

DATA LITERACY ESSENTIALS

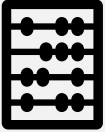
HWS 2021/22

Data Literacy Essentials: Daten erheben

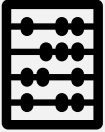
Dipl.-Soz. Markus Herklotz

[Webseite zu den Data Literacy Essentials](#)

Outline

	Forschungsinteresse
	Datentypen und Erhebungsmethoden
	Auswahlverfahren
	Fragebogendesign
	Resümee: Potentiale und Grenzen

Outline

	Forschungsinteresse
	Datentypen und Erhebungsmethoden
	Auswahlverfahren
	Fragebogendesign
	Resümee: Potentiale und Grenzen

Welche Faktoren beeinflussen in welchem Maße die Effektivität der Online-Lehre bei Studierenden der Uni Mannheim?

- Die **Empirie** erfordert eine **solide theoretische Basis**, dafür müssen **Probleme formuliert, Begriffe definiert** und **Hypothesen aufgestellt** werden.
- Darauf aufbauend folgt die **Operationalisierung**, also die **Messbarmachung** der in der Hypothesen enthaltenen Sachverhalte.
- Insbesondere Begriffe wie hier „Effektivität“, die sehr unterschiedlich ausgelegt werden können, erfordern eine genaue Definition und **Sorgfalt** bei der Operationalisierung.
- Diesen Prozess verkürzen wir hier auf **beispiel- und ausschnittshafte Anwendungsfälle**.
- *Vertiefung: Häder (2019), Kapitel 3.*

Anstoß für das Untersuchungsdesign

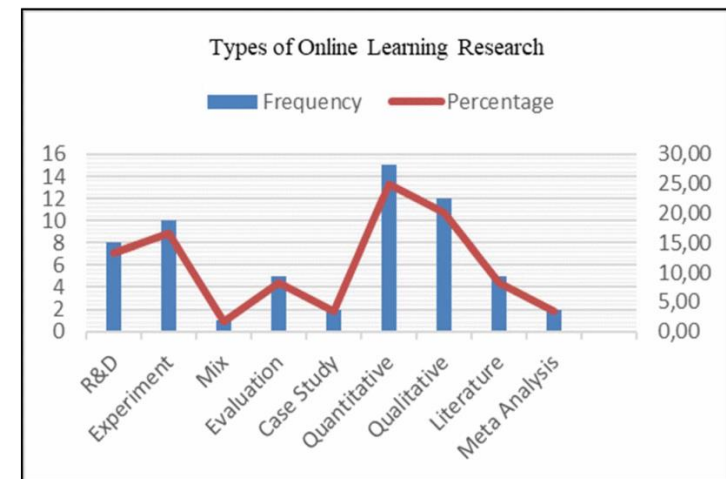


HWS 2021/22

Meta-Analysis of Online Learning Implementation in Learning Effectiveness

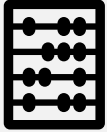
<https://ieeexplore.ieee.org/abstract/document/9276557>

- **meta-analysis** of 60 research articles on online learning
- The effectiveness and quality of online learning **are influenced by several factors**, including
 - the quality of teachers in making **online learning materials** and compiling online learning **guidelines**,
 - **tools used** in online learning such as **electronic media** and **internet network support**, and
 - **student behavior** in implementing online learning.



D. Prestiadi, Maisyaroh, I. Arifin and A. N. Bhayangkara
(2021)

Outline

	Forschungsinteresse
	Datentypen und Erhebungsmethoden
	Auswahlverfahren
	Fragebogendesign
	Resümee: Potentiale und Grenzen

Quantitativer und qualitativer Ansatz

- **Quantitative Forschung**

- Meist große Anzahl von Untersuchungsobjekten
- Numerische Beschreibung (Quantifizierung) von inhaltlichen Dimensionen
- Eher standardisierte, geschlossene Instrumente
- Weniger Exploration, sondern Überprüfung von bestehenden Theorien und Ansätzen
- Auswertung von statistischen Kennzahlen und entsprechender statistischer Methoden im Vordergrund

- **Qualitative Forschung**

- Eher geringere Zahl von Untersuchungsobjekten mit Fokus auf inhaltlicher Tiefe
- Verbalisierung von inhaltlichen Dimensionen, bspw. einzelne Motive, Werte, Gefühle und Gründe
- Häufig zur Exploration von Themen- und Forschungsgebieten
- Eher offene, flexible Erhebungsinstrumente wie Gesprächsleitfäden
- Statistische Kennzahlen nicht im Vordergrund

Braunecker (2021), S. 22ff.

Poll:

Würden Sie unserem Forschungsinteresse „Welche Faktoren beeinflussen in welchem Maße die Effektivität der Online-Lehre bei Studierenden der Uni Mannheim?“ mit einem qualitativen Ansatz nachgehen?

Designed and found data

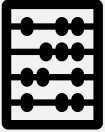


HWS 2021/22

Designed Data Mit Forschungsabsicht „entworfen“ i.d.R. strukturiert 			Organic data „Organisch“, nicht „entworfen“ i.d.R. unstrukturiert 				
 Befragung	 Experiment	 Administrativ	 Trans- aktion	 Soziale Netzwerke	 Internet of things	 Satelliten- daten	 Smartphone Sensoren
Erhebung einer vordefinierten Gruppe mittels Fragebogens Befragung von Studierenden zu ihren Erfahrungen mit Online-Lehre	Systematische Variation der Bedingungen zwischen einer oder mehreren Gruppen Eine Gruppe erhält Materialien vorab, die andere nach der Online-Vorlesung	Für administrative Zwecke erhoben Noten, Studiendauer, Studiengang, Geburtsdatum, ...	z.B. Käufe in Online-Shops	Clicks, views, likes, shares & comments	Miteinander über das Internet kommunizierende Geräte	z.B. Satellitenbilder des ESA Copernicus Programmes	GPS, Beschleunigungssensor, ...
			Learning Analytics Data, d.h. Daten die auf Lernplattform aufgezeichnet werden (Anzahl an Downloads & Abrufen, Aufenthaltsdauer auf bestimmten Seiten, ...)				

Schema aufbauend auf Groves (2011), Rigobon (2015), Kreuter (2017)

Outline

	Forschungsinteresse
	Datentypen und Erhebungsmethoden
	Auswahlverfahren
	Fragebogendesign
	Resümee: Potentiale und Grenzen

Grundgesamtheit und Stichprobe: Für wen treffen wir Aussagen und wen erheben wir?

• **Grundgesamtheit**

- Basis einer empirischen Untersuchung, sagt aus wofür die Ergebnisse gelten
- Genaue Beschreibung notwendig

• **Vollerhebung**

- Einbezug aller Mitglieder der Grundgesamtheit
- Dafür muss Grundgesamtheit überschaubar, bekannt und zugänglich sein
- Direktes Abbild der Grundgesamtheit, jedoch Berücksichtigung von Verweigerungen und dadurch Verzerrungen im Datensatz

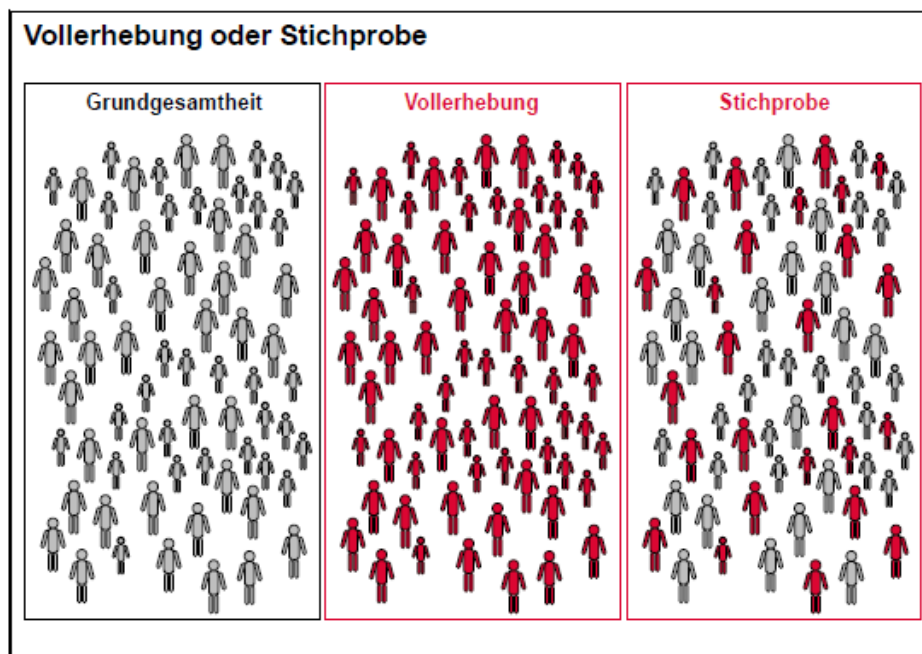


Abbildung 10: Vollerhebung oder Stichprobe

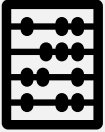
Braunecker (2021), S. 44

Teilerhebung (Stichprobe)

- Erhebung eines Teils der Grundgesamtheit stellvertretend für alle
- Repräsentativität: exaktes, lediglich verkleinertes Abbild der Grundgesamtheit
- Wenn Grundgesamtheit unbekannt oder schwer erreichbar ist
- Zufallsstichprobe, Quotenstichprobe, ...

 **ggf. Bereinigung und Gewichtung des Datensatzes**

Outline

	Forschungsinteresse
	Datentypen und Erhebungsmethoden
	Auswahlverfahren
	Fragebogendesign
	Resümee: Potentiale und Grenzen

Skalen: Wie messe ich?

Skalen

- im *weiteren* Sinne: Instrumente zur empirischen Erhebung eines bestimmten Sachverhalts,
- im *engeren* Sinne: numerische Relative zur Abbildung eines empirischen Relativs



Skalen-Niveau		Definition	Beispiel
Nominal		Klassifikation (Gleichheit oder Ungleichheit, keine weitere Relation)	Studiengang
Ordinal		Rangordnung, aber keine Abstände interpretierbar	Bildungsabschluss
Metrisch	Intervall	Rangordnung, Abstände zwischen Messwerten interpretierbar	Geburtsjahr
	Verhältnis/ Ratio	Rangordnung, Abstände und natürlicher Nullpunkt	Gehalt

Skalierungsverfahren

- Messbarmachung von latenten Sachverhalten wie etwa Einstellungen und Meinungen
- Fortsetzung der Operationalisierung (Bspw. Likert-Skalierung, Thurstone Skalierung, ...)

Likert-Skala	(1) Stimme völlig zu	(2) Stimme eher zu	(3) Stimme teilweise zu	(4) Stimme eher nicht zu	(5) Stimme überhaupt nicht zu
--------------	----------------------------	--------------------------	----------------------------	--------------------------------	-------------------------------------

vgl. Häder (2019), S. 96ff.

Einige Hinweise zum Wording: Wie stelle ich Fragen und Antworten?

- Frageformulierungen **kurz, verständlich** und **präzise**
- Antwortkategorien **disjunkt, erschöpfend** und **präzise**
- Auf **doppelte Verneinungen** verzichten
- Vermeidung von **Mehrdimensionalität**
- **Indirekte Fragen** vermeiden
- Keine **Suggestivfragen**
- **Fragen nach Motiven** und **nach lange Zeit zurückliegenden** Sachverhalten vermeiden

Häder (2019), S. 245ff.

Eine Frage, drei Varianten

<https://wiki.bib.uni-mannheim.de/limesurvey/index.php/677956?lang=de>

Eine Frage, drei Varianten

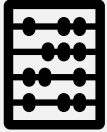
Inwiefern stimmen Sie der folgenden Aussage zu oder nicht zu?	(1) Stimme völlig zu	(2) Stimme eher zu	(3) Stimme teilweise zu	(4) Stimme eher nicht zu	(5) Stimme überhaupt nicht zu
Das Lehrmaterial wurde nicht rechtzeitig zur Verfügung gestellt.	Doppelte Verneinung			☐	☐
Das Lehrmaterial wurde rechtzeitig zur Verfügung gestellt und war gut gegliedert.	Double barreled		☐	☐	☐
Das Lehrmaterial wurde rechtzeitig zur Verfügung gestellt.	Präzise genug?		☐	☐	☐

Pretests

„If you don't have the resources to pilot test your questionnaire, don't do the study“
(Sudman und Bradburn 1982)

Testen auf Verständlichkeit der Fragen, Varianz der Antworten, Übersichtlichkeit des Fragebogens, Eventuelle Schwierigkeiten bei Fragen, theoretische Aussagekraft, Feldbedingungen testen (vgl. Häder (2019), S. 411ff.)

Outline

	Forschungsinteresse
	Datentypen und Erhebungsmethoden
	Auswahlverfahren
	Fragebogendesign
	Resümee: Potentiale und Grenzen

Resümee: Potenziale und Grenzen



HWS 2021/22

Designed Data Mit Forschungsabsicht „entworfen“ i.d.R. strukturiert 	Administrativ 	Organic data „Organisch“, nicht „entworfen“ i.d.R. unstrukturiert 
Befragung von Studierenden zu ihren Erfahrungen mit Online-Lehre	Noten, Studiendauer, Studiengang, Geburtsdatum,	Learning Analytics Data, d.h. Daten die auf Lernplattform aufgezeichnet werden (Anzahl an Downloads & Abrufen, Aufenthaltsdauer auf bestimmten Seiten, ...)

- Verschiedene Datentypen ersetzen sich nicht zwangsläufig, sondern können sich ergänzen, betrachten Sachverhalte ggf. aus einer anderen Perspektive
- Wichtigkeit sich der Potenziale und Grenzen bewusst zu sein und zu reflektieren
- Daher besonders dringlich, die Empirie auf eine solide theoretische Basis zu stellen, sorgfältig zu operationalisieren und genau zu definieren wofür die Aussagen getroffen werden

- Häder M. (2019). Empirische Sozialforschung. Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-26986-9_1
- Braunecker, C. (2021). How to do empirische Sozialforschung : Eine Gebrauchsanleitung (UTB; 5595). Wien.
- D. Prestiadi, Maisyaroh, I. Arifin and A. N. Bhayangkara, "Meta-Analysis of Online Learning Implementation in Learning Effectiveness," *2020 6th International Conference on Education and Technology (ICET)*, 2020, pp. 109-114, doi: 10.1109/ICET51153.2020.9276557.
- Peng/Matsui (2017): The Art of Data Science. <https://bookdown.org/rdpeng/artofdatascience/>
- Kreuter, F. (2018): Data Sciences as Essential Job Skill! A Social Science, Economics, and Public Policy Training Perspective. In Paderborn Symposium on Data Science Education at School Level 2017: Biehler, R., Budde, L., Frischemeier, D. et al. (Eds.) (2018), The Collected Extended Abstracts (pp. 21-26). Paderborn: Universitätsbibliothek Paderborn.
- Groves, R. M. (2011). "Designed data" and "organic data." Directors Blog, May 31. U.S. Census Bureau. Retrieved January 28, 2015 (<http://1.usa.gov/15NDn8w>).
- Rigobon, R. (2015). Discussion on applications and issues with using commercial data in research. BEA Expert Meeting on Exploiting Commercial Data for Official Economic Statistics, November 19, 2015. Washington D.C. Slides retrieved from http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_170272.pdf
- Sudman, Seymour, und Norman M. Bradburn. 1982. *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.
- Beispiele Organic Data:
 - <https://www.sciencedirect.com/topics/computer-science/transactional-data>
 - <https://www.datasciencegraduateprograms.com/social-media/>
 - <https://www.analyticssteps.com/blogs/data-science-iot-how-and-why>
 - <https://towardsdatascience.com/deep-learning-with-satellite-data-b78b20708de>
 - https://www.researchgate.net/publication/354380196_SHARING_DATA_COLLECTED_WITH_SMARTPHONE_SENSORS_WILLINGNESS_PARTICIPATION_AND_NONPARTICIPATION_BIAS

This Data Literacy Essential was developed
as part of BERD – Business, Economic and Related Data.

BERD@BW

funded by



Baden-Württemberg

MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST



BERD@NFDI funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 460037581