

Mixed Methods – Potenziale für die erziehungswissenschaftliche Interventionsforschung

Udo Kelle

*Gastvortrag auf der 2. Frankfurter Tagung zu Videoanalysen
in der Unterrichts- und Bildungsforschung
am 20. Februar 2013*

*Prof. Dr. Udo Kelle
Methoden empirischer Sozialforschung
Helmut-Schmidt-Universität Hamburg*

22043 Hamburg

kelle@hsu-hh.de



Gliederung

- Erziehungswissenschaftliche Forschung und ihre Ziele
- Mixed Methods als Alternative zu „Paradigmenkriegen“
- Methodenintegration zur Identifizierung und Bearbeitung von Methodenproblemen quantitativer Forschung
- Ein empirisches Beispiel
- Konklusion



Ziele erziehungswissenschaftlicher Forschung (tentativ)

(vgl. u.a. Ackermann et al. 2010, S. 8; Völter 2008; Bock, Miethe & Ritter 2010)

- Erziehungs- und Bildungsprozesse beschreiben, erklären und verstehen
- pädagogische Handlungspraxis anleiten

Erziehungswissenschaftliche Forschung ist ohne einen (wie auch immer gearteten, deskriptiven und/oder normativen) Bezug zu Bildungs- und Erziehungszielen schlechterdings nicht möglich, das heißt: sie braucht immer eine handlungswissenschaftliche Begründung.



Ein Typus erziehungswissenschaftlicher Forschung: „Programmevaluation“

(...)

*ist die systematische Anwendung empirischer Forschungsmethoden zur **Bewertung** des Konzeptes, des Untersuchungsplanes, der Implementierung und der Wirksamkeit sozialer **Interventionsprogramme**.*

(FREEMAN & ROSSI 1993)



Erziehungswissenschaftliche Forschung

- heterogen
(WITTGENSTEIN: „*Familienähnlichkeit*“)
- multi-, inter- und transdisziplinär
(Pädagogik, Psychologie, Soziologie ...)
- multimethodisch
(experimentelle Forschung, „Quasiexperimente“, Fallstudien)



In der Methodologie: „*Paradigm wars*“ (GAGE 1989)

Für qualitative Verfahren sind keinerlei methodische Standards und Gütekriterien „jenseits subjektiver Evidenzerlebnisse formulier- und überprüfbar.“ (SCHNELL, HILL, ESSER 1999, S. 110)

Quantifizierung „mit ihrem Labyrinth der verschiedenen logischen, mathematischen und technischen Anhänge“ führt zu einem „verminderten Verständnis der empirischen sozialen Welt“ (FILSTEAD 1979, S. 31)
„Ich glaube sowieso nicht an Zahlen...“ (eine Kollegin, mündlich, 1990)

***Die Alternative:
„romantischer
Holismus“?***

Die „Mixed Methods“ Bewegung z.Bsp. TASHAKORRI und TEDDLIE 2003, 2010

„... what we call the fundamental principle of mixed methods research. According to that fundamental principle, methods should be mixed in a way that has complementary strengths and non-overlapping weaknesses (BREWER & HUNTER 1989; TASHAKORRI & TEDDLIE 1998). It involves the recognition that all methods have their limitations as well as their strengths.“
(JOHNSON & TURNER 2003, p. 299)

**Stärken und Schwächen (Grenzen)
qualitativer und quantitativer Methoden?**

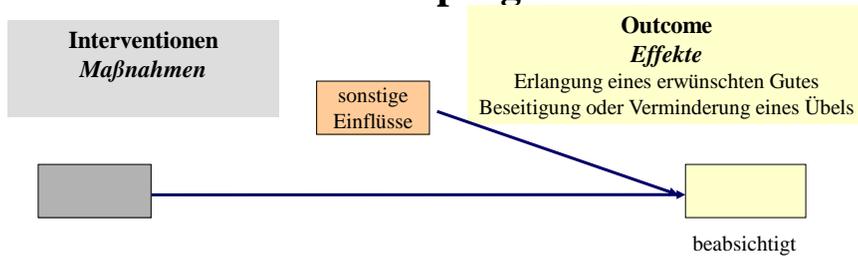


Stärken quantitativer Methoden

- die **Standardisierung der Datenerhebung**
(z. Bsp. durch Fragebögen)
 - ➔ *weithin akzeptierte Standards von Wissenschaftlichkeit (Replizierbarkeit, Überprüfbarkeit, Objektivität)*
 - ➔ *Modell hypothesen- und theoriegeleiteter Forschung kommt zur Anwendung*
- die Erhebung **großer Fallzahlen** und die **kontrollierte Ziehung von Stichproben**
 - ➔ *stichprobentheoretisch begründbares Konzept von Generalisierbarkeit, „Repräsentativität“*

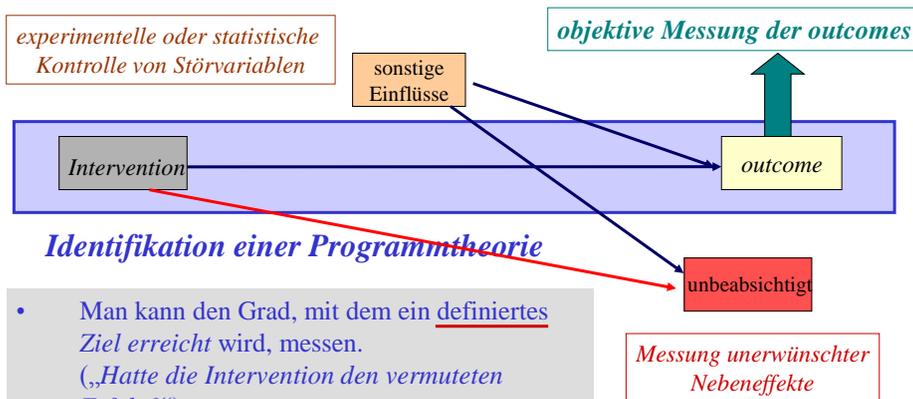


Interventionsprogramme



... sind der geplante und gezielte Eingriff politisch mächtiger Akteure (des Staates, von Organisationen) in bestehende Handlungsabläufe und –routinen, die (tatsächlich oder vermeintlich) im Einflussbereich dieser Akteure liegen, mit dem **Ziel**, ein erwünschtes Gut zu erlangen oder ein bestehendes Übel zu beseitigen oder zu mindern.

Stärken quantitativer Evaluationsdesigns



Identifikation einer Programmtheorie

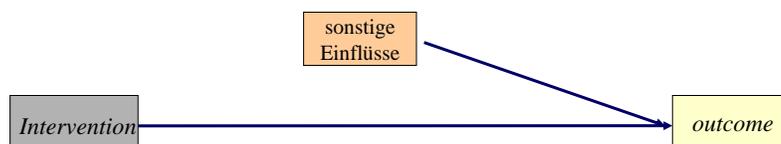
- Man kann den Grad, mit dem ein definiertes Ziel erreicht wird, messen. („Hatte die Intervention den vermuteten Erfolg?“)
- Man kann die Stärke erwünschter und unerwünschter Nebeneffekte feststellen („Wurden vermutete positive Nebeneffekte erzielt? Wie stark waren die vermuteten Schäden?“)

unbekannte Nebeneffekte ??

Schwächen (Grenzen) quantitativer Methoden

„Die vorherrschende Methodologie geht implizit davon aus, die Forscher hätten in genügendem Maße inhaltsreiche Vorstellungen über die untersuchten Wirklichkeitsbereiche zur Verfügung; (...) **Sachverhalte, über die der Forscher keine Vorstellungen hat, weil er den betreffenden Wirklichkeitsbereich nicht umfassend kennt, können nämlich in seinen Hypothesen gar nicht auftauchen, werden also auch nicht getestet und fehlen folglich im wissenschaftlichen Bild dieses Wirklichkeitsbereichs.**“
(GERDES 1979, S.5)

Schwächen (Grenzen) quantitativer Methoden



- Unzureichende Hypothesen, die **relevante Effekte (Nebeneffekte!) einer Intervention nicht berücksichtigen**

Merkmale qualitativer Methoden

- die *Offenheit und geringe Strukturierung der Datenerhebung* (z. Bsp. durch offene Interviews)
- die *intensive Analyse* von *einzelnen* oder *wenigen Fällen*
- *Hypothesenentwicklung* in der Auseinandersetzung mit den Daten



Stärken qualitativer Methoden

- die *Offenheit und geringe Strukturierung der Datenerhebung* (z. Bsp. durch offene Interviews)
- die *intensive Analyse* von *einzelnen* oder *wenigen Fällen*
- *Hypothesenentwicklung* in der Auseinandersetzung mit den Daten



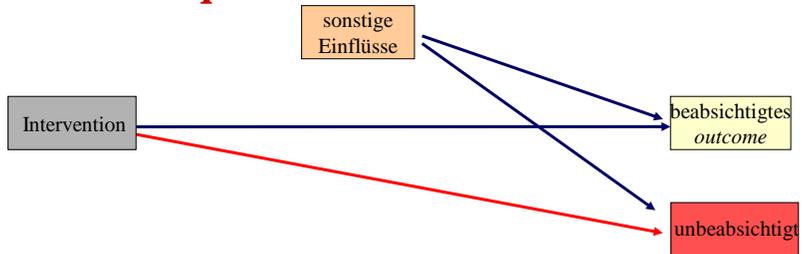
Exploration der Perspektiven, Wissensbestände und Handlungsorientierungen der Befragten



Entdeckung bislang unbekannter Phänomene (etwa: dem Forscher nicht bekannte Handlungsregeln im Feld)

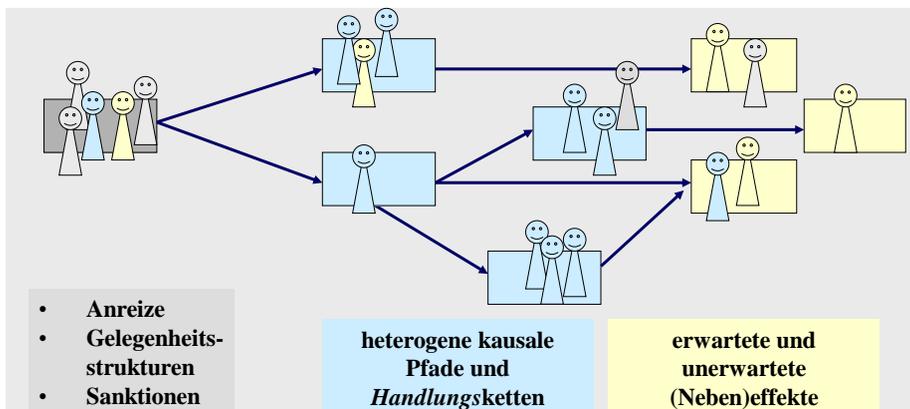
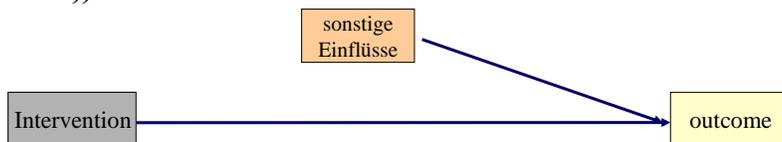


Schwächen quantitativer Evaluationsmethoden



- Unzureichende Hypothesen, die *relevante Effekte (Nebeneffekte!)* eines treatments nicht berücksichtigen
- *mangelndes Wissen über die „kausalen Pfade“*, über die eine Intervention wirkt (z.Bsp. Faktoren für „compliance“)

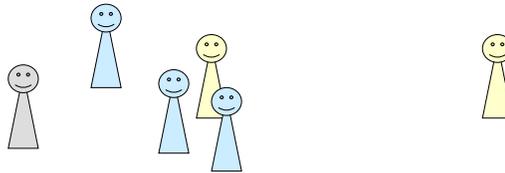
„Kausale Pfade“ einer Intervention



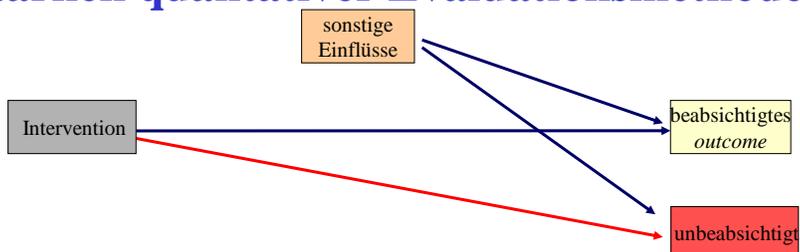


Beteiligte („stakeholder“)

- verfolgen *eigene Ziele*, die nicht notwendigerweise mit den Programmzielen übereinstimmen,
- verhalten sich dabei *kreativ* und findig,
- verfügen über *lokales Wissen* über Wege zur Umsetzung von Programmzielen,
- fühlen sich durch die Maßnahmen, die zur Erreichung von Programmzielen beitragen, möglicherweise *beeinträchtigt*



Stärken qualitativer Evaluationsmethoden



- Die Identifikation von Kriterien für den „Erfolg“ einer Intervention („Was wird bei unterschiedlichen Akteuren als Erfolg gewertet?“)
- Die Identifikation erwünschter und unerwünschter Nebenwirkungen („Was hat die Maßnahme noch für Folgen? Richtet die Maßnahme Schaden an?“)
- Die Beschreibung der „kausalen Pfade“, durch die eine Intervention wirkt („Wie und auf welchen Wegen wirkt eine Maßnahme?“)
- Die Identifikation von „Methodenartefakten“ („Werden mit der quantitativen Untersuchung relevante Phänomene erfasst? Bilden die Messungen überhaupt das ab, was sie abbilden sollen?“)

Ein empirisches Beispiel: Von der Messung von Qualität zur Qualität von Messungen

Bitte kreuzen Sie die Antwort an, die Ihre Einschätzung am besten wiedergibt.		stimmt							
		stimmt eher	↓	↓	↓	↓	↓	↓	↓
		stimmt eher nicht	↓	↓	↓	↓	↓	↓	↓
		stimmt nicht	↓	↓	↓	↓	↓	↓	↓
1. Das Seminar verläuft nach einer klaren Gliederung.	<input type="checkbox"/>								
2. Der Dozentin/Dem Dozenten scheint der Lernerfolg der Studierenden gleichgültig zu sein.	<input type="checkbox"/>								

Das quantitative „Messparadigma“ Qualitätskriterien:

- **Objektivität:** Ist das Ergebnis der Messung unabhängig vom Beobachter?
- **Reliabilität (Wiederholbarkeit und Konsistenz):** Führt die wiederholte Messung derselben Sachverhalte zu denselben Resultaten?
- **Validität:** Ist das Resultat der Messung ein guter Indikator für das untersuchte Phänomen?



Praxisprobleme

„Obwohl die Ergebnisse im Schnitt positiv ausfallen (...), bleibt Lehrevaluation umstritten, zudem trägt ihre Durchführung nicht immer zu einer Verbesserung der Lehre bei.“ (Rindermann 2003).

- **Evaluationsmüdigkeit** von Studierenden angesichts ausbleibender Folgen von Evaluationen („... das bringt doch sowieso nichts“)
- Vorbehalte und **Befürchtungen von Lehrenden** (Furcht vor Kontrolle, Bewertung, öffentlichem ranking...)
- **Besorgnisse** hinsichtlich der Wirkung von „**Biasvariablen**“ wie z.B. *Vorinteresse* oder *Benotungserwartung*
- Unklarheit und **unterschiedliche Auffassungen** davon, was „Qualität in der Lehre“ ist



Fragen nach dem Nutzen und der Qualität von Lehrevaluation



Validität (Gültigkeit)

Misst das Instrument tatsächlich das, was es messen soll?

Bitte kreuzen Sie die Antwort an, die Ihre Einschätzung am besten wiedergibt.

	nicht sinnvoll beantwortbar				
		stimmt		↓	
		stimmt eher		↓	
		stimmt eher nicht		↓	
	stimmt nicht		↓		
	↓	↓	↓	↓	↓
1. Das Seminar verläuft nach einer klaren Gliederung.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Der Dozentin/Dem Dozenten scheint der Lernerfolg der Studierenden gleichgültig zu sein.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Der/Die Dozent/in gestaltet das Seminar interessant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Der/Die Dozent/in verhält sich den Studierenden gegenüber freundlich und respektvoll.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Das Seminar ist vermutlich für die spätere Berufspraxis sehr nützlich. ...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Das Seminar gibt einen guten Überblick über das Themengebiet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Der/Die Dozent/in geht auf Fragen und Anregungen der Studierenden ausreichend ein.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Der/Die Dozent/in gibt zu wenig erklärende und weiterführende Information zu den behandelten Themen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Der/Die Dozent/in verdeutlicht zu wenig die Verwendbarkeit und den Nutzen des behandelten Stoffes.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Das Seminar ist eine gute Mischung aus Wissensvermittlung und Diskussion.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Sind die Items eines Fragebogens sinnvolle Indikatoren für das untersuchte Konzept?



Statistische Validierung

Komponentenmatrix(a)

	Komponente	
	1	2
gute Aussichten auf sicheren Arbeitsplatz	-,749	,296
Studium, wegen Unkenntnis von Alternativen	,640	,338
Entwicklung eines eigenen Stils	,638	,221
ein Studium ist für späteren Beruf erforderlich	-,548	,480
als Student/in werden Vorstellungen vom Leben verwirklicht	,235	,621
Studium ermöglicht größeres Fortkommen	-,348	,589
Selbsterfahrung	,416	,450

Extraktionsmethode: Hauptkomponentenanalyse.
a 2 Komponenten extrahiert

- Suche nach Übereinstimmungen (*Korrelationen*) zwischen Items (*Konsistenz* bzw. *Reliabilität*)
- Statistische Assoziationen (*Korrelationen*) mit einem anderen quantitativ messbaren Kriterium (z.Bsp. Prüfungserfolg der Studierenden)



- Werden die Items von den Befragten so **verstanden**, wie dies beabsichtigt wurde?
- Haben die Befragten überhaupt das **Wissen**, um die Items richtig zu beantworten?
- Haben die Befragten genügend Gründe (**Motivation**), den Fragebogen (ehrlich) zu beantworten?

Die Befragten müssen bereit und in der Lage sein, gültige Antworten zu geben.
(Fowler 1995)

Quantitative Teilstudie

Standardisierte Befragung zu „Einstellungen von Studierenden zu Lehrveranstaltungsevaluationen“

- **qualitative Pilotstudie** mit Studierenden (n = 141) zur Exploration von negativer und positiver Konsequenzen aus Lehrveranstaltungsevaluationen
- **Fragebogenkonstruktion** auf der Grundlage der „*Theory of Planned Behavior*“ (AJZEN 1991, 2002, 2005)
- **standardisierte Erhebung** der Einstellungen und Intentionen von Studierenden (n = 549) an 3 Universitäten und 1 Fachhochschule in Niedersachsen und Hessen

Qualitative Teilstudie

Qualitative Befragung zur „Verständlichkeit von Items zur Lehrveranstaltungsevaluation“

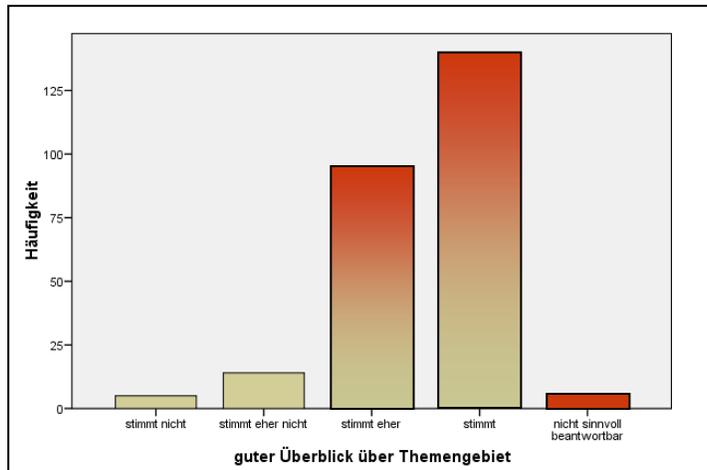
- kognitive Interviews mit Studierenden (n = 21) mit Hilfe der „Think aloud“ Technik und mit Probingfragen zur Interpretation und Beantwortung von Items eines standardisierten, statistisch validierten Fragebogens zur Evaluation von Lehrveranstaltungen
- kategorienbildende Auswertung der qualitativen Textdaten
- Entwicklung von Umformulierungsoptionen für die Items



Item: Die Vorlesung/ das Seminar gibt einen guten Überblick über das Themengebiet

*„je, nachdem, welchen Dozenten sie haben, ist es auch sehr einseitig ausgerichtet. Zum Beispiel [...] der Herr B. [...], der ist halt ein großer Anhänger der Verhaltenstherapie und die Tiefenpsychologie wird so ganz ausgeblendet. Aber am Anfang merken sie da gar nicht. [...] durch Praktika oder dadurch, dass man etwas mehr Zugang zur Materie kriegt, merkt man, dass es vielleicht doch ein bisschen einseitig ist.“
(Interview Nr. 7, 4.10.2007, Abs. 30)*





- Befragte, die bei der Beantwortung dieser Frage nicht über die nötigen Informationen und Kriterien verfügen, kreuzen relativ **selten** „**nicht sinnvoll beantwortbar**“ an
- Befragte, die bei der Beantwortung dieser Frage nicht über die nötigen Informationen und Kriterien verfügen, antworten **häufig** mit „stimmt eher“ oder „stimmt“



Item: Die Vorlesung/ das Seminar ist vermutlich für die spätere Berufspraxis sehr nützlich

Ist hier gemeint: die „eigene Berufspraxis“

„Wenn ich in den speziellen Bereich der Psychologie gehe, zu der das Seminar eben gehört, ist das Seminar auf jeden Fall nützlich, weil ich dort sehr viel gelernt hab. [...] Ja, wäre so ein „stimmt eher“, weil ich nicht so genau weiß, wo ich lang will.“ (Interview Nr. 6, 24.9.2007, Abs. 43)

Item: Die Vorlesung/ das Seminar ist vermutlich für die spätere Berufspraxis sehr nützlich

... oder ist gemeint: „Nützlichkeit für die Berufspraxis allgemein“??

*„Unter diesem Aspekt stelle ich mir auf jeden Fall vor, oder erwarte ich, dass, wenn halt viel Berufspraxis drin sein soll, dass halt sehr viel erzählt wird, wie ist das im Alltag, wie wird das umgesetzt [...]. Und sie hat auch viel über die Ausbildung erzählt. Man würde das jetzt so und so lernen.“
(Interview Nr. 4, 22.8.2007, Abs. 25.27)*

Streuung der Frage:

„Das Seminar ist vermutlich für die spätere Berufspraxis sehr nützlich.“

Veranstaltungs- nr.	stimmt nicht (1)	stimmt eher nicht (2)	stimmt eher (3)	stimmt (4)	Mittel- werte	Standardab- weichung
3	3	13	7	3	2,38	0,852
4	5	13	11	8	2,59	0,985
5	9	5	4	3	2,05	1,117
6	1	16	18	6	2,71	0,750
7	4	12	15	6	2,62	0,893
8	13	14	9	4	2,10	0,982

Ergebnisse

- Dass Studierende nicht über Beurteilungskriterien zur Beantwortung einer Frage verfügen, lässt sich an standardisierten Daten oft nicht ablesen, sondern erfordert qualitative Interviews.
- Hohe Streuungen auf Veranstaltungsebene können ein erster Indikator für unterschiedliche Beurteilungskriterien der Befragten sein.
- Diese Probleme zeigen sich in statistischen Kontrollprozeduren wie Faktorenanalysen kaum.
- Mit Hilfe eines Methodenmix von quantitativen und qualitativen Verfahren lassen sich Validitätsprobleme identifizieren und beheben und validere Instrumente entwickeln.



Stärken und Schwächen (Grenzen) quantitativer Methoden

- Modell **hypothesen- und theoriegeleiteter** Forschung ist anwendbar
- **standardisierte („harte“) Daten**, die quantifizierbar sind und klassische Forderungen nach Objektivität und Zuverlässigkeit erfüllen
- stichprobentheoretisch begründbares Konzept von „**Verallgemeinerung**“
- „**Messung**“ von **Zielerreichung** anhand vorab definierter Kriterien möglich.

- Unzureichende Programmtheorien, die **relevante Effekte (Nebeneffekte!) nicht berücksichtigen**
- Irrelevanz der untersuchten Zielvariablen für die Akteure im Feld und deshalb **mangelnde compliance**
- unzureichender lebensweltlicher Bezug der Messinstrumente und deshalb **Verzerrungen durch Methodenartefakte** (bspw. nonresponse und soziale Erwünschtheit)



Stärken und Schwächen (Grenzen) qualitativer Methoden

- Die Identifikation von Kriterien für den „Erfolg“ einer Intervention („Was wird bei unterschiedlichen Akteuren als Erfolg gewertet?“)
- Die Identifikation erwünschter und unerwünschter Nebenwirkungen („Was hat die Maßnahme noch für Folgen? Richtet die Maßnahme Schaden an?“)
- Die Beschreibung der „kausalen Pfade“, durch die eine Intervention wirkt („Wie und auf welchen Wegen wirkt eine Maßnahme?“)
- Die Identifikation von „Methodenartefakten“ („Werden mit der quantitativen Untersuchung relevante Phänomene erfasst? Bilden die Messungen überhaupt das ab, was sie abbilden sollen?“)

- **kleine Fallzahlen**, mit dem Risiko, auf untypische Fälle zu fokussieren
- **ressourcenintensive** Datenerhebung und -auswertung
- in der Grundlagenforschung **bewährte Verfahren** (etwa: hermeneutische Feinanalyse) in der Interventionsforschung **nur beschränkt einsetzbar**
- **Tonaufzeichnung** und **Transkription** der Daten oft nur begrenzt möglich
- bei der Datenauswertung **divergierende Interpretationen**