

Zusammenfassend sollte die vorausgehende, kurze Diskussion zeigen:

1. Die Beantwortung der Frage, ob eine geplant eingesetzte Intervention nachweisbar zu dem erhofften Ergebnis beiträgt, erfordert für die **Beweisführung** einen beträchtlichen Aufwand.
2. Die Hauptschwierigkeit ergibt sich nicht durch den Stichprobenumfang von Studien oder die Anzahl der inhaltlich einschlägigen, durchgeführten Untersuchungen, sondern durch die erfolgreiche Kontrolle und weitmögliche Ausschaltung weiterer „störender“ Einflüsse und die dafür erforderliche **methodische Qualität** der Untersuchungen, auf denen die Verlässlichkeit der Erklärungsmodelle beruht.
3. Heute stehen umfangreiche Technologien zur Sicherung wissenschaftlicher Beweiskraft zur Verfügung. Fachkompetente Beratung in Anspruch zu nehmen, kann hier sehr hilfreich sein, um Fehler und Kosten zu vermeiden.

Anmerkungen.

A1. Nachdrücklich aufgegriffen wurde die Themenstellung des Faltblattes in verschiedenen Fachpublikationen und Resolutionen der großen amerikanischen Psychologenorganisation APA, auf die teilweise auch in der vorliegenden Ausarbeitung zurückgegriffen wurde (in "American Psychologist" 58, 2003, u.a. die Beiträge von C. E. Thoresen).

A2. Ohne Zweifel wird jeder sehr intensiv mit den wissenschaftlichen Grundlagen seines Faches vertraute Leser Ergänzungen und vielleicht auch kritische Einwände gegenüber den hier ausgewählten Qualitätskriterien vortragen wollen; ob die vorgelegte Ausarbeitung einem sehr allgemeinen wissenschaftlichen Qualitätskriterium, der Objektivität i.S. einer Übereinstimmung unterschiedlicher fachkundiger Beurteiler entspricht, sei als Problem hier nur erwähnt, auch diese Frage kann nach wissenschaftlich kontrollierten Verfahrensregeln untersucht werden.

A3. Auch wissenschaftliche Studien werden "produziert", sie entstehen im Kontext von Forschungseinrichtungen. Auf die Wichtigkeit der Einflüsse von Forschungseinrichtungen als Organisationen auf die "Prozessqualität" von Einzelstudien kann hier lediglich hingewiesen werden.

A4. Qualitätssicherung bei Erfassung von Schlüsselvariablen beruht auch immer auf vorausgehenden, sorgfältigen Analysen zum Untersuchungsgegenstand; qualitative und quantitative Forschungsansätze ergänzen einander und stehen nicht – wie in der Fachliteratur teilweise behauptet – in einem grundsätzlichen Spannungsverhältnis zueinander.

A5. Die bei Humanuntersuchungen besonders häufig erhobenen Befragungsdaten - sie sind meist einfacher zu gewinnen als Verhaltensdaten oder etwa physiologische Reaktionsdaten - sind besonders vielfältigen Störeinflüssen ausgesetzt. Zur deren Kontrolle und (weitmöglichen) Ausschaltung stehen umfangreiche Skalenentwicklungen (etwa zur Erfassung "sozial erwünschter" Antworttendenzen) oder anspruchsvollere (multivariate) statistische Prüfverfahren zur Verfügung.

Das Verhalten der Versuchsteilnehmer wird vor allem auch unbewusst verändert; ein berichtigtes Beispiel dafür waren die scheinbar ungewöhnlichen Rechenfähigkeiten eines Pferdes („Kluger Hans“- Studie, von Oskar Pfungst 1907 veröffentlicht). Zur Ausschaltung in experimentellen Studien bietet das Verfahren des Blindversuchs - hier kennt der Proband die Ziele der Untersuchung nicht - keine Gewähr. Eine wirkungsvollere Kontrolle bietet der Doppel-Blindversuch, bei dem auch der Untersuchungsleiter nicht weiß, unter welcher Verfahrensvariante ein Versuchspartner teilnimmt. Erkauft wird die experimentelle Vorgehensweise allerdings oft durch eine alltagsferne Künstlichkeit; auch dagegen wurden aber Verfahrenstechniken entwickelt (z.B. "Quasi-biotische" Experimentalanordnungen in der Werbepsychologie).

A6. Für biometrische Signifikanzprüfungen wurden völlig unterschiedliche Verfahren entwickelt, die in umfangreichen Lehrbüchern dargestellt sind und in der Ausbildung oft zu den unbeliebten Studienabschnitten zählen. Inhaltlich wird vielfältig unterschieden, vor allem zwischen korrelativen vs. varianzanalytischen Prüfverfahren (letztere zur Bewertung von Gruppenunterschieden bzgl. Mittelwert oder Zentraltendenz), parametrischen vs. verteilungsfreien Verfahren (letztere setzen keine Normalverteilung in der Grundgesamtheit voraus), Voraussetzungen des Skalenniveaus, auf dem die Daten erfasst sind (Nominal-, Ordinal-, Intervallskalen), etc.

A7. Der Titel "Meta-Analyse" von Sammelarbeiten ist präzise definiert und ausschließlich bezogen auf quantitativ- integrierende Verfahren.

A8. Jede Fachdisziplin mit etabliertem Fachniveau verfügt über eine oder mehrere elektronische Publikationsverzeichnisse als Datenbanken, aus denen nach Suchbegriffen alle international verfügbaren, qualitativ akzeptierten Studien abgerufen werden können, etwa ZPID oder DIMDI.

Überreicht durch:



Herausgeber: Vorstand der BDP-Sektion Gesundheits-, Umwelt- und Schrift-psychologie (GUS)
Sektionsvorsitzender: PD Dr. Rudolf Günther (ViSdP)
Ganghofer Str. 28, 72764 Reutlingen

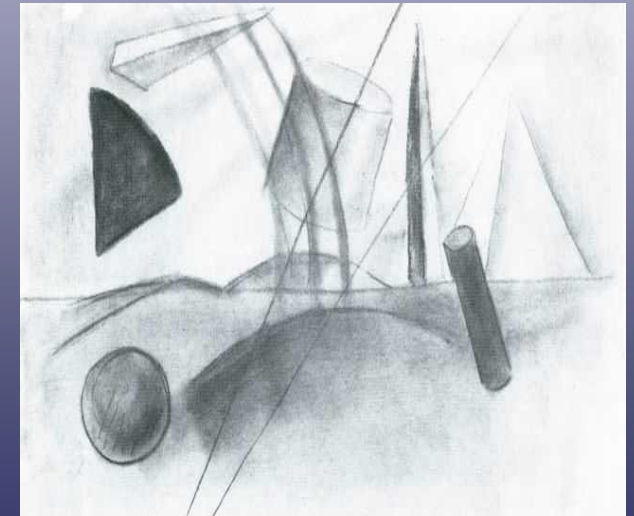
www.bdp-gus.de/gus/

Abbildung: Macke, August: Formen in den Raum komponiert (1914). @ DigiBib 4.01.231



**BERUFSVERBAND
DEUTSCHER
PSYCHOLOGINNEN und
PSYCHOLOGEN e.V.**

Qualitätskriterien für wissenschaftlich aussagekräftige empirische Untersuchungen



Zweck des Faltblatts

Psychologinnen und Psychologen stehen sehr oft in fachlichem Kontakt zu jenen Berufsgruppen, in denen Fragen der empirischen Forschung nicht zu den selbstverständlichen Kernkompetenzen ihrer Berufsausbildung und –praxis zählen. Wegen der inhaltlichen Offenheit und Interessenbreite der psychologischen Fachrichtungen existieren diese Kontakte zu besonders vielen anderen Fachdisziplinen. Insbesondere für diese will der Berufsverband deutscher Psychologinnen und Psychologen (BDP) eine Hilfestellung geben.

Allerdings legen die folgenden Ausführungen den Schwerpunkt auf Fragen der Produktion und Bewertung von Facherkenntnissen, während weitere, umfangreiche Aufgabenfelder der Qualitätssicherung (Anforderungen für die Anwendung vorhandener Verfahren und Erkenntnisse in der Berufspraxis, Qualitätssicherung in der Ausbildung, in der kontinuierlichen Fortbildung und für die Weiterentwicklung von qualitätsgesicherten Praxiskennnissen) außer Betracht bleiben. (Anm.1).

Ausgangsüberlegung und einige Hauptkriterien

Als Ausgangspunkt der folgenden Diskussion versuchen wir, gedanklich die Fragestellung zu verfolgen:

"Bietet eine Maßnahme M Aussicht, ein Ziel Z zu erreichen?"
Beispiele: "Regelmäßige Wellness-Urlaube fördern körperliches und psychisches Wohlbefinden"; "Religiöse Überzeugungen können bei erkrankten Menschen die Heilungschancen verbessern"; "Bruststillen fördert die Intelligenzentwicklung des Babys"

Jede wissenschaftliche Antwort (Anm.2) beginnt dazu mit den Fragen: Was weiß man? Welche Erkenntnisse liegen vor? Unter dieser Zielsetzung werden wissenschaftliche Untersuchungen durchgeführt und veröffentlicht. Um eine Antwort auf die genannte Frage zu gewinnen, werden verfügbare Erkenntnisse zusammengetragen, wobei zwei elementare Regeln beachtet werden:

- (a) Es sollen alle für die Frage relevanten einschlägigen Erkenntnisse - also die relevanten durchgeführten Untersuchungen - berücksichtigt werden, die aber
- (b) ausreichende Qualitätsvoraussetzungen aufweisen. **Nicht berücksichtigt** werden also Studien oder auch Praxisberichte, die entscheidende wissenschaftliche **Qualitätsmängel** aufweisen.

In das Blickfeld gerät damit die Frage, welche Kriterien sind dabei besonders wesentlich für die Qualitätsbewertung? (Anm.3)

1. Kontrolle elementarer Störgrößen. Ist in der Studie angemessen kontrolliert, ob der Hauptbefund durch elementare demographische und soziale Störgrößen erklärbar ist?

Beispiel: Tätigkeiten im Garten: Sie setzen die Verfügbarkeit über eine entsprechende Gartenfläche voraus, was möglicherweise einkommensabhängig ist. Menschen mit höherem Einkommen beschreiben insgesamt ihre Lebenssituation positiver. Findet sich höheres Wohlbefinden bei häufigerer Gartentätigkeit innerhalb von einkommensgleichen Untersuchungsgruppen?

2. "Repräsentativität." Je stärker die Zusammensetzung der Stichprobe sich von den Merkmalen in der Grundgesamtheit unterscheidet, umso höher ist das Risiko für mögliche Verzerrung der Ergebnisse. Eine oft übersehene Fehlerursache ergibt sich bei allen Studien, an denen Menschen freiwillig teilnehmen oder nicht teilnehmen. Wird das Problem in der Veröffentlichung zur Studie überhaupt angesprochen? Zahlenmäßig präzise erläutert?

3. Kontrolle der Kausalrichtung. Rein "korrelative" Studien ermöglichen grundsätzlich keine Entscheidung über die Richtung von Kausalwirkungen.

Beispiel: Musikhören erhält kognitive Leistungsfähigkeit im Alter: Eine festgestellte, positive Beziehung könnte auch dadurch zu erklären sein, dass kognitiv leistungsfähige Menschen insgesamt aktiver sind und deshalb auch häufiger Musik hören.

4. Operationalisierung. Für eine angemessene Erfassung der Schlüsselvariablen ist eine adäquate Operationalisierung zwingend notwendig. Dies bringt z.B. die Fragen mit sich: Gibt es bereits wissenschaftlich geprüfte Skalen oder aber weshalb wurde ein ad hoc entwickeltes Erhebungsinstrument eingesetzt? Wurde dazu ein Pretest durchgeführt (Anm. 4)?

Beispiel: Psychologische Entspannungsverfahren vermindern Schlafstörungen: Ist es ausreichend, nach der zeitlichen Dauer und Häufigkeit zu fragen, um die Wirksamkeit zu erfassen? Und nicht zusätzlich nach der Art des Entspannungsverfahrens oder der Intensität der erzielten Entspannungswirkung?

5. Kontrolle der Untersuchungsbedingungen. Eine beträchtliche Zahl der Fehlerquellen kann durch einen experimentellen Versuchsaufbau recht zuverlässig ausgeschaltet werden; dabei bedient man sich vor allem echter Zufallszuweisungen der Einzelfälle auf unterschiedliche Untersuchungsbedingungen (darunter meist eine der normalen Alltagsrealität entsprechende "Kontrollbedingung"). Ferner bedient man sich vor allem der Methodik des Blindversuchs. Nicht bei jeder Fragestellung ist aber ein experimentelles Untersuchungsdesign machbar oder aussagekräftig (Anm.5).

6. Statistische Inferenzprüfungen. Signifikanzprüfungen dienen der Absicherung der Ergebnisse gegen Zufallseinflüsse ("Null-Hypothese") auf Grundlage der mathematischen Wahrscheinlichkeitsrechnung. Signifikanzberechnungen beruhen oft auf Voraussetzungen (z.B. Skalenniveau, Voraussetzung der Normalverteilung, Varianzhomogenität) und können daher fehlerhaft sein. Bei einer großen Zahl von durchgeführten Signifikanzprüfungen sinkt die Aussagekraft des "Signifikanz-Niveaus". Mit steigender Stichprobengröße werden auch geringe Unterschiede statistisch signifikant, die nicht mehr alltagsrelevant sein müssen (Anm.6).

Sichtbare Kriterien zur Bewertung des fachlich-wissenschaftlichen Entwicklungsstandes

Insgesamt ist vor allem darauf hinzuweisen, dass die Zahl der denkbaren Störeinflüsse so hoch ist, dass ein wissenschaftlich zu prüfender Sachverhalt niemals durch eine einzelne, auch noch so sorgfältig angelegte Studie verlässlich bewertet werden kann; hierdurch ergeben sich für sämtliche wissenschaftlichen Disziplinen zusätzliche Qualitätsanforderungen bei der vergleichenden Bewertung unterschiedlicher Untersuchungen und Ergebnisse zum gleichen Problem.

Dies ist auch eine Grundlage, um ein erstes Gesamtbild über das erreichte Qualitätsniveau in einem Fach- oder Forschungsgebiet zu gewinnen:

1. Liegen Sammelreferate zum einschlägigen Erkenntnisstand vor? Handelt es sich dabei um qualitative Reviews oder um Meta-Analysen? (Anm.7)
Falls Sammelreferate (qualitative Reviews oder Metaanalysen) zum aktuellen Forschungsstand Vorliegen:
2. Werden unter den Qualitätskriterien für die Berücksichtigung von Studien insbesondere die Qualitätsschwellen (v.a. für die Publikation in den Zeitschriften) explizit benannt?
3. Sind bei den meist interdisziplinär verankerten Publikationen Arbeiten aus allen einschlägigen Fachdisziplinen berücksichtigt? (Anm.8)
4. Auf wie viele unabhängig durchgeführte Einzeluntersuchungen zum Forschungsgebiet stützen sich diese?
5. Lassen sich zu den Haupthypothesen bestätigende Befundlagen erkennen? (Bei Meta- Analysen: Mit jeweils welchen Effektstärken?) Differenzierung der Befundbewertung - für welche wesentlichen Fragestellungen lassen sich ausreichend belegbare Bewertungen ableiten/nicht ableiten? (Als Ursache kommt jede der vorausgehend genannten Voraussetzungen in Betracht)