

Der Einfluss von Visualisierungen auf die Güte von Likert-Skalen oder wie Umfragen unbewusst das Antwortverhalten von Teilnehmenden beeinflussen können

Teresa Tewardt*, Lisa Stinken-Rösner*

*Universität Bielefeld, Physik und ihre Didaktik, Universitätsstraße 25, 33615 Bielefeld
ttewardt@physik.uni-bielefeld.de

Kurzfassung

Bei Fragebogenerhebungen mit jüngeren Lernenden werden die Antwortniveaus von Likert-Skalen häufig durch Icons veranschaulicht. Dies soll den Lernenden ein Gefühl für die Abstufungen der Antwortniveaus vermitteln (Ouwehand et al., 2021). Die Verwendung von Visualisierungen birgt jedoch gleichzeitig die Gefahr, dass das Antwortverhalten der Teilnehmenden verzerrt wird, z.B. durch eine emotionale Konnotation der Visualisierungen. Die vorliegende Studie geht der Frage nach, inwiefern verschiedene Formen der Visualisierungen zu Unterschieden im Antwortverhalten von Lernenden führen. Hierfür wurden N = 635 Lernende im Alter von 10-12 Jahren, nach dem Besuch des Schülerlabores „teutolab-physik“, befragt. Die Fragebögen erheben das Interesse beim Experimentieren (Fechner, 2009) und unterscheiden sich ausschließlich in den unterstützenden Visualisierungsformen. Die Zuordnung der verschiedenen Visualisierungen in Form von Smileys, Daumen, Handybalken oder einem Farbverlauf zu den Teilnehmenden erfolgte randomisiert. Die Auswertung ergab signifikante Differenzen im Antwortverhalten der Teilnehmenden, in Abhängigkeit von der Visualisierungsform.

1. Likert-Skalen in der physikdidaktischen Forschung

Likert-Skalen sind die mit Abstand am häufigsten verwendeten psychometrischen (Kurz-)Skalen in den Sozialwissenschaften (Döring & Bortz, 2016). Anhand mehrerer Indikatoren (Items) wird ein Konstrukt erfasst, welches von außen nicht direkt beobachtbar ist (ebd.). Dafür beurteilen die Teilnehmer*innen alle Items entlang vorgegebener mehrstufiger Antwortniveaus. Die Niveaustufen repräsentieren, wie stark die teilnehmende Person einem Item zustimmt oder es ablehnt. Es liegt zunächst ein ordinale Skalenniveau vor. Geht man zusätzlich davon aus, dass die Abstände zwischen den Antwortoptionen gleich groß und mittels Zahlen darstellbar sind, kann eine Intervallskalierung angenommen werden. In diesem Fall wird der Skalenwert über das arithmetische Mittel bestimmt.

Die physikdidaktische Forschung als Teil der empirischen Bildungswissenschaft greift oftmals auf Likert-Skalen als Erhebungsinstrumente zurück (Döring & Bortz, 2016). Bei Befragungen von Lernenden, die anonym oder mit wenigen personenbezogenen Daten durchgeführt werden, sind Likert-Skalen sehr bildgebend (ebd.). Durch das Ankreuzen der Antworten können keine Rückschlüsse auf Teilnehmende anhand von Schriftmustern gezogen werden, was wiederum die Anonymität dieser gewährleistet. Typische Einsatzfelder von Likert-Skalen in der physikdidaktischen Forschung sind: Erhebung des Interesses (z.B. Riese & Kulgemeyer et al., 2015), Selbstwirksamkeitserwartungen von Lehrkräften oder Lernenden (z.B. Rabe, Krey & Meinhardt, 2012), Erhebung von

Fähigkeiten und Performanz Lernender (z.B. Schreiber, Theyßen & Dickmann, 2016).

1.1. Herausforderung: Antwortniveaus

Es ist bekannt, dass Lernende Unterstützung bei der Einordnung ihrer eigenen Meinung in die vorgegebenen Antwortniveaus von Likert-Skalen benötigen (Jamieson, 2004; Ouwehand et al., 2021). Neben der verbalen Erläuterung des Testinstrumentes wird eine visuelle Unterstützung empfohlen (Ouwehand et al., 2021). Insbesondere die Visualisierung der konsistenten Abstände zwischen den Niveaustufen, welche die Voraussetzung für die Annahme der Intervallskalierung darstellt, sollen durch Visualisierungen unterstützt werden. Typische Beispiele sind der Einsatz von Smileys (Pantiri, 2023), Daumen (Sullivan, 2013) oder Farbskalen (Stinken, 2017). Einige dieser Visualisierungen haben jedoch den Nachteil, dass sie eine emotionale Konnotation besitzen, welche gegebenenfalls das Antwortverhalten der Teilnehmer*innen beeinflussen kann. Reynolds-Keefe et al. haben bereits 2009 eine Studie zu Visualisierungshilfen von Likert-Skalen durchgeführt. Hierbei wurden Smileys, Sonne/Wolke und Worte mit 136 Grundschüler*innen auf ihre Validität hin geprüft. Diese Studie ergab keine Varianzen im Antwortverhalten der Lernenden.

Die systematische Untersuchung des Einflusses unterschiedlicher Visualisierungen auf das Antwortverhalten an einer größeren Stichprobe mit älteren Lernenden, stellt ein Forschungsdesiderat dar. Diese Studie soll eine erste Einschätzung des Forschungsgebietes ermöglichen und an die bisherigen Ergebnisse (Reynolds-Keefe et al., 2009) anschließen.

2. Forschungsfrage und Hypothese

Aufgrund des oben beschriebenen Forschungsinteresses und daraus abgeleitetem Desiderat kann folgende Forschungsfrage formuliert werden:

Inwiefern können Differenzen im Antwortverhalten auf unterschiedliche Visualisierungshilfen zurückgeführt werden?

Hierbei wird die Nullhypothese angenommen, dass verschiedene Visualisierungen keine Unterschiede im Antwortverhalten der Teilnehmer*innen hervorrufen. Dementsprechend wird weiterhin angenommen, dass die interne Konsistenz der Skala unabhängig von der unterstützenden Visualisierungsform ist (vgl. Reynolds-Keefe, 2009).

3. Studiendesign

Zur Beantwortung der Forschungsfrage, wurde ein Studiendesign geplant, um entsprechende Daten zu verschiedenen Visualisierungshilfen zu erheben. Im Folgenden wird die Entscheidung, die Daten in einer Laborstudie zu erheben erläutert und das methodische Vorgehen begründet.

3.1. Setting: Das „teutolab-physik“

Das „teutolab-physik“ ist ein Schülerlabor der Universität Bielefeld, in dem Lernende eigenständig physikalische Experimente durchführen und gemeinsam Erklärungsansätze formulieren. Dies geschieht unter der Anleitung und Betreuung von geschulten Studierenden und abgeordneten Lehrkräften.

Der Besuch einer Schulklasse im „teutolab-physik“ folgt einem standardisierten Ablauf für jede Altersstufe. Alle Lernenden führen die gleichen Experimente, unterstützt durch dieselben Personen, durch. Es kann daher davon ausgegangen werden, dass alle Schüler*innen vergleichbare Erfahrungen während ihres Besuches sammeln, selbst wenn sie von unterschiedlichen Schulen kommen und an unterschiedlichen Tagen das „teutolab-physik“ besuchen.

Durch die breite Zielgruppe des Angebots im „teutolab-physik“ sind verschiedene Bildungsniveaus vertreten, was die Voraussetzung für eine große und vielfältige Stichprobe schafft. Die Kombination aus Stichprobengröße, Diversität, Kontrolle und möglicher Varianz macht das Schülerlabor zu einer besonders geeigneten Umgebung für die vorliegende Untersuchung.

3.2. Methodik

Zur Beantwortung der Forschungsfrage wurde, passend zum Setting der Erhebung, die psychometrische Kurzsкала zum situativen Interesse beim Experimentieren (Fechner, 2009) als exemplarisches Instrument ausgewählt und um verschiedene Visualisierungshilfen ergänzt. Die Skala besteht aus sechs Items, die auf einer 5-stufigen Likert-Skala von den Befragten beantwortet werden:

- Beim Experimentieren habe ich mich wohl gefühlt.

- Beim Experimentieren habe ich über nichts anderes nachgedacht.
- Die Experimente haben mir Spaß gemacht.
- Ich freue mich auf die nächsten Experimente.
- Die Experimente waren langweilig.
- Beim Experimentieren ist die Zeit sehr schnell vergangen.

Ausgehend vom Originalinstrument wurden vier Versionen mit zusätzlichen Visualisierungen entwickelt. Daraus ergaben sich vier Experimental- und eine Kontrollgruppe.

Die Kontrollgruppe erhielt die Kurzsкала vollkommen ohne Visualisierungen, die Niveaustufen waren mit Worten beschriftet („stimmt gar nicht“, „stimmt kaum“, „stimmt etwas“, „stimmt ziemlich“, „stimmt sehr“). Die vier Experimentalgruppen bekamen den Fragebogen ebenfalls mit wörtlicher Beschriftung der Niveaustufen und Visualisierungshilfen zur zusätzlichen Unterstützung. Bei den vier Visualisierungshilfen handelte es sich um Daumen, Smileys, Handybalcken und einen Farbverlauf. Diese wurden randomisiert in den Schulklassen nach dem Besuch des „teutolab-physik“ verteilt und von den Schüler*innen bearbeitet. Die verschiedenen Visualisierungsformen sind in Abbildung 1 dargestellt.

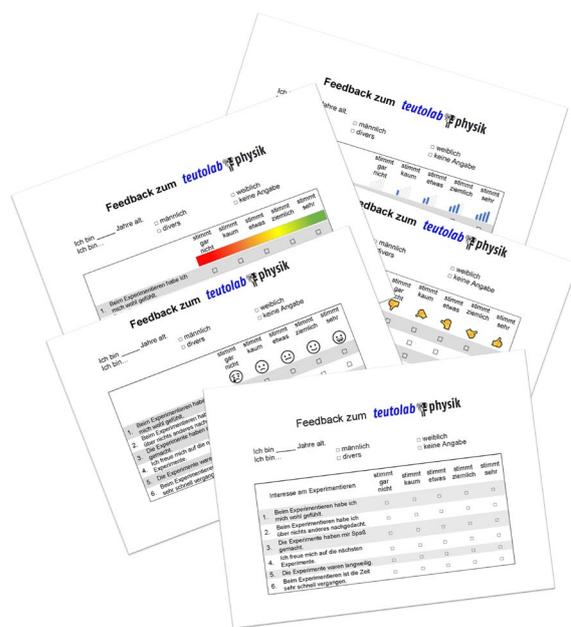


Abb. 1: Verwendete Fragebögen der Erhebung.

Die Erhebung fand über einen Zeitraum von drei Monaten statt. Die erhobenen Daten wurden im Anschluss digitalisiert, in SPSS eingepflegt und mittels statistischer Analysen ausgewertet. Aufgrund der hohen Standardisierung der Erhebung im „teutolab-physik“, welche bereits in Abschnitt 3.1. erläutert wurde, werden die erhobenen Daten als konsistent und vergleichbar eingestuft. Fehlende Werte oder Kreuze, die zwischen zwei Niveaustufen lagen wurden aus der Statistik ausgeschlossen. Dies war bei fünf Fragebögen der Fall.

3.3. Stichprobe

Es wurden 635 Lernende nach dem Besuch des „teutolab-physik“ befragt. Diese waren zwischen zehn und zwölf Jahren ($\bar{X} = 11,3$ Jahre) alt und besuchten die 5. oder 6. Klasse einer weiterführenden Schule.

Die Stichprobengröße ($N = 630$ bei fünf ausgeschlossenen Fragebögen, vgl. 3.2) verteilt sich auf insgesamt fünf Gruppen. Die Kontrollgruppe ohne Visualisierungshilfen bestand aus $n = 89$ Lernenden. Die vier Experimentalgruppen erhielten Fragebögen mit zusätzlichen Visualisierungen in Form von Smileys ($n = 135$), Daumen ($n = 136$), Handybalken ($n = 137$) oder Farbverläufen ($n = 133$), diese sind auf Abbildung 1 dargestellt.

4. Auswertung und Ergebnisse

Die Daten der Experimentalgruppen wurden jeweils mit der Kontrollgruppe mittels eines t-Test für unabhängige Stichproben verglichen. Zudem wurde die interne Konsistenz der Skala getrennt innerhalb der fünf Gruppen ermittelt.

Während der Datenerhebung konnte beobachtet werden, dass Fragebögen mit zusätzlichen Visualisierungshilfen in Form von Smileys unter den Schüler*innen am beliebtesten waren. Nach der zufälligen Zuteilung der Fragebögen innerhalb der Lerngruppe, wurde vermehrt von Schüler*innen der Wunsch geäußert, auch einen Fragebogen mit Smileys haben zu wollen. Teilweise ging dies so weit, dass Kinder versucht haben, diese Fragebögen ihren Mitschüler*innen wegzunehmen. Skalen ohne Visualisierungshilfen wurden von den Lernenden weniger euphorisch kommentiert.

4.1. Auswirkung der Visualisierungen auf das Antwortverhalten

Beim Vergleich der Mittelwerte von Abbildung 2 fallen geringe Unterschiede zwischen den vier Experimentalgruppen mit zusätzlichen Visualisierungsformen und der Kontrollgruppe auf. Bei den Visualisierungsformen Smileys und Daumen beurteilen die Lernenden ihre Erlebnisse im Schülerlabor leicht positiver. Ein t-Test für unabhängige Stichproben zeigt, dass zwischen Smileys und der Kontrollgruppe ($t(135) = 1,9, p = .026$), sowie Daumen und der Kontrollgruppe ($t(136) = 2,5, p = .005$) signifikante Unterschiede vorliegen. Zwischen den anderen Experimentalgruppen und der Kontrollgruppe konnten keine signifikanten Unterschiede ermittelt werden.

4.2. Auswirkung der Visualisierungen auf die interne Konsistenz der Skala

Tabelle 1 zeigt die interne Konsistenz der genutzten Skala für die Kontrollgruppe sowie die der vier Experimentalgruppen.

Die Ergebnisse zeigen, dass die Fragebögen ohne Visualisierungshilfen die beste interne Konsistenz aufweisen. Diese ist als gut einzustufen (Blanz, 2015). Eine ebenfalls gute interne Konsistenz konnte, für die Fragebögen mit Handybalken als zusätzliche Visualisierungshilfe, nachgewiesen werden (ebd.).

Tab. 1: Cronbachs Alpha der verschiedenen Visualisierungshilfen.

Visualisierung	Cronbachs Alpha
keine	.814
Handybalken	.801
Farbverlauf	.780
Smileys	.749
Daumen	.699

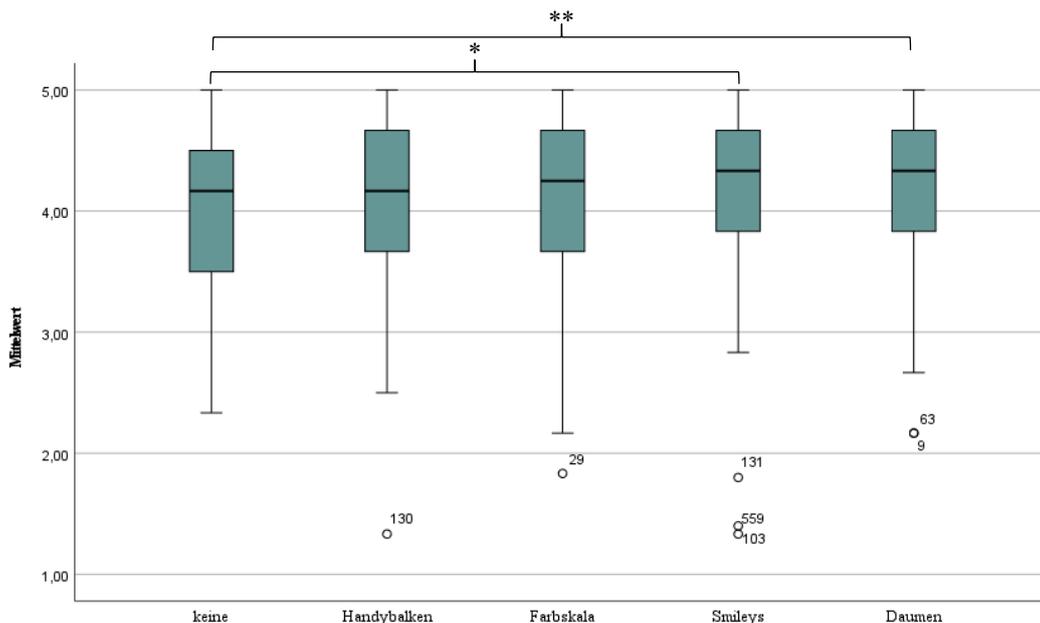


Abb. 2: Boxplots der Ergebnisse.

Die interne Konsistenz für die Visualisierungen Smileys und Farbverlauf liegen in einem akzeptablen Bereich (Blanz, 2015). Die interne Konsistenz der Skala für die zusätzliche Visualisierungshilfe Daumen ist als fragwürdig einzustufen (ebd.). Die interne Konsistenz der Kurzsкала scheint somit von den Visualisierungshilfen beeinflusst zu werden.

Auf Grundlage der Ergebnisse der Studie kann die Forschungsfrage (siehe Abschnitt 2.) wie folgt beantwortet werden. Die angenommene Nullhypothese, dass verschiedene Visualisierungen keine Unterschiede im Antwortverhalten der Teilnehmer*innen hervorrufen wird verworfen. Es wird die Gegenhypothese aufgestellt, dass die Differenzen im Antwortverhalten auf die verwendeten Visualisierungen zurückgeführt werden können. Zusätzlich wird die interne Konsistenz der Skala maßgeblich von den Visualisierungshilfen beeinflusst.

5. Diskussion

Die Erhebung des situativen Interesses am Experimentieren erfolgte einmalig nach dem Besuch des Schülerlabors „teutolab-physik“. Ohne vorherige Erhebung des experimentellen Interesses können die Verschiebungen im Antwortverhalten nicht ausschließlich auf die Visualisierungshilfen zurückgeführt werden. Dies wirkt sich limitierend auf die Generalisierbarkeit der Aussagen dieser Studie aus.

Zusätzlich können Einschränkungen in der Aussagekraft der Studie durch die, im Vergleich zu den vier Experimentalgruppen, kleinere Kontrollgruppe begründet sein. Des Weiteren ist die Verschiebung des Antwortverhaltens in den positiven Bereich kritisch zu betrachten, da an dieser Stelle der Deckeleffekt Einfluss nehmen kann.

Die Kurzsкала zum situativen Interesse am Experimentieren wurde für diese Studie ausgewählt, da sie bereits Teil des Projektes war, durch dessen Vorstellung auf einer Tagung die vorliegende Fragestellung inspiriert wurde. Dabei wurde die Kurzsкала von Fechner (2009) von den Vortragenden um Smileys ergänzt. Ein weiterer Faktor für die Wahl dieser Skala ist, dass sie zum Grundkonzept des „teutolab-physik“ passt. So wurde sichergestellt, dass der Fragebogen von den Teilnehmenden sinnvoll in den Kontext des Besuches eingeordnet werden konnte.

Abschließend lässt sich zusammenfassen, dass die Ergebnisse darauf hindeuten, dass die Visualisierungshilfen mit Smileys und Daumen das Antwortverhalten positiv beeinflussen, was auf die emotionale Konnotation zurückzuführen sein könnte. Farbverläufe suggerieren Antwortmöglichkeiten zwischen den Niveaustufen der Likert-Skala, da sie ein Spektrum abbilden. Zusätzlich ist der Farbverlauf für Menschen mit Farbsehschwäche ungeeignet. Gleichzeitig sinkt die interne Konsistenz der Skalen. Vergleichbare interne Konsistenzen, zum exemplarischen Instrument, ohne signifikante Differenzen im Antwortverhalten, konnten für die Visualisierungshilfen Farbskala, die

jedoch nicht barrierefrei ist, und Handybalken nachgewiesen werden.

Die Handybalken werden daher als geeignet eingestuft, da diese weder an Emotionen, noch an das Farbsehvermögen der Lernenden geknüpft sind. Die interne Konsistenz der Kurzsкала wird durch die visuelle Unterstützung mit Handybalken am geringsten gesenkt. Darüber hinaus stellen die Handybalken klar abgegrenzte Stufen dar und verleiten nicht zum Ankreuzen zwischen den Antwortniveaus.

5.1. Ausblick

Die Verwendung einer anderen Kurzsкала könnte Deckeleffekte reduzieren. Hierfür könnten Items verwendet werden, die weniger polarisierend ausfallen und insgesamt kritischer sind.

In dieser Studie wurden lediglich vier Visualisierungshilfen für Likert-Skalen einbezogen. Diese sind nicht die einzigen Visualisierungshilfen, die in der empirischen Bildungswissenschaft zu finden sind. Hierbei könnte der Pool der möglichen Visualisierungen erweitert werden.

Eine weitere Möglichkeit, die Forschung an dieser Stelle aufzugreifen, wäre, die Untersuchung der emotionalen Konnotation der Ergebnisse für Smileys und Daumen im Vergleich zu Handybalken und Farbverlauf. Hierbei würde der Fokus darauf liegen, ob die Unterschiede in der internen Konsistenz der verschiedenen Visualisierungen tatsächlich emotionsgebunden zu begründen sind.

Die gezielte Manipulation der Niveaustufen wird teilweise bereits im Marketing eingesetzt. Empirische Forschungsergebnisse zur Urteilverzerrung liegen jedoch nicht vor.

6. Literatur

- Blanz, M. (2021). *Forschungsmethoden und Statistik für die Soziale Arbeit: Grundlagen und Anwendungen*. Kohlhammer Verlag.
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. (6. Aufl.). Springer Verlag.
- Fechner, S. (2009). *Effects of context-oriented learning on student interest and achievement in chemistry education*. Logos Verlag.
- Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, 38(12), 1217–1218. <https://doi.org/10.1111/J.1365-2929.2004.02012.X>
- Laerhoven, H., Zaag-Loonen, H. & Bhf, D. (2004). A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta Paediatrica*, 93(6), 830–835. <https://doi.org/10.1111/j.1651-2227.2004.tb03026.x>
- Ouwehand, K., Kroef, A. V. D., Wong, J. & Paas, F. (2021). Measuring cognitive load: Are there more valid alternatives to likert rating scales?. *Frontiers in Education*, 6, 702616.

- Pantiri, G., Wilhelm, T., Burkhardt, L. M., Wenzel, V., Lühken, A. & Katzenbach, D. (in Druck). Inklusiver NaWi-Unterricht: Erprobung von Lernstationen zu Farben. In H. v. Vorst (Hrsg.), *Frühe naturwissenschaftliche Bildung*.
- Rabe, T., Krey, O. & Meinhardt, C. (2012). Physikdidaktische Selbstwirksamkeitserwartungen zukünftiger Physiklehrkräfte I. In S. Bernholt (Hrsg.), *Inquiry-based Learning - Forschendes Lernen* (S. 635–637). IPN-Verlag.
- Reynolds-Keefer, L., Johnson, R., Dickenson, T. & McFadden, L. (2009). Validity issues in the use of pictorial Likert scales. *Studies in Learning, Evaluation, Innovation and Development*, 6(3), 15–24.
- Riese, J., Kulgemeyer, C., Zander, S., Borowski, A., Fischer, H. E., Gramzow, Y., Reinhold, P., Schecker, H. & Tomczyszyn, E. (2015). Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. *Zeitschrift für Pädagogik*, 61, 55–79.
- Schreiber, N., Theyßen, H. & Dickmann, M. (2016). Wie genau beurteilen Schülerinnen und Schüler ihre eigenen experimentellen Fähigkeiten? – Ein Ansatz zur praktikablen Diagnostik experimenteller Fähigkeiten im Unterrichtsalltag. *PhyDid A – Physik und Didaktik in Schule und Hochschule*, 1(15), 49–63.
- Stinken, L. (2017). *"Ich hoffe du weißt das zu schätzen?!" Eine Erhebung der Schätzkompetenz in der Sekundarstufe I*. Logos Verlag.
- Sullivan, G. & Artino, A. (2013). Analyzing and interpreting data from likert-type scales. *Journal of graduate medical education*, 5(4), 541-542. <https://doi.org/10.4300/JGME-5-4-18>

Danksagung

Wir bedanken uns bei den Mitarbeiter*innen des „teutolab-physik“ und bei den Schüler*innen für ihre rege Teilnahme an der Studie. Die Unterstützung und das Engagement des Schülerlabors sowie die Beiträge der Lernenden haben die Forschung maßgeblich vorangetrieben.