

## Vertrauenswürdigkeit von Messungen als Brücke zu Messunsicherheiten

### Ausgangssituation und Fragestellungen

Der Themenkreis Messunsicherheiten nach GUM (ISO/IEC Guide 98-3:2008) wird nicht nur an der Universität von Studierenden, die am Anfang ihrer akademischen Laufbahn stehen, oft stiefmütterlich behandelt, sondern wird in den Schulen im Physikunterricht leider nur sehr oberflächlich thematisiert oder ganz ausgelassen (vgl. Nagel et al., 2021). Um dieser wesentlichen experimentellen Kompetenz in Schulen mehr Aufmerksamkeit zukommen zu lassen, bedurfte es praktikabler Unterrichtskonzeptionen. Solche haben Loidl (2021) für die Sekundarstufe I und Bärenthaler-Pachner (2022) für die Sekundarstufe II mit dem methodischen Rahmen des Design-Based Research (vgl. Haagen-Schützenhöfer, 2015) entwickelt und untersucht. Zentrales domänenspezifisches Designprinzip in beiden Konzeptionen ist die Beurteilung der Vertrauenswürdigkeit von Messungen, Messgeräten und Ergebnissen. Dies hat sich als in der Praxis gut funktionierende und von den Schülerinnen und Schülern beider Altersstufen gut akzeptierte Brücke aus der Alltagswelt zum Fachbegriff der Messunsicherheiten herausgestellt. Allerdings äußert Loidl (2021) für die Sekundarstufe I basierend auf den Ergebnissen der Pretesting-Interviewstudien und Akzeptanzbefragungen mit den Schülerinnen und Schülern der Sekundarstufe I die Bedenken, es könnte das Konzept zur Vertrauenswürdigkeit in dieser Altersgruppe sehr stark anthropozentrisch ausgeprägt sein. Sie schlägt aufgrund der Dominanz alltagsweltlicher Präkonzepte in dieser Altersgruppe jedenfalls weitere Untersuchungen vor. Dem wurde mit der Entwicklung eines quantitativen Befragungsinstruments zum Konzept der Vertrauenswürdigkeit nach der Four Building Blocks Methode nach Wilson (2005) Rechnung getragen.

Die zentralen Fragestellungen an die Entwicklung des Befragungsinstruments sind:

- Was bedeutet der Begriff Vertrauenswürdigkeit für Schülerinnen und Schüler der Sekundarstufe I
- Welches durch alltagsweltliche Präkonzepte indizierte gedankliche Konstrukt steckt hinter dem Begriff Vertrauenswürdigkeit?
- Ist das Begriffskonstrukt in geeigneter Weise anknüpfungsfähig für den Vertrauenswürdigkeitsbegriff von Messungen?
- Wie gut lässt sich das Konstrukt mit dem Begriff der Messunsicherheit verbinden?
- Werden Medien, Personen, Institutionen/ etc. die Eigenschaften vertrauenswürdig oder nicht vertrauenswürdig zugesprochen?
- (In Abhängigkeit von der vorherigen Frage) Welche Eigenschaften müssen Medien/Personen/Institutionen besitzen, damit sie als vertrauenswürdig eingestuft werden?

### Four Building Blocks als methodischer Rahmen zur Fragebogenentwicklung

Die Four-Building-Blocks-Methode (nach Wilson, 2005) stellt den methodischen Rahmen der Studie dar, mit dessen Hilfe es ermöglicht werden soll, ein Messinstrument zu erstellen, indem verstanden wird, welche Konstrukte und Dimensionen hinter der „Vertrauenswürdigkeit“ bei Sek I Schülerinnen und Schüler liegt.

Beim ersten Block wird eine „Construct Map“ z.B. durch Literaturrecherche basierend auf den unterschiedlichen Ausprägungen (Dimensionen) eines Konstrukts erstellt.

Beim zweiten Block, dem „Item Design“ muss erarbeitet werden, wie das theoretische Konstrukt mit Situationen aus dem Leben der Befragten zusammenhängt, woraus die einzelnen Items entstehen. Hier kommen umfangreiche Pretesting-Methoden wie Expertenbefragung, Expertendiskussion, 2-stufiges Brainstorming mit 6 Schulklassen und 14 Think-Aloud-Interviews (vgl. Krüger et al. 2014) inkl. Überprüfung der Intercoder-Reliabilitäten (Fleiss'  $\kappa$ , vgl. O'Connor & Joffe 2020) zum Einsatz.

Beim dritten Block „Outcome Space“ wird das beobachtbare Verhalten, in unserem Fall die Antworten auf die Items, kategorisiert und mit bestimmten Werten versehen. Variablen werden definiert und mit Skalen versehen.

Mit dem vierten Block „Measurement Model“ wird eine Möglichkeit gefunde, die Werte, welche man beim Block 3, dem Outcome Space erhält, wieder auf das Konstrukt der Construct Map (oder ihrer Weiterentwicklung) zurückzuführen. Ein möglicher Ansatz ist eine Fokussierung auf die Items und ihre Beziehungen zu dem Konstrukt, gemessen durch die Faktoranalyse (Backhaus et al. 2006), was mit einer quantitativen Stichprobe von 273 Schülerinnen und Schülern der Schulstufe 5-8 (Alter 10-15 Jahre) aus 12 Schulklassen in 4 verschiedenen Schulen durchgeführt wurde. Ungeeignete Items werden so identifiziert und ggf. entfernt. Die überarbeiteten Skalen werden sodann mit der Berechnung von Cronbachs  $\alpha$  auch auf ihre Reliabilität untersucht sowie Abgrenzungen zu anderen Konstrukten mit dem t-Test auf signifikante Unterschiede untersucht (Janssen & Laatz, 2007).

### **Fragebogendesign**

Die ursprünglich 8 Dimensionen, die sich in der Literaturrecherche der Construct Map ergaben, wurden im Zuge der qualitativen Pretestings zu 2 (eigentlich 4) Dimensionen reduziert:

- menschlich/Bezugspersonen  $\leftrightarrow$  menschlich/digital (z.B. social media)
- künstlich/dinghaft (z.B. Messgeräte, Maschinen)  $\leftrightarrow$  künstlich/nicht dinghaft (z.B. Internet)

Innerhalb dieser Dimensionen wurde der Focus in der Fragestellung der 22 fünfstufig Likertskalierten Items auf die Lebensrealitäten der Sek I - Schülerinnen und Schüler gelegt und bewusst auch auf die Beurteilung der Vertrauenswürdigkeit von Messgeräten, Messprozessen, Messmethoden und messenden Personen gesetzt. Zudem wurde die Abgrenzung der Vertrauenswürdigkeit zu gegenteiligen bzw. ähnlichen Konzepten oder solchen, die nichts mit einander zu tun haben, mit 11 nominalskalierten Items getestet.

### **Ergebnisse**

Die Entwicklung des Fragebogens wird durch die Auswertung der Ergebnisse sowie Überprüfung der Konstrukt-Validität und -Reliabilität im Zuge der quantitativen Befragung abgeschlossen. Die Faktoranalyse (Eignungsprüfung mittels KMO-Kriterium: 0,71 bzw. 0,72; Rotationsmethode der Strukturmatrix: Oblimin mit Kaiser-Normalisierung (vgl. Backhaus et al. 2006)) ergab in der Dimension „menschlich/Bezugspersonen“ drei deutliche Faktoren:

- vertrauenswürdige Personen,
- nicht vertrauenswürdige Personen,
- messende Personen.

In der Dimension künstlich/dinghaft versus künstlich/nicht dinghaft wurden diese beiden Ausprägungen exakt als Faktoren wiedergefunden:

- physische Informationsquellen (= künstlich/dinghaft),
- nicht-physische Informationsquellen (= künstlich/nicht dinghaft).

Wobei hier zwei Items aus dem Fragebogen entfernt werden mussten, da sonst die Konstruktvalidität nicht ausreichend gegeben wäre.

Alle so zusammengefassten Skalen haben die gleiche Reliabilität im Cronbach  $\alpha$ -Test von  $\alpha = 0,74$ .

Erste Ergebnisse der deskriptiven Analyse der Antworten in der quantitativen Befragung zeigen, dass das Konzept der Vertrauenswürdigkeit bei den Sek I – Schülerinnen und Schülern deutlich ausgeprägt ist. Abb. 1 zeigt die Ergebnisse der Abgrenzung der Vertrauenswürdigkeit zu gegenteiligen bzw. ähnlichen Konzepten oder solchen, die nichts mit einander zu tun haben. Man kann deutlich erkennen, dass ähnliche Konzepte überwiegend als ähnlich beurteilt werden, gegenteilige als gegenteilig und bezugslose Konzepte als ebensolche.

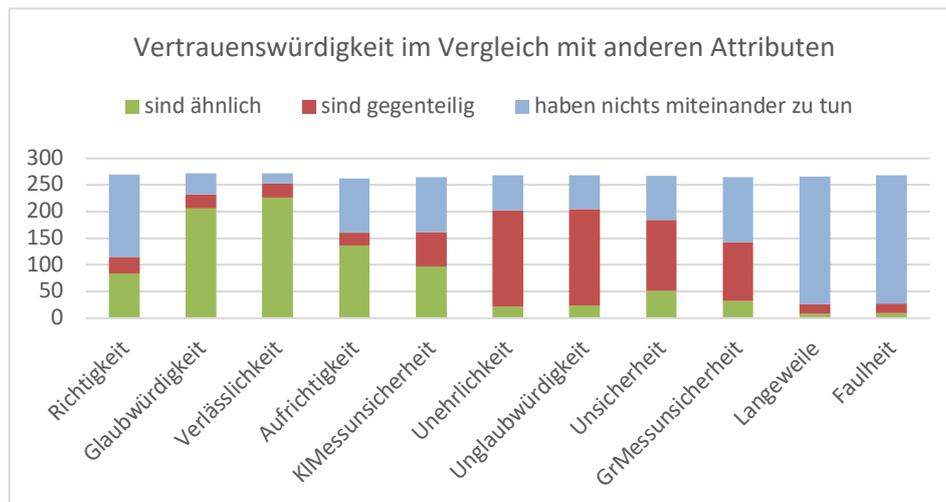


Abb. 1: Abgrenzung Vertrauenswürdigkeit zu anderen Konzepten

Auffällig ist, dass bei den drei Begriffen aus dem Konzept der Messunsicherheiten (Richtigkeit, kleine und große Messunsicherheit) die Einschätzung der Bezugslosigkeit dominiert. Abgesehen davon ordnet der Rest der Befragten Ähnlichkeit bzw. Gegenteiligkeit dieser drei Begriffe signifikant richtig der Vertrauenswürdigkeit zu.

### Diskussion

Die Dimensions-Erweiterung bei menschlich/Bezugsperson durch den Faktor „messende Person“ könnte zum Schluss verleiten, dass die Vertrauenswürdigkeit „messender Personen“ nicht beurteilt werden kann. Detailliertere Ergebnisse einzelner Items zeigen aber, dass die Vertrauenswürdigkeit sehr wohl richtig bei messenden Personen beurteilt wird (z.B. „Kindergartenkind stoppt Zeit“ versus „Sportlehrerin stoppt Zeit“ – hier wird die Messung des Sportlehrers signifikant vertrauenswürdiger eingeschätzt). Zudem zeigt die Ähnlichkeits/Gegenteil-Abfrage die deutlich richtige Ausbildung des Konzepts der Vertrauenswürdigkeit. Somit kann die Entwicklung des Testinstruments als gelungen betrachtet werden. Außerdem erhält die lokale Theorie im Rahmen der Design-Based - Research Studien zur Entwicklung der Unterrichtskonzeption für die Einführung von Messunsicherheiten in der Sek I die empirische Bestätigung, dass mit „Vertrauenswürdigkeit“ als zentrales domänenspezifisches Designprinzip ein geeignetes Brückenkonzept zwischen Alltags- und Fachsprache gewählt wurde.

## Literatur

- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2006). *Multivariate Analysemethoden*. Berlin & Heidelberg: Springer Verlag.
- Bärenthaler-Pachner, R. (2022). *Entwicklung und Evaluation einer Lernumgebung zum Thema Messunsicherheit in der Sekundarstufe II*. Masterarbeit an der Universität Wien.
- Haagen-Schützenhöfer, C. (2015). *Kumulative Habilitationsschrift mit dem Schwerpunkt Lehr- und Lernprozesse im Anfangsoptikunterricht der Sekundarstufe I* (Habilitationsschrift, Universität Wien). Universität Wien.
- ISO/IEC Guide 98-3:2008 Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement (GUM:1995) (Corrected version 2010-11). [www.austrian-standards.at](http://www.austrian-standards.at) download am 2023-02-10
- Janssen, J. & Laatz, W. (2007). *Statistische Datenanalyse mit SPSS für Windows*. Berlin & Heidelberg: Springer
- Krüger, D., Parchmann, I. & Schecker, H. (Hrsg.). (2014). *Methoden in der naturwissenschaftsdidaktischen Forschung*. Springer Spektrum: Berlin Heidelberg.
- Loidl, H. (2021). *Entwicklung und Evaluation von Unterrichtseinheiten zum Thema Messunsicherheiten*. Masterarbeit an der Universität Wien.
- Nagel, C., Lux, B., Steindl, S. (2021) Die Thematisierung von Messunsicherheiten im Physikunterricht – Eine Umfrage. In: Neumann, S. & Nagel C. (Hrsg.) *Plus Lucis 4/2021: Messunsicherheiten - Sicher ist sicher!* VFPC: Wien. pp 7-11.
- O'Connor, C. & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods* , 19 , 160940691989922. <https://doi.org/10.1177/1609406919899220>
- Wilson, M. (2005). *Constructing Measures - An Item Response Modelling Approach* . New York: Taylor & Francis.