



Psychologische Forschungsmethoden

BSc Philosophie-Neurowissenschaften-Kognition WiSe 2022/23

BSc Psychologie WiSe 2022/23

Prof. Dr. Dirk Ostwald

(11) Alternative Messtheorien

Vorbemerkungen

Schwächen der Repräsentationstheorie

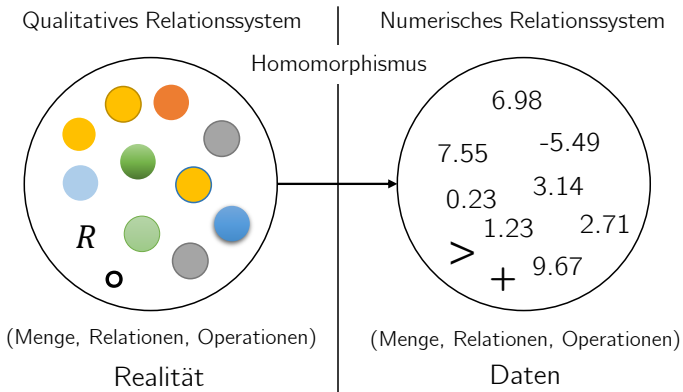
Alternative Messtheorien

Vorbemerkungen

Schwächen der Repräsentationstheorie

Alternative Messtheorien

Repräsentationstheorie des Messens



Vorbemerkungen

Repräsentationstheorie des Messens

Ordinalmessung

- Das qualitative Relationssystem $\mathcal{M} := (M, R)$ ist eine *strenge schwache Ordnung*
- Das numerische Relationssystem ist $\mathcal{N} := (\mathbb{R}, >)$
- Es existiert ein Homomorphismus f von \mathcal{M} nach \mathcal{N} und $(\mathcal{M}, \mathcal{N}, f)$ heißt *Ordinalskala*
- Die Menge der zulässigen Transformationen von f ist die Menge der *monotonen Funktionen*
- Größenvergleiche von Messwerten sind bedeutsam
- Größenvergleiche von Messwertdifferenzen oder Messwertverhältnisaussagen sind i.A. nicht bedeutsam
- Summen- und Mittelwertstatistiken sind i.A. nicht bedeutsam

Differenzmessung

- Das qualitative Relationssystem $\mathcal{M} := (M, D)$ ist eine *Differenzstruktur*.
- Das numerische Relationssystem ist $\mathcal{N} := (\mathbb{R}, \Delta)$
- Es existiert ein Homomorphismus f von \mathcal{M} nach \mathcal{N} und $(\mathcal{M}, \mathcal{N}, f)$ heißt *Intervallskala*
- Die Menge der zulässigen Transformationen von f ist die Menge der *positiv linear-affinen Funktionen*
- Größenvergleiche von Messwerten und Größenvergleiche von Messwertdifferenzen sind bedeutsam
- Messwertverhältnisaussagen sind i.A. nicht bedeutsam
- Summenstatistiken sind i.A. nicht bedeutsam, Mittelwertstatistiken sind bedeutsam

Extensivmessung

- Das qualitative Relationssystem $\mathcal{M} := (M, R, \circ)$ ist eine *extensive Struktur*
- Das numerische Relationssystem ist $\mathcal{N} := (\mathbb{R}, >, +)$
- Es existiert ein Homomorphismus f von \mathcal{M} nach \mathcal{N} und $(\mathcal{M}, \mathcal{N}, f)$ heißt *Verhältnisskala*
- Die Menge der zulässigen Transformationen von f ist die Menge der *Ähnlichkeitstransformationen*
- Größenvergleiche, Größenvergleiche von Messwertdifferenzen und Verhältnisaussagen sind bedeutsam
- Summen- und Mittelwertstatistiken sind bedeutsam

Vorbemerkungen

Schwächen der Repräsentationstheorie

Alternative Messtheorien

Die Repräsentationstheorie des Messens spielt in der Psychometrie keine Rolle

Skalennomenklatur nach Döring and Bortz (2016)

Psychometrische Skala

“Unter eine *psychometrischen Skala*, Testskala oder kurz: Skala versteht man einen Satz von Skalen-Items (Selbstaussagenfragen, Aussagen oder Aufgaben samt Antwortformat), die dazu dienen, gemeinsam ein latentes Merkmal zu messen (...). Eine psychometrische Skala muss erwiesenermaßen messtheoretischen Skalierungseigenschaften und Gütekriterien (Objektivität, Reliabilität, Validität) genügen”.

Likert Skala

“Eine Likert-Skala ist eine psychometrische Skala, die aus mehreren (z.B. 10 - 20) Aussagen besteht, die alle dasselbe Merkmal messen und auf Ratingskalen hinsichtlich des Grades der Zustimmung einzuschätzen sind. Die Likert-Items repräsentieren unterschiedliche Intensitäten des gemessenen Merkmals. Der Skalenwert der Likert-Skala berechnet sich als Summen- oder Durchschnittsscore der einzelnen Ratings”.

Rating Skala

“Ratingskalen als Antwortformate für mündliche oder schriftliche Fragen bzw. Aussagen geben (durch Zahlen, verbale Beschreibungen, Beispiele etc.) markierte Abschnitte eines Merkmalskontinuums vor, die die Urteilenden als gleich groß bewerten sollen (...).”

Schwächen der Repräsentationstheorie

Die Repräsentationstheorie des Messens spielt in der Psychometrie keine Rolle

Beispiel Lehrevaluation

Messung der subjektiven Wissenserweiterung durch Studiengangsmodule

Folgende Module haben mein Wissen erweitert

	Trifft voll zu (1)	Trifft zu (2)	Trifft nicht zu (3)	Trifft gar nicht zu (4)
Modul A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Modul B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Modul C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Modul D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Modul E	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Im Folgenden skizzieren wir, wie empirisch die Ordinal- oder Verhältnisskalenart dieses Messvorgangs nachgewiesen werden müsste. De-facto wird aber in der Praxis lediglich angenommen, dass es sich bei diesem Messvorgang um eine Messung mit Ordinal- oder Intervallskalenart handelt ("Messung per fiat"), die unten skizzierte Herangehensweise zur Überprüfung der Eigenschaften des qualitativen Relationssystems findet in der Regel nicht statt.

⇒ Die Repräsentationstheorie des Messens wird in der Anwendung ignoriert.

Die Repräsentationstheorie des Messens spielt in der Psychometrie keine Rolle

Messtheoretisch-fundierte Etablierung der Ordinalskalenart

Es sei $\mathcal{M} := (M, R)$ das qualitative Relationssystem mit

- der Menge der modulspezifischen Wissenserweiterungseinschätzungen $M := \{A, B, C, D, E\}$,
- der Wissenserweiterungseinschätzungrelation $R \subseteq M \times M$ mit der Bedeutung

$$(m, n) \in R \Leftrightarrow \text{Modul } m \text{ hat mein Wissen mehr erweitert als Modul } n. \quad (1)$$

Es sei $f : M \rightarrow \mathbb{R}$ die Auswahl einer der Antwortmöglichkeiten in einer Zeile obigen Fragebogens. Damit f ein Homomorphismus nach $\mathcal{N} := (\mathbb{R}, >)$ ist und $(\mathcal{M}, \mathcal{N}, f)$ eine Ordinalskala, muss R eine strenge schwache Ordnung sein, d.h. R müsste insbesondere asymmetrisch und transitiv sein.

Hinsichtlich der Asymmetrie müsste gewährleistet sein, dass es keine Modulpaare (m, n) gibt, bei denen sowohl Modul m wissenserweiternder als Modul n und gleichzeitig Modul n wissenserweiternder als Modul m eingeschätzt wird. Hinsichtlich der Transitivität müsste gewährleistet sein, dass wenn immer ein Modul m wissenserweiternder als ein Modul n und ein Modul n wissenserweiternder als Modul p eingeschätzt wird, Modul m auch wissenserweiternder als Modul p eingeschätzt wird.

Vor Durchführung einer Befragung mit obigem Fragebogen als Ordinalskala müsste also die Wissenserweiterungseinschätzungrelation durch Paarvergleiche empirisch validiert werden. Schließlich hätten bei Etablierung der Ordinalskalenart die Werte 1,2,3,4 über ihre Rangfolge hinaus keine Bedeutung, d.h. die Antwortmöglichkeiten könnten auch mit 10,11,16,20 oder 0.1,0.11,0.111,0.1111 repräsentiert sein.

Schwächen der Repräsentationstheorie

Die Repräsentationstheorie des Messens spielt in der Psychometrie keine Rolle

Messtheoretisch fundierte Etablierung der Verhältnisskalenart

Es sei $\mathcal{M} := (M, R, \circ)$ das qualitative Relationssystem mit

- der Menge der modulspezifischen Wissenserweiterungseinschätzungen $M := \{A, B, C, D, E\}$,
- der Wissenserweiterungseinschätzungrelation $R \subseteq M \times M$ mit der Bedeutung

$$(m, n) \in R \Leftrightarrow \text{Modul } m \text{ hat mein Wissen mehr erweitert als Modul } n, \quad (2)$$

- der Wissenserweiterungseinschätzungskombination $\circ : M \times M \rightarrow M$ mit der Bedeutung

$$m \circ n \text{ ist meine Wissenserweiterung durch Kombination der Module } m \text{ und } n. \quad (3)$$

Es sei $f : M \rightarrow \mathbb{R}$ die Auswahl einer der Antwortmöglichkeiten in einer Zeile obigen Fragebogens. Damit f ein Homomorphismus nach $\mathcal{N} := (\mathbb{R}, >, +)$ ist und $(\mathcal{M}, \mathcal{N}, f)$ eine Verhältnisskala, muss \mathcal{M} eine extensive Struktur sein, es müsste also beispielsweise die Monotonie

$$(m, n) \in R \Leftrightarrow (m \circ p, n \circ p) \in R \quad (4)$$

gelten. Es müsste also gewährleistet sein, dass wenn immer die Wissenserweiterung durch Modul m im Vergleich zu n als größer eingeschätzt wird, die Wissenserweiterung durch die Kombination von Modul m mit Modul p im Vergleich zur Kombination von Modul n mit Modul p größer eingeschätzt wird.

Vor Durchführung einer Befragung mit obigem Fragebogen als Verhältnisskala müsste also die Wissenserweiterungseinschätzungrelation und -kombination auf ihre Monotonieeigenschaft empirisch überprüft werden. Wenn dies möglich ist, bedeutet zum Beispiel die Antwort "trifft zu" gegenüber der Antwort "trifft nicht zu", dass die eingeschätzte Wissenserweiterung nur halb so groß ist.

Die Repräsentationstheorie des Messens ist in die Datenanalyse nicht integriert

Die Repräsentationstheorie des Messens modelliert keine Messfehler

Die klassische Repräsentationstheorie ist deterministisch, nimmt also die Abwesenheit von Messfehlern an. Dies macht die Repräsentationstheorie *per se* ungeeignet für die Datenanalyse, da ein einziger von den Vorhersagen der Repräsentationstheorie abweichender Datenpunkt ihre gesamte Validität in Frage stellen würde. Es gibt jedoch Ansätze, die Repräsentationstheorie probabilistisch zu erweitern und in die moderne Sicht der Datenanalyse als Unsicherheitsquantifizierung zu integrieren, so zum Beispiel bei vgl. Falmagne (1980) und Rossi (2014).

Die Repräsentationstheorie des Messens ist aus Sicht der Statistik prohibitiv

Die enge Beziehung von Skalenart und bedeutsamen Statistiken ist auf viel Kritik und Gegenwehr von Seiten der mathematischen Statistik gestoßen. Beispiele hierfür sind die Beiträge von Lord (1953), Gaito (1980), Borgatta and Bohrnstedt (1980), Davison and Sharma (1988), Velleman and Wilkinson (1993), Hand (1996), Norman (2010). In modernen datenanalytischen Lehrbüchern findet die Repräsentationstheorie des Messens nicht statt. In der Psychologie ist dies auch gut nachvollziehbar: wenn die Skalenart eines Messinstrumentes sowieso nur angenommen wird, dann kann auch die Datenanalyse die Repräsentationstheorie getrost ignorieren.

Schwächen der Repräsentationstheorie

Prävalenz der Repräsentationstheorie des Messens in der Psychologie

Trotz der genannten Schwächen finden sich Einführungen in die RTM in allen gängigen deutschen Lehrbüchern zum Them "Statistik für Psycholog:innen, meist aber ohne kritische Diskussion, genauere Definitionen oder etwa Beweise (vgl. Bortz and Schuster (2010), Kubinger, Rasch, and Yanagida (2011), Wirtz and Nachtigall (2012), Eid, Gollwitzer, and Schmitt (2017), Bühner and Ziegler (2017), Sedlmeier and Renkewitz (2018)). Es stellt sich also die Frage, warum die RTM scheinbar populär bleibt und zumindest der Begriff der Messtheorie auch seinen Weg in die Approbationsordnung für Psychotherapeut:innen und damit alle gängigen BSc und MSc Psychologie Programme gefunden hat.

Wissenschaftspolitischer Benefit des Messtheorie nach Stevens (1946)

Der wissenschaftspolitische Vorteil der Definition des Messens nach Stevens (1946), dass "Messen die Zuordnung von Zahlen zu Objekten anhand von Regeln ist" zeitgenössischen Psycholog:innen erlaubte, weiterhin "messen zu dürfen" und scheinbar gegenüber der Physik ein tieferes Verständnis dafür entwickelt zu haben, was "Messen" denn eigentlich bedeutet.

Michell (2021) schreibt dazu pointiert: "Ask psychologists what measurement is and the answer will be "assignment of numerals to objects or events according to rule" (Stevens, 1946, p. 677) or some variant of this meme. Ask psychologists for examples and nominal, ordinal, interval, and ratio scales will be mentioned—these are also Stevens' terms. Fragmented as psychology is, it stands united behind Stevens' theory of measurement. Theories of measurement purport to explain the position of measurement in science. In the 1940s, psychologists craved a theory meeting their aspirations and they felt besieged because some members of an expert committee investigating mental measurement (Ferguson et al., 1940), including Campbell, rejected its possibility. Stevens delivered his theory in the nick of time and the spirit of triumph pervading his announcement not only eased their minds; psychologists suddenly felt ahead of the game in understanding measurement. And Stevens smuggled in a metaphysical dividend, blending operationism and logical positivism, both of which, despite disclaimers to the contrary, are metaphysical doctrines.¹⁶ He gave psychologists an apparently incontestable definition of measurement and a plausible rationale for it, which was actually metaphysics, masquerading as avantgarde, scientific philosophy. Psychologists welcomed their medicine, thinking it was an antiphilosophical prophylaxis; never dreaming it was metaphysical bromide."

Vorbemerkungen

Schwächen der Repräsentationstheorie

Alternative Messtheorien

Alternative Messtheorien

Klassizistische Positionen

Klassizistische Positionen (Michell (1986), Michell (2021)) sind explizit realistisch und nehmen an, dass es in der Welt Attribute von Dingen gibt, die explizit "quantitativ" sind. Messen im klassizistisch-realistischem Sinne ist das Schätzen des Verhältnisses eines quantitativen Attributs in Bezug auf eine relevante Maßeinheit.

Die klassizistische Position differenziert zwischen realen Objekten und metaphysischen Attributen von Objekten (vgl. Mundy (1987), Swoyer (1987)). So ist zum Beispiel eine physische Konkatenierungsoperation Evidenz für den quantitativen Gehalt des Attributs "Länge", aber das Fehlen einer solchen Konkatenierungsoperation sei kein Beweis für die Nicht-Quantitativität eines Attributs. Letztlich sind klassizistische Positionen eine Rückwendung auf Messkonzepte, wie sie vor Beginn des 20. Jahrhunderts vertreten wurden und dabei recht fundamentalistisch.

Positivistisch-operationalistische Positionen

Der Operationalismus definiert wissenschaftliche Konzepte anhand der Operationen, die zu ihrer Messung nötig sind: "We mean by any concept nothing more than a set of operations, the concept is synonymous with the corresponding set of operations" (Bridgman (1927)). Ein Repräsentationsproblem stellt sich in diesem Kontext nicht, es wird nichts repräsentiert. Der Ansatz von Stevens (1946) war explizit operationalistisch, auch wenn er die Grundlage für die Repräsentationstheorie legte: jede Prozedur, die eine Zahl liefert wird als Messung verstanden, die homomorphe Repräsentation eines zugrundeliegenden realen qualitativen Relationssystems wird nicht gefordert.

Der Operationalismus und seine psychologische Ausprägung als Behaviourismus sind in der Psychologie umstritten. Einerseits ist es einsichtig, dass manche Fragebögen zum Beispiel zur Lebensqualität eine operationalistische Definition des zugrundeliegenden Konstrukts bieten, andere dagegen tatsächlich existierende nicht direkt beobachtbare Phänomene abbilden, man denke zum Beispiel an Tests kognitiver Fähigkeiten. Allgemein scheint die Fragebogenforschung recht operationalistisch geprägt.

Modell-basierter Realismus

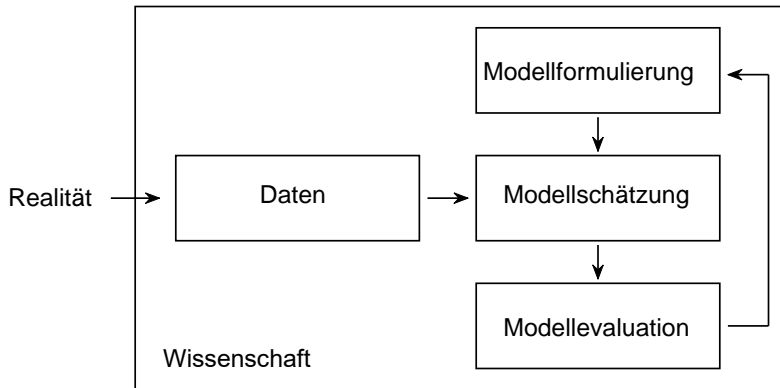
“In summary, a model-based realist perspective on measurement at the same time maintains something of each of the previous positions but rejects their most radical aspects:

- It accepts from realism the position that some properties do exist in the world, and are not just human constructs, but rejects the metaphysical claim that values of properties exist independently of our models.
- It accepts from non-realist empiricism (positivism, operationalism, representationalism) that empirical data can provide the evidential foundation for knowledge, but rejects the foundationalist claim that generic observation can have such a role, which is instead vested only in the empirical component of measurement systems, which are specifically designed for this purpose; moreover, it accepts that such evidence is always revisable.
- It accepts from pragmatism (and relativism) that measurement is a designed-on-purpose process and that models in measurement are unavoidable, but rejects the possible conclusion that “anything goes” and that the quality of measurement results can be evaluated only a posteriori, in terms of the effectiveness of their application.”

Aus Mari, Wilson, and Maul (2021), Measurement across the Sciences, Kapitel 4

Siehe auch Tal (2020) und Measurement at the Crossroads 2020 - 2022: Measuring and Modeling

Modell-basierter Realismus



Vorbemerkungen

Schwächen der Repräsentationstheorie

Alternative Messtheorien

Referenzen

- Borgatta, Edgar F., and George W. Bohrnstedt. 1980. "Level of Measurement: Once Over Again." *Sociological Methods & Research* 9 (2): 147–60. <https://doi.org/10.1177/004912418000900202>.
- Bortz, Jürgen, and Christof Schuster. 2010. *Statistik Für Human- Und Sozialwissenschaftler*. 7., vollständig überarbeitete und erweiterte Auflage. Springer-Lehrbuch. Berlin Heidelberg: Springer.
- Bridgman, P. W. 1927. *The Logic of Modern Physics*. New York: MacMillan.
- Bühner, Markus, and Matthias Ziegler. 2017. *Statistik Für Psychologen Und Sozialwissenschaftler*. 2., aktualisierte und erweiterte Auflage. Ps Psychologie. Hallbergmoos: Pearson Deutschland GmbH.
- Davison, Mark L., and Anu R Sharma. 1988. "Parametric Statistics and Levels of Measurement."
- Döring, Nicola, and Jürgen Bortz. 2016. *Forschungsmethoden Und Evaluation in Den Sozial- Und Humanwissenschaften*. Springer-Lehrbuch. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-41089-5>.
- Eid, Michael, Mario Gollwitzer, and Manfred Schmitt. 2017. *Statistik Und Forschungsmethoden: Mit Online-Materialien*. 5., korrigierte Auflage. Weinheim Basel: Beltz.
- Falmagne, Jean-Claude. 1980. "A Probabilistic Theory of Extensive Measurement." *Philosophy of Science* 47 (2): 277–96. <https://doi.org/10.1086/288933>.
- Gaito, John. 1980. "Measurement Scales and Statistics: Resurgence of an Old Misconception."
- Hand, D. J. 1996. "Statistics and the Theory of Measurement." *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 159 (3): 445. <https://doi.org/10.2307/2983326>.
- Kubinger, Klaus D., Dieter Rasch, and Takuya Yanagida. 2011. *Statistik in Der Psychologie: Vom Einführungskurs Bis Zur Dissertation*. Psychlehrbuch Plus. Göttingen Bern Wien: Hogrefe.
- Lord, Frederic M. 1953. "On the Statistical Treatment of Football Numbers." *American Psychologist* 8 (12): 750–51. <https://doi.org/10.1037/h0063675>.
- Mari, Luca, Mark Wilson, and Andrew Maul. 2021. *Measurement Across the Sciences: Developing a Shared Concept System for Measurement*. Springer Series in Measurement Science and Technology. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-65558-7>.
- Michell, Joel. 1986. "Measurement Scales and Statistics: A Clash of Paradigms." *Psychological Bulletin* 100 (3): 298–407.
- . 2021. "Representational Measurement Theory: Is Its Number Up?" *Theory & Psychology* 31 (1): 3–23. <https://doi.org/10.1177/0959354320930817>.
- Mundy, Brent. 1987. "The Metaphysics of Quantity." *Philosophical Studies* 51 (1): 29–54. <https://doi.org/10.1007/BF00353961>.
- Norman, Geoff. 2010. "Likert Scales, Levels of Measurement and the 'Laws' of Statistics." *Advances in Health Sciences Education* 15 (5): 625–32. <https://doi.org/10.1007/s10459-010-9222-y>.
- Rossi, Giovanni Battista. 2014. *Measurement and Probability: A Probabilistic Theory of Measurement with Applications*. Springer Series in Measurement Science and Technology. Dordrecht: Springer Netherlands. <https://doi.org/10.1007/978-94-017-8825-0>.
- Sedlmeier, Peter, and Frank Renkewitz. 2018. *Forschungsmethoden Und Statistik Für Psychologen Und Sozialwissenschaftler*. 3., aktualisierte und erweiterte Auflage. Hallbergmoos: Pearson.
- Stevens, S. S. 1946. "On the Theory of Scales of Measurement." *Science, New Series* 103 (2684): 677–80. <http://www.jstor.org/stable/1671815>.
- Swoyer, Chris. 1987. "The Metaphysics of Measurement." In *Measurement, Realism and Objectivity*, edited by John Forge, 235–90. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-009-3919-6_8.
- Tal, Eran. 2020. "Measurement in Science." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2020. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2020/entries/measurement-science/>.
- Velleman, Paul F., and Leland Wilkinson. 1993. "Nominal, Ordinal, Interval, and Ratio Typologies Are Misleading." *The American Statistician* 47 (1): 65. <https://doi.org/10.2307/2684788>.
- Wirtz, Markus Antonius, and Christof Nachtigall. 2012. *Statistische Methoden Für Psychologen. 1: Deskriptive Statistik / Markus A. Wirtz ; Christof Nachtigall*. 6., überarb. Aufl. Weinheim: Beltz [u.a.].