellbar. Nicht so gut auf andere Schulen anwendbar.

Fazit: Progamm A hatte größeren Impact (mehr Leute erzielten eine Leistungsverbesserung für längere Zeitdauer)

Aufwand und Kosten eines Programms

- Wichtig bei Evaluation = **Kosten eines Programms und ihre Beziehung zu seinen Effekten bzw. seinem Nutzen**.

- **Typische Fragen des Evaluators**: Wie effizient ist das Programm? Welche zusätzlichen Kosten entstehen z.B. einer Schule durch die Anwendung dieses Programms? Welche Vorteile bringt z.B. ein Sportprogramm an der Schule mit sich? (z.B. weniger Absenzen, da gesündere Schüler, weniger Kosten fürs Gesundheitswesen,...)

Auftraggeber von Evaluationen:

- Behörden (z.B. Ministerium, Krankenkassen,...)
- Entwickler von bestimmten Programmen (z.B. Stadtschulrat, AMS,...)
- Politiker (z.B. Gemeinde, Land, Bund,...)
- Kostenträger von Programmen (z.B. Stiftungen, Vereine,...)
- Bildungs- und Sozialwissenschaftler (z.B. Universität, Wirtschaft,...)

Evaluationen werden manchmal von Teams durchgeführt, wobei jedes Mitglied des Evaluatorenteams einen bestimmten Bereich untersucht und einen eigenen Beitrag zum **gemeinsamen Expertenurteil** leistet.

BEISPIEL: Programm-Evaluation als interdiziplinäre Maßnahme
3-jährige Evaluation eines speziellen Trainingsprogramms in einem Pflegeheim: 2 professionelle Evaluatoren, ein Physiotherapeut, 2 Trainer, 1 Krankenschwester, Kostenrechner, angelernte Hilfskräfte zur Datenerhebung

BORTZ: TAFEL 1.1

Frauenfeindlichkeit: Formulierung einer wissenschaftlichen Hypothese

Studentin plant Untersuchung zum Thema Frauenfeindlichkeit;

- **Ausgangspunkt** = Zeitungsartikel über „Frauenfeindlichkeit“ im Fernsehen (Inhalt: die meisten Sendungen vermitteln ein falsches Frauenbild).

- **Behauptung**: Falsches Frauenbild beeinflusst Reaktionen der Männer auf Frauen, daher: Fernsehende Männer sind frauenfeindlich.

- Frage: **Handelt es sich bei dieser Behauptung um eine wissenschaftliche Hypothese?**

    - ***1. Kriterium = empirischer Gehalt*** -> trifft nach Meinung der Studentin zu (Frauenfeindlichkeit = realer Sachverhalt, der sich empirisch untersuchen lässt)

    - ***2. Kriterium = Allgemeingültigkeit*** -> trifft nach Meinung der Studentin zu (sie denkt an alle fernsehenden Männer)

    - ***3. Kriterium = Konditionalsatz*** -> „Wenn Männer fernsehen, dann sind sie frauenfeindlich.“ Dafür aber wird es schwierig sein, VPn zu finden [weil sie auch nicht-fernsehende Männer bräuchte], daher: „Je häufiger Männer fernsehen, desto frauenfeindlicher sind sie.“

  ⇨ wissenschaftliche Hypothese lautet hiermit: „Je häufiger Männer fernsehen, desto frauenfeindlicher sind sie.“

- Frage: **Ist die Hypothese als Kausalhypothese interpretierbar?** [d.h. Fernsehen = Ursache für Frauenfeindlichkeit]

  - Dafür schauen, ob ***Je-Teil der Hypothese und Desto-Teil der Hypothese austauschbar*** sind -> „Je frauenfeindlicher Männer sind, desto häufiger sehen sie fern.“ Jetzt sieht man, dass die Hypothese keine strenge, gerichtete Kausalannahme beinhaltet (zeitliche Ursache [erst Fernsehen, dann Frauenfeindlichkeit] ist nicht zwingend!)

  - Man sieht daran auch, dass Frauenfeindlichkeit ***nicht monokausal*** durch das Fernsehen beeinflusst wird -> Frauenfeindlichkeit ***hat viele Ursachen*** -> es wird sich bei Untersuchung ***kein perfekter Zusammenhang zwischen Frauenfeindlichkeit und Fernsehen nachweisen lassen***

- **Fazit**: Studentin plant nun ihre Untersuchung mit dem Ziel, eine einfache Zusammenhanghypothese zu überprüfen.

BORTZ: KAPITEL 1.3

PRAKTISCHES VORGEHEN

Zur Lösung der wissenschaftstheoretischen Fragen gibt es kein Patentrezept, sondern verschiedene Lösungsansätze, deren Vertreter einander zum Teil heftig kritisieren.

Kritischer Rationalismus: = bekanntestes wissenschaftstheoretisches Rahmenmodell

Am Beginn des 20. Jhds. wurde **statistische Hypothesenüberprüfung** mittels **Signifikanztests** entwickelt -> hat sich heute in der empirischen Forschung weitgehend durchgesetzt.

- ***K. Pearson*** (1896): Korrelations- und Regressionsanalysen
- ***Fisher & McKenzie*** (1923): Varianzanalyse
- ***Popper*** (1934): Logik der Forschung

Wie kann man Forschungslogik des Kritischen Rationalismus bzw. Falsifikationismus und statistischen Signifikanztest als Methode der Hypothesenüberprüfung verbinden?

- ***Fisher***: Anhänger eines induktiven Modells;
  „**induktive Inferenz**“ = Schlussfolgerung von Stichproben auf Populationen, von Beobachtungsdaten auf Hypothesen.
  Zu Erkenntnisfortschritt kommt man durch **wiederholtes Widerlegen von Nullhypothesen** bzw. durch indirekten Nachweis von Effekten.
  Ausgangspunkt = immer statistische Hypothesen (d.s. **Wahrscheinlichkeitsmodelle**)

- ***Popper***: ihn interessieren auch die unterschiedlichen Abstraktionsebenen von Aussagen und die Ableitungsbedingungen zwischen Hypothesen.
  Falsifikation bezieht sich bei Popper auf Relation zwischen Hypothesen, Randbedingungen und Theorien – Fischer dagegen beschäftigt sich primär mit der Relation von Daten und statistischen Hypothesen.

statistische Hypothesen bzw. Wahrscheinlichkeitsaussagen sind

- ***weder falsifizierbar*** (Wahrscheinlichkeitsaussagen lassen grundsätzlich alle Ereignisse zu, schreiben ihnen nur unterschiedliche Auftretenshäufigkeiten zu, deshalb gibt es keine logisch falsifizierenden Ereignisse)
- ***noch verifizierbar*** (es lassen sich nicht alle Elemente in einer Population, über die eine Aussage getroffen werden soll, untersuchen)

ABER: Bei statistischen Hypothesen lässt sich die **Falsifizierbarkeit durch Festlegung von Falsifikationskriterein HERSTELLEN**
➔ Festlegung eines Signifikanzniveaus (Fisher) = **Vereinbarung einer Falsifikaltionsregel**.

Merke: Daten „sagen“ nicht, ob eine Hypothese „stimmt“, sondern **Daten bilden die Grundlage einer Entscheidung für oder gegen eine Hypothese**. Möglichkeit einer falschen Entscheidung kann NIE ausgeschlossen werden, soll aber minimiert werden.

STATISTISCHE HYPOTHESENÜBERPRÜFUNG

- **Ausgangspunkt** = eine **Theorie** (bzw. als Ersatz eine gut begründete Überzeugung).

- Aus ihr wird unter Festlegung von Randbedingungen eine **inhaltliche Hypothese** abgeleitet.

- Diese wird in eine **statistische Hypothese** umformuliert. Statistische Hypothese **sagt das Ergebnis einer empirischen Untersuchung voraus** (Prognose) und gibt durch den theoretischen Hintergrund eine Erklärung für den untersuchten Effekt.

1) ***Untersuchungsplanung***

BEISPIEL: **Arousal-Theorie**

Aus ihr kann man folgende **Forschungshypothese ableiten**:
⇨ Harmoniefolgen mit mittlerem Erregungspotential werden positiv bewertet.

Jetzt muss man die **Zielpopulation bestimmen** (z.B. alle erwachsenen Personen aus dem westeuropäischen Kulturkreis). Aus dieser Zielpopulation soll später eine **Stichprobe** gezogen werden.

Dann muss man sich überlegen, welche Art von Harmoniefolgen die VPn bewerten und wie sie ihre Urteile äußern sollen = **Festlegung eines Untersuchungsdesigns**.

➔ z.B. 2-Gruppen-Design:
- eine Gruppe hört zufällig ausgewählte Popmusik (= ***Kontroll-gruppe***)
- eine Gruppe hört nur Popmusik mit mittlerem Erregungspotential (= ***Experimentalgruppe***)

➔ Zusammenstellung der Musikstücke mit mittlerem Erregungspotential durch Fachleute (z.B. Musikwissenschaftler)

➔ Zuordnung der VPn zu den beiden Gruppen erfolgt zufällig (**Randomisierung**)

➔ Zur Erfassung der Einschätzung der Musikstücke Einsatz von **Rating-Skalen** (z.B. gefällt mir gar nicht – wenig – teils/teils – ziemlich – völlig)

➔ jede VP hört 20 Musikstücke, bewertet jedes auf einer Ratingskala; Musikstück erhält Punkte (z.B. gefällt mir gar nicht = 0 Punkte – gefällt mir völlig = 5 Punkte). Je höher die Punktzahl, umso positiver ist die Bewertung eines Musikstücks.

2) ***Statistisches Hypothesenpaar:***

Inhaltliche Unterschiedshypothese („Experimentalgruppe soll die Musikstücke positiver bewerten als die Kontrollgruppe“) wird **in eine statistische Mittelwertshypothese umgewandelt** („Mittelwert der Musikbewertungen in der Experimentalgruppe ist größer als Mittelwert in der Kontrollgruppe“, d.h. $\mu_1 > \mu_2$)

Charakteristikum der statistischen Hypothesenprüfung:

Es gibt IMMER ein Hypothesenpaar:

- ***Alternativhypothese (H_1)*** [sie entspricht üblicherweise der Forschungshypothese], kann sein:

  ➔ **gerichtet** [d.h. Unterschiede der Mittelwerte unter Experimentalbedingung sind größer oder kleiner als Unterschiede der Mittelwerte unter der Kontrollbedingung]

  ➔ **ungerichtet** [d.h. es gibt Unterschiede zwischen den Mittelwerte in der Experimentalbedingung und in der Kontrollbedingung]

- ***Nullhypothese (H_0)*** [sie widerspricht der Alternativhypothese]: drückt aus, dass Unterschiede, Zusammenhänge, Veränderungen, besondere **Effekte NICHT oder NICHT IN DER ERWARTETEN RICHTUNG auftreten** [d.h. es gibt keine Unterschiede zwischen den Mittelwerten in der Experimentalbedingung und in der Kontrollbedingung].

WICHTIG: **statistisches Hypothesenpaar** muss **VOR der Untersuchung** so formuliert werden, dass alle Ausgänge der Untersuchung abgedeckt sind.

MERKE:	Eine statistische Hypothese wird stets als **statistisches Hypothesenpaar**, bestehend aus **Nullhypothese (H_0)** und **Alternativhypothese (H_1)**, formuliert. Die Alternativhypothese postuliert dabei einen bestimmten Effekt, den die Nullhypothese negiert.

H_0 und H_1 stehen *zueinander* in einem komplementären Verhältnis; Folge: Bei Zurückweisung der H_0 kann „automatisch“ auf H_1 geschlossen werden.

3) ***Auswahl eines Signifikanztests***

Kriterien für die Auswahl sind z.B.
- Anzahl der Untersuchungsgruppen
- Anzahl der VPn
- Anzahl der unabhängigen und abhängigen Variablen
- Qualität der Daten

BEISPIELE:
- für Mittelwertsvergleich zwischen 2 Gruppen -> t-Test
- für Mittelwertsvergleich zwischen mehreren Gruppen -> Varianzanalyse
- für Zusammenhangshypothese -> Korrelationstest

WICHTIG: muss **VOR der Datenerhebung festgelegt** werden

4) ***Stichprobenergebnis***

Stichprobenergebnis [= ***deskriptives Ergebnis***] (z.B. X-quer$_1$ = 3,4 und X-quer$_2$ = 2,9) gibt „per Augenschein" erste Hinweise auf empirische Haltbarkeit der Hypothesen.
ABER: Es erlaubt **KEINE Einschätzung, ob Ergebnis auf Population generalisierbar ist**! Dafür braucht man Signifikanztest!

5) ***Berechnung der Irrtumswahrscheinlichkeit mittels Signifikanztests***

Signifikanztest fragt danach, ob Ergebnis durch die Nullhypothese erklärt werden kann -> über ein Wahrscheinlichkeitsmodell wird ein Wert ermittelt (= **Irrtumswahrscheinlichkeit**), der angibt, mit welcher bedingten Wahrscheinlichkeit das gefundene Ergebnis auftritt, wenn in der Population die Nullhypothese gilt.

MERKE: Die **Irrtumswahrscheinlichkeit** ist die bedingte Wahrscheinlichkeit, dass das empirisch gefundene Stichprobenergebnis zustande kommt, wenn in der Population die Nullhypothese gilt.

6) ***Signifikante und nicht-signifikante Ergebnisse:***

Vernachlässigbar **geringer Unterschied zwischen Experimental- und Kontrollgruppe** ergibt **hohe Irrtumswahrscheinlichkeit im Signifikanztest**
➔ ***nicht-signifikantes Ergebnis***. D.h. die Alternativhypothese gilt als nicht bestätigt [nähme man sie dennoch an, so hat man sich mit hoher Wahrscheinlichkeit geirrt!].

Je größer Mittelwertsunterschiede zwischen VG und KG, **desto schlechter vereinbar mit der Nullhypothese**. Wenn VG wirklich repräsentativ für die Population war, so ist Unterschied zwischen den Mittelwerten beider Gruppen mit hoher Wahrscheinlichkeit NICHT ZUFÄLLIG.

Fazit: Ist Stichprobenergebnis schlecht vereinbar mit der H0, so ergibt der **Signifikanztest eine geringe Irrtumswahrscheinlichkeit**
➔ ***signifikantes Ergebnis***. D.h. Verwerfung der H0 und Annahme der H1.

MERKE: Ein **signifikantes Ergebnis** liegt vor, wenn ein Signifikanztest eine **sehr geringe Irrtumswahrscheinlichkeit** ermittelt. Dies bedeutet, dass sich das gefundene Stichprobenergebnis nicht gut mit der Annahme vereinbaren lässt, dass in der Population die Nullhypothese gilt. Man lehnt deshalb die Nullhypothese ab und akzeptiert die Alternativhypothese.

ABER: **Restrisiko bleibt IMMER bestehen**, weil es trotzdem vorkommen könnte, dass in der Population die H0 gilt und aufgrund untypischer VPn die gefundenen Effekte reine Zufallsprodukte sind.

7) ***Signifikanzniveau***:

Um Restrisiko, dass H0 irrtümlich verworfen wird, zu minimieren, wird **Signifikanzniveau** festgelegt -> **meist wählt man 5%** [d.h. ist Irrtumswahrscheinlichkeit kleiner als 5%, so kann H1 angenommen werden]

MERKE: Bei α = 0.05 darf man NICHT sagen, dass die H1 mit 95%iger Wahrscheinlichkeit gilt, sondern:
„*die Wahrscheinlichkeit für das Untersuchungsergebnis für den Fall, dass die H0 gilt, beträgt 5%*".

Grund: **Irrtumswahrscheinlichkeit** ist eine **bedingte DATENWAHRSCHEINLICHKEIT,** aber **KEINE HYPOTHESENWAHRSCHEINLICHKEIT!**

In besonderen Fällen (z.B. wenn vom Ergebnis praktische Konsequenzen abhängen und Irrtum gravierende Folgen hätte) prüft man noch strenger, dann nimmt man ein **Signifikanzniveau von 1%.** [in Grundlagenforschung = 5% üblich]

–> ***α-Fehler-Wahrscheinlichkeit*** (= Irrtumswahrscheinlichkeit):
überprüft, **wie gut die Daten zur Nullhypothese passen**

–> ***β-Fehler-Wahrscheinlichkeit***:
überprüft, **wie gut sich Daten mit den in der Alternativhypothese formulierten Populationsverhältnissen vereinbaren lassen**. Wird verwendet im sogenannten „Hybrid"-Modell; Signifikanztest dazu von Neymann & Pearson.

ERKENNTNISGEWINN DURCH STATISTISCHE HYPOTHESENTESTS

Erkenntniszugewinn entsteht durch die Eliminierung falscher Theorien, d.h. durch deren **Falsifikation** (= Nachweis der Untauglichkeit einer Theorie durch kritische Empirie). Dem entspricht beim statistischen Hypothesenprüfen **ein nicht-signifkantes Ergebnis** (= Ergebnis, bei dem die aus einer Theorie abgeleitete Forschungshypothese als nicht bestätigt gilt). **Falsifizierende Untersuchungen** = Untersuchungen mit nicht-signifikanten Ergebnissen.

Empirische Falsifikation erfordert NICHT die Ablehnung der gesamten, der Hypothese zugrunde liegenden Theorie. Vor dem Verwerfen der „Kerntheorie" muss überprüft werden, ob die Ursache für das nicht-signifikante Ergebnis im „Schutzgürtel der Hilfstheorien" zu finden ist (d.h. gab es Untersuchungsfehler wie ungeeignete operationalisierte Indikatoren, ungenaue Messvorschriften, etc.)

Teile der Untersuchungsergebnisse können trotzdem hypothesenkonform sein, daher Replikation der Untersuchung! Erst wenn mehrere Replikationen wieder zu nicht-signifikanten Ergebnissen führen, wird Theorie schließlich aufgegeben werden müssen.

Falsifikation führt nicht nur ***zum Aussondern von Theorien***, sondern auch zu

- ***Modifikationen von Theorien***
  (z. B. durch **EXHAUSTION** = Theorienmodifikation, bei der der Wenn-Teil der Theorie durch eine oder mehrere „Und-Komponenten" erweitert werden)
- ***Neubildung von Theorien***

ABER: Scientific Community muss falsifizierende Untersuchungsbefunde auch in ausreichendem Maß bekannt machen (heute aber werden Publikationen mit signifikanten Ergebnissen bevorzugt publiziert...)

Signifikantes Ergebnis = ENTSCHEIDUNGSGRUNDLAGE für vorläufige Annahme der Forschungshypothese bzw. der geprüften Theorie.

ABER:

- **Forschungshypothese ist durch signifikantes Ergebnis NICHT bestätigt** (-> Induktionsschluss, bei dem aus begrenzter Anzahl theoriekonformer Ereignisse auf uneingeschränkte Gültigkeit der Theorie geschlossen wird; [wäre eine Verifikation, die aber nicht möglich ist!]
- **Es bleibt das statistische Restrisiko** von z.B. 5%, dass man sich geirrt hat

MERKE: **Statistische Signifikanz** liegt vor, wenn die empirisch ermittelte ***Irrtumswahrscheinlichkeit*** das konventionell festgelegte ***Signifikanzniveau (z.B. 1% oder 5%) unterschreitet***.

Statistische Signifikanz ist ein per Konvention festgelegtes Entscheidungskriterium für die ***vorläufige Annahme von statistischen Populationshypothesen***.

In Naturwissenschaft löst **ein einziges, dem Wenn-Dann-Satz widersprechendes Ereignis Zweifel an der Richtigkeit der Aussage** aus. Ist berechtigt z.B. in der Physik, wo man mit prinzipiell austauschbaren Objekten zu tun hat.

ABER: **bei sehr heterogenen Objekten** (z.B. Menschen, Schulklassen, Tiere,...) ist diese **Austauschbarkeit nicht gegeben** -> daher: ***Hypothesenprüfung***, die sich nicht auf ein einzelnes Untersuchungsobjekt bezieht, sondern ***auf eine STICHPROBE von Objekten, die für die POPULATION von Untersuchungsobjekten repräsentativ ist,*** für die die Hypothese Gültigkeit beansprucht.

in der Stichprobe können sich mehrere der Hypothese widersprechende Einzelereignisse befinden (Falsifikation erst notwenidg, wenn sie zu „häufig“ vorkommen bzw. wenn Abweichungen zu gravierend sind!)

Statische Hypothesen = Aggregathypothesen:

machen **Aussagen über die Tendenz von Gruppen und nicht über jeden Einzelfall**. D.h. individuelle Daten der VPn werden zu einem Gesamtwert zusammengefasst (aggregiert) -> erst über diesen Aggregatwert werden Aussagen gemacht.

ABER: **Mittelwert ist nur ein sehr grobes Maß** (vgl. Verzerrung der Mittelwerte durch Ausreißer!), daher VOR der Mittelung Stichprobe anschauen!

MERKE: Bei einem **Aggregatwert** handelt es sich um die **Zusammenfassung der Individualwerte einer Variablen über eine Gruppe** von Untersuchungspersonen bzw. Untersuchungsobjekten hinweg. Obwohl die Bildung von Aggregatwerten **statistisch immer möglich** ist, sollte jeweils **genau überprüft** werden, ob Aggregatwerte im Kontext der konkreten Untersuchung

a) inhaltlich sinnvoll interpretierbar sind und
b) die Merkmalverteilung innerhalb der Gruppe angemessen repräsentieren.

Ein **Zusammenhang** kann **statistisch signifikant**, aber **dennoch praktisch nicht von Bedeutung** sein, dann spricht man von einer ***Effektgröße***. Aufpassen muss man, denn:
- kleine Effekte können aufgrund ihrer Signifikanz überbewertet werden
- kleine Effekte können aber auch in ihrer Bedeutung unterschätzt werden (z.B. bei parapsychologischen Phänomenen)

BORTZ: KAPITEL 1.4

AUFGABEN DER EMPIRISCHEN FORSCHUNG

Der Wert einer **empirischen Methode** kann nicht für sich allein beurteilt werden, sondern nur daran gemessen werden, inwieweit die Methode **den inhaltlichen Anforderungen einer Untersuchung gerecht** wird.

Hauptaufgaben empirischer Forschung:
- Erkundung von Hypothesen
- Überprüfung von Hypothesen

1) ***Hypothesenprüfung und Hypothesenerkundung***

Empirische Untersuchungen = Untersuchungen, die **auf Erfahrung beruhen**.

Dazu gehören z.B.
- ***Einzelfallstudie***, die die Biographie eines Einzelnen beschreibt
- ***Experimentelle Untersuchung*** zur Wirksamkeit verschiedener Unterrichtsmethoden

ABER -> wichtige Unterschiede!

- Biographiestudie: sammelt Erfahrungen, aus denen sich Vermutungen ableiten lassen (z.B. über Bedeutung außergewöhnlicher Lebensereignisse, über Entwicklung von Einstellungen,...) -> kann weitere Untersuchungen veranlassen
  ⇨ **empirische Daten dienen der Formulierung einer Hypothese**

- Vergleichende Untersuchung verschiedener Untersuchungsmethoden: am Anfang steht gut begründete Hypothese; sie wird durch systematisch herbeigeführte Erfahrungen (= experimentelle Untersuchung) bestätigt oder verworfen
  ⇨ **empirische Daten dienen der Überprüfung einer Hypothese**

Funktionen empirischer Forschung:

- Empirische Forschung will allgemeingültige Erkenntnisse gewinnen -> Theorien und Hypothesen sind daher allgemein (für klar definierten Geltungsbereich) formuliert.
  Aufgabe der Sozialwissenschaften = durch empirische Untersuchung zu überprüfen, inwieweit sich die aus Theorien, Voruntersuchungen oder persönlichen Überzeugungen abgeleiteten Hypothesen in der Realität bewähren
  ⇨ ***hypothesenprüfende oder deduktive Funktion*** der empirischen Forschung (**Deduktion** = Ableitung des Besonderen aus dem Allgemeinen)

- Rascher zeitlicher Wandel -> viele Theorien sind bald wieder veraltert; daher: empirische Forschung muss sich auch um die Entwicklung neuer Theorien bemühen.
  Das erfordert einerseits geistige, andererseits aber auch empirische Arbeit -> reales Geschehen beobachten, beschreiben, protokollieren -> daraus können neue Hypothesen abgeleitet werden
  ⇨ ***hypothesenerkundende oder induktive Funktion*** der empirischen Forschung (**Induktion** = Schließen vom Einzelnen auf etwas Allgemeines)

MERKE: Eine **Hypothese** ist
-> bei **induktiver Vorgangsweise das Resultat** und
-> bei **deduktiver Vorgangsweise der Ausgangspunkt**
einer empirischen Untersuchung

Ob **Untersuchung zur Erkundung oder Überprüfung einer Hypothese** durchgeführt wird, **richtet sich nach dem Wissensstand** in dem Forschungsfeld:

- Kenntnisse und Theorien sind vorhanden
  ➔ hypothesenüberprüfende Untersuchung

- keine Kenntnisse und Theorien sind vorhanden
  ➔ hypothesengenerierende Untersuchung

Strikte Dichotomie zwischen erkundenden und überprüfenden Untersuchungen in der Praxis selten; meist Anknüpfung an Bekanntes und daraus Entwickelung von Neuem.

2) ***Empirische Forschung und Alltagserfahrung***

Alltagstheorien und wissenschaftliche Theorien unterscheiden sich
- in den Fragestellungen
- in den Erkenntnismethoden
- in der Art der getroffenen Aussagen

Kriterien einer wissenschaftlichen Theorie:
- muss empirisch falsifizierbar
- allgemeingültig und
- konditional sein

Alltagstheorie:
- geht von subjektiven Erfahrungen aus
- auf Einzelfall zugeschnitten
- dient der Orientierung, Sinngebung und Selbstdefinition

Unterschiede zwischen empirischer Forschung und Alltagserfahrung:

a) **Systematische Dokumentation**

Sinneserfahrungen und ihre Verarbeitung sind **zunächst grundsätzlich subjektiv**. Damit sie zum Gegenstand wissenschaftlicher Auseinandersetzung werden können, müssen Umstände, unter denen die Erfahrungen gemacht wurden, wiederholbar gestaltet und genau beschrieben werden -> **Replikation muss möglich sein, damit andere diese Erfahrungen bestätigen können!**

Objektivität = intersubjektive Nachprüfbarkeit;
Voraussetzung = **Standardisierung des Vorgehens** durch methodische Regeln und vollständige Dokumentation von Untersuchungen (= **Transparenz**)

Alltagserfahrung:
- wird unsystematisch gesammelt
- wird in der Regel nicht dokumentiert
- gewonnene Überzeugung kann kaum nachvollzogen werden

b) **Präzise Terminologie**

Alltagserfahrungen werden **umgangssprachlich** mitgeteilt. Umgangssprache ist aber **oft zu ungenau**, um zweifelsfrei unterscheiden zu können, ob von verschiedenen Personen gemachte Erfahrungen identisch oder unterschiedlich sind.

In **wissenschaftlicher Forschung** dagegen bedient man sich eines **Vokabulars**, über dessen Bedeutung sich die **Vertreter eines Faches weitgehend geeinigt** haben. Wissenschaftliches Vokabular = auch oft ungenau, daher: **Erhöhung der Präzision durch Operationalisierung** (= Angabe von konkreten Verhaltensweisen z.B. für Leistungsmotivation)

c) **Statistische Analysen**

Wissenschaftliche Aussagen beanspruchen **Allgemeingültigkeit** -> beziehen sich auf ***Populationen*** (= Grundgesamtheiten). Vollerhebung einer Population ist nicht möglich, daher werden nur Ausschnitte aus der Population untersucht = ***Stichproben***; von diesen wird **auf die Population generalisiert**.
⇨ **Je repräsentativer die Stichprobe, umso gesicherter ist die Verallgemeinerung.**

Um Wahrscheinlichkeitsaussagen über Populationen auf der Basis von Stichprobendaten zu überprüfen, verwendet man die Methoden der ***Inferenzstatistik*** (vor allem Signifikanztests). Signifikanzergebnis kann als Entscheidungsgrundlage für die Bewertung von Populationshypothesen herangezogen werden, weil mit dem **Signifikanzniveau ein Falsifikationskriterium** für Wahrscheinlichkeitsaussagen eingeführt wird
⇨ Risiko einer **falschen Entscheidung** kann so **kalkuliert und minimiert** werden

Alltag: es werden Entscheidungen aufgrund von subjektiven Wahrscheinlichkeiten getroffen

d) **Interne und externe Validität**

➔ **zu strenge Definitions-, Operationalisierungs- oder Messvorschriften** bewirken, dass untersuchter Realitätsausschnitt verkürzt, unvollständig oder verzerrt erfasst wird -> **Gültigkeit der Erkenntnisse wird dadurch zweifelhaft**.

➔ **zu ungenaue Beschreibung** der untersuchten Merkmale oder Untersuchungsobjekte -> **Ergebnisse werden mehrdeutig**

ABER: **Besonders wichtig = eindeutig interpretierbare und generalisierbare Untersuchungsergebnisse**

- ***interne Validität***: Eindeutigkeit, mit der ein Untersuchungsergebnis inhaltlich auf die Hypothese bezogen werden kann

- ***externe Validität***: Generalisierbarkeit der Ergebnisse einer Untersuchung auf andere Personen, Objekte, Situationen und/oder andere Zeitpunkte

Merke: **interne und externe Validität sind die wichtigsten Qualitätsunterschiede zwischen empirischer Forschung und Alltagserfahrung**

Alltag: Gültigkeitsbeurteilung beruht auf wesentlich auf Intuition, Weltbild, anekdotischen Evidenzen, usw.; häufiges Zurückgreifen auf Pauschalisierungen

e) **Umgang mit Theorien**

Wissenschaftliche Theorien sind
- einem **permanenten Prozess der Überprüfung und Kritik** ausgesetzt
- von **Konsensfähigkeit** abhängig

Theorie hat nur dann Chance auf Verbreitung im wissenschaftlichen Publikationswesen, wenn sie bei Fachkollegen auf Akzeptanz stößt.

Peer Reviewing: Manuskripte durchlaufen Prozess der Begutachtung, werden von Fachkollegen entweder bedingt (d.h. es müssen Veränderungen vorgenommen werden) oder unbedingt zur Publikation empfohlen oder abgelehnt.

Funktion des wissenschaftlichen Diskurses = einseitige Sichtweisen und Voreingenommenheiten eigenen Theorien gegenüber aufzudecken, Alternativvorschläge zu erarbeiten, Theorien zu ergänzen, usw. Wichtig dabei =

Pluralität; Publikationschancen dürfen nicht nach Position und Status vergeben werden -> daher bei vielen Fachzeitschriften **anonymes Peer-Reviewing** (Begutachtende und Begutachtete kennen einander nicht)

Andere wichtige Plattform für kritische Diskussion von Forschungsarbeiten ist neben **Zeitschriftenwesen** auch das **Tagungs- und Kongresswesen**.

BORTZ: KAPITEL 3.1.1

EVALUATIONSFORSCHUNG UND GRUNDLAGENFORSCHUNG

Evaluationsforschung sollte sich **an den methodischen Standards der empirische Grundlagenforschung orientieren**. Diese dürfen NICHT zugunsten eines „auftraggebenfreundlichen" Untersuchungsdesigns oder Berichterstattung aufgegeben werden [darüber gibt es in der Literatur aber unterschiedliche Ansichten...]

Evaluationsstudie muss so abgefasst werden,
- dass sie ***für den Auftraggeber verständlich*** ist [d.h. keine ausufernde Fachterminologie verwenden]
- dass sie ***eine klare Entscheidung nahe legt*** [z.B. soll die Maßnahme fortgesetzt oder abgebrochen werden]

1) ***Gebundene und offene Forschungsziele***

Grundlagenforschung:

thematischer Rahmen ist vorgegeben; ABER: **zu starke Fixierung auf das intendierte Forschungsziel ist nicht gut**, weil viele Forschungsergebnisse auf wissenschaftlichem Neuland NICHT Produkte einer zielgerichteten Forschung sind, sondern durch Integration und Berücksichtigung von scheinbar irrelevanten oder auch fachfremden Ideen entstehen.

Evaluationsforschung:

ist in der Regel **Auftragsforschung**, d.h. ein Auftraggeber (z.B. Ministerium, Behörde, Unternehmen,...) stellt zur Begleitung und Bewertung einer geplanten oder durchgeführten Maßnahme Geld zur Verfügung [Evaluationsforschung = **Begleitforschung**] -> unterliegt daher anderen Limitierungen als die Grundlagenforschung.

Evaluationsforschung **= auftraggeberorientiert und zielorientiert**
(d.h. alles muss darauf ausgerichtet sein, die vom Auftraggeber gestellte Evaluationsfrage möglichst eindeutig und verständlich zu beantworten)

BEISPIEL: Welche Konsequenzen hat eine Begrenzung der Fahrgeschwindigkeit auf 100km/h?

- ➔ Gegenstand der Studie = alle Auswirkungen genau dieser Maßnahme
- ➔ Nicht dagegen: optimale Fahrgeschwindigkeit für Personen verschiedenen Alters, günstigste Richtgeschwindigkeit in Abhängigkeit von Verkehrsdichte und Witterungsbedingungen, mit der Fahrgeschwindigkeit verbundene psychische und physische Belastungen, usw. [das wären Aufgabenstellungen für Angewandte Forschung]

Grundlagenforschung: **Ziel = Generierung von Hintergrundwissen**, dessen **funktionaler Wert** nicht unmittelbar erkennbar sein muss und der deshalb **von untergeordneter Bedeutung** ist.

BEISPIELE:

- ➔ Gerontologische Studien über altersbedingte Beeinträchtigung des Reaktionsvermögens
- ➔ Studien über Signal- und Informationsverarbeitung unter Stressbedingungen
- ➔ Studien über wahrnehmungspsychologische Erkenntnisse zur Identifizierung von Gefahren

ad Wertfreiheit von Forschung:

Evaluationsforschung wird oft vorgeworfen, sie wäre nicht wertfrei, sondern:

- ➔ ***parteilich*** (ist so angelegt, dass gewünschtes Ergebnis mit hoher Wahrscheinlichkeit eintritt)
- ➔ ***abhängig von politisch-ideologischen Positionen*** des Evaluators bzw. Auftraggebers,
- ➔ ***abhängig von finanziellen Interessen*** der von der Evaluation betroffenen Gruppen
- ➔ Gefahr, dass ***Mitarbeiter die von ihnen eingeführten Maßnahmen positiver beurteilen***, als es den Fakten entspricht [Abhilfe: keine Selbstevaluation, sondern Fremdevaluation]

ABER: das alles kann es auch in der Grundlagenforschung geben...
⇨ DAHER: **Berücksichtigung der ethischen Normen der Wissenschaft**

weitere Unterschiede der Evaluationsforschung zur Grundlagenforschung:
➔ eher ***Zurückhaltung gegenüber Publikationen***
➔ ***Replikationsstudien selten***

2) ***Entscheidungszwänge und wissenschaftliche Vorsicht***

Evaluationsforschung:

Politiker, Führungskräfte, usw. = Entscheidungsträger -> von ihren Entscheidungen hängt oft sehr viel ab. Viele Entscheidungen sind aber risikobehaftet, da Informationsbasis lückenhaft und Folgen unabsehbar. DAHER: Teil der Verantwortung wird an einen **fachkompetenten Evaluator** abgegeben.

- ***retrospektive Evaluation***: es wird untersucht, ob in einer Maßnahme die angestrebten Ziele erreicht wurden.

- ***prospektive Evaluation***: es wird untersucht, ob eine geplante Maßnahme den (finanziellen) Aufwand rechtfertigen wird

Evaluator kann seiner **Ratgeberpflicht** nur nachkommen, wenn sich für oder gegen den Erfolg einer Maßnahme ausspricht. Kann dies umso leichter, je höher die interne Validität seiner Ergebnisse ist.

Grundlagenforschung:

➔ **keine Entscheidungszwänge aufgrund EINER EINZIGEN Studie** (z.B. darüber ob eine Theorie als Ganzes verworfen werden muss oder nicht)

➔ muss Ergebnisse vorsichtig und selbstkritisch interpretieren und so Wege aufzeigen, wie Geltungsbereich einer Theorie ausgeweitet oder gefestigt werden könnte.

3) ***technologische und wissenschaftliche Theorien***

wissenschaftliche Theorien:

➔ beinhalten **in sich schlüssiges Annahmengefüge** über Ursachen und Wirkungen eines Sachverhaltes

➔ werden **in empirischer Grundlagenforschung entwickelt** und **auf ihre Gültigkeit geprüft**

➔ **gute wissenschaftliche Theorie**: ist gekennzeichnet durch
- präzise Terminologie
- logisch konsistenter Informationsgehalt (Widerspruchslosigkeit)
- möglichst breite inhaltliche Tragweite
- begrenzte Anzahl von Annahmen (Sparsamkeit)

➔ Überprüfung wissenschaftlicher Theorien = **Aufgabe der Grundlagenforschung**

technologische Theorien:

➔ **knüpfen an den Output einer wissenschaftlichen Theorie** an

➔ bilden **Basis für Gewinnung von Regeln**, mit denen **wissenschaftliche Erkenntnisse praktisch nutzbar** gemacht werden können

➔ **primäres Erkenntnisinteresse = Formen des Handelns**, mit denen etwas hervorgebracht, vermieden, verändert oder verbessert werden kann

➔ **gute technologische Theorie**: soll
- wissenschaftliche Erkenntnisse in effiziente, routinisierbare Handlungsanweisungen umsetzen
- Wege der praktischen Nutzbarmachung von wissenschaftliche Erkenntnissen aufzeigen

➔ Überprüfung technologischer Theorien = **Aufgabe der Evaluationsforschung**

MERKE:

- ***wissenschaftliche Theorien:***
  -> dienen der Beschreibung, Erklärung und Vorhersage von Sachverhalten
  -> werden in Grundlagenforschung entwickelt

- ***technologische Theorien:***
  -> geben konkrete Handlungsanweisungen zur praktischen Umsetzung wissenschaftlicher Theorien
  -> fallen in Aufgabenbereich der angewandten Forschung bzw. Evaluationsforschung

BEISPIEL: **Politikverdrossenheit** ist auf die überwiegend negative Berichterstattung über politische Ereignisse in den Medien zurückzuführen („Videomalaise"; Robinson 1976)

➔ **Grundlagenforschung untersucht**: was ist „Politikverdrossenheit" überhaupt? wie sind mediale Stimuli beschaffen, die aversive Reaktionen auslösen? usw. Ergebnis: die Art, wie im Fernsehen über Politik berichtet wird, bestimmt das Ausmaß der Politikverdrossenheit

➔ **technologische Theorie**: politische Nachrichten müssen unterhaltsamer präsentiert werden („Infotainment"); Politiker müssen mehr von ihrer menschlichen Seite gezeigt werden; ausgewogenes Verhältnis zwischen guten und schlechten Nachrichten

⇨ ***präsentiert Handlungsoptionen***, ob sie befolgt werden oder nicht, ist abhängig vom Einzelnen und seinen persönlichen Überzeugungen und Zielsetzungen

Ist es wünschenswert, Politikverdrossenheit zu reduzieren, oder sollte man sie nicht eher langfristig steigern, damit eine Wendung forciert wird? -> Das ist eine ***Frage der Wertung*** (nicht wissenschaftlich, sondern ethisch zu begründen!)

4) ***Evaluationsforschung und Interventionsforschung***:

Evaluationsforschung: befasst sich mit Überprüfung der Wirkungen und Folgen einer Maßnahme oder Intervention. Voraussetzung: Es muss mindestens ein Entwurf der zu evaluierenden Maßnahme vorliegen.

Beispiel: Es wurde in Grundlagenforschung erkannt, dass bestimmte Verhaltensstörung auf traumatisches Kindheitserlebnis zurückgeht. Experten entwickeln nun aufgrund ihres Wissens in entsprechender technologischer Theorie eine Therapie dagegen. Alle Aktivitäten, die sie dazu heranziehen = Interventionsforschung.

MERKE:

* ***Interventionsforschung*** befasst sich auf der Basis technologischer Theorien mit der Entwicklung von Maßnahmen.

* ***Evaluationsforschung*** befasst sich mit der Bewertung dieser Maßnahmen.

ABER: ***In der Praxis***: Grenze = fließend; meist ist es so, dass **Interventions- und Evaluationsaufgaben in EINER Hand liegen** (z.B. Interventionsforscher hat genügend Methodenkenntnisse, um seine Maßnahme selber zu evaluieren)

Idealfall: **Parallelentwicklung von Maßnahme und ihrem Evaluationsplan**, damit man schon im Vorfeld erkennen kann, welche konkreten Besonderheiten der Maßnahme untauglich bzw. nur schwer evaluierbar sind.

WOTTAWA: WISSENSCHAFT, MORAL und die GRENZEN DER PLANBARKEIT

Definitionsversuche von „Evaluation“:

Evaluation = sehr vielfältiger Begriff -> daher: nur EINE umfassende Definition ist nicht möglich.

Definition von Suchman (1967):

Evaluation (Bewertung):

> Prozess der Beurteilung des Wertes eines Produkts, Prozesses oder eines Programms, was nicht notwendigerweise systematische Verfahren oder datengestützte Beweise zur Untermauerung einer Beurteilung erfasst.

Evaluation research (Evaluationsforschung):

> explizite Verwendung wissenschaftlicher Forschungsmethoden und Forschungstechniken für den Zweck der Durchführung einer Bewertung. Evaluationsforschung betont die Möglichkeit des Beweises anstelle der reinen Behauptung bezüglich des Wertes und Nutzens einer bestimmten sozialen Aktivität.

[Diese Definition konnte sich aber **noch nicht allgemein durchsetzen**. In der Praxis wird **Evaluation oft synonym verwendet mit Begriffen wie** Erfolgskontrolle, Effizienzforschung, Begleitforschung, Bewertungsforschung, Qualitätskontrolle, usw.]

Allgemeine Kennzeichen wissenschaftlicher Evalution:

1) Evaluation dient als **Planungs- und Entscheidungshilfe**, hat somit etwas zu tun mit der Bewertung von Handlungsalternativen.

2) Evaluation = **ziel- und zweckorientiert** (d.h. primäres Ziel = Überprüfung von praktischen Maßnahmen, um sie zu verbessern oder über sie zu entscheiden)

3) Evaluationsmaßnahmen sollen **dem aktuellen Stand wissenschaftlicher Techniken und Forschungsmethoden angepasst** sein.

Ethisch-moralische Verantwortung:

Evaluatoren verändern durch ihre Arbeit beratend die Lebensumstände anderer Menschen, erheben dabei den Anspruch aufgrund ihrer „Wissenschaftlichkeit“ über die entsprechenden Kompetenzen zu verfügen, die den „Laien“ fehlen...; daher: große Verantwortung!!!

Problematisch dabei könnte sein:

➔ ***Werthaltung des Evaluators*** [könnte zu Verzerrungen der Ergebnisse führen (z.B. Projekt zur Feststellung der Folgen von Sterbehilfe)]

➔ ***persönliche Vorlieben des Evaluators*** [z.B. seine Überzeugung von der Richtigkeit einer bestimmten Therapie]

DAHER:

- besondere Sorgfalt bei Projektplanung
- Explikation aller durchgeführten Festlegungen
- Offenlegung der eigenen Position des Evaluators

Zusammenfassungen:

DEFINITIONSVERSUCHE VON EVALUATION

jegliche Art der Festsetzung des Wertes eines Sache

systematische Anwendung sozialwissenschaftlicher Methoden

verwandte Begriffe: Erfolgskontrolle, Qualitätskontrolle Controlling

Allgemeine Kennzeichen wissenschaftlicher Evaluation:

➔ dient der **Planungs- und Entscheidungshilfe**
➔ ist **ziel- und zweckorientiert**
➔ soll dem **aktuellen Stand wissenschaftlicher Forschung angepasst** sein

Optimierung von Handlungsmodellen durch Evaluation
unverzichtbare Form wissenschaftsgestützten Lernens

Ziel der Evaluation:

⇨ **Übelminimierung** statt unrealistischer Ideallösung

ZIELSETZUNGEN BEI EVALUATIONSVORHABEN:

Grundlagen der Akzeptanz von Evaluationen:
Psychologische Voraussetzungen für die Entwicklung von Evaluation

1) ***Akzeptanz der Veränderbarkeit*** relevanter gesellschaftlicher Strukturen und Gestaltung gesellschaftlicher Verhältnisse unter rationalen Optimierungsaspekten
2) ***Bereitschaft***, sich ***zu Zwecken der Verbesserung des bestehenden Verhaltens*** („Hoffen auf Erfolg“) ***dem Risiko des Scheiterns*** („Furcht vor Misserfolg“) auszusetzen.
3) ***Bereitschaft der Entscheidungsträger zur Akzeptanz von Fakten***

Mögliche Nutzenerwartungen des Auftraggebers:

- Bewertung ohne detaillierte Zielsetzung
- Verantwortungsdelegation
- Durchsetzungs- / Entscheidungshilfe
- Optimierungsgrundlage

Einteilung der Evaluationsformen nach:

- Rahmenbedingungen
- grundsätzlicher Zielorientierung
- Zeitperspektive
- Nutzenüberlegungen
- Bearbeitungsformen
- Erfahrungsaufbereitung

WISSENSCHAFT ALS MÖGLICHE GRUNDLAGE VON EVALUATION

Grundvoraussetzungen:

1) Bereitschaft, an praxisorientierten Evaluationen teilzuhaben
2) Wissenschaftlicher Beitrag muss mehr Nutzen bringen als Kosten
3) Wissenschaftliche Evaluation sollte den nicht-wissenschaftlichen Alternativen überlegen sein

Evaluation als besondere Form empirisch-wissenschaftlichen Arbeitens

Evaluation = entscheidungsorientierte Forschung	**Wissenschaftliche Forschung = schlussfolgeorientierte Forschung**

Konsequenz: Rollenverständnis des Evaluators steht teilweise im Widerspruch zu den Ansprüchen der wissenschaftlichen Forschung; ABER: Wissenschaft als Basis und Hilfe für die Evaluationsarbeit!