

Johannes Schulz<sup>1</sup>  
Burkhard Priemer<sup>1</sup>

<sup>1</sup>Humboldt-Universität zu Berlin

## **Entwicklung eines Messinstruments zur Erfassung von Kompetenzen im Umgang mit Messunsicherheiten**

Für die Arbeit in Naturwissenschaften und Technik ist der Umgang mit Messunsicherheiten essentiell. Genauer formuliert (Möhrke et al. 2018): „Der Umgang mit Messunsicherheiten ist in den experimentellen Naturwissenschaften von zentraler Bedeutung, da erst sie ein Ergebnis bewert- und vergleichbar machen.“ Weiterhin ist bekannt, dass Schüler\*innen und Studierende oftmals Schwierigkeiten im Umgang mit Messunsicherheiten haben (Priemer & Hellwig 2018; Ludwig, Priemer & Lewalter 2018). Um diesen Schwierigkeiten zu begegnen, sind geeignete Instruktionen notwendig. Deren Wirksamkeit - operationalisiert durch Kompetenzen - muss wiederum evaluiert werden können. Mit diesem Beitrag soll dazu ein Testinstrument vorgestellt werden, das Kompetenzen im Umgang mit Messunsicherheiten auf Basis des Sachstrukturmodells nach Hellwig (2012) zielgenau erfassen kann.

### **Grundlagen und Forschungsfragen**

Hellwig (2012) stellt ein durch Expert\*innen aus dem Messwesen validiertes Sachstrukturmodell im Bereich Messunsicherheiten vor, das das Themenfeld in vier Dimensionen und zehn Konzepte gliedert (vgl. Abb. 1). Diese Konzepte teilen sich weiter in Unterasspekte auf.

<b>Grundsätzliche Existenz von Messunsicherheiten</b>	Ursachen der Messunsicherheit	<b>Einfluss auf das Messwesen</b>	Ziel der Messung
	Unterscheidung zw. Messunsicherheit und Messabweichung		Ergebnis der Messung
<b>Aussagekraft</b>	Verlässlichkeit der Messung und ihres Ergebnisses	<b>Erfassung von Messunsicherheiten</b>	Erfassung einer Unsicherheitskomponente bei direkter Messung
	Vergleich von Messwerten		Zusammensetzung der Messunsicherheit aus mehreren Komponenten
	Regression		Erweiterte Messunsicherheit

*Abb.1: Auszug aus dem Sachstrukturmodell nach Hellwig (2012)*

Auf dieser Basis wurde u.a. die folgende Forschungsfrage formuliert:

1. Inwiefern lassen sich die zehn Konzepte des (unreduzierten) Sachstrukturmodells nach Hellwig (2012) in Form von Kompetenzen beschreiben, operationalisieren und messen?

### **Methode**

In einem ersten Schritt wurden zur Beantwortung der Forschungsfragen zu den einzelnen Konzepten des Sachstrukturmodells Kompetenzen formuliert. Anschließend wurden die Kompetenzen durch Testaufgaben operationalisiert. Hier wurde auf eine möglichst breite inhaltliche Abdeckung der Konzepte nach Hellwig (2012) geachtet, sowie möglichst ein

großes Spektrum von Kontexten, Feldern der Physik, verschiedener gängiger Darstellungen und Repräsentationen integriert. In den Items wurde außerdem versucht, sowohl theoretische Inhalte als auch konkrete Experimentiersituationen abzudecken. In der Formulierung folgt die Entwicklung der Items weitestgehend den Empfehlungen von Haladyna (2004) und Downing und Haladyna (2006). Insgesamt wurden auf diese Weise 145 Testaufgaben formuliert, so dass zu jedem Konzept nach Hellwig (2012) 10-15 Items zur Verfügung stehen. In einem zweiten Schritt wurde dann zur Validierung eine Teilmenge von 35 Testaufgaben sechs Expert\*innen vorgelegt. Diese sollten die Testaufgaben den formulierten Kompetenzen zuordnen. Zusätzlich gab es für die Expert\*innen die Möglichkeit, Testaufgaben als nicht zuordenbar zu markieren. In einem letzten Schritt wurden die Testaufgaben mit 1299 Studierenden in der Anfangsphase des Studiums pilotiert. Dabei enthielt ein Testheft jeweils Testaufgaben zu zwei verschiedenen Konzepten innerhalb einer Dimension nach Hellwig (2012). Diese Zusammensetzung war gleichmäßig permutiert, ebenso wurden die Testaufgaben in der Hälfte der Testhefte in umgekehrter Reihenfolge gestellt. Weiterhin wurde den Testaufgaben eine Seite mit Erläuterungen und den wesentlichen Regeln im Themenfeld vorangestellt, um eine einheitliche Sprachgrundlage zu schaffen und möglicherweise unklare Begriffe zu erläutern.

### Ergebnisse

Zur Auswertung des Expert\*innenratings wurden die Übereinstimmungen für die einzelnen Testaufgaben betrachtet, aber auch das Antwortverhalten der Expert\*innen auf mögliche Unterschiede untersucht. In letzterem traten keine Besonderheiten derart auf, dass die Rater\*innen systematisch voneinander abweichen würden. Zur Übereinstimmung der Zuordnung wurden die Kennwerte Fleiss Kappa, Congers Kappa und Lights Kappa herangezogen. Alle diese Werte deuten mit Werten über 0.86 auf eine sehr hohe Übereinstimmung der Rater\*innen hin.

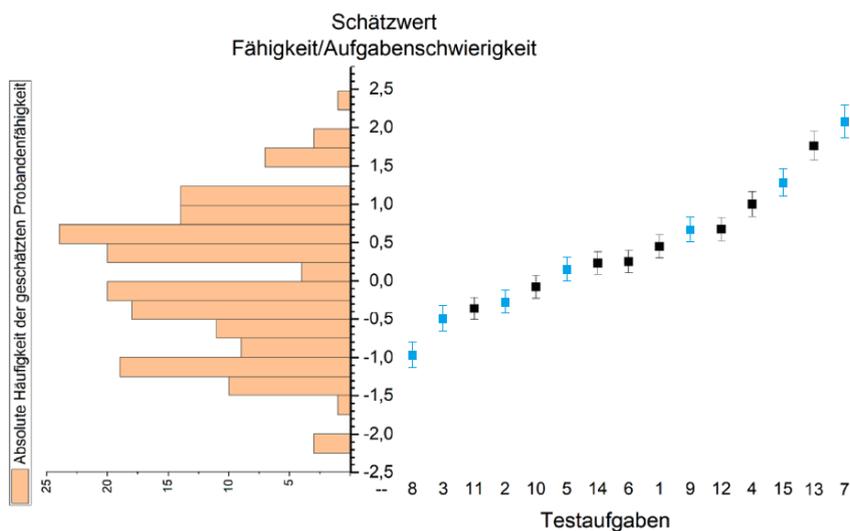


Abb. 2: Wright-Map zum Konzept „Ursachen der Messunsicherheit“

Die Ergebnisse der Erhebung mit Studierenden zu den einzelnen Konzepten des Sachstrukturmodells wurden mit Hilfe des dichotomen Rasch-Modells ausgewertet und Modellparameter wie z.B. die Aufgabenschwierigkeit, Reliabilität, MNSQ, Wright-Map bestimmt. Weiter wurden zur Eignung der Testaufgaben Wald-Tests in verschiedenen Teilungsgruppen (Mittelwert, Geschlecht, Fachsemester, Alter und Reihenfolge der Testaufgaben) geprüft und die lokale stochastische Abhängigkeit der Testaufgaben mit Hilfe der Q3-Parameter untersucht. Schlussendlich wurden auch die Distraktoren der einzelnen Testaufgaben auf mögliche Besonderheiten, wie z.B. eine besonders seltene Auswahl durch die Proband\*innen, untersucht. Auf Basis dieser Parameter wurde schlussendlich eine Empfehlung für die Verwendung der Testaufgaben erarbeitet. Ein Beispiel für eine solche Empfehlung für das Konzept „Ursachen der Messunsicherheit“ findet sich in Abb. 2, die blau markierten Testaufgaben stellen dabei die Empfehlung auf Basis der erhobenen Parameter dar, die Verteilung der inhaltlich repräsentierten Unterasspekte des Sachstrukturmodells bleibt dabei in dieser Auswahl erhalten.

Insgesamt konnten bei einer EAP-Reliabilität von 0.61 – 0.81 zu jedem Konzept des Sachstrukturmodells geeignete Testaufgaben identifiziert werden. Eine Übersicht über die Konzepte ist in Tab. 1 dargestellt.

*Tab. 1: Übersicht über die empfohlenen Testaufgaben zu den Konzepten*

<b>Testinstrument zum Konzept nach Hellwig 2012</b>	<b>Anzahl der empfohlenen Testaufgaben</b>	<b>EAP-Reliabilität</b>
Ursachen der Messunsicherheit des	7	0.61
Unterscheidung zwischen Messunsicherheit und -abweichung	8	0.81
Verlässlichkeit der Messung	7	0.76
Vergleich von Messwerten	10	0.80
Regression	9	0.61
Ziel der Messung	12	0.70
Ergebnis der Messung	7	0.72
Erfassung einer Unsicherheitskomponente bei direkter Messung	8	0.66
Zusammensetzung der Messunsicherheit	9	0.77
Erweiterte Messunsicherheit	9	0.64

### **Fazit**

Mit der vorgestellten Arbeit werden auf der Basis des Sachstrukturmodells nach Hellwig (2012) Kompetenzen formuliert und ein zugehöriges Testinstrument zur Verfügung gestellt. Dieses ist theoriegeleitet entwickelt und kann auf insgesamt zehn Skalen verschiedene Aspekte von Kompetenzen im Umgang mit Messunsicherheiten erfassen. Das Testinstrument liefert damit ein Verfahren, die Wirksamkeit von Lernumgebungen zu evaluieren, und legt eine Möglichkeit für mögliche Kompetenzstrukturen im Bereich der Messunsicherheiten nahe, aus denen in Zukunft möglicherweise ein Kompetenzmodell für diesen Bereich entwickelt werden könnte. Details zu der dargestellten Testentwicklung sind in Schulz (2022) zu finden.

**Literatur**

- Möhrke, P., Pampel, B., & Runge, B.-U. (2018). Konzepte Studierender zur Unsicherheit von Messdaten. *PhyDid B - Didaktik Der Physik - Beiträge Zur DPG-Frühjahrstagung*, 1. Abgerufen von <https://ojs.dpg-physik.de/index.php/phydid-b/article/view/872>
- Priemer, B., & Hellwig, J. (2018). Learning About Measurement Uncertainties in Secondary Education: A Model of the Subject Matter. *International Journal of Science and Mathematics Education*, 16(1), 45–68. <https://doi.org/10.1007/s10763-016-9768-0>
- Ludwig, T., Priemer, B., & Lewalter, D. (2018). Decision-making in uncertainty-infused learning situations with experiments in physics classes. In *Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018)*. Kyoto, Japan: Voorburg, The Netherlands: International Statistical Institute.
- Hellwig, J. (2012). *Messunsicherheiten verstehen: Entwicklung eines normativen Sachstrukturmodells am Beispiel des Unterrichtsfaches Physik* (Doktorarbeit). Ruhr-Universität, Bochum
- Haladyna, T.M. (2004). *Developing and Validating Multiple-choice Test Items* (3rd ed.). Routledge. doi: <https://doi.org/10.4324/9780203825945>
- Haladyna, T.M., & Downing, S.M. (Eds.). (2006). *Handbook of Test Development* (1st ed.). Routledge. <https://doi.org/10.4324/9780203874776>
- Schulz, J. (2022). *Entwicklung eines Testinstrumentes zur Erfassung von Kompetenzen im Umgang mit Messunsicherheiten* (Doktorarbeit), Humboldt-Universität, Berlin. doi: <http://dx.doi.org/10.18452/23957>