

Sebastian Stuppan¹
Markus Wilhelm^{1,2}
Katrin, Bölsterli Bardy¹
Markus, Rehm²

¹Pädagogische Hochschule Luzern
²Pädagogische Hochschule Heidelberg

Analysieren und Kategorisierung von MINT-Aufgabensets Entwicklung und Validierung

Ausgangslage

Aufgaben können im schulischen Lernen als Dreh- und Angelpunkt angesehen werden. Folglich steuern sie den Lehr-Lernprozess (Luthiger et al., 2018, S. 11; Reusser, 2014). Damit der in den Lehrplänen anvisierte Kompetenzaufbau der Schüler:innen erreicht werden kann, treten Sets von aufeinander aufbauenden Aufgaben in den Fokus (Büchter & Leuders, 2006). Ein Beispiel zur Erstellung von Aufgabensets ist das Luzerner Modell zur Entwicklung kompetenzfördernder Aufgabensets (LUKAS-Modell; Luthiger et al. (2018)). In einzelnen Fachdidaktiken und in der Allgemeinen Didaktik wurden verschiedene Kategoriensysteme zur Analyse oder Kategorisierung einzelner Aufgaben entwickelt und teilweise erprobt (Blömeke et al., 2006; Bölsterli Bardy & Wilhelm, 2018; Jordan et al., 2006; Luthiger et al., 2018; Maier et al., 2010; Neubrand, 2002). Derzeit fehlen empirisch validierte Instrumente zur Analyse von Aufgabensets. Somit erzielt dieses Forschungsprojekt, die Konstruktion und Validierung eines Messinstruments auf der Grundlage bestehender Kategoriensysteme.

Theoretischer Hintergrund

In den meisten pädagogisch-didaktischen Fachbereichen haben Aufgaben eine zentrale Bedeutung, da sie ein wegweisendes Steuerungsinstrument in Lehr-Lern-Situationen und wertvolle Elemente des Lernweges darstellen. Müller und Helmke (2008, S. 32) schreiben ihnen für die Unterrichtsentwicklung und Unterrichtsqualität eine zentrale Rolle zu. Während Lernaufgaben dem Aufbau von fachlichen und überfachlichen Kompetenzen dienen, sind Leistungsaufgaben zur Überprüfung von Bildungsstandards oder der Evaluation der Zielerreichung dienlich (Reusser, 2014, S. 79). Inzwischen sind Aufgabenstellungen in der empirischen Unterrichtsforschung sowie in der Forschung der meisten Fachdidaktiken ein zentraler Gegenstand geworden. Zudem basieren digitalisierte und individualisierte Lernangebote im Wesentlichen auf einer Zusammenstellung von qualitativ hochwertigen Aufgaben (Maier et al., 2014, S. 36). All diesen skizzierten Anwendungsbereichen ist gemein, dass sie eine differenzierte Analyse und Bewertung von Aufgabenstellungen unterstellen. Als eine mögliche Hilfe kann das LUKAS-Modell der Arbeitsgruppe um Luthiger (2018) dienen. Dem LUKAS-Modell zugrunde liegt die Idee, dass Aufgaben innerhalb des Lernprozesses einer Reihenfolge unterliegen und als Gesamtheit gezielt den anvisierten Kompetenzaufbau ermöglichen. Jede einzelne Aufgabe hat somit im Unterrichtsetting eine didaktische Funktion (z.B. Aufbau von Kompetenzen, Vertiefung, Ausdifferenzierung sowie Variantenbildung von Kompetenzen, etc.) und ist geprägt von Merkmalen (Kategorien), die wiederum in unterschiedlichen Ausprägungen (Skalen) vorliegen.

Fragestellung und Methode

Es existiert zurzeit kein operationalisiertes und validiertes Messinstrument, um Aufgaben analysieren und kategorisieren zu können. Mit dem bestehenden LUKAS-Kategoriensystem können Aufgaben nur textbasiert einzelnen Ausprägungen zugeordnet werden, wie beispielsweise bei Reinfried (2016). Hier setzt der erste Schritt der Studie mit folgender Fragestellung an: *1. Lassen sich aus den bestehenden Kategoriensystemen Items entwickeln?* Der erarbeitete Itempool wird im zweiten Schritt empirisch validiert und führt zu den nachkommenden Fragestellungen: *2. Stimmen die Items mit den theoretischen Konstrukten überein? 3. Wie zuverlässig können Rater:innen, MINT-Aufgaben beurteilen? 4. Lassen sich Skalen mit den Items bilden?* Es wird erwartet, dass sich die theoretischen Konstrukte mit einer Closed Card Methode korrekt zuordnen lassen. Mittels explorativer Faktorenanalysen sollte eine sparsame und eindeutige Faktorenlösung gefunden werden. Dies aus dem Grund, weil sich die theoretischen Konstrukte inhaltlich deutlich voneinander unterscheiden.

Für die Konstruktion und Validierung des Messinstruments wurden acht Schritte durchlaufen. Dabei wurden die von Bühner (2006), Busker (2014), Döring und Bortz (2016) sowie Mummendey und Grau (2014) beschriebenen Konstruktion- und Validierungsstrategien berücksichtigt. 1) Konzeptspezifikation: In einem ersten Teilschritt wurden theoriebasiert die Dimensionen und Qualitätskriterien von Aufgaben studiert und eingegrenzt. 2) Operationalisierung: In einem zweiten Teilschritt erfolgte mithilfe der theoretischen Grundlage eine Definition der Kategorien und der zu messenden Konstrukte. Aus den erarbeiteten Definitionen wurden Itemtexte formuliert und eine 5-stufige Ratingskala von «stimmt nicht» bis «stimmt sehr» ausgearbeitet. 3) Expert:innen-Befragungen: Mithilfe von Expert:innen-Befragungen in Anlehnung an die Delphi-Methode (Döring & Bortz, 2016, S. 420f.) wurde in einem nächsten Teilschritt der gesamte Itempool verfeinert und qualitativ überarbeitet. Die Autor:innen des LUKAS-Modells fungierten als Expert:innenteam ($N = 4$, davon 50% weiblich). 4) Expert:innen-Pretest und Revision: Zusammen mit dem Team der LUKAS-Expert:innen wurde ein Testmanual erstellt und anhand von drei Aufgabenratings pilotiert. 5) Itempool den theoretischen Konstrukten zuordnen: Die Inhaltsvalidität erfolgte, wie Döring und Bortz (2016, S. 446f.) vorschlugen, theoretisch-argumentativ und gestützt durch weitere Urteile von Fachpersonen. Die Fachpersonen aus den MINT Fächern kannten den theoretischen Hintergrund der ausgearbeiteten Kategorien und Konstrukte. Sie hatten mittels einer digitalen *Closed Card Sorting* Methode, jedes Item aus dem Itempool nach inhaltlichen Überlegungen einem theoretischen Konstrukt zugeordnet (Schilb, 2005; Wandke et al., 2009). 6) Raterschulung: Zum Analysieren der Aufgaben durchliefen zwei Fachdidaktiker eine mehrtägige Raterschulung. 7) Aufgaben Beurteilung: Die geschulten Fachdidaktiker:innen ($N = 2$) beurteilten insgesamt 146 Aufgaben. Die Aufgaben stammten aus dem «MINT unterwegs»-Projekt. 24% der Aufgaben wurden doppelt geratet (Brückmann & Duit, 2014, S. 190; Döring & Bortz, 2016, S. 575). 8) Faktorielle Validität, Item- und Reliabilitätsanalyse: Mit den 146 eingeschätzten Aufgaben konnte das Messinstrument statistisch geprüft werden. Zu diesem Zweck wurde die interne Konsistenz der Items eines Faktors mit dem Cronbach-Alpha-Koeffizienten bestimmt (Moosbrugger & Kelava, 2012, S. 130–137). Um von den empirisch erhobenen Daten auf die latenten Variablen (Faktoren) zu schließen, wurden innerhalb der Kategorien explorative Faktorenanalysen durchgeführt (Moosbrugger & Kelava, 2012, S. 226–334).

Validierung - Ergebnisse der explorativen Faktorenanalysen

Mittels Promax-rotierten Faktorenanalysen konnte von den empirisch erhobenen Daten auf die latenten Variablen (Faktoren) geschlossen werden (Moosbrugger & Kelava, 2012, S. 326–334). Durch die Zusammenfassung mehrerer Kategorien (angelehnt an Luthiger et al., 2018 und Neubrand, 2002) gewinnt das Instrument an Übersichtlichkeit. Die erste Faktoranalyse beinhaltet die Kategorien Kompetenzabbild und Lebensweltbezüge mit je drei Skalen. Die zweite Faktoranalyse besteht aus den Kategorien (Prä-)Konzepte, Wissensarten, Wissensaktivitäten und Repräsentationsformen mit einer bis vier Skalen. Die dritte Faktoranalyse schließt die Kategorien Offenheiten, Lernunterstützungen und Lernwege mit je drei Skalen mit ein. Die Tabelle 1 stellt zusammenfassend die drei Faktorenanalysen mit ihren jeweiligen Teststatistiken und der Gesamtvarianz dar.

Tabelle 1. Teststatistiken der Faktorenanalysen, Bartlett-Test, KMO-, MAP Test und Gesamtvarianz ($N = 146$)

Faktorenanalyse	χ^2	df	p	KMO	MAP	Gesamtvarianz
Kompetenzabbild und Lebensweltbezüge	4033.18	153	< .001	.869	6	91%
(Prä-)Konzepte, Wissensarten, Wissensaktivitäten und Repräsentationsformen	4740.22	630	< .001	.748	11	79%
Offenheiten, Lernunterstützungen und Lernwege	4967.43	435	< .001	.82	9	85%

Anmerkung. χ^2 = Chi-Quadrat Bartlett-Test; df = Freiheitsgraden; p = Signifikanzniveau; KMO = Kaiser-Meyer-Olkin Measure of Sampling Adequacy; MAP = Minimum-Average-Partial-Test

Sowohl die Bartlett-Test's als auch das Kaiser-Meyer-Olkin Kriterium weisen auf die Eignung der Items für Faktorenanalysen hin. Aufgrund des Screeplots (nicht dargestellt) und Minimum-Average-Partial-Test (MAP) wurden den empfohlenen Faktorenlösungen gefolgt. Die Reliabilitätsanalyse der einzelnen Skalen ergibt Cronbach's α -Werte zwischen .824 und .988 ($M = 0.93$, $SD = 0.039$). Die Itemschwierigkeit liegt mit einem Mittelwert von 0.5 ($SD = 0.17$) im guten Bereich.

Diskussion, Limitation und Ausblick

Die in diesem Beitrag vorgestellte Art von Messinstrument ist ein Novum in der Fachdidaktik. Erstmals können geschulte Fachdidaktiker:innen mit empirisch getesteten Skalen Aufgaben in einem großen Umfang analysieren und in Form von Aufgabenprofilen darstellen und so auch Aufgaben in einem Aufgabenset differenziert analysieren und einteilen (Stuppan et al., eingereicht). Einschränkend muss erwähnt werden, dass eine genauere Überprüfung der Faktorstruktur der Skalen mit einer konfirmatorischen Faktorenanalyse aussteht. Die Beurteilung durch Schüler:innen wird in zukünftigen Vorhaben aufgenommen, geht es bei der Kategorie Lebensweltbezüge im Wesentlichen um ihre Erfahrungswelt. Aber auch Lehrpersonen müssen bei zukünftigen Forschungsprojekten in den Fokus genommen werden (Blömeke et al., 2006). In einem nächsten Schritt werden die Aufgaben mittels Clusteranalysen nach ihren didaktischen Funktionen gruppiert, mit den theoretischen Aufgabentypen des LUKAS-Modells (Luthiger et al., 2018) verglichen und die Ergebnisse diskutiert.

Literatur

- Blömeke, S., Risse, J., Müller, C., Eichler, D., & Schulz, W. (2006). Analyse der Qualität von Aufgaben aus didaktischer und fachlicher Sicht. Ein allgemeines Modell und seine exemplarische Umsetzung im Unterrichtsfach Mathematik. *Unterrichtswissenschaft*, 34(4), 330–357.
- Bölsterli Bardy, K., & Wilhelm, M. (2018). Von kompetenzorientierten zu kompetenzfördernden Aufgaben im Schulbuch. *Erziehung & Unterricht*, 168(1/2), 121–129.
- Brückmann, M., & Duit, R. (2014). Videobasierte Analyse unterrichtlicher Sachstrukturen. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 189–201). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37827-0_16
- Büchter, A., & Leuders, T. (2006). Ein Aufgabenmodell für die Praxis. Einschätzung, Auswahl und Entwicklung von Mathematikaufgaben. *Praxis der Naturwissenschaften - Chemie in der Schule*, 55, 16–20.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. Pearson Studium.
- Busker, M. (2014). Entwicklung eines Fragebogens zur Untersuchung des Fachinteresses. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 269–281). Springer Spektrum.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. vollständig überarbeitete, aktualisierte und erweiterte Auflage). Springer.
- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., Löwen, K., Brunner, M., & Kunter, M. (2006). *Klassifikationsschema für Mathematikaufgaben. Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt*. Max-Planck-Institut für Bildungsforschung.
- Luthiger, H., Wilhelm, M., Wespi, C., & Wildhirt, S. (Hrsg.). (2018). *Kompetenzförderung mit Aufgabensets: Theorie - Konzept - Praxis* (1. Auflage). hep, der bildungsverlag.
- Maier, U., Bohl, T., Kleinknecht, M., & Metz, K. (2014). Allgemeine Didaktik und ein Kategoriensystem der überfachlichen Aufgabenanalyse. In P. Blumschein (Hrsg.), *Lernaufgaben – Didaktische Forschungsperspektiven* (S. 35–51). Verlag Julius Klinkhardt.
- Maier, U., Kleinknecht, M., Metz, K., Schymala, M., & Bohl, T. (2010). *Entwicklung und Erprobung eines Kategoriensystems für die fächerübergreifende Aufgabenanalyse*. Univ., Lehrstuhl für Schulpädagogik.
- Moosbrugger, H., & Kelava, A. (Hrsg.). (2012). *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Auflage). Springer.
- Müller, A., & Helmke, A. (2008). Qualität von Aufgaben als Merkmale der Unterrichtsqualität – verdeutlicht am Fach Physik. In J. Thonhauser (Hrsg.), *Aufgaben als Katalysatoren von Lernprozessen: Eine zentrale Komponente organisierten Lehrens und Lernens aus der Sicht von Lernforschung, allgemeiner Didaktik und Fachdidaktik* (S. 31–46). Waxmann.
- Mummendey, H. D., & Grau, I. (2014). *Die Fragebogen-Methode: Grundlagen und Anwendung in Persönlichkeits-, Einstellungs- und Selbstkonzeptforschung*. Hogrefe Verlag.
- Neubrand, J. (2002). *Eine Klassifikation mathematischer Aufgaben zur Analyse von Unterrichtssituationen. Selbsttätiges Arbeiten in Schülerarbeitsphasen in den Stunden der TIMSS-Video-Studie*. Franzbecker.
- Reusser, K. (2014). *Aufgaben – Träger von Lerngelegenheiten und Lernprozesse im kompetenzorientierten Unterricht*. Seminar, 4/2014, 77-101.
- Schilb, S. (2005). Card Sorting – Techniken im Überblick. *i-com*, 4(1/2005). <https://doi.org/10.1524/icom.4.1.49.61671>
- Stuppan, S., Wilhelm, M., Bölsterli Bardy, K., & Künzle, R. (eingereicht). *Messinstrument zur Analyse und Kategorisierung von MINT-Aufgaben – Konstruktion und Validierung*. 5. Tagung Fachdidaktiken, Lugano.
- Wandke, H., Kain, S., & Struve, D. (2009). *Mensch & Computer 2009: 9. fachübergreifende Konferenz für interaktive und kooperative Medien - Grenzenlos frei*. Oldenbourg Verlag.