

# Psychologie kompakt

## Grundlagen und Forschungsperspektiven

Rainer W. Alexandrowicz  
Thorsten-Christian Gablonski  
Judith Glück (Hg.)

**Kapitel 5**  
**Wissenschaftliche Methoden  
psychologischer Forschung**  
Rainer W. Alexandrowicz & Judith Glück



Lernunterlage zum Aufnahmeverfahren Psychologie an der Universität Klagenfurt. Dieser Text enthält das neue Kapitel 5, das in nebenstehend abgebildeten Buch ausgetauscht wurde.

Alexandrowicz, R. W., Gablonski, T.-C. & Glück, J. (2014).  
*Psychologie kompakt. Grundlagen und Forschungsperspektiven.*  
Wien: facultas. ISBN: 978-3-7089-1122-9



# Inhaltsverzeichnis

|          |   |            |
|----------|---|------------|
| <b>5</b> | <b>Wissenschaftliche Methoden psychologischer Forschung</b> | <b>99</b>  |
|          | <i>Rainer W. Alexandrowicz &amp; Judith Glück</i>           |            |
| 5.1      | Einleitung . . . . .  | 99         |
| 5.1.1    | Ein Ankerbeispiel . . . . .                                 | 99         |
| 5.1.2    | Psychologie als Wissenschaft . . . . .                      | 100        |
| 5.1.3    | Empirie versus Theorie . . . . .                            | 101        |
| 5.2      | Der prototypische Ablauf eines Forschungsprojektes .        | 102        |
| 5.2.1    | Die Literaturrecherche . . . . .                            | 103        |
| 5.2.2    | Die Fragestellung . . . . .                                 | 107        |
| 5.2.3    | Die Methodik der Studie . . . . .                           | 108        |
| 5.2.4    | Die Stichprobe . . . . .                                    | 120        |
| 5.2.5    | Die Auswertung und Ergebnisdarstellung . . .                | 125        |
| 5.2.6    | Die Diskussion . . . . .                                    | 135        |
| 5.3      | Schlussbetrachtung . . . . .                                | 136        |
|          | <b>Literaturverzeichnis</b>                                 | <b>139</b> |



# 5 Wissenschaftliche Methoden psychologischer Forschung

*Rainer W. Alexandrowicz & Judith Glück*

## 5.1 Einleitung

Die zahlreichen Erkenntnisse, die in den verschiedenen Kapiteln dieses Sammelbandes präsentiert werden, beruhen auf den Ergebnissen psychologisch-wissenschaftlicher Forschung. Wie „funktioniert“ psychologische Forschung? Dieses Kapitel soll einige grundsätzliche Prinzipien vermitteln, die in diesen Forschungsarbeiten verfolgt wurden und die wissenschaftliche Psychologie auszeichnen. Diese Grundprinzipien psychologischer Forschungsmethodik bilden ein Fundament für die unterschiedlichen Teildisziplinen des Faches und sind daher für jede weitere Beschäftigung mit dem vielfältigen Themenkanon der Psychologie unabdingbar, sei es forschend oder anwendend. Im Psychologiestudium werden diese Prinzipien im Rahmen der Methodenausbildung vermittelt.

### 5.1.1 Ein Ankerbeispiel

Da theoretische Ausführungen schwierig zu lesen sind und Gefahr laufen, aufgrund ihrer Allgemeinheit unverständlich zu bleiben, soll

die Darstellung der verschiedenen Aspekte anhand des folgenden Ankerbeispiels erfolgen:

Wir wollen die hypothetische Forschungsfrage nach der Wirksamkeit von *blended*<sup>1</sup> *learning* im Kontext universitärer Lehre untersuchen. *Blended learning* gilt als Überbegriff für Lernformen, in denen konventioneller Unterricht mit technischen Unterstützungen angereichert wird. Auch an Universitäten werden solche Lernformen immer häufiger eingesetzt, indem beispielsweise klassische Vorlesungen durch Online-Elemente wie Quizzes, Übungsmaterialien und weiterführenden Informationen ergänzt werden. Führen solche Elemente tatsächlich zu besserem Lernerfolg? Verschiedene Varianten, wie man sich dieser Fragestellung wissenschaftlich annähern kann, und methodische Aspekte, die dabei zu beachten sind, werden an diesem Beispiel dargestellt. Zunächst erfolgt eine kleine Begriffsklärung, danach werden fünf zentrale Elemente eines wissenschaftlichen Forschungsbeitrags beschrieben und anhand von zwei prototypischen Studienvarianten durchgespielt.

### 5.1.2 Psychologie als Wissenschaft

Wir verstehen Psychologie als empirische Wissenschaft. Der erste Teil des Begriffes, empirisch, leitet sich ab aus dem griechischen *ἐμπειρία* (*empeiria*, Erfahrung) und offenbart damit die Verankerung von (psychologischer) Erkenntnis in erfahrbaren Phänomenen. Eine Definition ist allerdings nur hilfreich, wenn sie auch eine Abgrenzung zulässt – in unserem Fall wäre das jene zur Alltagsvermutung, häufig gestützt durch subjektive Gewissheit. Umgelegt auf unser Beispiel hieße das: Auch wenn es durchaus plausibel scheint, dass *blended learning* einen positiven Effekt auf die Vermittlung von Lehrstoff an der Universität hat, bedarf es eines empirischen Belegs dafür.

Der zweite Teil, Wissenschaft, formuliert (unter anderem) den methodischen Anspruch, bei der Überprüfung unserer Vermutung systematisch, transparent und intersubjektiv nachvollziehbar vorzugehen und

---

<sup>1</sup> *the blend*: Mischung; *to blend*: mischen

überprüfbare Aussagen zu tätigen. Damit erfüllt die obige Vermutung noch nicht die Kriterien von Wissenschaftlichkeit, da keine dieser Forderungen erfüllt ist. Eine detaillierte Auflistung von Aspekten wissenschaftlicher Integrität bietet beispielsweise die Schweizer Akademie der Wissenschaften ([Akademien der Wissenschaften Schweiz, 2008](#)).

### **Exkurs: Schule versus Uni (I)**

An dieser Stelle scheint uns ein Hinweis für all jene, die neu auf der Universität sind, angebracht: Während Schulstoff in der Regel in Form von Fakten vermittelt wird, sollen Studierende ein Bewusstsein dafür entwickeln, dass viele der präsentierten Inhalte letztlich Theorien darstellen (wenngleich in der Regel gut abgesicherte). Ein klassisches Beispiel wäre das Kugelmodell des Atoms im Chemieunterricht: Tatsächlich ist der aktuelle Wissensstand im subatomaren Bereich ungleich komplexer, sodass sich die Frage stellt, ob das Kugelmodell vielleicht sogar hinderlich für ein vertieftes Verständnis ist.

Im Studium soll die Fähigkeit erlangt werden, die zentralen Theorien des eigenen Faches zu verstehen, wissenschaftlich zu hinterfragen und unter Anwendung geeigneter Methoden gegebenenfalls zu widerlegen beziehungsweise zu neuen, besseren Theorien weiterzuentwickeln.

### **5.1.3 Empirie versus Theorie**

Psychologische Forschung bezieht sich in ihrem Erkenntnisgewinn einerseits auf die Empirie (hier im Sinne einer „erfahrbaren Welt“) und andererseits auf die Theorie (also eine im Denken verankerte und schlussendlich auf Verstehen abzielende Struktur). Erkenntnisprozesse laufen dabei in beide Richtungen ab: Aus einer (abstrakten) Theorie leiten wir (konkrete) Hypothesen ab, die wir hinsichtlich ihrer empirischen Belegbarkeit untersuchen – dieser Vorgang wird Deduktion genannt –, oder uns liegen empirisch gewonnene Beobachtungen vor,

aus denen wir eine Theorie formen – dieser Vorgang wird Induktion genannt<sup>2</sup>.

Für unsere Vermutung zu den positiven Auswirkungen von *blended learning* auf den Lernerfolg böte sich ein deduktiver Ansatz zur Überprüfung an: Wir könnten beispielsweise auf dem theoretischen Konzept der Andragogik aufbauen, das sich mit dem lebenslangen Lernen Erwachsener (im Gegensatz zur Pädagogik, die auf Kinder fokussiert) beschäftigt. Dieser Theorie zufolge haben Erwachsene (unter anderem) ein Bedürfnis nach selbstgesteuertem Lernen (siehe z. B. Knowles, 1975, 1984). Da *blended learning* (vorläufig noch ohne klare Definition verwendet!) im Allgemeinen mehr Möglichkeiten der eigenständigen Erarbeitung von Inhalten einschließt als traditionelle Lernformen, können wir die Erwartung ableiten, dass es ein wirksames Mittel zur Vermittlung von Wissen und Fertigkeit für Erwachsene darstellt. Darauf aufbauend gilt es, uns eine Vorgangsweise zu überlegen, um diese Erwartung empirisch zu überprüfen.

Ein induktiver Ansatz läge vor, hätten wir im Rahmen einer Serie von Unterrichtseinheiten mehrmals die Beobachtung gemacht, dass die Verwendung von Online-Materialien, die bestimmte Stoffinhalte visualisieren und zu Hause ausprobiert werden können, zu besserem Verständnis dieser Inhalte beigetragen hat. Auf dieser Beobachtung aufbauend könnten wir eine Theorie über die positiven Auswirkungen formulieren und diese anhand einer wissenschaftlichen Studie überprüfen.

## 5.2 Der prototypische Ablauf eines Forschungsprojektes

Nachfolgend wird eine Reihe zentraler Schritte, die in nahezu jedem wissenschaftlichen Projekt eine Rolle spielen, vorgestellt. Es ist nicht möglich, einen universell gültigen Ablauf zu formulieren, doch spielen

---

<sup>2</sup> Auf eine dritte Form der Schlussfolgerung – die Abduktion – wird hier nicht eingegangen.



die hier genannten Elemente in den meisten Forschungsprojekten eine Schlüsselrolle. Sie bilden auch das Grundgerüst einer wissenschaftlichen Abschlussarbeit, also einer Bachelorarbeit, Masterarbeit oder Dissertation.

### 5.2.1 Die Literaturrecherche

In jedem Fall ist zu Beginn einer Forschungsarbeit im Rahmen einer umfassenden Literaturrecherche zu klären, welche Erkenntnisse bislang zum gewählten Thema vorhanden sind. Nur wenn man den aktuellen Forschungsstand zu einem Thema kennt, lässt sich eine sinnvolle Fragestellung formulieren, die zur Erweiterung dieses Forschungsstandes beiträgt. Wir unterscheiden zwischen wissenschaftlichen und nicht-wissenschaftlichen Quellen. Wissenschaftliche Quellen finden wir primär in Fachzeitschriften, die Beiträge nur nach einem erfolgreich absolvierten Reviewprozess veröffentlichen.

#### Exkurs: Der Reviewprozess

Einen wissenschaftlichen Artikel zeichnet – in Abgrenzung zu populär- bzw. pseudowissenschaftlichen Texten – aus, vor der Veröffentlichung einen wissenschaftlichen Begutachtungsprozess (*review*) durchlaufen zu haben. Ein Forschungsteam hat die Ergebnisse seiner Studie verschriftlicht und als Manuskript bei einer Fachzeitschrift eingereicht. Dort entscheidet der Herausgeber oder die Herausgeberin (*editor*) nach einer ersten Durchsicht, ob der Text thematisch zur Ausrichtung der Zeitschrift passt, *a prima vista* die nötige wissenschaftliche Seriosität aufweist und allfälligen Formalkriterien (z. B. Länge und Gestaltung) genügt. Fällt dieses Urteil positiv aus, dann wird das eingereichte Manuskript zur Begutachtung an typischerweise zwei oder drei Personen mit facheinschlägiger Expertise (*reviewer*) weitergegeben. Diesen fällt die Aufgabe zu, den Text nach wissenschaftlichen Kriterien de-

tailliert zu beurteilen. Als Qualitätsmerkmal gilt *double-blinded reviewing*, bei dem weder die *reviewer* wissen, wer die Arbeit verfasst hat, noch das Forschungsteam erfährt, wer begutachtet hat. Aufgrund der gutachterlichen Empfehlungen erfolgt schließlich die editoriale Entscheidung über *acceptance* (Beitrag wird veröffentlicht), *revision* (Beitrag muss überarbeitet werden) oder *rejection* (Beitrag wird abgelehnt). Dieser Prozess soll wissenschaftliche Mängel, unbelegte bzw. unhaltbare Behauptungen oder Fehlschlüsse aufdecken.

Eine wissenschaftliche Literaturrecherche ist ein komplexes Unterfangen, das mehrere Schritte umfasst.

**Erste Orientierung.** Je nach Vorkenntnissen kann eine Orientierungsphase erforderlich sein, in der wir erst jene Terminologie identifizieren, unter der das betreffende Thema wissenschaftlich abgehandelt wird. Im ersten Schritt mag hierbei zur Orientierung bei völlig neuer Thematik auf universelle Suchmaschinen – wie etwa [DuckDuckGo](#), [ixquick/Startpage](#) oder [Google](#) – zurückgegriffen werden. In deren Suchergebnissen findet sich allerdings eine erhebliche Vermengung wissenschaftlicher mit nicht- oder pseudo-wissenschaftlichen Informationen (teilweise durch teure Suchmaschinenoptimierung in den Vordergrund gerückt). Wir können daher grundsätzlich festhalten: „Das Internet“ (im Sinne genereller Suchmaschinen) liefert keine wissenschaftlich zuverlässigen Quellen.

Eine möglicherweise strukturiertere Einführung in ein neues Themengebiet mag [WIKIPEDIA](#) bieten, da dort im Idealfall ein Diskurs über jeden Eintrag geführt wird – doch zeigt die Erfahrung (und die offen abrufbare Begriffshistorie), dass auch hier Auffassungsunterschiede bis hin zu Grundsatzdiskussionen oder gar Ideologiestreit (ganz abge-

sehen von Vandalismus<sup>3</sup>) eine Verzerrung der Informationen bewirken können. Um es daher unmissverständlich auszudrücken: Google und WIKIPEDIA sind keine wissenschaftlich zitablen Informationsquellen.

**Spezialisierung.** Wesentlich verlässlichere Informationen erhalten wir aus akademischen Datenbanken und Suchmaschinen. Für das Fach Psychologie wären hier vor allem die Datenbanken [psycINFO](#) oder [psycARTICLES](#) zu nennen, im klinisch/medizinischen Kontext spielen beispielsweise die [medline](#) (mit der öffentlich zugänglichen Komponente [PubMed](#)) oder die [Cochrane Library](#) eine wichtige Rolle. Diese Datenbanken bilden einschlägige Fachliteratur ab, d. h. Artikel, die in wissenschaftlichen Zeitschriften publiziert wurden. Auf die Datenbanken kann in der Regel über die [Bibliothekshomepage der Universität](#) zugegriffen werden. Im Gegensatz zu den allgemeinen Suchmaschinen kann [Google Scholar](#), eine spezielle Suchmaschine zur Auffindung wissenschaftlicher Literatur, nützlich sein. Hier werden nicht nur Zeitschriftenartikel, sondern auch Bücher und Buchkapitel gelistet, oft sogar direkt verlinkt. Da hier jedoch keine Auswahl nach Qualitätskriterien (wie bei einem Journal) erfolgt, kann die Qualität dieser Beiträge stark variieren.

### **Exkurs: Verfügbarkeit wissenschaftlicher Artikel**

Es war lange Zeit ein großes Problem, dass wissenschaftliche Artikel von Bibliotheken (über-)teuer(-t) eingekauft werden mussten. Aus dem Spannungsverhältnis von einerseits Forschung, die durch öffentliche Gelder finanziert wurde, und andererseits kommerzialisiertem Zugriff auf deren Ergebnisse, hat sich das Prinzip des *open access* (etwa: freier Zugriff) entwickelt (siehe z. B. die [Berliner Erklärung](#)). Dabei müssen Forschende bzw. deren Universitäten oder Forschungsinstitutionen den Verlagen die Unkosten

---

<sup>3</sup> vgl. [Wikipedia über Vandalismus](#), [Süddeutsche Zeitung über den Streich mit Gutenbergs Vornamen](#), [Spiegel über das Einfrieren von wikipedia-Beiträgen](#) oder die interessante Visualisierung von [Stefaner, Taraborelli und Ciampaglia \(2011\)](#), abrufbar unter [notabilia.net](#).

für die Produktion ersetzen (*publication fees* oder *article processing charges*, APC), im Gegenzug sind die Beiträge frei online verfügbar. Das hat enorm zur Verbreitung von wissenschaftlichen Ergebnissen beigetragen. Leider entstanden im „Kielwasser“ dieser Entwicklung auch Pseudo-Journale, die nur darauf abzielen, die *publication fees* zu kassieren, ohne einen seriösen Reviewprozess zu gewährleisten oder in irgendeiner Form an der Qualität der Beiträge interessiert zu sein (sie werden auch *predatory journals* – Raubtierjournale – genannt). Es liegt an uns, Quellen kritisch zu hinterfragen – was unter anderem auch eine fundierte Methodenausbildung erfordert, um die Qualität der publizierten Beiträge selbst beurteilen zu können.

**Feinschliff.** Die zumeist wichtigsten Quellenangaben finden wir oft nicht in den Datenbanken und Suchmaschinen, sondern in den dort identifizierten Artikeln selbst. Ein zentrales Merkmal eines wissenschaftlichen Artikels besteht darin, alle Aussagen zu belegen – einzig der empirische Beleg in Form nachvollziehbarer Ergebnisse im Wechselspiel mit zitablen Quellen zählt: Wurde durch eine fundierte und gut geplante Studie ein Erkenntnisgewinn erzielt, der den aktuellen Wissensstand zum Thema erweitert? Haben wir also Publikationen identifiziert, die unser Thema betreffen, dann finden wir darin auch jene wissenschaftlichen Quellen, auf die sich diese Artikel stützen – sie liefern auch häufig die wertvollsten Hinweise auf wichtige Erkenntnisse für die eigene Studie. So kann man sich von einigen relevanten Artikeln schrittweise „weitertasten“, bis man einen guten Gesamtüberblick über das betreffende Thema erlangt hat. Suchmaschinen wie psycARTICLES sind hier in doppelt hilfreich: Sie ermöglichen nicht nur, das Literaturverzeichnis eines gefundenen Artikels zu durchsuchen, sondern auch jene Artikel zu identifizieren, die den betreffenden Artikel ihrerseits zitieren.

Es sei noch darauf hingewiesen, dass der Rechercheprozess niemals als abgeschlossen gesehen werden kann. Praktisch wöchentlich erschei-

nen neue Artikel, die weitere wichtige Aspekte für das eigene Projekt enthalten können. Letztlich endet die Recherchephase erst mit der Einreichung des Manuskriptes oder der Abgabe der wissenschaftlichen Arbeit.

### 5.2.2 Die Fragestellung

Auf Basis der Literaturrecherche haben wir uns auf eine Themenstellung festgelegt, die interessante neue Erkenntnisse erwarten lässt. Dies umfasst auch eine klare Definition der zu untersuchenden Phänomene. In unserem Beispiel der Auswirkungen von *blended learning* auf den Wissenserwerb hätte die Literaturrecherche zunächst wohl vor allem ergeben, dass der Begriff reichlich unpräzise ist. [Moskal, Dziuban und Hartman \(2013\)](#) sehen ihn beispielsweise als Sammelbegriff für jegliche Art der Kombination von traditionellen Unterrichtsmethoden (face to face) mit modernen technischen Methoden („online learning, mobile technologies and resources that exist in the cloud“; S. 15). Damit bleibt unklar, ob die technischen Komponenten nur begleitend oder als primäres Vermittlungsinstrument zur Anwendung kommen.

Wir wollen vereinfachend folgende Festlegung vornehmen: Der Begriff *blended learning* soll hier im Sinne von computer aided instruction (CAI; z. B. [Mann, 2009](#)) verwendet werden. Konkret wollen wir untersuchen, ob die Einbeziehung zweier spezifischer Online-Elemente in einer Vorlesung an der Universität – (1) Quizzes und Befragungen via Smartphone während der Vorlesungen und (2) Selbsttests und Links zu weiterführenden Informationen für die Aufarbeitung zuhause – den Lernerfolg der Studierenden positiv beeinflusst.

Die ursprüngliche Formulierung unserer Alltagsvermutung besagte, dass *blended learning* „einen positiven Effekt auf den Lernerfolg hat“. Damit ist im Grunde noch recht wenig ausgesagt. Wir müssen konkretisieren, in welcher Form ein allfälliger „positiver Effekt“ beobachtet werden kann. Hierzu bedarf es einer Operationalisierung (sinngemäß: Messbarmachung; s. Abschnitt 6.3.1: wie kann man den Lernerfolg messen? Wir wollen zwei Fragestellungen empirisch untersuchen:

1. Lässt sich mittels CAI der Wissensstand steigern?
2. Lässt sich mittels CAI die Lernmotivation steigern?

Diese Beispiele sollen eine wichtige Eigenschaft der Fragestellung illustrieren: Es wird tatsächlich eine Frage formuliert. Nicht als Fragestellung sind Formulierungen geeignet, die mit „Über...“ oder „Zur...“ beginnen („Über die Vorteile von *blended learning*“; „Zur Brauchbarkeit von *blended learning*“ u. dgl.). Solche vagen Aussagen lassen nicht erkennen, worauf das Forschungsvorhaben konkret abzielt. Sie drücken lediglich aus, dass ein Projekt zum Thema *blended learning* durchgeführt wurde. Aber nicht alles, was als Frage formuliert wird, ist auch automatisch eine gute Fragestellung – „Wie geht es Studierenden mit *blended learning*?“ wäre beispielsweise sehr unklar formuliert. Ist mit „wie geht es“ das Wohlbefinden, der Lernerfolg oder die Motivation – oder alle drei – gemeint? Wissenschaftliche Fragestellungen beziehen sich auf mess- oder beobachtbare Variablen (beispielsweise den Wissenserwerb oder die Lernmotivation) und sind eindeutig beantwortbar.

### 5.2.3 Die Methodik der Studie

Nachdem wir das für die eigene Fragestellung relevante Wissen anhand entsprechender Belege aufgearbeitet und daraus unsere konkrete Fragestellung abgeleitet haben, ergibt sich aus dieser, welche Methoden zu ihrer Beantwortung zur Auswahl stehen. Da uns ein umfassendes Inventar zur Verfügung steht, erfordert diese Entscheidung fundierte Kenntnisse über ein breites Spektrum anwendbarer Methoden. Welche davon wir auch wählen, ihre Auswahl ist in jedem Fall genau zu beschreiben und ihre Wahl zu begründen. In jedem Fall gilt: Die Wahl der Methode ergibt sich aus der Fragestellung. Es gibt keine Methode, die generell gut oder für alle Fragestellungen geeignet ist.

### Quantitativ, qualitativ oder mixed

Das umfangreiche Methodeninventar der Sozialwissenschaften (die neben Psychologie auch Fächer wie beispielsweise Soziologie oder Politikwissenschaften einschließen) wird typischerweise in qualitative und quantitative Verfahren unterteilt. Nachfolgend werden wesentliche Charakteristika dieser Ansätze anhand von Beispielen illustriert.

**Der quantitative Ansatz.** Ein wesentliches Element quantitativer Methoden zur Beschreibung psychologischer Phänomene und deren Zusammenspiel liegt in der Extraktion von Datenstrukturen. In Fragestellung 1 wollen wir beispielsweise die Auswirkung von *blended learning* auf den Wissensstand untersuchen. Der Wissensstand könnte etwa quantifizierend erfasst werden, indem jeder Testperson 20 Fachbegriffe vorgegeben werden, die sie den zutreffenden Erklärungen aus einer Liste zuordnen muss. Die Anzahl korrekter Zuordnungen soll eine Quantifizierung des Wissensstandes einer einzelnen Person darstellen. Der Mittelwert über mehrere Personen wäre nun eine mathematische Abstraktion (ein „in eine Zahl gegossenes Abbild“) des Wissensstandes der gesamten Gruppe. Der Vergleich der Mittelwerte unterschiedlich unterrichteter Gruppen (z. B. mit und ohne *blended learning*; s. Abschnitt 5.2.3, S. 113) stellte somit eine quantifizierende Aussage über die Struktur dar, die den Daten innewohnt – hier der Mittelwertsunterschied.

Die Auswirkungen von *blended learning* auf die Lernmotivation könnte man quantitativ erfassen, indem man die Lernenden eine Reihe von Fragen zu ihrer Motivation mithilfe von Antwortskalen (z. B. von „stimme gar nicht zu“ bis „stimme sehr zu“) beantworten lässt. Die Antwortkategorien können mit Punkten (z. B. 0 für „gar nicht“ bis 4 für „sehr“) codiert und anschließend summiert werden – eine solche Herangehensweise kann jedoch nur vorgenommen werden, wenn zuvor die Adäquatheit einer solchen Verrechnung hinreichend belegt werden konnte (vgl. Abschnitt 6.2.2); Die hierfür erforderlichen Methoden werden wir im Rahmen der Testtheorie diskutieren.

Wir wollen an dieser Stelle zwei wichtige Punkte herausarbeiten: (1) Die Entscheidung, die Anzahl korrekt zugeordneter Fachbegriffe als Quantifizierung des Wissensstandes heranzuziehen, wurde vorab getroffen (darauf werden wir im nachfolgenden Unterabschnitt zurückkommen). Diese Methode kann auf große Gruppen mit vertretbarem Aufwand angewendet werden, was das Erkennen struktureller Merkmale erleichtert. Warum hingegen eine Person einen bestimmten Begriff der korrekten Definition zuordnen konnte und einen anderen nicht, wird so noch nicht erfasst. Auf diese individuelle Komponente soll im Rahmen der Fragestellung nicht eingegangen werden. (2) Allein der Vergleich zweier Mittelwerte ist nur der erste Schritt auf dem Weg zu einer wissenschaftlichen Aussage, in Abschnitt 5.2.5 (S. 126) werden wir diesen Ansatz vertiefen.

**Der qualitative Ansatz.** Ein wesentliches Merkmal einer qualitativen Herangehensweise besteht in der Fokussierung auf das Individuum. Personen wird die Möglichkeit gegeben, sich umfangreich und differenziert darzustellen. Das kann beispielsweise im Rahmen eines Interviews oder einer Gruppendiskussion (einer sog. Fokusgruppe) erfolgen. Die Personen könnten auch ein Tagebuch führen, etwa bei Fragestellungen, die auf eine Entwicklung oder einen Verlauf abzielen. Ganz allgemein lässt sich festhalten, dass qualitative Methoden sprachliches oder auch bildliches Material in verbaler Form analysieren. Bei der Entscheidung für ein Interview oder eine Gruppendiskussion wird das Gespräch aufgezeichnet und anschließend verschriftlicht (transkribiert). Diese Transkripte bilden in der Folge die Grundlage für die weitergehende Auswertung: Es können beispielsweise charakteristische Aussagen identifiziert werden, die sich zu einem aussagekräftigen Gesamtbild zusammenfügen lassen (s. Abschnitt 5.2.5, S. 130).

In einer qualitativen Studie zu den Auswirkungen von *blended learning* an der Universität könnte man Studierende in Fokusgruppen darüber diskutieren lassen, welche Vor- und Nachteile sie im Vergleich zu traditionellen Vorlesungsformaten wahrnehmen, ihre Aussagen protokollieren und darin zentrale Themen identifizieren. Diese Heran-



gehensweise erlaubte beispielsweise differenzierte Erkenntnisse, welche Elemente von *blended learning* besser angenommen werden oder welche Gestaltungsmerkmale zu berücksichtigen sind. Fragestellung 2 ist besonders prädestiniert für eine qualitative Herangehensweise: Für die Lernmotivation spielen eine Reihe emotionaler Faktoren eine Rolle, beginnend mit der höchst individuellen Interessensstruktur, früheren Lernerfahrungen, Einstellung zum Medium der Vermittlung (hier: Lehrperson vs. Computer), u.v.m.

Auf eine potentielle Problematik bei Fokusgruppen soll hier hingewiesen werden: dem möglichen Einfluss von Gruppendruck: Manche Menschen schließen sich lieber den Sichtweisen dominanter Gruppenmitglieder an als ihnen zu widersprechen, auch wenn sie eigentlich anderer Meinung sind (vgl. dazu das Experiment von Asch, Abschnitt 10.2, S. 255). Dies kann durch parallel durchgeführte Einzelinterviews (mit den gleichen oder auch anderen Personen) oder eine Fragebogenerhebung überprüft werden.

Analog zur Problematik, dass ein paar spontan ausgedachte Fragen noch kein wissenschaftlich tragfähiges Instrument ergeben, gilt auch hier: Einfach Personen in einen Raum zu setzen und diskutieren zu lassen genügt ebenfalls noch nicht den Ansprüchen von Wissenschaftlichkeit. Ein wesentlicher Aspekt ist beispielsweise, dass bei der Anleitung der Diskussionen, Extraktion von Inhalten aus den transkribierten Materialien und deren anschließender Strukturierung anhand inhaltlicher Merkmale subjektive Entscheidungen der Durchführenden erforderlich sind. Daher bedarf es klarer Richtlinien, wie ein verzerrender Einfluss der Forschenden hintanzuhalten ist. Ein weiterer wichtiger Unterschied zur quantitativen Herangehensweise besteht im großen Aufwand der Transkription und (zwar computerunterstützten, aber dennoch händisch vorzunehmenden) Analyse. Deshalb ist diese Methode vorläufig nur auf kleine Gruppen anwendbar. (Eine automatisierte Sprach- oder Schrifterkennung würde hier enorm weiterhelfen, doch scheint eine solche gegenwärtig nicht in Reichweite.)

**Der *mixed methods* Ansatz.** Da der quantitative und der qualitative Ansatz unterschiedliche Erkenntnisziele verfolgen, kann es für viele Fragestellungen besonders gewinnbringend sein, diese nebeneinander oder aufeinander aufbauend einzusetzen und die verschiedenen Erkenntnisse zu integrieren. Zwei Varianten der Integration spielen dabei eine besondere Rolle.

**Das *Vorstudienmodell*.** Wollen wir eine Thematik beforschen, zu der noch wenig Wissen vorliegt und deren Struktur daher noch unklar ist, dann ist es sinnvoll, zunächst im Rahmen einer qualitativen Studie strukturelle Merkmale herauszuarbeiten. Eine solche Herangehensweise hat sich in unserem Beispiel bereits abgezeichnet: Die qualitative Teilstudie kann uns zeigen, welche Elemente von *blended learning* sich besonders effektiv einsetzen lassen und angenommen werden, um im darauffolgenden quantitativen Schritt den Unterschied der so optimierten Unterrichtsmethode zu einer „klassischen“ Lehrmethode im großen Maßstab und statistisch abgesichert zu erfassen.

**Das *Vertiefungsmodell*.** Wenn wir zu einer Thematik arbeiten, die bereits eine lange Forschungstradition aufweist und zu der schon viele Befunde vorliegen, widmen wir uns meist schwierigen Detailfragen. Häufig verfügen wir in solchen Fällen schon über ausgefeilte Messmethoden (z. B. psychologische Tests), die mit komplexen statistischen Verfahren untersucht werden. Dabei kommt es aber immer wieder vor, dass wir auf unerwartete Ergebnisse stoßen, die im Rahmen des theoretischen Modells zunächst nicht nachvollziehbar sind. In solch einem Fall kann es sich als sehr ergiebig erweisen, in einer gezielten qualitativ angelegten Folgestudie zu untersuchen, ob es sich nur um eine atypische Datenkonstellation handelt, oder ob etwa ein wichtiges strukturelles Merkmal übersehen wurde. Doch auch wenn keine Widersprüchlichkeiten aufgetreten sind kann eine qualitative Folgestudie wichtige Erkenntnisse liefern: Angenommen, wir hätten in der quantitativen Teilstudie relevante Unterschiede zwischen den beiden Lehrmethoden gefunden, dann ließe sich in einem nachfolgenden qualitativen Schritt untersuchen, ob sich die Studierenden nur die Definitionen eingeprägt haben, oder ob sie dieses erworbene Wissen auch

eigenständig auf andere Bereiche übertragen können. Dafür könnte man sie beispielsweise kleine Essays zu solchen Übertragungsfragen verfassen lassen und dann untersuchen, sich wie die verschiedenen Unterrichtsformen auf qualitative Aspekte der Essays – Inhalt, Struktur, kritische Reflektion – auswirken.

### Die Wahl eines geeigneten Studiendesigns

Auch die Wahl des Studiendesigns hängt in erster Linie von der konkreten Fragestellung ab. Grundlegende Studiendesigns der empirischen Psychologie werden in den Basislehrveranstaltungen zur psychologischen Methodenlehre vermittelt. Aus der Vielzahl möglicher Studiendesigns wollen wir im Folgenden zwei typische Studiendesigns exemplarisch durchspielen.

**Ein experimenteller Zugang.** Auf den ersten Blick scheint die Frage, ob *blended learning* den Wissenserwerb und die Lernmotivation steigert, leicht beantwortbar. Wir brauchen doch nur vor und nach einer Vorlesungseinheit, in der *blended learning*-Elemente eingesetzt wurden, den Wissensstand und die Motivation der Studierenden zu messen – oder? So einfach ist es leider nicht: vielleicht hätten die Studierenden genauso viel und genauso motiviert gelernt, wenn die Vorlesung ganz klassisch als Frontalunterricht abgehalten worden wäre. Wir müssen also eigentlich überprüfen, ob *blended learning* eine größere Steigerung von Wissen und Motivation bewirkt als traditionelle Unterrichtsmethoden. Dafür bietet sich ein experimenteller Zugang an. Wir könnten zwei Gruppen untersuchen: Eine wird nach „klassischer“ Methode unterrichtet und die andere unter Einsatz von *blended learning*. Die erste wird als Referenz- oder Kontrollgruppe (KG) bezeichnet und die zweite als Experimental- oder Versuchsgruppe (VG). Beide Gruppen werden entsprechend ihrer Versuchsbedingung unterrichtet, wobei streng darauf geachtet wird, dass sich nur die Lehrmethode, nicht aber die Inhalte unterscheiden: die KG erhält traditionellen Unterricht, die VG *blended learning*. Danach wird der Wissensstand beider Gruppen

mit einem geeigneten Instrument erhoben.

Der Vergleich der beiden Gruppen ist allerdings nur sinnvoll, wenn wir davon ausgehen können, dass sie sich nicht schon vor Beginn der Studie in Eigenschaften wie der Lernfähigkeit, dem Interesse, dem Vorwissen oder anderen möglichen Störeinflüssen unterschieden haben. Wenn sich Studierende beispielsweise selbst entscheiden können, ob sie die traditionelle oder die *blended learning*-Vorlesung besuchen, könnte das die Ergebnisse in verschiedene Richtungen verzerren. Besteht die Möglichkeit, sich die Gruppe selbst auszusuchen, dann kann ein allfällig beobachteter Unterschied nicht mehr ausschließlich auf den Effekt von *blended learning* attribuiert werden, sondern wird auch durch die persönlichen Präferenzen – und dadurch z. B. erhöhtes Engagement – verursacht sein.

Da es praktisch unmöglich ist, alle möglicherweise relevanten Rahmenbedingungen zu berücksichtigen (bedenke: es könnten auch unbekannte Einflussgrößen eine Rolle spielen), wählen wir eine Technik, die ohne explizite Kontrolle möglicher Störeinflüsse die Vergleichbarkeit der Gruppen erwarten lässt: die Randomisierung (englisch: *random* = zufällig). Damit ist die zufällige Zuordnung von Personen zur VG und zur KG gemeint, beispielsweise per Münzwurf. Man kann mit hinreichender Plausibilität davon ausgehen, dass sich bei randomisierter Zuordnung relevante Einflüsse in beiden Gruppen hinreichend ausgeglichen wiederfinden, sodass keine systematische Verzerrung mehr zu erwarten ist. Randomisierung ist ein Schlüsselement psychologischer Experimente, in denen Gruppenvergleiche vorgenommen werden. Allerdings kann eine Randomisierung selten unter Alltagsbedingungen vorgenommen werden – etwa, wenn echte Lehrveranstaltungen untersucht werden sollen, um auch die Notwendigkeit einer abschließenden Beurteilung berücksichtigen zu können. Das ermöglicht das im Folgenden beschriebene Design.

**Eine Feldstudie.** Werden zwei echte Lehrveranstaltungen zur Untersuchung ausgewählt, dann liegen wesentlich realitätsnähere Bedingungen vor als im reinen Experiment. Manche besuchen eine Lehrver-

anstellung nur, weil sie im Curriculum vorgeschrieben ist, andere aus Interesse am Stoff; ebenso kann der Notendruck wie bereits erwähnt eine Rolle spielen. In so einem Fall stehen uns technische (statistische) Möglichkeiten zur Verfügung, unerwünschte Effekte (wir sprechen dann von Störvariablen) herauszufiltern (herauszupartialisieren). Dazu müssen diese für jede Person ebenso erfasst werden: Bezogen auf das zuvor eingeführte Beispiel der Interessenslage könnte dies beispielsweise mit einem geeigneten Interessenstest für jede Person erfolgen. Verfahren, die solche raffinierte Auswertungsschritte zulassen, werden wir im Rahmen der Statistik III kennenlernen.

Eine wichtige Störvariable in diesem Design könnte die Lehrperson darstellen: Lassen wir zwei verschiedene Personen die VG und die KG unterrichten, dann fließen auch deren didaktische Fähigkeiten in die Messung ein. Unterrichtet eine Person nach beiden Methoden, dann mag ihre persönliche Präferenz für eine der beiden Unterrichtsformen durchschlagen. Medizinische Studien zur Wirksamkeit von Medikamenten werden aus diesem Grund „doppelblind“ (*double blinded*; vgl. Abschnitt 5.2.1) durchgeführt: Weder das medizinische Personal noch die Versuchspersonen wissen, ob sie mit dem zu testenden Wirkstoff (Verum) oder einem Placebo ohne Wirkstoff behandelt werden. Würde das nicht so gemacht, dann könnten die Behandelnden ihre Überzeugungen betreffend der Un-/Wirksamkeit des untersuchten Medikaments unwillentlich weitergeben.

**Der qualitative Zugang.** Legt die Forschungsfrage eine qualitative Methodik nahe, so lässt sich diese ebenso als Experiment oder als Feldstudie realisieren, Methodik und Design lassen sich auf vielfältige Art kombinieren. Ein qualitatives Experiment wäre beispielsweise die Durchführung von Fokusgruppen, allerdings nach randomisierter Zuordnung zu einer VG (*blended learning*) und einer KG (klassischer Unterricht). Dann ließe sich anhand der Transkripte feststellen, welche Aspekte der jeweiligen Unterrichtsmethode wirksam waren bzw. gut oder schlecht angenommen wurden. Beim klassischen Vortrag könnte sich zum Beispiel herausstellen, dass die Foliengestaltung bzw. das

Tafelbild eine Schlüsselrolle spielt, während beim *blended learning* die Selbstlernanteile nur wirksam sind, wenn sie auch mit einer aktiven Übungsmöglichkeit verbunden sind.

In einer qualitativen Feldstudie könnte ähnlich verfahren werden, allerdings werden Personen aus echten Lehrveranstaltungen zum Fokusgruppeninterview eingeladen. Das böte den Vorteil, dass Störvariablen nicht wie bei der quantitativen Herangehensweise vorab bekannt sein müssen, um sie miterfassen zu können (s.o.); sie könnten vielmehr im Rahmen der Diskussion überhaupt erst identifiziert werden. Es könnte sich beispielsweise herausstellen, dass die klassische Methode besser ankommt, wenn der Vortrag mit klarer Stimme und ohne „ähh ... ähh“ gehalten wird; dass ein bestimmter online angebotener Text im *blended learning* Unterricht zu schwierig oder unverständlich war; dass das online-Forum zur Diskussion unter Studierenden zu umständlich zu bedienen war, weshalb es trotz Interesses nicht genutzt wurde, usw.

**Weitere Studiendesigns.** Neben den hier beschriebenen Varianten gibt es noch viele weitere, teilweise sehr spezialisierte Studiendesigns. Zu nennen wäre etwa die Meta-Analyse – die „Analyse von Analysen“ –, die darauf ausgerichtet ist die in zahlreichen Einzelstudien dargelegten Erkenntnisse zu einem Gesamtbild zu integrieren. Ferner wäre auch die Einzelfallstudie (*case study*) zu nennen, im Rahmen derer eine besonders interessante Beobachtung eben eines Einzelfalles beschrieben wird. Ein solcher ist berichtenswert, wenn er weitere Forschungsaktivitäten erwarten lässt. Weiters kann ein Personenkollektiv über einen längeren Zeitraum mehrfach untersucht werden. Für solche Längsschnittuntersuchungen wurden eine Reihe spezialisierter Auswertungsmethoden entwickelt.

### „Instrumente“: Wie wir psychologische Eigenschaften erfassen

In Abschnitt 5.2.3 haben wir eine simple Methode zur Erfassung des Wissensstandes – Zählen korrekt zugeordneter Fachbegriffe – gewählt, flankiert vom Hinweis, dass eine solche Vorgangsweise noch nicht

die Ansprüche an Wissenschaftlichkeit erfüllt. Hier soll auf diesen Aspekt etwas näher eingegangen werden. Es gibt einen eigenen Fachbereich, der auf diese Frage spezialisiert ist, die Testtheorie – wenn wir uns spezifisch mit psychologischen Testverfahren auseinandersetzen – bzw., allgemeiner, die Psychometrie, wenn wir die Thematik aus einer mathematischen Perspektive betrachten. Wie diese Definition bereits nahelegt, handelt es sich dabei um einen quantitativen Zugang. Die Testtheorie ermöglicht es festzustellen, wie man feststellen kann, ob ein Test oder Fragebogen ein gutes Maß für die zu messende Eigenschaft – in unserem Beispiel den Lernerfolg oder die Lernmotivation – darstellt und die Psychometrie sagt uns, welcher mathematischer Verfahren wir uns dabei sinnvollerweise bedienen können.

**Die Frage der Dimensionalität.** Ein zentraler Aspekt bei der Entwicklung eines Tests oder Fragebogens liegt in der Zusammenstellung der Fragen. Vorweg: Spontan irgendwie interessant klingende Fragen hinzuschreiben lässt keinen wissenschaftlich verwertbaren Informationsgewinn erwarten. Fragen wir beispielsweise danach, wie gut den Lernenden eine Unterrichtseinheit gefallen hat, könnte dies eher eine Sympathiebekundung die vortragende Person betreffend hervorrufen als eine objektive Einschätzung der Rolle, die *blended learning* im Lernprozess gespielt hat.

In unserem Beispiel der Erfassung des Wissensstandes ist es nur sinnvoll, die richtigen Antworten zu einem Gesamtwert zusammenzuzählen, wenn die einzelnen Aufgaben bzw. Fragen auch die gleiche Fähigkeit oder Eigenschaft bzw. das gleiche Konvolut von Fähigkeiten oder Eigenschaften erfassen. Wir nennen ein Erhebungsinstrument eindimensional, wenn es nur eine solche Fähigkeit (eben: eine Dimension) abbildet. Wenn beispielsweise ein Test, der eigentlich das Verständnis von Fachbegriffen erfassen soll, Aufgaben enthält, die man auch durch logisches Denken oder durch Anwendung von Allgemeinwissen beantworten kann, dann misst der Test eben nicht nur Fachwissen, sondern auch logisches Denken oder Allgemeinwissen. Folglich misst ein solches Instrument mehrdimensional, die Verrech-

nung zu einem einzigen Punktwert bildete dann von Person zu Person unterschiedlich stark ausgeprägte Komponenten ab und wäre daher nicht eindeutig interpretierbar (umgangssprachlich: Äpfel mit Birnen vergleichen); das Instrument wäre dann nicht zur Messung des Wissensstandes in unserer Studie geeignet. Im Rahmen der Vorlesung zur Testtheorie werden wir statistische Verfahren kennenlernen, die eine Überprüfung der Dimensionalität – und damit der Einsetzbarkeit eines Messinstrumentes – erlauben.

**„Testen“ vs. „befragen“.** Die Erfassung des Wissensstandes durch eine prüfungsähnliche Vorgangsweise ist ein Grundbaustein des psychologischen Tests. Ein solcher ist vorzugsweise für Merkmale geeignet, die im Sinne einer Fähigkeit verstanden werden können. Schwieriger ist hingegen die Erfassung von Merkmalen, die als Teil der Persönlichkeit aufgefasst werden; dazu zählen Eigenschaften wie z. B. Extraversion, Werthaltungen oder Interessen. Für solche Bereiche werden üblicherweise Fragebögen eingesetzt, in denen verschiedene Facetten des interessierenden Merkmals häufig auf abgestuften Antwortskalen (z. B. „stimme sehr/etwas/wenig/gar nicht zu“) einzuschätzen sind (zur Verrechnung solcher Antwortformate siehe Abschnitt 6.2.2 in diesem Buch). In unserem Beispiel könnten wir Fragen bezüglich der eigenen Einstellung zum Stoffgebiet und zur Unterrichtseinheit, bevorzugten Lernstilen oder Persönlichkeitsmerkmalen stellen.

Bei Selbsteinschätzungen dieser Art ist allerdings zu bedenken, dass sie auf das subjektive Empfinden (im Gegensatz zur objektivierbaren Leistung) abzielen. Der Wert der Antworten hängt daher auch vom nötigen Ausmaß an Selbstreflexion sowie der Bereitschaft ab, diese offenzulegen. Zudem kann eine Person – vor allem bei persönlich heiklen Themen oder wenn sie generell unsicher ist – auch dazu tendieren, sozial erwünschte Antworten zu geben. Wir erkennen schon anhand dieser kurzen Einführung, dass die Fragebogenmethode ebenfalls eine Reihe kritischer Fragen aufwirft.



**Der qualitative Zugang.** Es ist sehr wichtig, die Fragen, die in einer qualitativen Studie gestellt werden, im Vorfeld genau zu planen und zu erproben. Wenn man beispielsweise eine Fokusgruppe mit Studierenden durchführen möchte, sollte man sich überlegen, welche Impulsfragen in welcher Reihenfolge gestellt werden sollen. Die Frage „Was sind aus Ihrer Sicht die größten Vorteile von computerunterstütztem Unterricht?“ ist zum Beispiel ziemlich suggestiv, da sie nahelegt, dass es auf jeden Fall große Vorteile gibt. „Wie finden Sie computerunterstützten Unterricht?“ ist weniger suggestiv, aber etwas vage; dadurch lädt sie zu sehr kurzen Antworten („Gut!“) ein. Eine relativ neutrale Formulierung wäre „Wo sehen Sie Vorteile und Nachteile von computerunterstütztem Unterricht im Vergleich zu traditionellem Unterricht?“. Da die Frage komplex ist, könnte es aber passieren, dass die Studierenden im Eifer der Diskussion die ursprüngliche Frage vergessen und nur von positiven (oder auch nur von negativen) Auswirkungen von *blended learning* sprechen. Dann sollte die Person, die die Diskussion leitet, eine entsprechende Nachfrage (z. B. „Gibt es aus Ihrer Sicht auch Nachteile?“) in petto haben. Da es in unserer Studie speziell um Wissenserwerb und Lernmotivation geht, ist es wichtig, diese beiden Aspekte auch direkt anzusprechen. Man könnte die Studierenden beispielsweise bitten, über konkrete Erfahrungen zu berichten, welchen Lernstoff sie mit Hilfe von *blended learning* sehr gut erlernen konnten und wann das nicht so gut gelungen ist. Die gleiche Frage könnte auch für traditionelle Unterrichtsformen gestellt werden.

Wichtig ist es, bei der Erhebung qualitativer Daten möglichst objektiv vorzugehen. Wenn die Forschenden beispielsweise schon mit einer starken Überzeugung („*blended learning* ist super!“) ihre Studie beginnen, dann müssen sie sehr darauf achten, dass diese Überzeugung nicht die Ergebnisse der Studie verzerrt. Selbst wenn die Fragen neutral formuliert sind, könnte es passieren, dass die Diskussionsleitenden bewusst oder unbewusst die Antworten der Studierenden beeinflussen, beispielsweise indem sie auf positive Aussagen über *blended learning* mit einem Lächeln und auf negative Aussagen mit einem verkniffenen Gesichtsausdruck reagieren.

**Ein technischer Aspekt.** Heutzutage bedienen sich viele Studien online-basierter Techniken, bei denen Fragen am Bildschirm durch Anklicken einer Antwortoption oder Ausfüllen eines Freitextfeldes beantwortet werden. Der Vorteil einer solchen Methode besteht in der unmittelbar verfügbaren elektronischen Speicherung der Antworten, die sofort mit entsprechender Software qualitativ wie quantitativ weiterverarbeitet werden können. Die Befragten können (bei entsprechender Gestaltung) den Fragebogen auch am Mobiltelefon oder *tablet*-artigen Geräten ausfüllen, wodurch sie zeitlich und räumlich flexibler werden. Doch die mobile Handhabung kann auch dazu führen, dass die Fragen im dicht gedrängten Autobus oder in feucht-fröhlicher Runde beim Wirten um's Eck (wenn nicht gar auf dessen stillem Örtchen) beantwortet werden. Ob dies wertvolle Rückschlüsse zulässt, sei dahingestellt – wir wissen es schlicht nicht!

#### 5.2.4 Die Stichprobe

Nachdem wir die geeigneten Untersuchungsinstrumente ausgewählt haben, müssen wir uns einerseits überlegen, *wer* befragt werden soll (damit definieren wir die Samplingstrategie), und andererseits, *wieviele* Personen zur Beantwortung der Forschungsfrage erforderlich sind.

**Die Samplingstrategie.** In der Regel werden quantitative Studien mit dem Ziel durchgeführt, eine Aussage über eine sogenannte *Population* zu treffen – auf diese wird in Abschnitt 6.1.2 näher eingegangen. Da es nicht möglich ist, diese zu befragen, ziehen wir daraus eine *Stichprobe*. Wir müssen daher darauf achten, dass die Personen unserer Stichprobe einen Rückschluss auf die Population zulassen. Die Standardmethoden, die wir zur statistischen Auswertung unserer quantitativen Studien nutzen, gehen von der Annahme aus, dass eine sogenannte *Zufallsstichprobe* gezogen wurde. Eine solche zeichnet sich dadurch aus, dass alle Personen der Population die gleiche (oder eine bekannte) Wahrscheinlichkeit hatten, in unsere Stichprobe aufgenommen zu werden. Tatsächlich ist das Zufallsprinzip (im Sinne der

Ziehung aus einer Urne wie bei Fernsehshows) aus mathematischer Sicht die beste Methode, eine für statistische Auswertungen geeignete Stichprobe zu erhalten.

Leider ist diese aus theoretischer Sicht optimale Methode in der Praxis aus zwei zentralen Gründen sehr schwierig umzusetzen: Erstens können wir die metaphorische Urne nicht füllen, d. h. eine Liste aller Personen der Zielpopulation steht in der Regel nicht zur Verfügung. Es wurden jedoch gute Ersatzstrategien entwickelt (sog. komplexe Stichprobendesigns), die wir im Rahmen der Statistik I diskutieren werden (vgl. [Alexandrowicz, Bacher & Wancata, in Druck](#)). Zweitens ist auch bei geeigneter Annäherung an eine Zufallsstichprobe nicht gewährleistet, dass alle gezogenen Personen auch mitmachen. Bei Studien, die aus der Allgemeinbevölkerung ziehen, kann eine Zielperson verzo-gen oder verstorben sein. Auch wenn wir sie erreichen, kann sie die Teilnahme an der Studie ablehnen. Problematisch wird dieser Aspekt, wenn bestimmte Teilgruppen der Population eine grundsätzlich geringere Bereitschaft zur Teilnahme haben und es daher zu systematischen Ausfällen kommt. Man spricht bei einer solchen systematischen Verzerrung der Stichprobe von einem *Bias*, genauer: einem *Stichprobenbias* (englisch: *bias/biased* = verzerrt). Leider sind solcherart verzerrte Ergebnisse nicht erkennbar! Ebenso haben bestimmte Personen grundsätzlich höhere Bereitschaft zur Teilnahme an einer Untersuchung als andere. [Bortz und Döring \(2006\)](#) fassen Faktoren zusammen, die laut Studien zu erhöhter Teilnahmebereitschaft beitragen können. Sie nennen unter anderem: bessere schulische Ausbildung, höhere Intelligenz, mehr Bedürfnis nach sozialer Anerkennung, höhere Geselligkeit, Frauen eher als Männer, geringere autoritäre Orientierung oder Tendenz zu konformem Verhalten (S. 73). Wir werden im Rahmen der Statistik III Techniken kennenlernen, die es uns erlauben, bestimmte Formen der Nicht-Teilnahme auf statistischem Wege zu berücksichtigen.

Grundsätzlich problematisch sind triviale Ersatzstrategien: Wurde beispielsweise als Erhebungsinstrument eine Onlinebefragung gewählt, so könnte man den Link einfach an alle versenden, die im eigenen Adressbuch abgespeichert sind, evtl. flankiert vom Hinweis, den Link

beliebig weiterzuschicken. Auch wenn auf diese Art eine große Gruppe erreicht werden kann, so ist gänzlich unbekannt, wer aller die Chance zur Teilnahme gehabt hätte. Vielmehr ist davon auszugehen, dass das eigene Adressbuch schon eine sehr spezifische Auswahl aufweist – etwa Personen der eigenen Altersgruppe, Personen mit ähnlich gelagerten Interessen, bei Studierenden überwiegend andere Studierende, usw. Selbst wenn die Zielpopulation tatsächlich die Studierendenpopulation sein sollte, so ist immer noch von einer Vorselektion hinsichtlich Studienrichtung oder Studiensemester auszugehen. Auch die Platzierung des Links in Foren und „sozialen Medien“ bewirkt eine Vorauswahl von Personen, die diese Plattformen „bevölkern“ (ganz abgesehen vom Ausschluss aller, die keinen Zugang zum Internet haben). In diesem Zusammenhang wird bisweilen auch von einer „Freiwilligenstichprobe“ gesprochen, wobei die Bezeichnung „Selbstselektion“ (engl: *self-selected samples*) eher zutrifft, da Freiwilligkeit immer zu gewährleisten ist. Wainer (1986) zeigt auf, wie schwierig es ist, aus einer solchen Stichprobe sinnvolle Schlussfolgerungen zu ziehen und welche kapitalen Fehlschlüsse dabei möglich sind.

Gänzlich ungeeignet ist die Ausübung von Druck: Im Rahmen von Erhebungen an Studierenden könnte das Ausfüllen eines Fragebogens zur Bedingung für den Zeugniserwerb einer Lehrveranstaltung gemacht werden. Da hier keine freiwillige Teilnahme vorliegt, ist eine solche Vorgangsweise schon allein aus ethischen Gründen unzulässig. Hinzu kommt der Aspekt, dass ein Fragebogen, der unter Zwang auszufüllen ist, kaum verwertbare Antworten erwarten lässt (Das gleiche gilt im Übrigen für sogenannte „Pflichtfelder“, wie sie mit dem Aufkommen von Onlineerhebungen gerne eingesetzt werden – eine Person, die eine Frage nicht beantworten will, kreuzt bestenfalls irgendetwas an, schlimmstenfalls das Gegenteil dessen, was ihrer eigentlichen Antwort entspräche; mit beidem ist uns nicht geholfen).

In unserem Ankerbeispiel zu *blended learning* geht es um universitäre Lehre, womit sich die Zielpopulation auf Studierende beschränkt. Hier wäre sogar die Ziehung einer Zufallsstichprobe denkbar, da Universitäten über vollständige Listen aller Studierenden verfügen. Es ist

technisch problemlos möglich, aus dieser Liste tatsächlich eine Zufallsauswahl zu erstellen (wobei auch geeignete Vorkehrungen zur Wahrung des Datenschutzes zu treffen sind) – allerdings ist fraglich, ob eine Uni ein solches Vorgehen unterstützt. Da wir aber wissen, dass sich Studierende regelmäßig in relativ konstanten Gruppen in Lehrveranstaltungen einfinden, könnten wir *zufällig* aus dem Vorlesungsverzeichnis eine geeignete Anzahl von Lehrveranstaltungen ziehen und dort unser Erhebungsinstrument vorgeben (z. B. den Link zum Onlinefragebogen über eine Lernplattform verteilen). Die *willkürliche Auswahl* von Lehrveranstaltungen (z. B. solche, die man selbst besucht, die möglichst groß sind oder zu deren Vortragenden leicht Kontakt herzustellen ist) ist hingegen wieder problematisch, da so bestimmte Gruppen erneut systematisch ausgeschlossen werden. Wenn wir beispielsweise – der leichten Erreichbarkeit wegen – nur Psychologiestudierende untersucht haben, dann sind die Ergebnisse unserer Studie nicht notwendigerweise auf Studierende der Germanistik oder der technischen Mathematik übertragbar (gerade wer Psychologie studiert, setzt sich mit jenen Themen, die in psychologischen Studien untersucht werden, vertieft auseinander und unterscheidet sich damit bereits von anderen). Schränken wir hingegen das Thema weiter ein – etwa auf die Eignung von *blended learning* für entwicklungspsychologische Themen –, dann wäre die Befragung in einer einschlägigen Lehrveranstaltung dieses Faches durchaus zu überlegen.

Damit haben wir zwar die „Frage nach der geeigneten Urne“ behandelt, wir müssen aber immer noch berücksichtigen, dass manche einfach nicht mitmachen wollen. In unseren Fragestellungen nach der Auswirkung von *blended learning*-Methoden auf den Wissenserwerb und die Lernmotivation müssen wir beispielsweise davon ausgehen, dass generell hochmotivierte Personen auch eine systematisch höhere Teilnahmebereitschaft aufweisen – womit erneut eine Verzerrung der Ergebnisse nicht ausgeschlossen werden kann.

Wir sehen, die Ziehung einer tatsächlich belastbaren Stichprobe ist ein überaus schwieriges Unterfangen. Letztlich gibt es keine universell optimale Methode, sondern es gilt, Möglichkeiten zu erörtern und Vor-

und Nachteile jeder Variante gegeneinander abzuwägen; Kompromisse sind dabei praktisch unausweichlich (siehe hierzu auch Abschnitt 5.2.5). Der beste Schutz vor Unzulänglichkeiten einer Studie (nicht nur ihre Stichprobe betreffend) ist deren Replikation: Gelingt es, einen Befund an einer neuen, unabhängig (und am besten anders) gewonnenen Stichprobe zu replizieren, dann ist dies der stärkste Indikator dafür, tatsächlich ein bedeutsames Ergebnis gefunden zu haben. Abschließend vielleicht noch ein Wort der Beruhigung: Bei Bachelorarbeiten, die mehr im Zeichen des Trainierens von Forschungsfertigkeiten stehen, ist die Frage der Stichprobengewinnung sicherlich mit größerer Gelassenheit zu behandeln als im Rahmen von Studien, die auf eine Publikation in einer Fachzeitschrift abzielen.

Bei qualitativen Studien werden – wie bereits in Abschnitt 5.2.3 (S. 109) angesprochen – typischerweise weniger Personen befragt als bei quantitativen. Umso wichtiger ist es, sich gut zu überlegen, welche Personen man befragen möchte. Je nach gewählter qualitativer Methode gibt es dabei unterschiedliche Herangehensweisen. Ein Beispiel wäre das *theoretical sampling* (theoretische oder besser gezielte Stichprobenzusammenstellung), bei dem die Personenauswahl auf theoretischen Vorüberlegungen beruht. Wir könnten beispielsweise versuchen, für unsere Studie einerseits sehr computer- und internetaffine Personen zu finden, andererseits Personen, die mit Computern wenig anfangen können. Vermutlich haben diese beiden Gruppen sehr unterschiedliche Sichtweisen auf *blended learning*. Wir könnten auch Studierende mit unterschiedlichen Vorkenntnissen im Fachgebiet, das in unserem Experiment unterrichtet wird, oder Studierende aus unterschiedlichen Studienrichtungen auswählen.

**Der angestrebte Stichprobenumfang.** Zur Stichprobengröße geistern eine Reihe von Faustregeln herum, die allesamt heuristisch bzw. nur selten inhaltlich begründet sind. Tatsächlich stehen uns jedoch – bei quantitativer Herangehensweise – sehr konkrete Techniken zur Verfügung, die uns auf Basis einer inhaltlichen Begründung und bestimmter wahrscheinlichkeitstheoretischer Annahmen eine optimale

Stichprobengröße festzulegen erlauben. Wichtig in diesem Zusammenhang ist der Hinweis, dass die Planung der Stichprobengröße die Ziehung einer Zufallsstichprobe voraussetzt. Wie das alles im Detail funktioniert, wird Thema der Statistik III.

Im qualitativen Ansatz lässt sich diese Methode zur Festlegung des optimalen Stichprobenumfangs nicht einsetzen, da die technischen Voraussetzungen nicht gegeben sind und auch das Erkenntnisziel ein anderes ist (vgl. Abschnitt 5.2.5, S. 130) – hier streben wir ja keine mathematische Abstraktion, sondern eine inhaltliche Strukturierung beispielsweise der Aussagen aus den Fokusgruppeninterviews an.

**Ethische Aspekte.** Unabhängig von der Methodik der Studie ist es wichtig, Studien nach ethischen Gesichtspunkten zu planen. Die Teilnahme an psychologischen Studien ist immer freiwillig, und die Teilnehmenden werden im Regelfall vorher über die Zielsetzungen der Studie aufgeklärt. In manchen Fällen mag es unabdingbar sein, das eigentliche Thema zu verschleiern oder Versuchspersonen hinsichtlich des Studienziels zu täuschen, da sonst ein zu untersuchen beabsichtigter Aspekt nicht mehr erfasst werden kann – eine solche Vorgangsweise ist jedoch nur zu vertreten, wenn der Täuschung auch ein adäquater Erkenntnisgewinn gegenübersteht. Weiters müssen die Daten, wenn irgend möglich, in anonymisierter Form ausgewertet und an einem sicheren Ort gespeichert werden. An unserem Institut werden alle geplanten Studien durch einen eigenen Ethikrat begutachtet, um die Einhaltung ethischer Richtlinien sicherzustellen.

### 5.2.5 Die Auswertung und Ergebnisdarstellung

In quantitativen Studien kommen zur Auswertung in der Regel einschlägige Statistikprogramme zum Einsatz. Hier sind vor allem **R** und **SPSS** zu nennen, die in den Sozialwissenschaften sehr verbreitet sind und auch in den verschiedenen Statistikveranstaltungen (v.a. den Statistik-Übungen) unterrichtet werden. Bei qualitativen Studien werden inhaltsanalytische Auswertungen durchgeführt, wofür ebenfalls

Softwarepakete wie etwa [MaxQDA](#) oder [QCAmap](#) zur Verfügung stehen.

### Quantitative Auswertungen

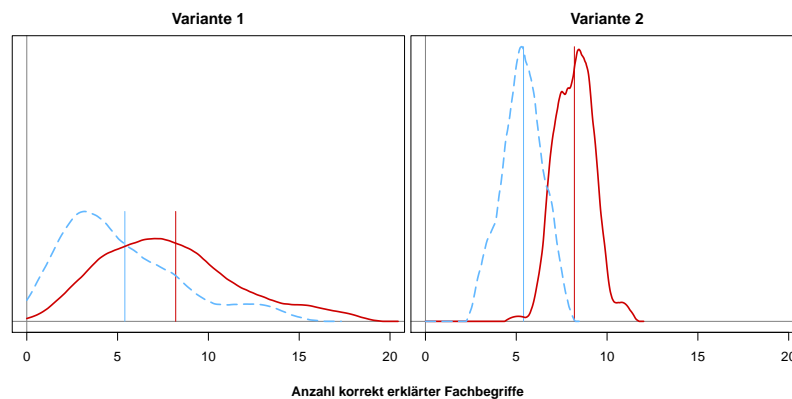
Hier wollen wir an zwei Beispielen demonstrieren, wie eine quantitative Auswertung aussehen könnte.

**Die statistische Hypothesenprüfung.** Wir wollen Fragestellung 1 betrachten, in der es darum ging, ob sich mittels CAI der Wissensstand steigern lässt. Als (sehr einfaches) Maß für den Wissensstand hatten wir für jede teilnehmende Person die Anzahl korrekt zugeordneter Fachbegriffe gewählt. Als Design hatten wir den Vergleich einer Versuchsgruppe, die unter Einbindung von Elementen des *blended learning* unterrichtet wurde mit einer Kontrollgruppe nach klassischer Unterrichtsmethode gewählt. Zur Beantwortung formulieren wir die Forschungsfrage in eine statistische Hypothese um. Wir könnten beispielsweise im Rahmen der Literaturrecherche Hinweise auf die besondere Wirksamkeit von *blended learning* gefunden haben. Dann formulieren wir als Hypothese 1

$H_1$ : Die VG erzielt eine höhere durchschnittliche Punktzahl als die KG.

Nun könnten wir beispielsweise beobachtet haben, dass in der Versuchsgruppe im Durchschnitt 7.2 von 20 Fachbegriffen korrekt zugeordnet wurden und in der Kontrollgruppe nur 5.4. Damit schneidet in unserer Stichprobe die VG im Mittel besser ab als die KG. Doch können wir daraus auch schließen, dass *blended learning* generell (also über unsere Stichprobe hinaus) überlegen ist? Nicht notwendigerweise, denn wir müssen auch die Verteilung der Werte berücksichtigen. [Abbildung 5.1](#) zeigt zwei mögliche Ergebnisvarianten, die beide einen Mittelwertsunterschied von 2.8 Punkten aufweisen. Wir sehen, in der linken Variante sind beide Gruppen breit gestreut und überlappen sehr stark, der Mittelwertsunterschied von 2.8 Fachbegriffen nimmt sich daher gering aus. Im Gegensatz dazu sind die Verteilungen in der Variante 2 deutlich schmaler, der gleiche Mittelwertsunterschied





**Abbildung 5.1:** Zwei mögliche Varianten, wie die Punkteverteilungen von KG (blau strichliert) und VG (rot durchgezogen) aussehen könnten. Die senkrechten Linien zeigen die Mittelwerte der beiden Gruppen; beide Varianten haben dieselben Gruppenmittelwerte.

scheint daher in diesem Fall bedeutsamer. Zur sinnvollen Beurteilung eines Mittelwertsunterschiedes muss also immer die Streuung innerhalb der beiden Gruppen angegeben werden.

Neben der Heterogenität (Streuung) der Werte pro Gruppe spielt auch der Stichprobenumfang bei der Beurteilung eines Mittelwertsunterschiedes eine Rolle. Untersuchen wir beispielsweise nur 15 Personen pro Gruppe, dann würde ein einzelner „Superstar“ in der KG deren Gruppenmittelwert stark ansteigen lassen (und wir würden dann keinen nennenswerten Mittelwertsunterschied mehr beobachten) – bei 150 hingegen würde sich ein einzelner stark atypischer Wert nicht merklich auf den Mittelwert auswirken. Deshalb wird zur statistischen Beurteilung eines Mittelwertsunterschiedes neben der Streuung auch der Stichprobenumfang berücksichtigt.

Im Kern basiert diese Beurteilung auf einer Vorgangsweise, die die Wahrscheinlichkeitsrechnung miteinbezieht. Der Vorgang lässt sich wie folgt skizzieren: Wir haben zuvor in der  $H_1$  einen zu prüfenden Unterschied (Mittelwert VG größer als Mittelwert KG) formuliert. Was wäre, wenn dieser Unterschied nicht vorliegt, die beiden Gruppen also

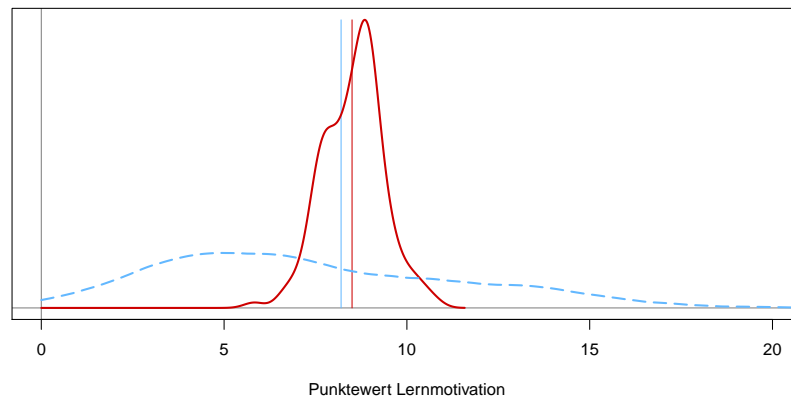
im Grunde gleich leistungsfähig – also konventioneller und *blended learning* Unterricht gleich wirksam – wären? Es wäre nicht zu erwarten, dass die Mittelwerte der beiden Gruppen dann exakt gleich groß wären, die Zufallsauswahl von Personen und deren randomisierte Zuordnung zu Gruppen bewirkt nicht ein exaktes „Ausnivellieren“, irgendein Unterschied wird immer auftreten. Die Frage ist daher, ist der beobachtete Unterschied (nach Berücksichtigung von Streuung und Gruppengröße) groß genug, um einen generellen Unterschied in der Wirksamkeit der beiden Unterrichtsmethoden anzunehmen? Zu diesem Zweck wird der oben formulierten Hypothese eine sogenannte Nullhypothese gegenübergestellt, die davon ausgeht, dass es eigentlich keinen Unterschied zwischen den beiden Gruppen gibt:

$H_0$ : Die beiden Gruppen erzielen die gleiche durchschnittliche Punktezahl.

Mittels statistischer Testverfahren kann man nun feststellen, ob der von uns gefundene Unterschied zwischen den beiden Mittelwerten statistisch aussagekräftig („signifikant“) ist. Dabei wird errechnet, wie wahrscheinlich es wäre, den beobachteten oder einen noch größeren Mittelwertsunterschied – unter Berücksichtigung von Stichprobenumfang und Streuung – rein zufällig zu finden, auch wenn in Wirklichkeit die Nullhypothese zutreffen würde<sup>4</sup>. Diese Wahrscheinlichkeit lässt sich errechnen und als Entscheidungsgrundlage zwischen  $H_0$  und  $H_1$  heranziehen: Ist sie (zu) klein, dann wäre es unwahrscheinlich, den beobachteten Mittelwertsunterschied zu erhalten, obwohl eigentlich kein tatsächlicher Unterschied zwischen den beiden Lehrmethoden existiert. Das spricht gegen die Nullhypothese, die wir daraufhin verwerfen. Ab wann die Wahrscheinlichkeit „zu“ klein ist und viele weitere Details und Varianten dieser Vorgangsweise werden wir im Rahmen der Statistik II diskutieren.

---

<sup>4</sup> Zum Vergleich: auch eine faire Münze wird bei 10 Würfeln nicht immer genau 5 Mal Kopf und 5 Mal Zahl zeigen – probieren Sie es aus! Gäbe es diese natürliche Variation nicht, dann gäbe es auch kein Glücksspiel.



**Abbildung 5.2:** Beispiel für zwei unterschiedlich stark streuende Gruppen. KG: blau trichliert, VG: rot durchgezogen; Die senkrechten Linien zeigen die Mittelwerte der beiden Gruppen.

**Kontrolle von Drittvariablen.** In Abschnitt 5.2.3 (S. 113) hatten wir die Möglichkeit kennengelernt, den Einfluss von Hintergrundvariablen mitzubedenken. Es könnte sich beispielsweise in Fragestellung zwei zur Lernmotivation herausgestellt haben, dass sich die Mittelwerte von VG und KG nicht in relevantem Ausmaß unterscheiden. Eine nähere Betrachtung der Werte hätte jedoch ergeben, dass die Kontrollgruppe (konventioneller Unterricht) wesentlich stärker streuende Werte aufweist als die Versuchsgruppe (vgl. Abbildung 5.2)

Offensichtlich berichten manche Studierende bei traditionellem Unterricht sehr hohe und gleichzeitig andere sehr niedrige Lernmotivation. Dieses Ergebnis lässt sich anhand der Daten des Experimentes zunächst nicht direkt erklären – aber möglicherweise durch die Methode des Herauspartialisierens: Wenn in der Studie beispielsweise auch Persönlichkeitseigenschaften gemessen wurden, dann könnten wir jetzt überprüfen, ob die Lernmotivation in der Kontrollgruppe vielleicht mit diesen zusammenhängt. Vielleicht berichten eher extravertierte Personen, die in Vorlesungen gerne Fragen stellen und Wert auf die direkte Interaktion mit den Vortragenden legen, im traditionellen

Lehrformat eine hohe Lernmotivation, während eher introvertierte Personen sich durch dieses Format weniger motiviert fühlen.

### Qualitative Auswertungen: Identifikation von Themen

In unseren Fokusgruppen wurden die Studierenden gefragt, welche Vor- und Nachteile sie bei *blended learning* im Vergleich zu traditionellen Lehrformaten sehen. Die Gruppendiskussionen wurden transkribiert und liegen nun in schriftlicher Form vor. Es gibt viele verschiedene Möglichkeiten, solche Daten auszuwerten; allen gemeinsam ist, dass man die angesprochenen Themen identifiziert. Ähnlich wie bei der Theoriebildung können wir auch hier deduktiv oder induktiv vorgehen. Bei der deduktiven Vorgangsweise werden bereits im Vorfeld auf Basis theoretischer Überlegungen bestimmte Themen festgelegt, nach denen man dann gezielt in den Transkripten sucht. Wir könnten beispielsweise, wie zu Beginn angesprochen, erwarten, dass viele Studierende einen Vorteil des *blended learning* in der Möglichkeit stärkerer Selbstbestimmtheit beim Lernen sehen, und dann gezielt nach entsprechenden Aussagen suchen. Bei der (häufiger verwendeten) induktiven Vorgangsweise „lassen wir die Daten sprechen“, indem wir ohne bestimmte Vorannahmen die Aussagen der Studierenden in Themen zusammenfassen. Vielleicht würden sie weniger über ihre Kenntnis von Fachbegriffen sprechen als darüber, wie sie die direkte Interaktion mit den Lehrenden empfunden haben und wodurch sie zum Lernen motiviert wurden. Ob *blended learning* dem traditionellen Unterricht vorzuziehen ist, wäre somit vor allem davon abhängig, wie die Vergleichsbedingung des traditionellen Unterrichts genau aussieht. Bezüglich des computerunterstützten Unterrichts würde sich vielleicht zeigen, dass den Studierenden bestimmte Aktivitäten am Computer Spaß machen, sie andere aber als langweilig empfinden.

Wichtig ist es bei der Auswertung qualitativer Daten, nach klaren und nachvollziehbaren Regeln vorzugehen. Bei den meisten Methoden ist die Objektivität der Auswertung ein explizites Ziel: die Ergebnisse der Auswertung sollten nicht davon abhängig sein, wer die Auswer-

tung durchgeführt hat. Es muss also genau überlegt und dokumentiert werden, nach welchen Regeln die Textstellen zur Identifikation von Themen ausgewählt wurden: beispielsweise sollte man nur jene Stellen auswählen, in denen die Studierenden explizit über Motivation gesprochen haben und nicht jene, in denen sie allgemein über Präferenzen gesprochen haben („ich mag einfach keine Powerpoint-Vorträge“). Nach deren Identifikation sollen die einzelnen Themen möglichst genau beschrieben werden. Idealerweise sollten diese Beschreibungen so genau sein, dass andere Forschende in der Lage wären, ähnliche Daten in Bezug auf die gleichen Themen auszuwerten. Sie sollten konkrete Zuordnungsregeln und Textbeispiele enthalten. Eine solche Beschreibung könnte etwa lauten:

Thema „Kontakt mit den Vortragenden“: Alle Aussagen, die sich darauf beziehen, dass die Befragten (a) gerne direkt mit den Lehrenden sprechen (z. B. „Ich finde es gut, wenn man direkt Fragen stellen und Antworten bekommen kann“, „Ich kann dann nachher noch hingehen und fragen, wenn mir etwas unklar ist“), oder (b) den Beziehungsaufbau durch die gemeinsame Arbeit schätzen (z. B. „Ich mag es, wenn so eine gemeinsame Arbeitsatmosphäre entsteht“, „Ich finde es gut, wenn Lehrende auch Humor haben und wir gemeinsam lachen“). Nicht zuzuordnen sind Aussagen über die Persönlichkeit der Lehrenden (z. B. „Ich mag Vortragende, die jeden Inhalt spannend darstellen können“).

Bei der Entwicklung solcher Regeln ist es wichtig, möglichst objektiv vorzugehen. Gerade wenn die Forschenden sehr gerne zeigen möchten, dass *blended learning* die „bessere“ Unterrichtsform ist, ist es wichtig, bewusst auf Objektivität zu achten und auch negative Aussagen einzubeziehen.

Qualitative Methoden analysieren also keine Zahlen, sondern typischerweise verbale Daten in Textform, wie hier die Aussagen der Studierenden. Auch visuelle Daten, wie beispielsweise die Gesichtsausdrücke der Beteiligten an der Gruppendiskussion (wenn eine Video-

aufnahme erstellt wurde), können qualitativ ausgewertet werden. In den Lehrveranstaltungen zu qualitativen Methoden und speziell zur Textanalyse werden Sie verschiedene Methoden der Auswertung qualitativer Daten kennenlernen.

### Die Ergebnisdarstellung

Die Ergebnisse der Auswertungen werden in Fachartikeln oder wissenschaftlichen Abschlussarbeiten in einem eigenen Abschnitt dargestellt, in dem die einzelnen Forschungsfragen nacheinander beantwortet werden. Dabei können Abbildungen und Tabellen zur Veranschaulichung verwendet werden. Allerdings steht es uns nicht frei, einen Artikel nach Gutdünken zu gestalten, vielmehr arbeiten wir nach dem von der [American Psychological Association \(APA\) \(2010\)](#) herausgegebenen Manual. Einige der dort verankerten Richtlinien sind in diesem Kapitel bereits wiedergegeben, eine Reihe weiterer werden Sie in verschiedenen Lehrveranstaltungen kennenlernen. Die Motivation für eine solche Richtlinie besteht darin, dass durch die dort vorgenommene Standardisierung von Artikeln einerseits nichts Wichtiges „übersehen“ werden kann, und es andererseits Forschenden leichter gemacht wird, sich rasch einen Überblick über die Unmenge an Beiträgen zu verschaffen, die praktisch täglich neu erscheinen. Müssten wir uns erst in der individuellen Struktur eines jeden Artikels zurechtfinden, wäre die Sichtung der zahllosen Beiträge kaum bewältigbar.

**Ergebnisse der quantitativen Studie.** Bei der Darstellung der Ergebnisse quantitativer Studien werden die relevanten deskriptivstatistischen („beschreibenden“) Kennwerte (etwa Mittelwerte und Streuungsmaße in den verglichenen Gruppen) sowie die Kennwerte der betreffenden Hypothesentests angegeben. Wichtig ist es, in jedem Fall alle relevanten Kennwerte auch anzugeben – das ermöglicht die Berücksichtigung einer Studie in einer späteren Metaanalyse (s. Abschnitt 5.2.3, S. 113). Eine solche Metaanalyse könnte beispielsweise prüfen, ob sich ein allfälliger Effekt von *blended learning* zwischen

schulischer und universitärer Lehre unterscheidet (etwa aufgrund der größeren Selbstständigkeit, derer es auf der Uni bedarf). Nicht nur als Basis für Metaanalysen, sondern in jedem Fall ist es essentiell, in der Ergebnisdarstellung auch die Samplingstrategie (s. Abschnitt 5.2.4) detailliert offenzulegen, um so das Lesepublikum in die Lage zu versetzen, sich selbst ein Bild über die Angemessenheit der dargestellten Resultate und Schlussfolgerungen machen zu können.

### Exkurs: Die Replikationsdebatte

In den letzten Jahren ist insbesondere die experimentelle Psychologie zunehmend in die Kritik geraten, weil sich herausstellte, dass die veröffentlichten Ergebnisse mancher, oft recht spektakulärer Experimente nicht in gleicher Form gefunden wurden, wenn andere Forschungsteams die betreffende Studie genauso durchführten, wie es in den Fachartikeln beschrieben wurde. Ein Grund dafür ist vermutlich der sogenannte *publication bias*: Studien werden eher dann veröffentlicht, wenn die Ergebnisse besonders spektakulär sind. Uninteressante Ergebnisse (bei denen z. B. kein Gruppenunterschied gefunden wird) laufen hingegen Gefahr „schubladiert“ zu werden, was zu einer Verzerrung des Wissensstandes über das betreffende psychologische Phänomen führt. Deshalb bedarf es der Replikation eines Befundes, um auszuschließen, dass es sich nur um einen Zufallsfund handelt (etwa bedingt durch einen *bias*, wie in Abschnitt 5.2.4 dargestellt). Um das Risiko, Unspektakuläres nicht zu berichten, zu minimieren, gibt es mittlerweile die Möglichkeiten, geplante Studien vorab offiziell registrieren zu lassen, sodass nachvollzogen werden kann, ob eine Studie wegen mangelndem Interesse (oder unerwünschter Resultate!) nicht publiziert wurde. (vgl. hierzu [dieses Interview](#)).

**Ergebnisse der qualitativen Studie.** Da qualitative Ergebnisse in Textform vorliegen, eignen sie sich kaum für eine numerische Darstellung, abgesehen vielleicht von Nennungshäufigkeiten der identifizierten Kategorien. In der Regel geht es eher darum, die Einzigartigkeit unterschiedlicher Sichtweisen verschiedener Personen herauszuarbeiten. Typischerweise werden allgemeine Definitionen der einzelnen identifizierten Themen gegeben und diese dann anhand von Beispielaussagen illustriert. Das könnte beispielsweise für das oben beschriebene Thema so aussehen:

„Kontakt mit den Vortragenden“: Der am häufigsten genannte Vorteil traditionellen Unterrichts war der persönliche Kontakt mit den Vortragenden. Hier sprachen sieben der 24 Befragten an, dass sie gerne mit den Vortragenden persönlich sprechen würden, etwa indem sie in der Vorlesung oder danach Fragen stellten. Eine Teilnehmerin sagte beispielsweise, „Ich finde es gut, wenn man direkt Fragen stellen und Antworten bekommen kann. Wenn das nicht möglich ist und ich etwas nicht verstehe, bleibe ich hängen, weil ja das Weitere dann oft darauf aufbaut.“ Ein Teilnehmer gab an, dass er sich ungern direkt in der Vorlesung zu Wort melde, aber „Ich kann dann nachher noch hingehen und fragen, wenn mir etwas unklar ist“. Es gab aber auch insgesamt drei Personen, die gerade den persönlichen Kontakt als potentiell unangenehm bezeichneten, z. B. sagte ein Teilnehmer, „Wenn man dann plötzlich etwas gefragt wird, das finde ich nicht gut, ich bekomme dann immer so einen Schreck, dass ich gar nicht denken kann.“

Natürlich stellt die Zusammenfassung der Ergebnisse einer größeren Studie mit vielleicht hunderten Seiten auf sehr wenige Seiten Ergebnisabschnitt im Artikel eine große Herausforderung dar. Wie kann man einerseits die Individualität und Unterschiedlichkeit der individuellen Perspektiven beibehalten, andererseits aber die Komplexität der Ergebnisse auf einen guten Gesamtüberblick reduzieren? Hinzu kommt noch der Anspruch der möglichst objektiven, also nicht durch



die eigenen Ziele der Forschenden verfälschten Darstellung. Eine wirklich gute qualitative Ergebnisdarstellung kann – wenn die Ergebnisse kompliziert sind – ausgesprochen schwierig sein.

### 5.2.6 Die Diskussion

Nach der neutralen Darlegung der Ergebnisse folgt abschließend der wichtigste Schritt eines Forschungsprojektes, die Diskussion. Während in der Ergebnisdarstellung nüchtern alle Resultate dargelegt wurden, erfolgt in der Diskussion deren Bewertung und Einbettung in bestehendes Wissen und den theoretischen Rahmen. Wie passen unsere Ergebnisse zu denen früherer Studien? Wie lassen sich unterschiedliche Ergebnisse erklären; was trägt unsere Studie zum allgemeinen Wissensstand bei? Vielleicht decken sich die Ergebnisse unserer experimentellen Studie im Allgemeinen mit denen anderer Studien, was den Wissenserwerb betrifft, aber die Erkenntnisse zur Lernmotivation, gerade im Zusammenhang mit Persönlichkeitseigenschaften, sind neu. Oder unsere Ergebnisse zum Wissenserwerb decken sich nur mit einem Teil der anderen Studien – dann könnte man Überlegungen darüber anstellen, ob das damit zu tun hat, wie wir den Wissenserwerb gemessen haben.

Neben der Einbettung in bestehendes Wissen bietet die Diskussion auch Raum für weiterführende Überlegungen: Zum Zeitpunkt der Erstellung sind wir im exklusiven Besitz der neuen Erkenntnisse, die unsere Studie hervorgebracht hat. Es ist uns daher an dieser Stelle auch gestattet, eigene Überlegungen hinsichtlich deren Bedeutung auszudrücken und die neuen Erkenntnisse zu bewerten.

#### **Exkurs: Schule versus Uni (II)**

Erfahrungsgemäß bereitet in den ersten eigenen Arbeiten die Erstellung der Diskussion die größten Schwierigkeiten. Dies mag daran liegen, dass im schulischen Kontext der Aspekt der Einbettung eines Befundes in einen wissenschaftlichen Kontext nicht unbedingt im Vordergrund

steht. Daher enden Schulaufsätze häufig mit einer „Zusammenfassung“. Es sei ganz explizit festgehalten: Der Diskussionsteil einer wissenschaftlichen Arbeit ist keine Zusammenfassung. Vielmehr ist es die Einbettung des „Neuen“ in das „Alte“ bzw. eine Gegenüberstellung des „Gefundenen“ mit dem „Bekanntem“. Die Diskussion ist die Darlegung der sich daraus ergebenden Gemeinsamkeiten (z. B. im Sinne einer Erhärtung einer Theorie) und Unterschiede (z. B. Widersprüche zu den Erkenntnissen anderer Forschungsarbeiten oder Unerwartetem im Kontext einer bestehenden Theorie). Das mag zu Beginn des Studiums schwierig – jedenfalls aber ungewohnt – erscheinen, doch keine Sorge, Sie werden in zahlreichen Lehrveranstaltungen an die Feinheiten einer überzeugenden Diskussion herangeführt.

### 5.3 Schlussbetrachtung

Mit den hier vorgestellten Begriffen und Beispielen sollte ein grober (und keineswegs vollständiger) Überblick darüber vermittelt werden, welche Vielfalt an Herangehensweisen uns zur Verfügung steht, und wo eine Grenze zur nicht (mehr) wissenschaftlichen Herangehensweise gezogen werden kann. Es wurde ferner versucht herauszuarbeiten, dass die Wahl der Methode massive Auswirkung auf das Spektrum möglicher Erkenntnisse haben kann und daher keineswegs leichtfertig (oder ideologisch geleitet) vorgenommen werden darf. Es sollte vermittelt werden, dass Forschungsfrage und Forschungsmethode eng verzahnt sind, da sich die Methode aus der Fragestellung ergibt. Wir haben an einigen ausgewählten Beispielen gesehen, dass sich die diffuse Frage, wie sich *blended learning* auswirkt, in Form unterschiedlicher Fragestellungen ausformulieren lässt, die jede mit einem anderen Studiendesign sinnvoll beforscht werden kann und die in der Folge auch zu unterschiedlichen Erkenntnissen führen wird.

Im Rahmen der Ausbildung in der Methodenlehre ist es uns ein großes Anliegen, dass Studierende fundiertes Wissen über ein breites Spek-

trum möglicher Methoden erwerben. Wir müssen dabei zwei Ebenen berücksichtigen: Auf der Metaebene ist zu beurteilen, ob eine gewählte Methode überhaupt geeignet ist, eine Forschungsfrage adäquat zu beantworten, und auf der Methodenebene ist zu beurteilen, ob die gewählte Methode auch *lege artis* angewendet wurde. Nur dann sind wir in der Lage, selbst wissenschaftlich sinnvoll zu arbeiten und den tatsächlichen Gewinn publizierter Ergebnisse (z. B. auch im Rahmen eines Reviewprozesses) zu beurteilen.



## Literaturverzeichnis

- Akademien der Wissenschaften Schweiz. (2008). *Wissenschaftliche Integrität: Grundsätze und Verfahrensregeln*. Zugriff auf [www.akademien-schweiz.ch/dms/D/Publikationen/Richtlinien\\_Empfehlungen/Wiss\\_Integritaet/Richtlinien.pdf](http://www.akademien-schweiz.ch/dms/D/Publikationen/Richtlinien_Empfehlungen/Wiss_Integritaet/Richtlinien.pdf)
- Alexandrowicz, R. W., Bacher, J. & Wancata, J. (in Druck). Sampling and Weighting of the Austrian Psychiatric Prevalence Survey (APPS). *Neuropsychiatrie*. doi: 10.1007/s40211-019-0305-6
- American Psychological Association (APA). (2010). *Publication Manual of the American Psychological Association* (6. Aufl.). Washington, DC: American Psychological Association.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer.
- Knowles, M. (1975). *Self-Directed Learning*. Chicago: Follet.
- Knowles, M. (1984). *Andragogy in Action*. San Francisco: Jossey-Bass.
- Mann, B. L. (2009). Computer-Aided Instruction. In B. W. Wah (Hrsg.), *Wiley encyclopedia of computer science and engineering*. NY: Wiley.
- Moskal, P., Dziuban, C. & Hartman, J. (2013). Blended learning: A dangerous idea? *Internet and Higher Education*, 18, 15–23.
- Stefaner, M., Taraborelli, D. & Ciampaglia, G. L. (2011). *Notabilia – Visualizing Deletion Discussions on Wikipedia*. Zugriff auf <http://notabilia.net>
- Wainer, H. (1986). *Drawing Inferences from Self-Selected Samples*. NY: Routledge.



# Index

- article processing charges; APC, 106
- Asch, Solomon, 111
- Berliner Erklärung, 105
- bias, 121, 133
- Bibliothek, 105
- blended learning, 100
- case-study, 116
- Cochrane Library, 105
- computer aided instruction; CAI, 107
- Datenschutz, 125
- Deduktion, 101, 130
- Dimensionalität, 117
- double blind, 115
- double blinded review, 104
- editor, 103
- Einzelfallstudie, 116
- empirische Wissenschaft, 100
- ethische Aspekte, 125
- Experiment, 113
- Feldstudie, 114
- Fokusgruppe, 110
- Fragestellung, 108
- Google Scholar, 105
- Gruppendiskussion, 110
- Gruppendruck, 111
- Herauspartialisieren, 115, 129
- Heterogenität, 127
- Hypothesenprüfung, 126
- Induktion, 102, 130
- Kontrollgruppe (KG), 113
- Längsschnittuntersuchung, 116
- Literaturrecherche, 103
- MaxQDA, 126
- medline, 105
- mehrdimensional, 118
- Meta-Analyse, 116, 132
- Mittelwertsunterschied, 126
- mixed methods, 112
- Nullhypothese, 128
- Objektivität, 131
- online-Erhebung, 120
- open access, 105
- Operationalisierung, 107
- Placebo, 115
- predatory journals, 106
- psychARTICLES, 105
- psychINFO, 105
- psychologischer Test, 112

## Index

---

- psychologisches Experiment, 113
- publication bias, 133
- publication fee, 106
- PubMed, 105
  
- QCAmap, 126
- qualitative Methoden, 110
- quantitative Methoden, 109
  
- R, 125
- Randomisierung, 114
- Raubtierjournale, 106
- registrierte Studien, 133
- Replikationsdebatte, 133
- review, 103
- reviewer, 103
  
- Selbsteinschätzung, 118
- signifikant, 128
- SPSS, 125
- Stichprobe, 120
- Stichprobenumfang, 124
- Streuung, 126, 127
- Studiendesign, 113
- Suchmaschinen, 104
- Summenbildung, 117
  
- Täuschung, 125
- Teilnahmebereitschaft, 121
- Testtheorie, 117
- Textanalyse, 132
- theoretical sampling, 124
- Transkription, 110
  
- Universitätsbibliothek, 105
  
- Versuchsgruppe (VG), 113
- Verteilung, 126
- Vertiefungsmodell, 112
- Verum, 115
- Verzerrung, 121
  
- Vorstudienmodell, 112
- Wahrscheinlichkeitsrechnung, 127
- Wikipedia, 104
  
- Zufallsstichprobe, 125



