

**Prof. Dr. Christoph Kulgemeyer**

AG Didaktik der Physik

Fakultät für Naturwissenschaften



# Wie erstelle ich einen Fragebogen? Eine Einführung in quantitative Methoden

---

# Quantitative Methoden - was ist das?

# Methode – was ist das?

---

- „Methode“ ist ein Sammelbegriff mit vielschichtiger Bedeutung (z.B. Didaktik / Methodik, Forschungsmethodik, Methodik der Datenanalyse...)
- Probate „Arbeitsdefinition“: Eine (wissenschaftliche) Methode ist ein nachvollziehbares Verfahren zur Generierung von Wissen
- In der Naturwissenschaftsdidaktik hat man es mit zwei Datenarten zu tun: qualitative (in der Regel Sprache) und quantitative Daten. Zum Umgang mit beiden Datenarten gibt es Methoden!

# Methode – was ist das?

---

- „Methode“ ist ein Sammelbegriff mit vielschichtiger Bedeutung (z.B. Didaktik / Methodik, Forschungsmethodik, Methodik der Datenanalyse...)
- Probate „Arbeitsdefinition“: Eine (wissenschaftliche) Methode ist ein nachvollziehbares Verfahren zur Generierung von Wissen
- In der Naturwissenschaftsdidaktik hat man es mit zwei Datenarten zu tun: qualitative (nicht-mathematisierbar, in der Regel Sprache) und quantitative Daten. Zum Umgang mit beiden Datenarten gibt es Methoden!
- Die Naturwissenschaftsdidaktik hat ein gewisses „Standardrepertoire“ an Methoden, die häufig vorkommen und immer einem bestimmten Zweck dienen – Methoden müssen dem Ziel angemessen sein!

# Ausgangslage: ein gängiges Problem

---

- Wir wollen testen, ob eine neuartige Unterrichtsmethode zu besseren Ergebnissen führt.
- Z.B. Vergleich Frontalunterricht gegen Gruppenpuzzle bei der Einführung des Kraftbegriffs – Vergleich zweier Gruppen
- Was könnte man erheben? Woran macht man den Erfolg fest?
  - Motivation
  - Fachwissen
  - Interesse
  - etc.
- Veränderungen durch die Methode müssen sich messen lassen - dazu braucht es Verfahren

# Was ist ein Test?

---

- Test: Entweder ein statistisches Verfahren, um Hypothesen auf Signifikanz prüfen zu können - oder ein Messinstrument
- „Ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad einer individueller Merkmalsausprägung“ (Lienert & Raatz, 1998)
- Verschiedene Arten von Tests:
  - **Leistungstest**: Leistung nach Kriterien als richtig oder falsch beurteilen - z.B. Fachwissen
  - **Persönlichkeitstest**: Ausprägung eines Persönlichkeitsmerkmals - z.B. Interesse

# Erzeugung quantitativer Daten

---

- In der Regel werden quantitative Daten aus zwei verschiedenen Instrumenten erzeugt:
  - Leistungstests (Messung von Fähigkeiten, Norm existiert)
  - Fragebögen (Messung von Einstellungen bzw. affektiven Merkmalen, ohne Norm)
  
- Möglich ist auch das Erzeugen quantitativer Daten auf Basis von:
  - Beobachtungen (z.B. Videomitschnitte)
  - Interviews (z.B. Leitfadeninterviews)
  - Dokumenten (z.B. Lerntagebücher, Portfolios)
  
- ... In der Regel können dann nur direkt abzählbare Oberflächenmerkmale ausgewertet werden oder eine qualitative Datenanalyse geht voraus – z.B. qualitative Inhaltsanalyse

# Überblick über quantitative Methoden

---

- Häufig existieren in der Naturwissenschaftsdidaktik drei Arten von Datenquellen:
  - Texte
  - Videos/ Audiomitschnitte
  - Tests/ bzw. Fragebögen



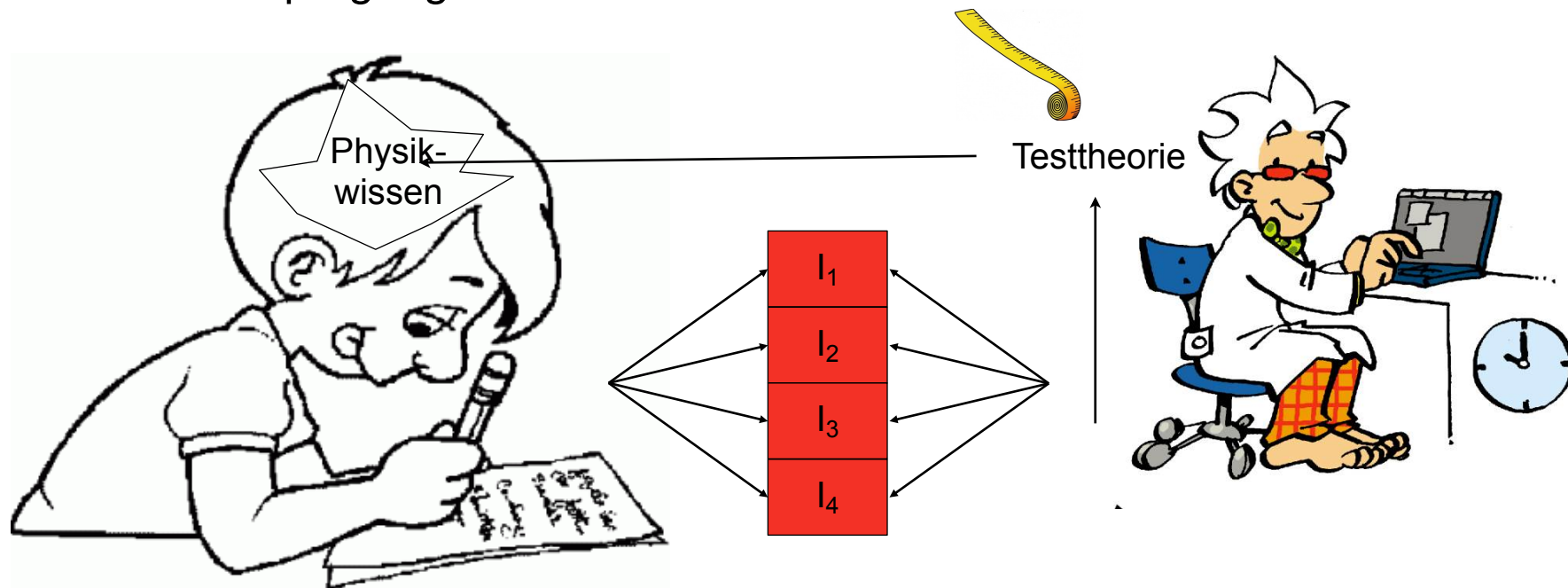
# Überblick über quantitative Methoden

---

- Zur Analyse quantitativer Daten werden zumeist statistische Tests verwendet:
- Klassische Statistik:
  - Deskriptive Statistik (Mittelwert, Standardabweichung, Verteilungen)
  - Skalenbildung (Reliabilitätsanalyse, Faktorenanalyse)
  - Zusammenhangstests (Korrelationen, Regressionsanalysen)
  - Unterschiedstests (T-Tests, Varianzanalysen)
  - Typisierende Verfahren (Clusteranalyse)
- Probabilistische Statistik:
  - Rasch-Modelle (Kontinuierliche Skalen)
  - Latente Klassenanalysen (Typisierende Skalen)

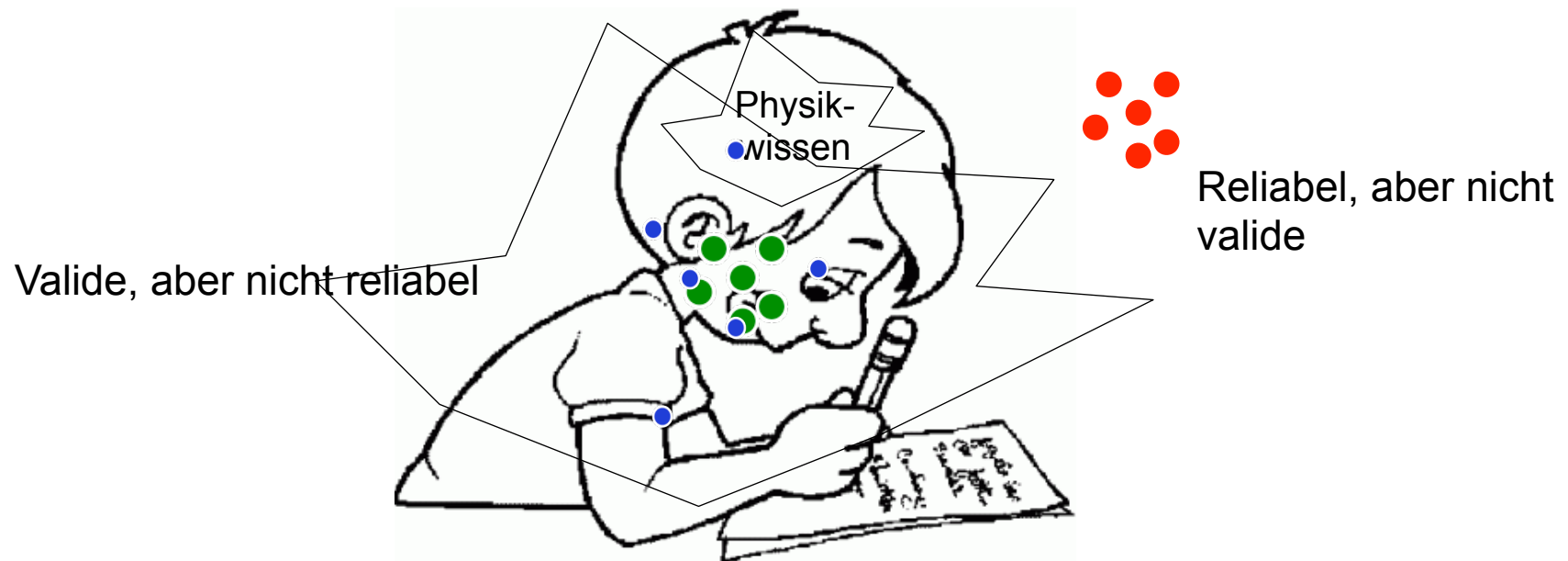
# Was ist Testtheorie?

- Die Testtheorie beschäftigt sich mit der Frage, wie aus einer Anzahl Verhaltensbeobachtungen von Versuchsperson in bestimmten Situationen auf die „wahre“ Ausprägung eines Persönlichkeitsmerkmals geschlossen werden kann.
- Dabei werden die „Verhaltensbeobachtungen“ (z.B. Itemantworten) als manifeste Variable bezeichnet und die Merkmalsausprägung als latente Variable



# Woher weiß ich, ob mein Test gut ist?

- **Validität:** Aus meinen Messwerten kann ich begründet Schlüsse über das Merkmal ziehen
- **Reliabilität:** Das Messinstrument soll genau messen – und zwar jedes Item (=Testaufgabe)
- **Objektivität:** Das Messergebnis soll unabhängig von äußeren Bedingungen sein (z.B. Testauswerter, Testdurchführerin)



---

# Entwicklung von Fragebögen

# Unterscheidung Fragebogen ↔ Test

---

- Mit einem Leistungstest will man Leistungen objektiv messen
  - Es gibt **richtige und falsche Antworten**
- Mit einem Fragebogen (Persönlichkeitstest) will man subjektive Selbstauskünfte erheben oder die Wahrnehmung von Sachverhalten erfassen
  - Meinungen, Einstellungen
  - Verhaltensweisen
  - Erinnerungen, Erfahrungen, Beobachtungen
- Es gibt **keine richtige oder falsche Antworten** (auch wenn man Antworten z.B. moralisch unterschiedlich bewerten kann)

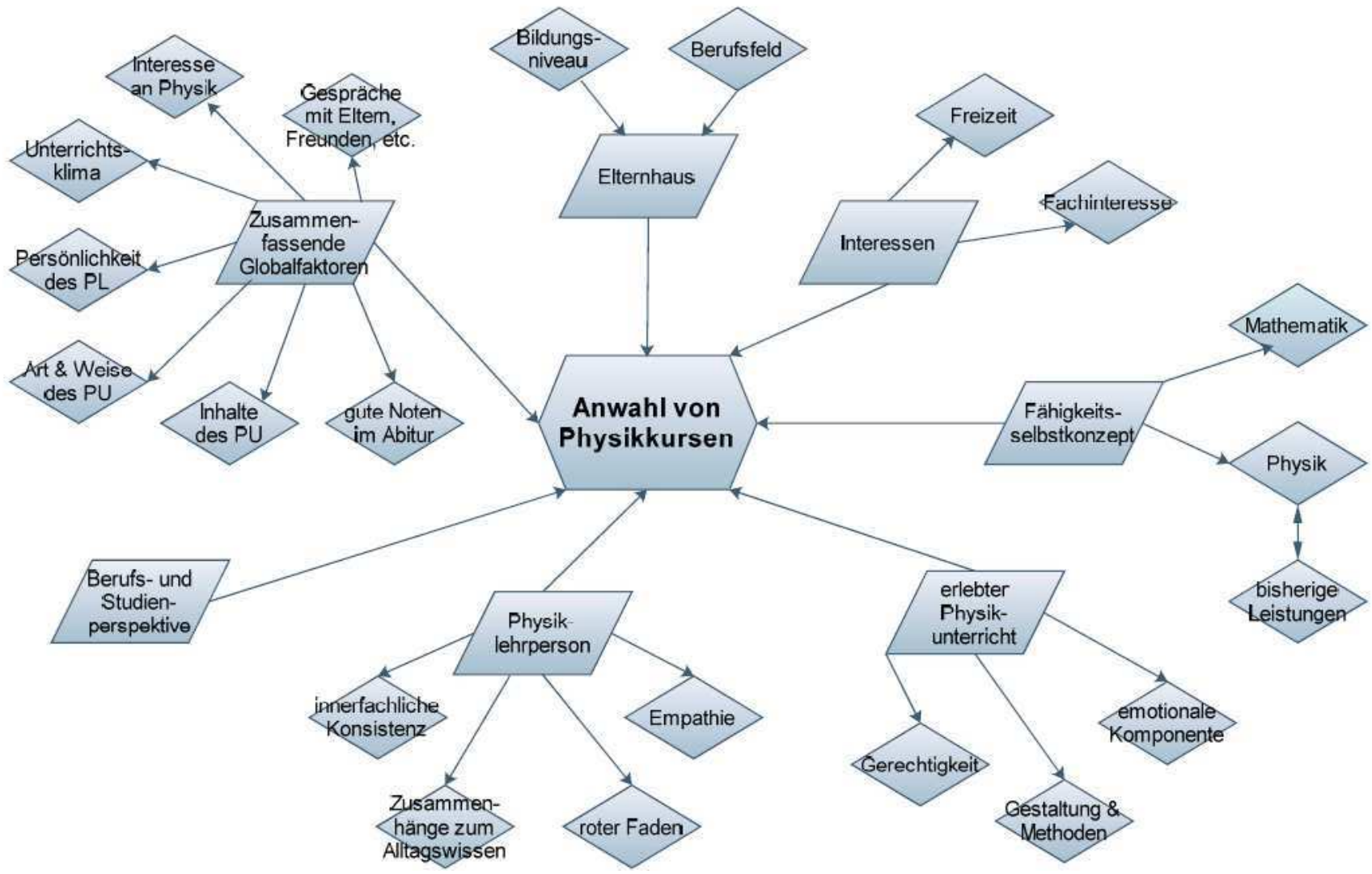
# Grundüberlegungen

---

- Was genau will ich herausfinden?  
Worauf kommt es mir besonders an?  
Was ist „nice to know“ aber nicht essentiell wichtig?  
(weglassen)
  - Wer sind meine Probanden? Welche Merkmale haben sie?
    - Wissen, Erfahrungen
    - sprachliches Niveau, Lesefähigkeit
    - Konzentrationsfähigkeit
    - Bereitschaft zur Mitwirkung an der Befragung
  - Gibt es in der Literatur bereits Instrumente, die ich einsetzen kann?
    - erspart Entwicklungsarbeit
    - ersetzt jedoch nicht die Vorerprobung und eine erneute Prüfung der Gütekriterien (Validität, Objektivität, Reliabilität)
    - DaQS - Datenbank zur Qualität von Schule  
Fragebögen, Tests und Skalen
- DIPF- Deutsches Institut für Internationale Pädagogische Forschung

# Generierung von Fragebogenitems

- Facettenanalyse / Mindmapping / Brainstorming



# Itemformate

---

- Offene Fragen  
besser: „Fragen mit offenem Antwortformat“
- Geschlossene Fragen  
auch: „Fragen mit gebundener Beantwortung“
  - Fragen mit 2 Antwortoptionen  
„ja“/ „nein“, „richtig“/ „falsch“
  - Rating-Skalen (mehr als 2 Antwortoptionen), auch „Likert-Skalen“ genannt  
„stimmt vollkommen“ ... „unentschieden“, ... „ist völlig falsch“
  - Polaritätsprofile  
„Ordnen Sie den Physikunterricht zwischen den folgenden Begriffspaaren ein:

kalt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
warm						
männlich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
weiblich						
...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	...



# Fragen mit offenem Antwortformat

---

- Vorteile
  - schneller zu entwickeln (das erscheint zumindest zunächst so)
  - besonders geeignet zur Exploration eines neuen Forschungsfeldes (über das man bisher wenig weiß)
- Probleme
  - Validität, Objektivität und Reliabilität (schwieriger sicherzustellen als bei geschlossenen Fragen)
  - aufwändige Auswertung (Codiermanuale auf Basis qualitativer Inhaltsanalyse)
  - werden von Probanden oft nur knapp bearbeitet (Schreibhemmung, empfundener Aufwand, ...)
- Empfehlung
  - Setzen Sie offene Fragen nur dort ein, wo es wirklich notwendig und sinnvoll ist!

# Offene Fragen: Praktische Umsetzung

---

- Geben den erwarteten Umfang der Antwort an
  - „Nehmen Sie in einigen Sätzen / in Stichworten Stellung zu ...“
  - „Listen Sie in Spiegelstrichen auf, ...“
  - Umkastelung des Antwortbereichs
- Beispiel
  - „Was halten Sie von der Einführung integrierten naturwissenschaftlichen Unterrichts in den Klassen 5 und 6?“
  - „In den Klassen 5 und 6 soll integrierter naturwissenschaftlicher Unterricht eingeführt werden. Listen Sie in Stichworten Argumente pro und contra auf! Unterstreichen Sie die aus Ihrer Sicht drei wichtigsten Argumente!“

<b>pro</b>	<b>contra</b>
-	-
-	-
-	-

# Offene Fragen: Praktische Umsetzung

---

- halboffenes Antwortformat: Geben Sie einen einführenden Text vor, zu dem Stellung genommen wird.
- Beispiel
  - „Herr Merz und Herr Kamp führen in der Pause das folgende Gespräch über die geplante Einführung naturwissenschaftlichen Unterrichts in den Klassen 5 und 6. Nehmen Sie dazu aus Ihrer Sicht in einigen Sätzen Stellung!
    - Merz: Ich fühle mich fachlich nicht qualifiziert, um ... (insges. ca. 1/3 Seite Gesprächstext)

Ich meine, ...

# Geschlossene Fragen

---

- Vorteile
  - einfach und objektiv auszuwerten
  - statistische Kennwerte einfacher zu berechnen
- Nachteile
  - Zu solchen Aspekte des Themas, die nicht in Fragen abgebildet sind, erhält man keine Informationen (anders als das bei offenen Fragen möglich ist).
  - Man weiß nicht genau, welche Überlegungen die Probanden mit ihrer Kreuzen verbinden.
- Empfehlung
  - Wo immer es möglich ist, sollten geschlossene Fragen verwendet werden

# Rating-Skalen (Likert-Skalen)

---

- Viele Fragebögen zu Personenmerkmalen verwenden Rating-Skalen.
- Die Probanden sollen einer Aussage zustimmen oder sie ablehnen, z.B. „Ich freue mich auf den Physikunterricht“
- Für die Antwort gibt es ein gestuftes Schema
  - Bsp. 4-stufig:
    - 1: „trifft voll zu“ | „2: trifft eher zu“ | 3: „trifft eher nicht zu“ | 4: „trifft gar nicht zu“
  - Bsp. 5-stufig:
    - 1: „trifft voll zu“ | „2: trifft eher zu“ | 3: „unentschieden“ | 4: „trifft eher nicht zu“ | 5: „trifft gar nicht zu“

# Rating-Skalen: Konstruktionsfragen

---

- Soll man 4- oder 5-stufige Antworten verwenden?
  - 5-stufig
    - Antworttendenz zur Mitte
    - „unentschieden“ schwer von „weder noch“ oder „weiß nicht“ abgrenzbar
  - 4-stufig
    - Mittelwert nicht klar interpretierbar
    - Probanden werden zu Tendenzentscheidungen gezwungen, die sie vielleicht eigentlich nicht treffen möchten.
- Mögliche Erweiterung bei beiden Skalen
  - 6: „Ich möchte die Frage nicht beantworten“ | „Ich habe dazu keine Meinung“ | „weder/noch“
  - Problem: Diese Option erzeugt systematisch „fehlende Werte“

# Rating-Skalen: Statistische Probleme

---

- Es handelt sich um ordinale Daten. Häufig werden die Daten jedoch als a) intervallskaliert oder zumindest b) als äquidistante Ränge angenommen. Zumindest werden sie so behandelt, obwohl das streng genommen nicht zulässig ist.
- Was eigentlich alles aufgrund der Datenqualität „ordinal“ nicht geht:
  - Mittelwerte und Standardabweichungen berechnen  
--> Median oder Modus angeben
  - Skalenreliabilitäten auf Grundlage von Produkt-Moment-Korrelationen oder Varianzen berechnen  
--> Rangkorrelationen zugrundelegen

# Rating-Skalen: Statistische Probleme

---

- Deshalb wird häufig bewusst über die Probleme hinweggesehen: „Augen zu und durch“. (Selbst Bortz drückt sich um klare Aussagen.)
- Wenn man den Probanden nahelegt, dass die Stufen äquidistant sein sollen, z.B. durch eine grafische Skala mit äquidistanten Abständen von 1 bis 5, kann man Rating-Skalen mit nicht zu schlechtem Gewissen als intervallskaliert handhaben.



# Ratingskalen – Was man bedenken sollte (vgl. Bühner, 2011)

---

- Vermeiden Sie unterschiedliche Polungen der Items  
möglichst keine negativen Formulierungen
- keine doppelten Verneinungen
- Kurze Formulierungen
- Jedem Item nur einen Sachverhalt zuordnen
  - ~~„Integrierter NaWi-U ist wichtig und schon recht verbreitet.“~~
- Verallgemeinerungen vermeiden
  - ~~„Ich freue mich immer auf den Physikunterricht.“~~
- Bezugs-Zeiträume genau angeben
  - „Ich fahre regelmäßig zu Lehrerfortbildungen.“
  - „Ich habe in den letzten 12 Monaten an einer Lehrerfortbildung teilgenommen.“
- Häufigkeiten konkretisieren
  - ~~„Im Physikunterricht haben wir viel experimentiert.“~~
  - „Wir haben fast in jeder Woche ein Schülerexperiment gemacht.“

# Ratingskalen – Was man bedenken sollte (vgl. Porst, 1998)

---

- „Befragungspersonen gehen davon aus, daß die ihnen vorgelegte Skala sinnhaft konstruiert ist, und sie nehmen an, daß sie in ihren Ausprägungen tatsächliche Verteilungen der Population wider-spiegeln. Die Informationen, die sie aus dem Wertebereich einer Skala erschließen, nutzen sie gleich mehrfach: zum einen ziehen sie den Wertebereich als Bezugsrahmen für ihre eigene Verhaltenshäufigkeit heran (und geben zum Beispiel höhere Frequenzen für ihr eigenes Verhalten an, wenn die Skala höhere Häufigkeiten vorgibt), zum zweiten entnehmen sie ihrer eigenen Platzierung auf der Skala Informationen über die relative Häufigkeit ihres Verhaltens verglichen mit dem Verhalten anderer und berücksichtigen dies, wenn sie komparative Urteile bilden; und schließlich ziehen sie bei nicht eindeutigen Fragen die Skalen als Interpretationshilfen für den vermeintlichen Sinn der Frage heran.“

# Ratingskalen – Was man bedenken sollte (vgl. Porst, 1998)

---

- Es macht einen Unterschied wie man nach Häufigkeiten fragt

- Wie häufig setzen Sie Schülerexperimente pro Schuljahr ein

0 mal	1-5 mal	6-10 mal	11-20 mal	über 20 mal
0 mal	1-3 mal	4-6 mal	7-10 mal	über 10 mal

# Layout

---

- An der Sorgfalt des Layouts messen Probanden die Wichtigkeit und Professionalität der Befragung
- Das Layout soll den Probanden bei der Bearbeitung des FB helfen
  - Einheitlichkeit des Layouts über alle Seiten
  - Hervorhebungen von Frageblöcken (Rahmen, graue Hinterlegung)
  - Unterscheidung zwischen Erläuterungen und Fragen
  - gut lesbare Texte und Abbildungen
  - Hilfslinien bei Ratingskalen
  - etc.
- Deckblatt mit
  - Institution, die die Befragung durchführt
  - Kontaktadresse
  - kurzer Einführungstext und Dankformel
  - Hinweis auf Anonymisierung
  - ggf. persönlicher Code