



Methoden der Evaluationsforschung

MSc Klinische Psychologie und Psychotherapie

SoSe 2024

Prof. Dr. Dirk Ostwald

(1) Evaluationsforschung

Definition und Funktion

Standards

Validität

Selbstkontrollfragen

Definition und Funktion

Standards

Validität

Selbstkontrollfragen

Definition und Funktion

Approbationsordnung für Psychotherapeutinnen und Psychotherapeuten (2020) Anlage 2 (zu § 8 Nummer 2)

Inhalte, die im Masterstudiengang im Rahmen der hochschulischen Lehre zu vermitteln und bei dem Antrag auf Zulassung zur psychotherapeutischen Prüfung nachzuweisen sind.

Vertiefte Forschungsmethodik

Die studierenden Personen

(a) wenden komplexe und multivariate Erhebungs- und Auswertungsmethoden zur Evaluierung und Qualitätssicherung von Interventionen an, (b) nutzen und beurteilen einschlägige Forschungsstudien und deren Ergebnisse für die Psychotherapie (c) Planen selbständig Studien zur Neu- oder Weiterentwicklung der Psychotherapieforschung oder der Forschung in angrenzenden Bereichen, führen solche Studien durch, werten sie aus und fassen sie zusammen, (d) bewerten wissenschaftliche Befunde sowie Neu- oder Weiterentwicklungen in der Psychotherapie inhaltlich und methodisch in Bezug auf deren Forschungsansatz und deren Aussagekraft, so dass sie daraus fundierte Handlungsentscheidungen für die psychotherapeutische Diagnostik, für psychotherapeutische Interventionen und für die Beratung ableiten können.

Zur Vermittlung der Inhalte der vertieften Forschungsmethodik sind bei der Planung der hochschulischen Lehre mindestens 6 ECTS-Punkte vorzusehen und die folgenden Wissensbereiche abzudecken:

(a) multivariate Verfahren und Messtheorie, (b) Evaluierung wissenschaftlicher Befunde und deren Integration in die eigene psychotherapeutische Tätigkeit.

Evaluation

“action of appraising or valuing,” from French *évaluation*, noun of action from *évaluer* “to find the value of,” from *é-* “out” (see *ex-*) + *valuer*, from Latin *valere* “be strong, be well; be of value, be worth” (from Proto-Indo-European *wal-* “to be strong”) by 1755. Meaning “job performance review” attested by 1947.

Online Etymology Lexicon

- Evaluationsforschung = Evaluation mit wissenschaftlichen Methoden
- Evaluationsforschung \neq Forschung zur Evaluation
- Evaluation von psychologischen Interventionen/Psychotherapien

Hauptquelle: “Grundlagen der Evaluationsforschung” Holling et al. (2009), Kapitel 1

Einordnung

- Wissenschaftliche Begleituntersuchungen von sozialpolitischen Reformen in den USA der 1960er
- Heute eigenständige Disziplin mit Anwendung in vielen gesellschaftlichen Bereichen
- ⇒ Evaluationen in Bildung, Gesundheit, Verkehr, Umwelt, Städtebau, Justiz, ...
- [Deutsche Gesellschaft für Evaluation](#)
- Synonyme "Controlling", "Qualitätskontrolle", "Erfolgskontrolle", "Effizienzforschung", ...

Lehrbuchdefinitionen

Suchman (1967)

- Evaluation als “Prozess der Bewertung eines Sachverhalts”, Evaluationsforschung als “die explizite Verwendung wissenschaftlicher Forschungsmethoden zur Bewertung eines Sachverhalts.”

Rossi, Lipsey, and Freeman (2004)

- Evaluationsforschung als “systematische Anwendung empirischer Forschungsmethoden zur Bewertung des Konzepts, des Untersuchungsplans, der Implementierung und der Wirksamkeit sozialer Interventionsprogramme”.

Bortz and Döring (2006)

- Evaluationsforschung als “alle forschenden Aktivitäten, bei denen es um die Bewertung des Erfolges von gezielt eingesetzten Maßnahmen oder um Auswirkungen von Wandel in Natur, Kultur, Technik und Gesellschaft geht”.

Definition und Funktion

Lehrbuchdefinitionen

Wottawa and Thierau (1998)

- “(Evaluatorische Tätigkeiten) haben etwas mit Bewerten zu tun. Evaluation dient als Planungs- und Entscheidungshilfe und hat etwas mit der Bewertung von Handlungsalternativen zu tun. Evaluation ist ziel- und zweckorientiert. Sie hat primär das Ziel, praktische Maßnahmen zu überprüfen, zu verbessern oder über sie zu entscheiden”

Hager (2000)

- Evaluationsforschung als “die wissenschaftlich fundierte, empirische und hypothesenorientierte Forschung unter systematischer Anwendung sozialwissenschaftlicher Forschungsmethoden. Die Ergebnisse diese Forschung bilden die wesentliche, wenn auch nicht die einzige Grundlage einer wissenschaftlichen Evaluation oder Bewertung der Konzeption, Ausgestaltung, Umsetzung und des Nutzens sozialer und psychologischer Interventionsprogramme.”

Attkinson and Broskowski (1978)

- Programmevaluation als “ein Prozess der Durchführung vernunftgeleiteter Beurteilung eines Programms [umfangreiche Intervention] hinsichtlich Aufwand, Effektivität und Angemessenheit auf der Grundlage systematischer Datenerhebung und Datenanalyse.”

Evaluation von Psychologischen Interventionen

Psychologische Intervention

Jede Art von außen gesteuerter, zielorientierter und systematischer Beeinflussung von Personen mithilfe von Lernerfahrungen, insbesondere im Gespräch und angeleitetem Selbststudium.

Vgl. Hager (2000)

Entwicklung und Evaluation psychologischer Interventionen

- Klinische Psychologie und Psychotherapie
- Arbeits- und Organisationspsychologie
- Pädagogische Psychologie

⇒ Im Bereich KliPP: Intervention = Psychotherapie

Funktionen der Evaluation von Psychologischen Interventionen

Erkenntnisfunktion

- Wirkt eine Intervention und wenn ja wie?

Optimierungsfunktion

- Welche Stärken, Schwächen, oder Nebenwirkungen hat eine Intervention?

Kontrollfunktion

- Wird die Intervention korrekt umgesetzt und wie ist ihre Kosten-Nutzen-Bilanz?

Entscheidungsfunktion

- Soll eine Intervention gefördert, weiterentwickelt, genutzt werden oder nicht?

Legitimationsfunktion

- Jede Evaluation dient auch der Legitimation der Intervention nach außen

Vgl. Bortz and Döring (2006)

Definition und Funktion

Standards

Validität

Selbstkontrollfragen

Allgemeine Standards der Evaluationsforschung

- Joint Committee on Standards for Educational Evaluation (1981)
- Neufassung als Program Evaluation Standards (1994)

⇒ Leitlinien für die Beurteilung, Planung, Durchführung und Vergabe von Evaluationen.

Standards der Deutschen Gesellschaft für Evaluation (DeGEval)

- Nützlichkeit
- Durchführbarkeit
- Genauigkeit
- Fairness

⇔ Standards der guten wissenschaftlichen Praxis

Nützlichkeit

(N1) Identifizierung der Beteiligten und Betroffenen

- Die am Evaluationsgegenstand oder an der Evaluation Beteiligten sowie die von Evaluationsgegenstand oder Evaluation Betroffenen sollen vorab identifiziert werden, damit deren Interessen und Informationsbedürfnisse geklärt und so weit wie möglich bei der Anlage der Evaluation berücksichtigt werden können.

(N2) Klärung der Evaluationszwecke

- Es soll deutlich bestimmt sein, welche Zwecke mit der Evaluation verfolgt werden, so dass die Beteiligten und Betroffenen Position dazu beziehen und die Evaluierenden einen klaren Arbeitsauftrag verfolgen können.

(N3) Kompetenz und Glaubwürdigkeit des Evaluators/der Evaluatorin

- Wer Evaluationen durchführt, soll fachlich und methodisch kompetent sein, damit für die Evaluation und ihre Ergebnisse ein Höchstmaß an Glaubwürdigkeit und Akzeptanz erreicht wird.

(N4) Auswahl und Umfang der Informationen

- Auswahl und Umfang der erfassten Informationen sollen die adäquate Beantwortung der zu untersuchenden Fragestellungen zum Evaluationsgegenstand ermöglichen und die Informationsbedürfnisse der Auftraggebenden und weiterer Beteiligter und Betroffener berücksichtigen.

Nützlichkeit

(N5) Transparenz von Werthaltungen

- Werthaltungen der Beteiligten und Betroffenen, die sich in deren Perspektiven und Annahmen manifestieren und einen Einfluss haben auf die Evaluation und Interpretation ihrer Ergebnisse, sollten transparent dokumentiert werden, um Evaluationsergebnisse besser einordnen zu können.

(N6) Vollständigkeit und Klarheit der Berichterstattung

- Evaluationsberichte sollen alle wesentlichen Informationen zur Verfügung stellen und für ihre Adressatinnen und Adressaten verständlich und nachvollziehbar sein.

(N7) Rechtzeitigkeit der Evaluation

- Evaluationsvorhaben sollen so rechtzeitig begonnen und abgeschlossen werden, dass ihre Ergebnisse in anstehende Entscheidungs-, Verbesserungs- oder sonstige Nutzungsprozesse einfließen können.

(N8) Nutzung und Nutzen der Evaluation

- Planung, Durchführung und Berichterstattung einer Evaluation sollen die Beteiligten und Betroffenen dazu ermuntern, die Evaluation mitzutragen und ihre Ergebnisse zu nutzen.

Durchführbarkeit

(D1) Angemessene Verfahren

- Evaluationsverfahren, einschließlich der Verfahren zur Beschaffung notwendiger Informationen, sollen so gewählt werden, dass einerseits die Evaluation professionell und den Erfordernissen entsprechend umgesetzt wird und andererseits der Aufwand für die Beteiligten und Betroffenen in einem adäquaten Verhältnis zum intendierten Nutzen der Evaluation gehalten wird.

(D2) Diplomatisches Vorgehen

- Evaluationen sollen so geplant und durchgeführt werden, dass eine möglichst hohe Akzeptanz der verschiedenen Beteiligten und Betroffenen in Bezug auf Vorgehen und Ergebnisse der Evaluation erreicht werden kann.

(D3) Effizienz von Evaluation

- Der Aufwand für Evaluation soll in einem angemessenen Verhältnis zum Nutzen der Evaluation stehen.

Fairness

(F1) Formale Vereinbarungen

- Die Rechte und Pflichten der an einer Evaluation beteiligten Parteien (was, wie, von wem, wann getan werden soll und darf) sollen schriftlich festgehalten werden.

(F2) Schutz individueller Rechte

- Evaluationen sollen so geplant und durchgeführt werden, dass Rechte, Sicherheit und Würde der in eine Evaluation einbezogenen Personen geschützt sind.

(F3) Umfassende und faire Prüfung

- Evaluationen sollen die Stärken und die Schwächen des Evaluationsgegenstandes möglichst fair und umfassend prüfen und darstellen.

(F4) Unparteiische Durchführung und Berichterstattung

- Die Evaluation soll unterschiedliche Sichtweisen von Beteiligten und Betroffenen auf Gegenstand und Ergebnisse der Evaluation beachten. Der gesamte Evaluationsprozess sowie die Evaluationsberichte sollen die unparteiische Position der Evaluierenden erkennen lassen.

(F5) Offenlegung von Ergebnissen und Berichten

- Evaluationsergebnisse und -berichte sollen allen Beteiligten und Betroffenen soweit wie möglich zugänglich gemacht werden.

Genauigkeit

(G1) Beschreibung des Evaluationsgegenstandes

- Sowohl das Konzept des Evaluationsgegenstands als auch seine Umsetzung sollen genau und umfassend beschrieben und dokumentiert werden.

(G2) Kontextanalyse

- Der Kontext des Evaluationsgegenstandes soll ausreichend umfassend und detailliert analysiert sowie bei der Interpretation von Ergebnissen berücksichtigt werden.

(G3) Beschreibung von Zwecken und Vorgehen

- Zwecke, Fragestellungen und Vorgehen der Evaluation, einschließlich der angewandten Methoden, sollen so genau dokumentiert und beschrieben werden, dass sie nachvollzogen und beurteilt werden können.

(G4) Angabe von Informationsquellen

- Die im Rahmen einer Evaluation genutzten Informationsquellen sollen hinreichend genau dokumentiert werden, damit die Verlässlichkeit und Angemessenheit der Informationen eingeschätzt werden können.

(G5) Valide und reliable Informationen

- Erhebungsverfahren und Datenquellen sollen so gewählt werden, dass die Zuverlässigkeit der gewonnenen Daten und ihre Gültigkeit bezogen auf die Beantwortung der Evaluationsfragestellungen nach fachlichen Maßstäben sichergestellt sind. **Die fachlichen Maßstäbe sollen sich an den Gütekriterien der empirischen Forschung orientieren.**

Genauigkeit

(G6) Systematische Fehlerprüfung

- Die in einer Evaluation gesammelten, aufbereiteten, analysierten und präsentierten Informationen sollen systematisch auf Fehler geprüft werden.

(G7) Angemessene Analyse qualitativer und quantitativer Informationen

- **Qualitative und quantitative Informationen einer Evaluation sollen nach fachlichen Maßstäben angemessen und systematisch analysiert werden, damit die Fragestellungen der Evaluation beantwortet werden können.**

(G8) Begründete Bewertungen und Schlussfolgerungen

- Die in einer Evaluation getroffenen wertenden Aussagen sollen auf expliziten Kriterien und Zielwerten basieren. **Schlussfolgerungen sollen ausdrücklich und auf Grundlage der erhobenen und analysierten Daten begründet werden, damit sie nachvollzogen und beurteilt werden können.**

(G9) Meta-Evaluation

- Meta-Evaluationen evaluieren Evaluationen. **Um dies zu ermöglichen, sollen Evaluationen in geeigneter Form dokumentiert, archiviert und soweit wie möglich zugänglich gemacht werden.**

Definition und Funktion

Standards

Validität

Selbstkontrollfragen

Four Types of Validity in the Donald Campbell tradition

We use the term *validity* to refer to the approximate truth of an inference. When we say something is valid, we make a judgement about the extent to which relevant evidence supports that inference as being true or correct.

Shadish, Cook, and Campbell (2001), vgl. Holling et al. (2009)

Statistical conclusion validity

- The validity of inferences about the correlation (covariation) between treatment and outcome.

Internal validity

- The validity of inferences about whether observed covariation between A (the presumed treatment) and B (the presumed outcome) reflects a causal relationship from A to B as those variables were manipulated or measured.

Construct validity

- The validity of inferences about the higher order constructs that represent sampling particulars.

External validity

- The validity of inferences about whether the cause-effect relationship holds over variation in persons, settings, treatment variables, and measurement variables.

Threats to statistical conclusion validity

Reasons why inferences about covariation between two variables may be incorrect

1. Low statistical power
2. Violated assumptions of statistical tests
3. Fishing and the error rate problem
4. Unreliability of measures
5. Restriction of range
6. Unreliability of treatment implementation
7. Extraneous variance in the experimental setting and experimental units
8. Inaccurate effect size estimation
9. ... and many more

Shadish, Cook, and Campbell (2001), vgl. Holling et al. (2009)

Threats to internal validity

Reasons why inferences that the relationship between two variables is causal may be incorrect.

1. Ambiguous temporal precedence
2. Systematic selection of respondent characteristics
3. History of events occurring concurrently with the treatment
4. Naturally occurring maturation changes over time
5. Regression to the mean after display of extreme values
6. Attrition
7. Repeated testing effects
8. Changes of instrumentation
9. Additive and interactive effects of threats to internal validity
10. ... and many more

Shadish, Cook, and Campbell (2001), vgl. Holling et al. (2009)

Threats to construct validity

Reasons why inferences about constructs that characterize study operations may be incorrect.

1. Inadequate explication of constructs
2. Construct confounding by additional constructs
3. Mono-Operation- and Mono-Method-Bias
4. Respondent reactivity and motivation/demotivation
5. Experimenter expectancies
6. Novelty and disruption effects
7. ... and many more

Shadish, Cook, and Campbell (2001), vgl. Holling et al. (2009)

Threats to external validity

Reasons why inferences about the generalizability of study's result may be incorrect.

1. Dependency on the experimental unit specificities
2. Dependency on the treatment specificities
3. Dependency on the outcome measure
4. Dependency on the setting specificities
5. ... and many more

Shadish, Cook, and Campbell (2001), vgl. Holling et al. (2009)

Definition und Funktion

Standards

Validität

Selbstkontrollfragen

1. Erläutern Sie den Begriff der Evaluationsforschung.
2. Nennen und erläutern Sie fünf Funktionen der Evaluation psychologischer Interventionen.
3. Nennen Sie die vier übergeordneten Standards der Deutschen Gesellschaft für Evaluation und erläutern Sie diese anhand von Beispielen.
4. Nennen Sie vier Formen der Validität von Evaluationen nach Shadish, Cook, and Campbell (2001) und erläutern sie mögliche Gefährdungen dieser Validitätsformen anhand von Beispielen.

Referenzen

- Bortz, Jürgen, and Nicola Döring. 2006. "Besonderheiten der Evaluationsforschung." In *Forschungsmethoden und Evaluation*, 95–136. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-33306-7_3.
- Hager, Willi, ed. 2000. *Evaluation psychologischer Interventionsmaßnahmen: Standards und Kriterien: ein Handbuch*. 1. Aufl. Aus dem Programm Huber: Psychologie-Lehrbuch. Bern Göttingen: Huber.
- Holling, Heinz, Wolfgang Bilsky, Ingwer Borg, David Cairns, Britta Colver, Kai S. Cortina, Nicola Döring, et al. 2009. *Grundlagen und statistische Methoden der Evaluationsforschung*. 1st ed. Göttingen: Hogrefe Verlag.
- Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman. 2004. *Evaluation: A Systematic Approach*. 7th ed. Thousand Oaks, CA: Sage.
- Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Suchman, Edward Allen. 1967. *Evaluative Research: Principles and Practice in Public Service & Social Action Programs*. New York: Russel Sage Foundation.
- Wottawa, Heinrich, and Heike Thierau. 1998. *Lehrbuch Evaluation*. Zweite, vollständig überarbeitete Auflage. Aus dem Programm Huber: Psychologie Lehrbuch. Bern Göttingen Toronto Seattle: Verlag Hans Huber.