

Entwicklung und Pilotierung eines Kompetenztests zur Untersuchung von Messkompetenzen in der Primarstufe

Naturwissenschaftlicher Sachunterricht in der Primarstufe zielt drauf ab, dass Lernende naturwissenschaftliche Denk-, Arbeits- und Handlungsweisen erwerben (GDSU, 2013). Das Experimentieren ist dabei zentrales Element in physikunterrichtlichen Kontexten (Tesch & Duit, 2004) und ermöglicht Lernenden, naturwissenschaftliche Phänomene eigenständig zu untersuchen. Um Experimente erfolgreich durchzuführen, ist der Erwerb spezifischer Kompetenzen erforderlich. Diese strukturieren den gesamten Experimentierprozess und sind in grundlegenden Kompetenzerwartungen des Sachunterrichts verankert (GDSU, 2013). Etablierte Modelle zur Beschreibung von Experimentierkompetenzen unterscheiden die drei übergeordneten Phasen Planung, Durchführung und Auswertung, in denen jeweils unterschiedliche Teilkompetenzen erforderlich sind (Nawrath et al., 2011; Schreiber et al., 2009). Eine grundlegende Arbeitsweise des Experimentierens ist das Messen. Dies ist in jeder der drei Phasen elementar und lässt sich weiter in verschiedene Teilkompetenzen operationalisieren (Murer et al., 2025). Der Erwerb und die Förderung von Messkompetenzen ist bereits im Sachunterricht relevant, da dies Lernende unterstützt, subjektive Sinneseindrücke von objektiv gemessenen physikalischen Größen zu differenzieren (GDSU, 2013). Dennoch sind Messkompetenzen in experimentellen Kontexten im Primarbereich bisher kaum systematisch untersucht oder erfasst. Vor diesem Hintergrund zielt der vorliegende Beitrag auf die Entwicklung und Pilotierung eines Kompetenztests zur Erfassung von Messkompetenzen ab.

Entwicklung des Instruments Zur systematischen Erfassung von Messkompetenzen in sachunterrichtlichen Experimentierprozessen der Primarstufe werden in einem ersten Schritt die erforderlichen Teilkompetenzen modelliert. Hierzu wird auf die Unterteilung des Experimentierprozesses in die Phasen Planung, Durchführung und Auswertung durch bestehende Modelle experimenteller Kompetenz (Schreiber et al., 2009) aufgebaut. Dies wird durch eine Analyse typischer kognitiver und praktischer Handlungsschritte im Experimentierprozess, die zur kompetenten Auswahl und Handhabung von Messgeräten sowie zur Auswertung von aufgenommenen Messwerten notwendig sind, erweitert: In der Planungsphase (P) sind Kenntnisse über Größen und Messgeräte notwendig, um zur Forschungsfrage passende Messgeräte auszuwählen. In der Durchführungsphase (D) werden Messwerte präzise abgelesen und systematisch dokumentiert. In der Auswertungsphase (A) werden die aufgenommenen Messwerte beispielsweise durch die Darstellung in einer Tabelle visualisiert, um diese übersichtlich und nachvollziehbar darzustellen. Insgesamt ergeben sich neun messbezogene Teilkompetenzen (Tab. 1), die das theoretische Konstrukt und damit die Grundlage für die Aufgabenentwicklung des Kompetenztests bilden (Kauertz, 2014). Einige dieser Teilkompetenzen lassen sich dabei klar einer Phase des Experimentierprozesses zuordnen, beispielsweise das richtige Ablesen und Dokumentieren von Messwerten. Andere Teilkompetenzen, etwa die Auswahl geeigneter Messgeräte, sind phasenübergreifend relevant. Eine Übersicht ist in Tabelle 1 dargestellt.

Tab. 1: Messbezogene Teilkompetenzen im Experimentierprozess

Messbezogene Teilkompetenz	P	D	A
TK1: Die Schüler:innen kennen Messgeräte und können sie benennen.	✓	×	×
TK2: Die Schüler:innen können angeben, mit welchen Messgeräten sie verschiedene Größen messen.	✓	×	×
TK3: Die Schüler:innen kennen Einheiten von Größen und können diese angeben.	✓	×	✓
TK4: Die Schüler:innen wählen geeignete Messgeräte in konkreten Experimentiersituationen aus.	✓	✓	×
TK5: Die Schüler:innen wählen geeignete Messgeräte bezüglich des Messbereichs aus.	✓	✓	×
TK6: Die Schüler:innen können Messwerte von verschiedenen Messgeräten zur gleichen Größe ablesen.	×	✓	×
TK7: Die Schüler:innen können Messwerte unter Beachtung des Parallaxenfehlers korrekt ablesen.	×	✓	×
TK8: Die Schüler:innen können abgelesene Messwerte dokumentieren.	×	✓	×
TK9: Die Schüler:innen können Messwerte korrekt in eine Tabelle übertragen.	×	×	✓

Die Operationalisierung dieser Teilkompetenzen erfolgt durch Festlegung spezifischer Merkmale, die in ein Aufgabenmodell überführt und für die Itementwicklung genutzt werden (Kauertz, 2014). Aus der Modellierung des theoretischen Konstrukts mit den neun Teilkompetenzen ergeben sich analog neun Aufgaben im Instrument (Abb. 1). Für die Fragenentwicklung innerhalb dieser Aufgaben werden bereits bestehende Fragen aus den FISS (COMPEAT, 2025a) und SISS (COMPEAT, 2025b) Studien angepasst, durch neu konzipierte Fragen ergänzt, und in das Aufgabenmodell integriert. Inhaltlich orientieren sich die Fragen an den für den Sachunterricht zentralen Messgrößen Temperatur, Masse, Länge und Zeit sowie den entsprechenden Messgeräten und Einheiten. Zur Sicherstellung der

Abb. 1: Beispielaufgaben zu den Kompetenzen 2, 3 und 4 (vgl. Tab. 1)

inhaltlichen Passung, Verständlichkeit und Eindeutigkeit dient ein Expert:innenrating (Kauertz, 2014). Durch die inhaltliche Differenzierung zu den Messgrößen ergeben sich insgesamt 28 Fragen. Ein Kodierleitfaden ermöglicht die systematische Zuordnung der Antworten zu dichotomen bzw. polytomen Kategorien. Auf dieser Grundlage werden die 28 Fragen weiter ausdifferenziert und in insgesamt 71 Items überführt. Bei einer Pilotgruppe von $N_K = 32$ ergibt sich für die Interrater-Reliabilität des zu den Items gehörigen Kodierleitfadens ein Cohens Kappa von $\kappa = .91$ und damit eine sehr hohe Übereinstimmung (Cohen, 1960).

Pilotierung und Ergebnisse Die Pilotierung des Kompetenztests erfolgt mittels Rasch-Modellierung, bei der die Itemschwierigkeiten (β) in Bezug auf die Personenfähigkeiten (θ) untersucht werden (Kauertz, 2014). Der Test wurde in zwei unabhängigen Gruppen erprobt: Die erste Gruppe ($N_1=21$) ist eine dritte Klasse einer Bielefelder Grundschule. Die Kinder dieser Gruppe haben ohne zeitliche Einschränkung den gesamten Test bearbeitet, ihnen

unbekannte Begriffe markiert und in einem anschließenden Gruppeninterview Schwierigkeiten benannt. Die zweite Gruppe ($N_2=69$) sind Schüler:innen vierter Klassen von Bielefelder Grundschulen. Diese haben randomisiert jeweils ca. ein Viertel des Tests im Anschluss an den Besuch im Schülerlabor teutolab-physik der Universität Bielefeld bearbeitet. Die Auswertung findet im Rahmen der Pilotierung gemeinsam statt ($N=90$), d.h. im Schnitt

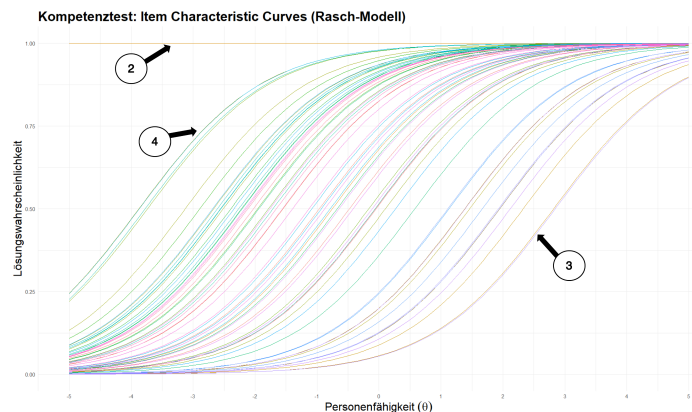


Abb. 2: ICC Plots nach Rasch-Modell

wurde jede Frage von 38 Kinder bearbeitet. Die Ergebnisse der Rasch-Modellierung zeigen sowohl in globaler als auch in lokaler Ebene insgesamt eine gute Modellpassung (Abb. 2). Sie weisen mit einer Varianz von 0.621 und einer Expected A Priori (EAP) Reliabilität von 0.644 auf eine hinreichende Differenzierung und moderate Schätzung der Fähigkeiten hin. Gleichzeitig werden Items identifiziert, deren Itemschwierigkeiten (β) außerhalb des für praktische Anwendungen informativen Bereichs von ± 3 Logits liegen. Items mit $\beta < -3$ oder $\beta > 3$ liefern nur geringe Informationen, da sie nahezu von allen bzw. keinen Teilnehmenden gelöst werden. Im entwickelten Instrument betrifft dies insbesondere drei Items aus Aufgabe 2 (TK2), die in der Pilotgruppe durchgängig gelöst wurden und daher keine Passung zum Rasch-Modell aufweisen (Abb. 2). Weitere auffällige Items finden sich in den Aufgaben zu TK3 und TK4 (Abb. 1). Nach der Rasch-Modellierung weisen die Items aus Aufgabe 3 eine vergleichsweise hohe Schwierigkeit auf und wurden von wenigen Teilnehmenden gelöst, während die Items aus Aufgabe 4 eine eher geringe Itemschwierigkeit zeigen und entsprechend häufig gelöst wurden (Abb. 2).

Diskussion und Ausblick Trotz der begrenzten Gruppengrößen liefert die Pilotierung bereits wichtige Hinweise auf die Eignung des Instruments. Insbesondere wird deutlich, dass das Testinstrument das relevante Fähigkeitsspektrum insgesamt gut abdeckt und grundsätzlich eine differenzierte Auswertung von Messkompetenzen im Primarbereich ermöglicht, auch wenn einzelne Items in der Pilotierung nur begrenzte Information liefern. Die auffälligen Items, deren Schwierigkeitsparameter außerhalb des optimalen Bereichs liegen, werden im Hinblick auf Aufgabenmerkmale und Kodierung sowie mit Rücksicht auf Hinweise aus dem Gruppeninterview der ersten Teilgruppe näher untersucht. Dabei wird insbesondere geprüft, ob Abweichungen durch die Wahl des Operators, die Formulierung der Aufgabenstellung oder die Nutzung von Piktogrammen und Bildern bedingt sind. Insgesamt deuten die Ergebnisse auf ein grundsätzlich tragfähiges Instrument hin, das nach gezielter Überarbeitung und Optimierung voraussichtlich auch im Prä-Post-Design geeignet ist, Kompetenzentwicklungen differenziert zu erfassen.

Literatur

- Center for Comparative Analysis of Educational Achievement (COMPEAT). (2025a). *First international science study 1976*. Universität Göteborg. <https://www.gu.se/en/center-for-comparative-analysis-of-educational-achievement-compeat/studies-before-1995/six-subject-survey-1970/first-international-science-study-1970> [17.10.2025]
- Center for Comparative Analysis of Educational Achievement (COMPEAT). (2025b). *Second international science study 1984*. Universität Göteborg. <https://www.gu.se/en/center-for-comparative-analysis-of-educational-achievement-compeat/studies-before-1995/second-international-science-study-1984> [17.10.2025]
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- GDSU (2013). *Perspektivrahmen Sachunterricht*. Verlag Julius Klinkhardt.
- Kauertz, A. (2014). Entwicklung eines Rasch-skalierten Leistungstests. In: D. Krüger, I. Parchmann & H. Schecker (Hrsg.). *Methoden in der naturwissenschaftsdidaktischen Forschung*. Springer. https://doi.org/10.1007/978-3-642-37827-0_27
- Murer, L., Metzger, S., Vorholzer, A., Bonetti, A. & Gut, C. (2025). Vergleich verschiedener Erhebungsmethoden zur Erfassung von Kompetenzen im Bereich des naturwissenschaftlichen Messens bei Tests mit Realexperimenten. *Zeitschrift für Didaktik der Naturwissenschaften*, 31(8). <https://doi.org/10.1007/s40573-025-00184-9>
- Nawrath, D., Maisyenko, V. & Schecker, H. (2011). Experimentelle Kompetenz - Ein Modell für die Unterrichtspraxis. *Praxis der Naturwissenschaften - Physik in der Schule*, 60(6), 42–49.
- Schreiber, N., Theyßen, H. & Schecker, H. (2009). Experimentelle Kompetenz messen?! *PhyDid A – Physik und Didaktik in Schule und Hochschule*, 3(8).
- Tesch, M. & Duit, R. (2004). Experimentieren im Physikunterricht - Ergebnisse einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 51–69.