Ein Messinstrument zur Analyse und Typisierung von MINT-Aufgaben – Entwicklung, Validierung und Anwendung

Von der Pädagogischen Hochschule Heidelberg
zur Erlangung des Grades
Doktor der Philosophie (Dr. phil.)
genehmigte Dissertation von

Sebastian Stuppan

aus
Schwyz (Schweiz)

2023

Erstgutachter:	Prof. Dr. Markus Wilhelm
Zweitgutachter:	Prof. Dr. Markus Rehm
Fach:	Chemie und ihre Didaktik
Tag der Mündlichen Prüfung:	27.6.2023

Für Sabine, Anna, Eva und Paula

Abstract

Aufgaben spielen eine bedeutende Rolle im Hinblick auf den Kompetenzaufbau und die Kompetenzentwicklung. Damit ein entsprechender Kompetenzaufbau nachhaltig initiiert und strukturiert werden kann, treten Sets von aufeinander aufbauenden Aufgaben in den Fokus. Ziel des Dissertationsvorhabens ist es, der Fachdidaktik, der Lehrpersonenbildung sowie der Bildungsforschung ein empirisch validiertes und erprobtes Messinstrument zur Verfügung zu stellen, mit dem die Aufgabenqualität analysiert und anschließend zentralen Aufgabentypen im Lernprozess (Konfrontation, Erarbeitung, Übung, Synthese und Transfer) zugeordnet werden können. Die neun Merkmale des Aufgaben-Analyse-Instruments (AAI) sind: *Kompetenzabbild*, *Lebensweltbezug*, *Lernendenvorstellung*, *Wissensart*, *Wissensaktivität*, *Repräsentationsform*, *Offenheit*, *Lernunterstützung* und *Lernweg*. Das operationalisierte AAI ist aus bestehenden und teilweise erprobten Kategoriensystemen entwickelt und unter Berücksichtigung unterschiedlicher Testgütekriterien mehrfach validiert worden. Die zur Validierung verwendeten Aufgaben ($N = 146$) stammen aus dem *„MINT unterwegs“*-Projekt (*Teilstudie I*). Das AAI wurde in der Folge in drei Anwendungsstudien erprobt (*Teilstudien II* bis *IV*). In der ersten Anwendung werden MINT-Aufgaben gemäß den Autor:innen den entsprechenden Aufgabentypen zugewiesen (z.B. $n_{\text{Konfrontationsaufgaben}} = 25$) und mit dem AAI von geschulten Rater:innen beurteilt. Anschließend werden die Aufgabenbeurteilungen einem theoretischen Expert:innenvorschlag gegenübergestellt. Die Ergebnisse dieser Teilstudie lassen den Schluss zu, dass Qualitätsmerkmale in Form von Aufgabenprofilen mit dem AAI gemessen und in Bezug zu einer theoretischen Norm gestellt werden können. Die ausgewiesenen Abweichungen der Analyse und des Expert:innenvorschlags in Bezug auf die Qualitätsmerkmale, aufbereitet in Form eines Ampelsystems, können dazu genutzt werden, Aufgaben gezielt zu überarbeiten. In der zweiten Anwendungsstudie werden mit dem AAI erfasste MINT-Aufgaben ($n = 58$) aufgrund ihrer beurteilten Qualitätsmerkmale mittels einer Clusteranalyse gruppiert. Die Befunde dieser Studie zeigen, dass eine Zuordnung möglich ist, sich die Aufgabengruppen jedoch nicht in allen Merkmalen signifikant unterscheiden. Die gewonnenen Erkenntnisse können dazu dienen, Aufgaben bei der Entwicklung von Lernarrangements den verschiedenen Lernphasen zuzuordnen. Die dritte AAI-Anwendung untersucht, wie Lernende ($N = 805$) in ausgewählten MINT-Aufgaben ($N = 16$) den Lebensweltbezug, ein Merkmal des AAI, wahrnehmen und wie sich diese Beurteilungen

von der Setzung der Lehrmittelautor:innen unterscheiden. Die Ergebnisse dieser Teilstudie verdeutlichen, dass die Lehrmittelautor:innen die Authentizität in den Aufgaben generell höher beurteilen als die Lernenden. Um diesem Problem entgegenzuwirken, wäre ein iterativer Aufgabenentwicklungsprozess zusammen mit der Praxis, wie sie der Design-Based-Research-Ansatz in der Forschung verfolgt, in Betracht zu ziehen. Insgesamt zeigen die Befunde dieses Dissertationsvorhabens, dass das entwickelte und validierte AAI MINT-Aufgaben in neun Qualitätsmerkmalen messen kann. Die Resultate können in weiteren Analysen dazu genutzt werden, Aufgaben den verschiedenen Typen nach dem Lernprozessmodell zuzuordnen und gemäß einer theoretischen Norm zu überarbeiten. Somit lassen sich einerseits Schlüsse zur Aufgabenanalyse ziehen und andererseits Folgerungen für zukünftige Aufgabenentwicklungsprojekte diskutieren.

Inhaltsverzeichnis

1 EINFÜHRUNG IN DAS FORSCHUNGSPROJEKT 1
1.1 Problemstellung und übergeordnete Fragestellung 1
1.2 Theoretische Einführung 3
1.2.1 Kompetenzabbild 3
1.2.2 Lebensweltbezug 4
1.2.3 Lernendenvorstellung 5
1.2.4 Wissensart 6
1.2.5 Wissensaktivität 7
1.2.6 Repräsentationsform 8
1.2.7 Offenheit 8
1.2.8 Lernunterstützung 10
1.2.9 Lernweg 11
1.3 Studiendesign des Forschungsprojektes 12

2 ÜBERSICHT DER PUBLIKATIONEN 15
2.1 Messinstrument zur Analyse und Kategorisierung von MINT-Aufgaben – Konstruktion und Validierung 15
2.2 Quality of task sets – An instrument for analysing science tasks with different functions along the learning process 18
2.3 Empirische Unterscheidung von Aufgabentypen – eine explorative Studie 21
2.4 Überschätzen die Lehrmittelautor:innen den authentischen Lebensweltbezug von MINT-Aufgaben? Eine Studie zur Lernendenperspektive 23

3 GESAMTDISKUSSION DER PUBLIKATIONEN UND SCHLUSSFOLGERUNGEN 26
3.1 Übergreifende methodische Diskussion der AAI-Skalen 27
3.2 Zentrale Ergebnisse und Diskussion der einzelnen AAI-Skalen 29
3.3 Ausblick 42
3.4 Fazit 44

LITERATUR 45

ANHANG A: PUBLIKATION TEILSTUDIE I 56

ANHANG B: PUBLIKATION TEILSTUDIE II 58

ANHANG C: PUBLIKATION TEILSTUDIE III 60

ANHANG D: PUBLIKATION TEILSTUDIE IV 67

EIDESSTATTLICHE ERKLÄRUNG **72**

DANKSAGUNG **73**

1 Einführung in das Forschungsprojekt

1.1 Problemstellung und übergeordnete Fragestellung

Aufgaben gelten als Kern und Zentrum des kompetenzfördernden Unterrichts und sind Bestandteil jedes Fachunterrichts (Heins, 2017, S. 23). Aus Sicht von Lehr-Lern-Situationen sind sie ein wesentliches Steuerungsinstrument und betreffen wichtige Schritte des Lernweges. Reusser (2014) hebt die Relevanz von Aufgaben speziell in heterogenen Lerngruppen hervor. Aufgaben dienen gemäß Reusser (2014, S. 79) dazu, fachliche und überfachliche Kompetenzen aufzubauen, und können als Leistungsaufgaben der Überprüfung von Bildungsstandards oder der Evaluation der Zielerreichung dienen. Daher schreiben Müller und Helmke (2008, S. 32) Aufgaben eine wesentliche Rolle für Unterrichtsentwicklung und -qualität zu. Somit bilden Aufgaben international ein fachdidaktisches Thema der Unterrichtsentwicklung – dies insbesondere im Zeitalter der Kompetenzförderung (Bölsterli Bardy & Wilhelm, 2018; Brod, 2021). Folglich ist die Analyse von Aufgaben ein bedeutendes Feld für schulische Lehr-Lernprozesse (Kleinknecht et al., 2013; Lombardi et al., 2021; Reusser, 2014; Stäudel et al., 2012). Darüber hinaus stärkt die Forderung nach Individualisierung des Lernens durch Binnendifferenzierung des Unterrichts den Aufgaben-Diskurs (Maier et al., 2014). Aufgaben werden auf der Mikroebene des Unterrichts als kleinste Einheit des Kompetenzaufbaus und der Schulleistungsmessung betrachtet.

In der Allgemeinen Didaktik der 60er- und 70er-Jahre des vergangenen Jahrhunderts spielen Aufgaben noch kaum eine Rolle. Erst in Folge der Schulleistungsmessungen, vorwiegend mit dem Programme for International Student Assessment (PISA), oder mit Tests für das Erreichen nationaler Bildungsziele rücken Aufgaben vermehrt in den Fokus der empirischen Unterrichtsforschung und werden für die meisten Fachdidaktiken ein wichtiger Forschungsgegenstand (Maier et al., 2014, S. 35–36). So setzen sich unter anderem Adamina (2010), Bölsterli Bardy und Wilhelm (2018), Büchter und Leuders (2016), Heins (2017), Jordan et al. (2006), Leisen (2006), Neubrand (2002) sowie Winkler (2011) mit Systemen zur Klassifikation von Aufgaben unter fachdidaktischen Aspekten auseinander. In der

Allgemeinen Didaktik beschäftigen sich z.B. Blömeke et al. (2006), Maier et al. (2010) und Luthiger et al. (2018) mit Kriterien von Aufgaben. Aufgrund ihrer hohen Relevanz ist auch bei Lehramtsstudierenden die didaktische Funktion und Passung von Aufgaben im Professionalisierungsprozess zentral. Ebenso ist in den Lehrerfortbildungen die Frage der Aufgabenqualität immer wieder ein Thema. Digitalisierte und individualisierte Lernangebote basieren im Wesentlichen auf einer qualitativ hochwertigen Zusammenstellung von Aufgaben (Maier et al., 2014, S. 36). All diesen skizzierten Anwendungsbereichen von Aufgaben ist gemein, dass dort eine differenzierte Analyse und Bewertung der Aufgabenqualitäten benötigen wird.

Damit der in den Lehrplänen anvisierte Kompetenzaufbau der Lernenden unterstützt werden kann, treten Sets von aufeinander aufbauenden Aufgaben in den Fokus (Astleitner, 2006). Der dieser Dissertation zugrunde liegende kompetenzfördernde Lernprozessmodellansatz (Luthiger et al., 2018; Stuppan, Wilhelm, & Bölsterli Bardy, 2022; Stuppan et al., under review) unterscheidet Aufgaben in fünf Typen: Konfrontation, Erarbeitung, Übung, Synthese und Transfer. In bisherigen Kategoriensystemen können einzelne Aufgaben nominal beschrieben werden. Weil in der fach- als auch in der allgemeindidaktischen Lehr-Lernforschung empirisch validierte Instrumente zur Analyse von Aufgabensets fehlen, ist das Ziel dieser Entwicklungs-, Validierungs- und Anwendungsstudie, ein Aufgaben-Analyse-Instrument (AAI) auf der Grundlage bestehender Kategoriensysteme zu entwickeln, zu validieren und anzuwenden (Stuppan, Wilhelm, & Bölsterli Bardy, 2022; Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022).

Ausgehend von diesen Überlegungen, konzentrieren sich die einzelnen Beiträge auf folgende vier zentrale Fragestellungen:

1) Welche Items erfassen die Qualität von Aufgaben valide, reliabel und objektiv, so dass der Itemsatz zur Analyse und Klassifikation von Aufgaben (Aufgabenprofile) eingesetzt werden kann?

2) Inwiefern weichen die Qualitätsmerkmale von analysierten MINT-Aufgaben, gepoolt in Aufgabentypen, von denen von einem theoretischen Expert:innenvorschlag ab?

3) Können mit dem AAI erfasste Aufgaben den Aufgabentypen aus dem Lernprozessmodellansatz zugeordnet werden?

4) Wie unterscheiden sich die Beurteilungen der Lernenden von der Setzung der Lehrmittelautor:innen hinsichtlich des Lebensweltbezugs?

Diesen Fragen wird in den einzelnen ausgearbeiteten Publikationen (Kapitel: 2 Übersicht der Publikationen) im Detail nachgegangen. Daraus lässt sich folgende übergeordnete Fragestellung herleiten:

Welches Potenzial lässt sich aus der Entwicklung, Validierung und Anwendung des AAI ableiten?

Diese übergeordnete Fragestellung wird in der Gesamtdiskussion wieder aufgenommen, indem einerseits das AAI übergreifend, andererseits die neun Merkmale einzeln erörtert werden.

1.2 Theoretische Einführung

Wilhelm et al. (2014) haben in einem heuristischen Verfahren mithilfe von empirischen (Blömeke et al., 2006; Jordan et al., 2006; Maier et al., 2010; Neubrand, 2002) und theoretischen Arbeiten (Büchter & Leuders, 2016) relevante Merkmale für die Qualität von Aufgaben formuliert und daraus das LUKAS-Kategoriensystem (Luzerner Modell zur Entwicklung Kompetenzfördernder Aufgabensets, LUKAS) entwickelt (Luthiger et al., 2018). Um das Potenzial von Aufgaben nicht nur nominal, vielfach durch Verwendung von Textbausteinen, sondern in Form von Skalenwerten beschreiben zu können, wurde das AAI von Stuppan, Wilhelm, Bölsterli Bardy et al. (2022) auf der Grundlage des LUKAS-Kategoriensystems entwickelt und validiert. Mit dem AAI lässt sich das Aufgabenpotenzial in neun Skalen und entsprechenden Subskalen operationalisieren und erfassen. Im Folgenden werden die Qualitätsmerkmale aus dem AAI dargelegt, indem auf das Konstrukt der Skalen eingegangen und die theoretischen Grundlagen ausgeführt werden.

1.2.1 Kompetenzabbild

Die Skala *Kompetenzabbild* erfasst die Anzahl der Subkompetenzen eines Bildungsstandards, die mit einer Aufgabe entwickelt werden können (Flechsig, 2008,

S. 245; Lersch & Schreder, 2013, S. 50; Nehring et al., 2016; Schecker & Parchmann, 2006). Gemäß dem Bericht der Arbeitsgruppe HarmoS (Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule) besteht eine Kompetenz aus einer Fähigkeit/Fertigkeit, die mit einem Inhalt/Konzept verbunden ist (EDK, 2011). Mit dem *Kompetenzabbild* werden in der Aufgabe die Kompetenzbereiche eines Bildungsstandards erfasst und in den Subskalen *singulär*, *additiv* sowie *integrativ* ausgegeben.

Kompetenzabbild – singulär: die Aufgabe befasst sich nur mit einem Aspekt einer (Sub-)Kompetenz der Bildungsstandards (Lersch & Schreder, 2013). Hier wird nur eine Fähigkeit und/oder Fertigkeit entwickelt, verbessert, geübt oder angewendet.

Bezieht sich eine Aufgabe auf mehrere Aspekte einer oder mehrerer curricularer (Sub-)Kompetenzen, wird sie als *Kompetenzabbild – additiv* bewertet. In diesem Fall werden mehrere Fähigkeiten und/oder Fertigkeiten entwickelt, verbessert, geübt oder angewendet.

Kompetenzabbild – integrativ: die Aufgabe umfasst möglichst viele Aspekte einer oder mehrerer (Sub-)Kompetenzen der curricularen Bildungsstandards (Luthiger et al., 2018, S. 61). In diesem Fall werden viele Fertigkeiten und/oder Fähigkeiten entwickelt, verbessert, geübt und auf eine vernetzte Weise angewendet. Die theoretischen Grundlagen der Skala *Kompetenzabbild* bilden Brophy und Alleman (1991, S. 15), Flechsig (2008, S. 254) sowie Lersch und Schreder (2013, S. 50).

1.2.2 Lebensweltbezug

Blömeke et al. (2006, S. 337), Heinle et al. (2022), Klieme et al. (2003, S. 79) und Leisen (2006, S. 6) haben den Lebensweltbezug (Repräsentation einer authentischen Situation) als unter anderem didaktisches und fachliches Merkmal hoher Aufgabenqualität beschrieben. Mit der Skala *Lebensweltbezug* wird das Verhältnis zwischen dem domänenspezifischen Fachwissen (z.B. wissenschaftlichen Erkenntnissen) und der individuellen Erfahrungs- und Lebenswelt der Lernenden beschrieben (Fuhrmann, 2020; Maier et al., 2010, S. 37–38; Stäudel et al., 2012; Stein et al., 1996, S. 486). Hierzu werden die drei Subskalen *konstruiert*, *authentisch* und *real* in einer Aufgabe erfasst.

Eine Aufgabe wird als *Lebensweltbezug – konstruiert* bezeichnet, wenn es eine künstliche Verknüpfung zwischen dem Kontext der Aufgabe und der Lebenswelt der Lernenden gibt (Bohl et al., 2015; Brandl, 2016; Maier et al., 2010, S. 38).

Eine Aufgabe im Sinne von *Lebensweltbezug – authentisch* ist eine Aufgabe, bei der der Lebensweltbezug für die Lernenden im Moment noch konstruiert erscheint, jedoch im Kontext logisch und authentisch ist (Blömeke et al., 2006; Luthiger et al., 2018). Der Kontext der Aufgabe kann in Zukunft für die Lernenden relevant sein. Somit wird der Erfahrungshorizont reflektiert und in die Zukunft erweitert (Luthiger et al., 2018; Stuppan, Wilhelm, & Bölsterli Bardy, 2022). Die Aufgabe tritt mit den Lernenden über eine reale Fremd- oder Selbsterfahrung in Kontakt (Luthiger et al., 2018, S. 61).

Eine Aufgabe wird als *Lebensweltbezug – real* bezeichnet, wenn ihr Kontext eine Situation beschreibt, in der sich die Lernenden mit einem (echten) Problem beschäftigen müssen, das tatsächlich im Hier und Jetzt gelöst werden muss (Luthiger et al., 2018, S. 61). In einem *realen* Kontext gibt es kaum einen Unterschied zwischen der Aufgabe und der Lebenswelt der Lernenden (Maier et al., 2010, S. 38; Neubrand, 2002, S. 113). Die Skala hat folgende Autor:innengruppen als theoretischer Hintergrund: Brophy und Alleman (1991, S. 19), Blömeke et al. (2006, S. 337), Flechsig (2008, S. 254), Maier et al. (2014, S. 37–38), Neubrand (2002, S. 112–114) sowie Stein et al. (1996, S. 486)

1.2.3 Lernendenvorstellung

Kompetenzen werden auf den Grundlagen vorhandener Vorstellungen aktiv revidiert, um- oder ausgebaut (Gropengießer & Marohn, 2018; Morrison & Lederman, 2003; Riemeier, 2007; Wilhelm & Kalcsics, 2017, S. 88). Duit erläutert: „Vorstellungen bestimmen das Lernen, weil man das Neue nur durch die Brille des bereits Bekannten ‚sehen' kann" (Duit, 1995, S. 1). Deshalb sollen Aufgaben an das vorhandene Wissen und Können der Lernenden anknüpfen (Baalmann et al., 2004; Chen et al., 2020; Heeg et al., 2021; Luthiger et al., 2018, S. 62; Taber, 2019). In Anlehnung an von Aufschnaiter und Alonzo (2018), Beerenwinkel et al. (2007), Luthiger et al. (2018, S. 62) sowie Wilhelm et al. (2014) wird die Skala *Lernendenvorstellung* bei Aufgaben in die drei Subskalen *erheben*, *abgleichen* und *reflektieren* unterteilt.

Lernendenvorstellung – erheben bedeutet, dass die Lernenden ihre Vorstellungen bezüglich der Aufgabe mitteilen oder aufschreiben. Lösungsideen werden etwa aufgeschrieben, mündlich untereinander ausgetauscht oder in anderer Form kommuniziert (Heeg et al., 2021; Mietzel, 2007, S. 305–306).

Lernendenvorstellung – abgleichen führt dazu, dass die Vorstellungen der Lernenden (z.B. eine Annahme über eine Lösung oder eine Argumentation) mit der Lösung (z.B. wissenschaftliche Erkenntnisse oder Fachwissen) der Aufgabe abgeglichen werden (Hattie & Yates, 2015, S. 108–109).

Lernendenvorstellung – reflektieren bedeutet, dass die Lernenden über ihre Wissensveränderung (Prä-/Post-Konzept) reflektieren. Die Lernenden müssen unter anderem darüber nachdenken, warum ihre Vermutung mit der Lösung übereinstimmt oder davon abweicht (Herrington & Herrington, 2006; Luthiger et al., 2018, S. 62).

1.2.4 Wissensart

Die Skala *Wissensart* stellt in Anlehnung an Anderson und Krathwohl (2001, S. 69–88), Astleitner (2006, S. 36), Blömeke et al. (2006, S. 337) sowie Neubrand (2002, S. 107–112) die Grundelemente, die beim Lösen der Aufgabe von den Lernenden verlangt werden, in den Subskalen *Fakten*, *Prozeduren*, *Konzepte* und *Metakognition* dar.

Faktenwissen umfasst Wissen über Terminologie, Begriffe, spezifische Details und Fakten (Anderson & Krathwohl, 2001; Gagné et al., 1992; Jordan et al., 2006; Kauertz et al., 2010; Lenske et al., 2015).

Prozedurales Wissen beschreibt Wissen, das für das Handeln nutzbar ist. Dabei handelt es sich um ein Kontinuum von basalen Verhaltensweisen bis hin zu komplexen Handlungsweisen (z.B. Lösen unbekannter Probleme). Prozedurales Wissen kann kognitive Aspekte und/oder Handlungsabläufe (z.B. zur Durchführung eines Experiments) beinhalten (Anderson & Krathwohl, 2001; Artelt & Wirth, 2014; Blömeke et al., 2006; Brühwiler et al., 2017; Luthiger et al., 2018, S. 63).

Lernende wenden *konzeptuelles Wissen* beim Lösen von Aufgaben an, wenn sie ihr Wissen explizit vernetzen, klassifizieren oder strukturieren (Anderson &

Krathwohl, 2001; Gagné et al., 1992; Hiebert & Lefevre, 1986, S. 3–5; Jordan et al., 2006).

Lernende nutzen *metakognitives Wissen*, wenn sie explizit über ihr Wissen oder ihre Problemlösungsstrategien nachdenken (Anderson & Krathwohl, 2001; Bremerich-Vos, 2008; Den Elzen-Rump et al., 2008; Kleinert et al., 2021; Maier et al., 2010; Sjuts, 2022).

1.2.5 Wissensaktivität

Die Skala *Wissensaktivität* bezieht sich auf die Prozesse, die eine Aufgabenbearbeitung von den Lernenden einfordert (Anderson & Krathwohl, 2001; Luthiger et al., 2018, S. 63; Töpfer et al., 2022). Gemäß Maier et al. (2010, S. 32) laufen, je nachdem wie komplex oder neuartig das neue Wissen ist, unterschiedlich anspruchsvolle Prozesse ab. Dennoch, alle Wissensarten, nicht nur Prozeduren, können in jeder Nutzungsform vorkommen. So können sowohl Faktenwissen, prozedurales, konzeptionelles als auch metakognitives Wissen zur Interpretation (Verstehen) von Informationen transferiert werden. In Anlehnung an Anderson und Krathwohl (2001) beinhaltet die *Wissensaktivität* die drei Subskalen *reproduzieren*, *transferieren* und *kreieren*.

Wissensaktivität – reproduzieren bedeutet, dass ein Abruf von Wissen aus dem Gedächtnis erforderlich ist, um die Aufgabe zu lösen (Anderson & Krathwohl, 2001, S. 215; Vorholzer & von Aufschnaiter, 2020). Die Aufgabe hat für die Lernenden einen Wiederholungscharakter. Die Reproduktion kann sich auf alle vier Wissensarten beziehen und nicht nur auf das Erinnern von Faktenwissen.

In Anlehnung an Ellis (1965), Maier et al. (2010, S. 32) sowie Dori und Sasson (2013) liegt eine Transferleistung dann vor, wenn sich Aufgabensituation und gespeichertes Wissen unterscheiden und somit mehr als die bloße Reproduktion von Vorwissen erforderlich ist. Wird das vorhandene Wissen in einer veränderten, neuen, unbekannten Situation angewendet oder ist für die Lernenden nicht sofort ersichtlich, welche Wissensart zur Anwendung kommen soll, wird ebenfalls von *Wissensaktivität – transferieren* gesprochen.

Als *Wissensaktivität – kreieren* wird eine Situation (vielfach eine unbekannte Aufgabe) bezeichnet, die Lernenden kreative Leistungen im Sinne von Zusam-

menfügen oder Kombinieren von vorhandenen Wissenselementen abverlangt, sodass neues Wissen für die Lösungsidee entstehen kann (Anderson & Krathwohl, 2001, S. 215; T. R. Johnson & Krems, 2001; Krüger & Upmeier zu Belzen, 2021; Rieck, 2005, S. 13–14; Töpfer et al., 2022).

1.2.6 Repräsentationsform

Auf der Grundlage von Ainsworth (2006), Jordan et al. (2006, S. 55–56), Neubrand (2002, S. 120), Stein et al. (1996) und Wilhelm et al. (2014) wird mit der Skala *Repräsentationsform* die Komplexität einer Aufgabe erfasst, das heißt die Wirkung der Repräsentationsformen des Wissens in einer bestimmten Aufgabe. Maier et al. (2010, S. 40) stellen sich, um die komplexitätssteigernde Wirkung der Repräsentationsformen des Wissens in einer Aufgabe analysieren zu können, zwei Fragen: 1) In welcher Form wird das Wissen in der zu bearbeitenden Aufgabe den Lernenden präsentiert? 2) In welcher Repräsentationsform wird die Aufgabe bearbeitet bzw. die Lösung erstellt?

Die Skala *Repräsentationsform – Transformation* erfasst die Wirkung der Repräsentationsform auf einem Kontinuum von „keine Transformation“ bis „Transformation“. *Keine Transformation* bedeutet, dass die Form der Repräsentation in der Aufgabenstellung, in der Aufgabenbearbeitung und in der erarbeiteten Lösung gleich ist. *Transformation* bedeutet, dass sich die Darstellungsform (z.B. Texte, Abbildungen, Tabellen usw.) der Aufgabenstellung von der Darstellungsform der Aufgabenbearbeitung unterscheidet. Das heißt, die Lernenden müssen den Lerninhalt in eine andere Darstellungsform transformieren (Klein et al., 2018; Maier et al., 2010, S. 39–40; Neubrand, 2002, S. 120; Stein et al., 1996; Sumfleth & Nakoinz, 2019).

1.2.7 Offenheit

Die Skala *Offenheit* gibt Aufschluss darüber, inwieweit die Lernenden die Aufgabe selbstständig bearbeiten können. Mit der Skala *Offenheit* wird in einer Aufgabe die Offenheit 1) hinsichtlich des *Lösungsvorgehens* (ein Kontinuum von *frei* bis *erklärt)*, 2) hinsichtlich des *Resultats* (ein Kontinuum von *vielfältig* bis *einheitlich)* und 3) in Bezug auf die *Bearbeitung (*ein Kontinuum von *vorstrukturiert* bis *unstrukturiert)* erfasst. Die Skala greift auf Vorschläge der Autor:innengruppen

Blömeke et al. (2006), Braund und Reiss (2006), Jordan et al. (2006) sowie Neubrand (2002) zurück.

Die Subskala *Offenheit – Lösungsvorgehen* erfasst ein Kontinuum von *frei* bis *erklärt* in einer Aufgabe. *Frei* bedeutet: Diese Aufgabe lässt Varianten des Lösungsvorgehens zu (Hmelo-Silver et al., 2007; Maier et al., 2010, S. 36; Niss et al., 2007). Auf diese Weise ist die Aufgabe mehrschichtig und anspruchsvoll, während *erklärt* ausdrückt, dass die Aufgabe ein eindeutig definiertes Lösungsvorgehen beinhaltet. Der Prozess der Lösungsfindung ist unter anderem in einem Algorithmus (definierte Schritte, Steuerung von modernen Gebäudetechniken oder Verkehrsanlagen usw.) abzubilden.

Die Subskala *Offenheit – Resultat* (Ergebnis/Ziel) beschreibt ein Kontinuum von *vielfältig* bis *einheitlich* in einer Aufgabe. *Vielfältig* bringt mehrere mögliche Resultate oder Erklärungen hervor, während *einheitlich* zu einer möglichen Erklärung oder Lösung hinführt (Arnold et al., 2017; Rieck, 2005, S. 12; Schukajlow & Krawitz, 2020; Stein et al., 1996).

Die Subskala *Offenheit – Bearbeitung* definiert die Bearbeitungsstruktur der Aufgabenelemente (z.B. Texte oder Abbildungen) und die Übereinstimmung dieser Strukturen im Hinblick auf die Lösungsfindung (Aufgabenbearbeitung und Lösungsprozess) (Kleinert et al., 2021; Luthiger et al., 2018, S. 63–64; Maier et al., 2010, S. 36; Neubrand, 2002, S. 122–129; Plath, 2020). Das Kontinuum reicht von *vorstrukturiert* bis *unstrukturiert*. Die Aufgabe besteht in der Regel aus einem oder verschiedenen Elementen. *Vorstrukturiert* bedeutet, dass die Struktur der Aufgabe klar definiert und die Elemente in der Reihenfolge der Aufgabenbearbeitung angeordnet sind (Luthiger et al., 2018, S. 64). *Unstrukturiert* bedeutet, dass die Reihenfolge der Elemente in der Aufgabe nicht mit der Reihenfolge der Lösungsfindung übereinstimmt. Folglich müssen die Lernenden die Struktur beim Bearbeiten der Aufgabe selbst bestimmen.

Wenn eine Aufgabe *frei*, *vielfältig* und *unstrukturiert* ist, können die Lernenden den Ablauf der Aufgabe frei bestimmen und die Lösung ist offen. Außerdem entsprechen die Aufgabenelemente nicht der Reihenfolge, in welcher die Lernenden die Informationen verarbeiten müssen. Wie die Lernenden eine freie, vielfältige, unstrukturierte Aufgabe lösen, hängt unter anderem von den Lernvoraussetzungen

wie den Lernendenvorstellungen, ihrer Motivation oder ihren Lösungsstrategien ab (Arnold et al., 2017; Kleinert et al., 2021; Schukajlow & Krawitz, 2020; Ulrich, 2019, S. 79). Die Komplexität einer freien, vielfältigen und unstrukturierten Aufgabe lässt sich bis zu einem gewissen Grad durch die Qualität und Struktur der Informationen, die den Lernenden zur Verfügung gestellt werden, steuern (Luthiger et al., 2018, S. 63). Im Allgemeinen verlangt sie von den Lernenden jedoch mehr Kompetenzen ab, als eine erklärte, einheitliche und strukturierte Aufgabe.

1.2.8 Lernunterstützung

Damit die Lernenden eine Aufgabe individuell bearbeiten können, werden stellenweise Unterstützungsangebote präsentiert, die sie bei Bedarf in Anspruch nehmen können (Kleinert et al., 2021; Ruelmann et al., 2021; Shute, 2008). In Anlehnung an Astleitner (2006, S. 37–38), Luthiger et al. (2018, S. 65–66), Mietzel (2007, S. 381), Siemon et al. (2018) sowie Wilhelm et al. (2014) werden bei der Skala *Lernunterstützung* folgende drei Subskalen erhoben: *Hilfestellung*, *Austauschmöglichkeit* und *Lehrpersonenfeedback*.

Lernunterstützung – Hilfestellung beschreibt, dass zu Beginn oder während der Aufgabenbearbeitung den Lernenden Hinweise (Prompts) oder Hilfestellungen (Scaffolds) zur Verfügung gestellt werden (Arnold et al., 2017; Chen et al., 2020; Hild et al., 2020; Leisen et al., 1999; Luthiger et al., 2018, S. 65; van de Pol et al., 2010). Die Hilfestellungen werden oft schrittweise segmentiert angeboten, um die kognitive Belastung zu verringern und die begrenzte Kapazität des Arbeitsgedächtnisses nicht zu überlasten (Arnold et al., 2017; Kirschner et al., 2006; Mayer & Moreno, 2003; Sweller et al., 1998).

Lernunterstützung – Austauschmöglichkeit: den Lernenden wird durch die spezifische Aufgabe die Möglichkeit geboten, dass sie sich während der Aufgabenbearbeitung austauschen. Durch die Interaktionen werden kognitive Prozesse angeregt, die für das Behalten und Verstehen der Lerninhalte bedeutend sein können (Dunekacke et al., 2022; Hattie & Yates, 2015; D. W. Johnson & Johnson, 2009; Schroeder et al., 2007). Lernende profitieren nicht nur, wenn sie anderen etwas erklären, sondern ebenso, wenn sie etwas erklärt bekommen (Mietzel, 2007, S. 381; Schriebl et al., 2022).

Lernunterstützung – Lehrpersonenfeedback beschreibt, dass die Lehrkraft unmittelbar nach der Bearbeitung der Aufgabe oder zu einem späteren Zeitpunkt (z.B. in der nächsten Unterrichtsstunde) ein Feedback gibt. So wird die Aufgabenleistung bewertet und mit einer Norm verglichen, es werden Möglichkeiten zur Verbesserung der Leistung beschrieben oder unvollständige oder falsche Lösungen korrigiert. Ferner bietet das Lehrpersonenfeedback strukturelle Unterstützung und enthält motivierende Botschaften (Buholzer et al., 2020; Chin, 2006; Erhel & Jamet, 2013; Hattie & Wollenschläger, 2014; Hattie & Yates, 2015; Jurik et al., 2014; Kingston & Nash, 2011; Zhu et al., 2020).

1.2.9 Lernweg

In Anlehnung an Büchter und Leuders (2016, S. 111), Jonas (2020), Lee et al. (2018), Luthiger et al. (2018, S. 66), Niggli (2013, S. 35) sowie Wilhelm et al. (2014) werden für die Skala *Lernweg* drei Subskalen entwickelt: *kompensierend*, *profilbildend* und *selbstdifferenzierend*.

Aufgaben, die mit der Subskala *Lernweg – kompensierend* bewertet werden, können unterschiedliche Lernstände ausgleichen und die Möglichkeit bieten, fehlendes Wissen zu ergänzen und Unklarheiten aufzuarbeiten (Orientierung am Kompensationsprinzip). Werden unterschiedliche Lernvoraussetzungen kompensiert, treten anstelle der gesamten Schulklasse vermehrt einzelne Lernende oder Lerngruppen als Adressaten in den Vordergrund. Somit lässt die Aufgabe einen Ausgleich unterschiedlicher Lernvoraussetzungen wahrscheinlich werden (Friedrich & Mandl, 1992; Niggli, 2013, S. 36; Stebler et al., 2022).

Eine als *Lernweg – profilbildende* apostrophierte Aufgabe bietet ein Wahlangebot nach individuellen Interessen an, damit die Lernenden in ihrem eigenen Profil eine Kompetenzerweiterung erfahren (Götz & Nett, 2011; Luthiger et al., 2018, S. 66; Mötteli et al., 2022).

Lernweg – selbstdifferenzierend bedeutet, dass eine Aufgabe auf verschiedenen Abstraktionsniveaus, individuellen Lern- und Bearbeitungswegen gelöst werden kann sowie Vertiefungsmöglichkeiten zulässt (Affeldt et al., 2018; Baumann et al., 2018; Brauns & Abels, 2021; Büchter & Leuders, 2016, S. 111).

1.3 Studiendesign des Forschungsprojektes

Das vorliegende Dissertationsvorhaben besteht aus vier Teilstudien (einer Entwicklungs- und Validierungsstudie und drei Anwendungsstudien), die gesamthaft in dieser Arbeit zusammengefasst werden.

In *Teilstudie I* (*Entwicklung und Validierung*) wird das operationalisierte Aufgaben-Analyse-Instrument (AAI) entwickelt. Es beruht auf bestehenden und teilweise erprobten Kategoriensystemen (Blömeke et al., 2006; Büchter & Leuders, 2016; Luthiger et al., 2018; Neubrand, 2002; Wilhelm et al., 2014). Der entwickelte Itempool des AAI wird unter Berücksichtigung der Testgütekriterien mehrfach validiert (Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Die zur Validierung verwendeten Aufgaben ($N = 146$) stammen aus dem *„MINT unterwegs“*-Projekt. Weil diese Aufgaben nach dem Lernprozessmodell erstellt wurden (DVS, 2020; Wilhelm et al., 2014), sind sie für die Instrumentenvalidierung geeignet. *„MINT unterwegs“* ist ein Kooperationsprojekt der Pädagogischen Hochschule Luzern und der Dienststelle Volksschulbildung Luzern (DVS, 2020). Während acht Jahren (2016 bis 2023) sollen möglichst viele Schulklassen an entsprechenden Projektwochen teilnehmen und *„MINT unterwegs“* nutzen können. Schulklassen erhalten während einer Woche Experimentierboxen und Unterrichtsmaterialien zu Bildungsstandards (EDK, 2011) der Bereiche Mathematik, Informatik, Naturwissenschaften und Technik (MINT). In *Teilstudie II* (*Anwendung*) zur *Aufgabenqualität von MINT-Aufgaben*, werden alle Aufgabentypen aus dem *„MINT unterwegs“*-Projekt gemäß den Angaben der Lehrmittelautor:innen indiziert (z.B. $n_{\text{Konfrontationsaufgaben}} = 25$), gepoolt und mit dem AAI von geschulten Rater:innen beurteilt (Stuppan, Wilhelm, & Bölsterli Bardy, 2022). Die einzelnen AAI-Subskalenwerte der indizierten MINT-Aufgabentypen werden anschließend einem theoretischen Expert:innenvorschlag gegenübergestellt. Zur Visualisierung der Abweichungen dient ein „Ampelsystem“, das ganzzahlige Differenzen in Form einer Farbcodierung darstellt. Die gewonnenen Erkenntnisse dieser Teilstudie dienen dazu, die MINT-Aufgaben nach spezifischen Qualitätsmerkmalen aus dem AAI zu überarbeiten und eine erste Anwendung des Messinstruments aufzuzeigen (Stuppan, Wilhelm, & Bölsterli Bardy, 2022). In *Teilstudie III* (*Anwendung*) werden mit dem AAI erfasste MINT-Aufgaben ($n = 58$) aufgrund ihrer beurteilten Qualitätsmerkmale mittels einer Clusteranalyse gruppiert. Ziel ist es, die theoreti-

sche Annahme, dass sich Aufgabentypen aufgrund ihrer spezifischen Merkmalsausprägungen voneinander unterscheiden lassen, empirisch zu prüfen (Stuppan et al., under review). Gemäß dem Lernprozessmodell müssten sich beispielsweise Konfrontations- und Übungsaufgaben in zwei Gruppen einteilen lassen, da sie sich in ihren Merkmalsausprägungen stark unterscheiden. In der *Teilstudie IV* (*Anwendung*) werden exemplarisch ausgewiesene Aufgaben ($N = 16$) von Lernenden ($N = 805$) mit dem AAI-Merkmal Lebensweltbezug, einer gekürzten Version des Messinstruments, eingeschätzt und mit den Ratings der fachdidaktischen Lehrmittelautor:innen verglichen. Infolgedessen wird untersucht, ob die Setzungen der Lehrmittelautor:innen ökologisch valide sind und inwiefern sich die beiden Perspektiven unterscheiden (Stuppan et al., 2023). Basierend auf den vier dargelegten Teilstudien können finale Kurzskalen ausgewiesen und diskutiert werden, die als Rückkopplung zur Theoriepräzisierung genutzt oder in weiteren Entwicklungs- und Forschungsprojekten eingesetzt werden können.

Abbildung 1.1 stellt die iterative Konstruktion und Validierung (*Teilstudie I*, blaue Färbung), die Anwendung (*Teilstudie II* bis *IV*, grüne Färbung) und abschließende Schlussfolgerung (pink Färbung) dieses Dissertationsprojektes als Ablaufmodell dar. Hierzu werden die verschiedenen Strategien als einzelne *Schritte* (1. bis 11.) beschrieben. Die Spalte *Ebene und Ziele* zeigt auf, bei welchem Schritt welche Ebene (wie Konstrukt, Item, Messinstrument etc.) mit welcher Zielerreichung zum Einsatz kommt. Die Spalte *Vorgehensweise und Methode* beschreibt einerseits die Forschungsmethode (wie qualitativ oder quantitativ), andererseits wird prägnant beschrieben, wie die entsprechenden Schritte methodisch umgesetzt werden. Die Spalte *Gütemaße* oder *statistische Kennwerte* führt die entsprechenden Kennwerte, die zur Überprüfung der Zielerreichung eingesetzt werden, an. Die Farbcodierung dient zur visuellen Hervorhebung der einzelnen Teilstudien. Dabei wird jeweils eine Farbcodierung für eine Publikation verwendet, beispielsweise blauer Hintergrund, *Teilstudie I* Publikationstitel: *Messinstrument zur Analyse und Kategorisierung von MINT-Aufgaben – Konstruktion und Validierung*. Auf die finalen AAI-Skalen wird in dieser Arbeit im Kapitel Gesamtdiskussion eingegangen.

Ablaufmodell: Entwicklungs-, Validierungs-, Anwendungsstudien sowie Schlussfolgerung des Aufgaben-Analyse-Instruments

	Schritte	Ebene und Ziele	Vorgehensweise und Methode	Gütemaße / statistische Kennwerte
Teilstudie I		**Konstrukt, Item- & Instrumentebene**	theoriebasiert	
	1. Konzeptspezifikation	Entwurf der Aufgaben-Kategorien & -Faktoren	Spezifische Literaturrecherche & -studium, deduktives Vorgehen	Kriterien der Genauigkeit, Exklusivität, Exhaustivität
	2. Operationalisierung	Item-Konstruktion, Itempool & Fragebogenentwurf	qualitativ: Itemtexte formulieren & Fragebogenkonstruktion	Augenscheinvalidität, 5-stufige Bewertungsskala
	3. Experten:innen-Befragungen	Qualitative Überarbeitung des Itempools	Experten:innen-Befragungen	
	4. Experten:innenteam Pretest & Revision	Itemselektion & -revision	quantitativ: Testmanual ausarbeiten & MINT-Aufgabenratings	Panel Meeting, ICC(C,1)
	theoriebasierter & empirisch geklärter Itempool			
		Item-, Faktoren- & Instrumentebene		
	5. Itempool den Konstrukten zuordnen	Überprüfung der Inhaltsvalidität	Digitales Closed Card Sorting von Fachpersonen	Übereinstimmungskoeffizienten (in %)
	6. Fd Raterschulung	Hohe Urteilübereinstimmung	Theoriebasierte Einführung, Aufgabenratings mit Testmanual	Interrater-Reliabilität
	7. Aufgaben beurteilen & Instrument prüfen	Datenerhebung der gesamten Aufgaben-Stichprobe	Voneinander unabhängige Aufgabenratings durch geschulte Fd	
		Überprüfung der Durchführungsobjektivität	Interrater-Reliabilität	ICC(A,1), Konfidenzintervall, Cohen's Kappa
		Überprüfung der faktoriellen Validität	Explorative Faktorenanalyse	Korrelation, KMO-Wert, Bartlett-Test, MAP-Test
		Überprüfung der Reliabilitätsmessung	Reliabilitätsanalyse	Cronbach-Alpha-Koeffizienten
	8. Beurteilung von projektnahen Aufgaben	Konvergente Validierung	Aufgabenkategorisierungen vergleichen	Konstrukt-Korrelationen
	Validierte AAI-Skalen			
Teilstudie II		**Skalen- & Aufgabenprofilebene**		
	9. Aufgabenprofile vergleichen	Skalenwertvergleiche	Skalenwerte von Aufgaben vergleichen	
		Theoretische Aufgabenprofile	Expert:innen Vorschlag für Aufgabenprofil ausarbeiten	AD_M
		Aufgabenprofile der indizierten MINT-Aufgaben	Skalenwerte für indizierte MINT-Aufgaben berechnen	Aufgabentyp-Index, ICC(A,1),
		Differenz: theoretische und indizierte Aufgabenprofile	Differenzen der Skalenmittelwerte berechnen	Differenz als Farbskala (Ganzzahl Index)
	Aufgabenqualität			
Teilstudie III		**Skalen- & Aufgabenprofilebene**		
	10. Aufgabentypen unterscheiden	Homogene Aufgaben-Gruppen	Clusteranalysen	
		Anzahl Clusterlösungen bestimmen	Hierarchisch-divisive Clusteranalyse, Dendrogramm	Konnektivität, Dunn-Index, Silhouettenwert
		Definitive Clustereinteilung	Partitioning Around Medoids, Jaccard-Bootstrap, PCA	Jaccard-Ähnlichkeitswert
		Aufgabentypen & Merkmalsausprägungen	Signifikanztests, Effektstärke	*T*-Test mit Welch-Korrektur, Cohen's *d*
	Empirische Aufgabentypen			
Teilstudie IV		**Skalenebene**		
	11. Perspektiven der Aufgaben-Einschätzung	Items & Instruktion sind für die Zielgruppen verständlich	Fragebogenkonferenz, Nachfrage- & Paraphrasierungstechnik	
		Itemanalyse & -selektion	Fragebogen-Pretests, Durchführungsmodalitäten prüfen	CFA (Fit-Indizes: CFI, SRMR, GFI, RMSEA), Cronbachs α
		Lehrmittelautor:innen- & Lernenden-Rating	Fragebogenerhebung mittels Testmanual	Active Consent
		Differenz: Lehrmittelautor:innen & Lernenden	Subskalenmittelwerte als POMP, Signifikanztests, Effektstärke	LMM, *T*-Test, α-Fehler-Kumulierung, Cohen's *d*
	Lebensweltbezug Perspektiven			
	Schlussfolgerung bzgl. den finalen AAI-Skalen und Rückkopplung zur Theoriepräzisierung			

Legende: Aufgaben-Analyse-Instrument (AAI), Average absolute deviation (AD_M), Confirmatory Factor Analysis (CFA), Fachdidaktiker:innen (Fd), Hauptkomponentenanalyse (principal component analysis, PCA), Intraklassenkorrelation (ICC), linear mixed-effects model (LMM), Minimum-Average-Partial-Test (MAP-Test), Scores Represented as the Percent of Maximum Possible (POMP)

Abbildung 1.1. AAI-Ablaufmodell: Entwicklungs-, Validierungs-, Anwendungsstudien sowie Schlussfolgerung des AAI (Erläuterungen im Text)

2 Übersicht der Publikationen

Im folgenden Abschnitt werden die im Rahmen des vorliegenden Dissertationsprojekts verfassten Artikel im Einzelnen vorgestellt. Hierzu werden der jeweilige Artikeltitel, die Autorenschaft, der aktuelle Bearbeitungsstand, eine Zusammenfassung, Schlüsselwörter, eine Verortung zur übergeordneten Fragestellung: „Welches Potenzial lässt sich aus der Entwicklung, Validierung und Anwendung des AAI ableiten?“, und die Eigenleistungen des Erstautors aufgeführt.

2.1 Messinstrument zur Analyse und Kategorisierung von MINT-Aufgaben – Konstruktion und Validierung

Sebastian Stuppan[a,b], Markus Wilhelm[a,b], Katrin Bölsterli Bardy[a] & Roland Künzle[a]

[a] Pädagogische Hochschule Luzern, Schweiz

[b] Pädagogische Hochschule Heidelberg, Deutschland

Eingereicht bei Proceedings Tagung Fachdidaktiken 2022

Status: publiziert (double-blind peer review)

DOI: https://doi.org/10.33683/dida.22.05.86

Zusammenfassung

In einem kompetenzorientierten Unterricht haben Aufgaben eine entscheidende Bedeutung. Mit Hilfe von Aufgabensets werden Gelegenheiten geschaffen, damit Lernende ihre Kompetenzen individuell entwickeln und darin gefördert werden. Im Forschungsvorhaben wird ein Messinstrument zur Analyse und Kategorisierung von MINT-Aufgaben erarbeitet. Dieser Artikel beschreibt und diskutiert die Konstruktion und Validierung des Aufgaben-Analyse-Instruments (AAI).

Schlüsselwörter

Validierungsstrategie, Messinstrument, MINT - Aufgaben, Kategorisierung, Skalenentwicklung

Verortung zur übergeordneten Fragestellung

Es existiert zurzeit kein operationalisiertes und validiertes Messinstrument, um Aufgaben analysieren und kategorisieren zu können (Luthiger et al., 2018, S. 58; Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Mit dem LUKAS-Kategoriensystem können Aufgaben einzelnen Ausprägungen nur beschreibend zugeordnet werden. Als Beispiel ist die von Reinfried (2016) durchgeführte Aufgabenanalyse zu erwähnen. Weiter wurden bisher mit keinem Analysesystem die Sichtweisen der Lernenden direkt erfragt und analysiert (Nachtigall et al., 2018; Stuppan et al., 2023; Weiss & Müller, 2015).

Hier setzt die *Teilstudie I* mit folgender Fragestellung an: Welche Items erfassen die Qualität von Aufgaben valide, reliabel und objektiv, sodass der Itemsatz zur Analyse und Klassifikation (Aufgabenprofile) von Aufgaben eingesetzt werden kann?

Die theoretische Grundlage zur Entwicklung des Messinstruments liefert das LUKAS-Kategoriensystem (Luthiger et al., 2018; Wildhirt & Luthiger, 2020; Wilhelm et al., 2014). Bei der Item-Konstruktion werden die von Bühner (2006), Busker (2014), Döring und Bortz (2016) sowie Mummendey und Grau (2014) beschriebenen Konstruktionsprozesse von Messinstrumenten berücksichtigt. Durch den mehrstufigen Entwicklungs- und Validierungsprozess kann das Messinstrument in Anwendungs- und Forschungszwecken vielseitig eingesetzt werden. Deshalb werden unter anderem gewisse Itemtexte differenziert für Fachdidaktiker:innen und Lernende ausgearbeitet. Mögliche Anwendungen des Messinstruments werden in *Teilstudie II* bis *IV* erläutert.

Eigenleistungen des Erstautors

In Tabelle 2.1 sind die Eigenleistungen des Erstautors in Prozent sowie die Orientierung im Text ausgewiesen.

Tabelle 2.1. *Eigenanteile und Orientierung im Text des Artikels: Messinstrument zur Analyse und Kategorisierung von MINT-Aufgaben – Konstruktion und Validierung*

Eigenanteile	Orientierung im Text
Konzeption/Initiative = 85 % Theoriearbeit = 90 % Datenerhebung = 50 % Analyse = 100 % Schreibarbeit = 95 %	Der Beitrag wurde primär vom Erstautor verfasst. Die Schreibarbeiten der Mitautor:innen beschränken sich auf Korrekturen und Anmerkungen. Der Viertautor hat bei der Datenerhebung mitgearbeitet.

2.2 Quality of task sets – An instrument for analysing science tasks with different functions along the learning process

Sebastian Stuppan[a,b], Markus Wilhelm[a,b] & Katrin Bölsterli Bardy[a]

[a] Pädagogische Hochschule Luzern, Schweiz

[b] Pädagogische Hochschule Heidelberg, Deutschland

Eingereicht bei Progress in Science Education (PriSE)

Status: publiziert (double-blind peer review)

DOI: https://doi.org/10.25321/prise.2022.1330

Zusammenfassung (in EN)

In competency-oriented education, tasks in science subjects have increased in importance in recent years. In addition, it is suggested that to develop a competency, a set of tasks with different functional embeddings is required. An option for the arrangement of tasks is explained in the Learning Process Model and starts with the so-called "confrontation" task at the beginning of a new topic, challenging learners with a new problem. Next, learners build up the required concepts and skills with the help of development tasks to be able to solve the initial problem, followed by exercises to train and expand the competency. To sum up, learners solve the initial problem in a synthesis task and may then need to apply the competency in a transfer task. Each task type (e.g., confrontation task) is described by the weighting of the following nine scales: 1) chart of competencies, 2) relationship to daily life, 3) learners' conceptions, 4) knowledge dimension, 5) knowledge activities, 6) forms of representation, 7) task openness, 8) learning supports, and 9) learning paths. Each scale contains between one and four subscales that describe the task types by their weighting. The description in the form of scale values is taken up empirically.

This study combines existing task scales from research with the different functions of tasks along the learning process model. An expert panel worked in a ge-

neral-theoretical manner, and trained lecturers rated tasks with the instrument to analyse tasks.

In our study, we used the existing scales from the *Instrument to Analyse Tasks* (I-AT). To develop the experts' proposed scale values, we consulted four experts. We calculated the AD_M index (average absolute deviation) as a quality-control measure for the experts' level of agreement. According to the tasks' scale values as rated by trained lecturers ($N = 2$), we selected 25 of the 146 science education tasks from the project *MINT unterwegs* ("STEM on the move"). In the comparison, we calculated the score differences between the experts' scale values and the rated task.

The results show that it is possible to describe different task types of a task set with the IAT when the scale values are obtained from expert proposals. Moreover, the IAT scale values obtained from the expert proposals are quite similar to those from task ratings by trained lecturers.

This study indicates that experts can distinguish and characterize the Learning Process Model's various task types by weighting IAT scale proposals. Furthermore, it has been shown that tasks can be analysed with the IAT. When the results of the task analysis by trained lecturers are compared to the experts' proposals, the tasks can be revised reasonably and optimized for the learning process.

Schlüsselwörter

competency, science, task sets, learning, model, instrument, analysis, task quality

Verortung zur übergeordneten Fragestellung

In dieser Studie wird das konstruierte und validierte Aufgaben-Analyse-Instrument (AAI) eingesetzt, um MINT-Aufgaben zu charakterisieren. Die von den unabhängigen Rater:innen beurteilten Aufgaben werden nach den Typen des Lernprozessmodells indiziert und gepoolt. Ein Expertengremium arbeitet generell-theoretisch und schlägt für jeden Aufgabentyp aus dem Lernprozessmodell spezifische AAI-Skalenmittelwerte vor. Die von den Rater:innen beurteilten MINT-Aufgaben werden anschließend mit den theoretischen Expert:innen-Skalenwerten

verglichen und es werden Schlüsse bezüglich einer Aufgabenüberarbeitung gezogen. Diese Anwendungsstudie dient dazu, ein mögliches Einsatzgebiet des AAI aufzuzeigen. Die Ergebnisse machen deutlich, dass die MINT-Aufgaben bis auf wenige Subskalen (u.a. *Wissensart – Metakognition* oder *Wissensaktivität – kreieren*) den Expert:innen-Vorschlägen folgen. Weiter wird deutlich, dass mit dem AAI Aufgaben beurteilt und die ausgegebenen Skalenmittelwerte einer theoretischen Setzung gegenübergestellt werden können. Mit dieser Analyse können die untersuchten Aufgaben theoriegeleitet qualitativ überarbeitet werden.

Eigenleistungen des Erstautors

In Tabelle 2.2 sind die Eigenleistungen des Erstautors in Prozent sowie die Orientierung im Text ausgewiesen.

Tabelle 2.2. *Eigenanteile und Orientierung im Text des Artikels: Quality of task sets – An instrument for analysing science tasks with different functions along the learning process*

Eigenanteile	Orientierung im Text
Konzeption/Initiative = 85 % Theoriearbeit = 80 % Datenerhebung = 100 %[a] Analyse = 100 % Schreibarbeit = 85 %	Der Beitrag wurde vom Erstautor verfasst. Der Zweitautor war primär für den Abschnitt „Learning Process Model“ verantwortlich. Die Drittautorin hat in enger Zusammenarbeit mit dem Erstautor den Abschnitt „Interpretation of the findings“ verfasst. Weiter haben die Mitautor:innen durch Korrekturen und Anmerkungen zur Qualitätssteigerung beigetragen.

Anmerkung. [a] teilweise wurden die Daten aus der Studie 2.1 als Datenpool verwendet

2.3 Empirische Unterscheidung von Aufgabentypen – eine explorative Studie

Sebastian Stuppan[a,b], Markus Wilhelm[a,b] & Katrin Bölsterli Bardy[a]

[a] Pädagogische Hochschule Luzern, Schweiz

[b] Pädagogische Hochschule Heidelberg, Deutschland

Eingereicht bei Swiss Journal of Educational Research

Status: under review (double-blind peer review)

Zusammenfassung

Mit dem wechselnden Fokus von der Input- zur Outcomeorientierung und von der Schul- zur Unterrichtsqualität rückt seit ungefähr einem Jahrzehnt die Aufgabenqualität in den Forschungsfokus, angefacht durch den verstärkenden Diskurs um mehr Binnendifferenzierung im Unterricht. Aufgaben können heute als Dreh- und Angelpunkt schulischen Lernens betrachtet werden, die den Lehr-Lernprozess maßgeblich steuern. Damit der in den Lehrplänen anvisierte kumulative Kompetenzaufbau der Lernenden unterstützt werden kann, gewinnen aufeinander aufbauende Aufgaben an Bedeutung. Der in dieser Studie vorgestellte Lernprozessmodellansatz unterscheidet dazu fünf unterschiedliche Aufgabentypen zur Konfrontation, Erarbeitung, Übung, Synthese und zum Transfer beim Lernen. Inwiefern sich die theoretische Annahme dieser Aufgabentypen in Aufgabensets tatsächlich wiederfindet, wurde bis dato noch nicht nachgewiesen. Die vorliegende Studie teilt mittels Clusteranalysen zur Eruierung von Aufgabentypen einen Pool von 58 MINT-Aufgaben in Gruppen ein und verwendet als Clustervariablen folgende neun Skalen zur Aufgabenqualität eines empirisch entwickelten Aufgabenanalyseinstruments: Kompetenzabbild, Lebensweltbezug, Lernendenvorstellung, Wissensart, Wissensaktivität, Repräsentationsform, Offenheit, Lernunterstützung und Lernweg. Die Befunde der Studie zeigen, dass eine Typisierung von Aufgaben grundsätzlich möglich ist.

Schlüsselwörter

Clusteranalyse, Lernaufgaben, Lernprozess, MINT-Unterricht, Aufgabenqualität, Aufgabenanalyse

Verortung zur übergeordneten Fragestellung

In dieser zweiten Anwendungsstudie wird das AAI eingesetzt, um Aufgaben entsprechend ihren Merkmalsausprägungen zu gruppieren. Werden die aufgedeckten Clusterlösungen verglichen, zeigt sich, dass sich gewisse Merkmalsausprägungen in den gefundenen Aufgaben-Clustern signifikant unterscheiden. Die Aufgaben-Cluster gleichen größtenteils den im Lernprozessmodell vorgeschlagenen Aufgabentypen. Die Befunde dieser Teilstudie lassen den Schluss zu, dass Aufgaben mit dem AAI beurteilt und anschließend mittels eines Verfahrens zur Strukturentdeckung in Datenbeständen analysiert werden können.

Eigenleistungen des Erstautors

In Tabelle 2.3 sind die Eigenleistungen des Erstautors in Prozent sowie die Orientierung im Text ausgewiesen.

Tabelle 2.3. *Eigenanteile und Orientierung im Text des Artikels: Empirische Unterscheidung von Aufgabentypen – eine explorative Studie*

Eigenanteile	Orientierung im Text
Konzeption/Initiative = 90 % Theoriearbeit = 90 % Datenerhebung = - [a] Analyse = 100 % Schreibarbeit = 90 %	Der Beitrag wurde vom Erstautor verfasst. Die Mitautor:innen haben durch Korrekturen, Literaturhinweise und Anmerkungen zur Qualitätssteigerung beigetragen.

Anmerkung. [a] Die Daten aus der Studie 2.1 dienten als Datenpool

2.4 Überschätzen die Lehrmittelautor:innen den authentischen Lebensweltbezug von MINT-Aufgaben? Eine Studie zur Lernendenperspektive

Sebastian Stuppan[a,b], Katrin Bölsterli Bardy[a], Andrea Maria Schmid[a,b] & Markus Wilhelm[a,b]

[a] Pädagogische Hochschule Luzern, Schweiz

[b] Pädagogische Hochschule Heidelberg, Deutschland

Eingereicht bei Zeitschrift für Didaktik der Naturwissenschaften (ZfDN)

Status: publiziert (double-blind peer review)

DOI: https://doi.org/10.1007/s40573-023-00158-9

Zusammenfassung

In der Scientific Literacy und der Context-Based Science Education wird dem Lebensweltbezug eine hohe Bedeutung zugeschrieben. In vielen Forschungsprojekten zum Lebensweltbezug, ein Aspekt der Aufgabenqualität, werden schriftliche Aufgabenstellungen beurteilt und ausgewertet. Eine Schwierigkeit beim Analysieren von Aufgaben ist die Tatsache, dass die Bewertung von geschulten Beurteilenden anders ausfallen kann als durch die Zielgruppe im Unterricht. Die Dokumentenanalyse bildet das objektive Potenzial der Aufgaben aus Sicht der jeweiligen Beurteiler:innen ab. Die Einbettung der Aufgaben im Unterricht, das Niveau der Lösungen sowie die Aufgabeneinschätzung der Lernenden werden weggelassen. Studien, die sich mit der Analyse der schriftlichen Aufgaben und der Aufgabenbeurteilungen durch Lernende im Unterricht auseinandersetzen, existieren derzeit kaum. Um diese Forschungslücke zu schmälern, wird in diesem Vorhaben verglichen, wie fachdidaktische Lehrmittelautor:innen ihre entwickelten Aufgaben einschätzen und Lernende dieselben Aufgaben im Unterricht nach deren Bearbeitung wahrnehmen. Hierzu beurteilten Lernende ($N = 805$) ausgewählte MINT-Aufgaben ($N = 16$) auf deren Lebensweltbezug (min. 66, max. 182 Lernende pro Aufgabe) mithilfe des Aufgaben-Analyse-Instruments. Die Lernenden-Einschätzungen werden anschließend inferenzstatistisch der fachdidaktischen Per-

spektive der Lehrmittelautor:innen gegenübergestellt und diskutiert. Die Befunde dieser Studie zeigen, dass die Lernenden in den untersuchten MINT-Aufgaben die Lebenswelt divers wahrnehmen und es keine geschlechtsspezifischen Unterschiede gibt. Die Gegenüberstellung der Lernenden-Einschätzungen und der Setzung der Lehrmittelautor:innen macht deutlich, dass die Autor:innen in der Regel die Aufgaben authentischer beurteilen, als die Lernenden diese wahrnehmen. Aufgrund der doch deutlichen Diskrepanz zwischen den zwei Perspektiven wird diskutiert, welchen Mehrwert der Einbezug der Praxis in einem iterativen Verfahren beim Erstellen von Lernaufgaben haben könnte, um den Lebensweltbezug aus der Perspektive der Lernenden stringenter aufnehmen zu können.

Schlüsselwörter

Lebensweltbezug, Lernaufgaben, Einschätzungen, Aufgabenanalysen, Perspektiven, MINT-Lernmaterialien

Verortung zur übergeordneten Fragestellung

Neben der Aufgabenbeurteilung durch die Lehrmittelautor:innen wird in dieser AAI-Anwendungsstudie zusätzlich der Lebensweltbezug durch Lernende beurteilt, um die Fachdidaktik-Perspektive und die Realisierung von Aufgaben im Unterricht miteinander vergleichen zu können und dadurch die ökologische Validität der Setzung der Lehrmittelautor:innen zu überprüfen. Die Ergebnisse dieser Anwendungsstudie zeigen, dass sowohl Lehrmittelautor:innen als auch Lernende den Lebensweltbezug in MINT-Aufgaben mit dem AAI einschätzen können. Die heterogenen Perspektiven bezüglich Lebensweltbezugs bestätigen eine jüngst in der Forschung diskutierte Problemstellung (Akbaba et al., 2018; Fuhrmann, 2020; Schriebl et al., 2022; Weiss & Müller, 2015).

Eigenleistungen des Erstautors

In Tabelle 2.4 sind die Eigenleistungen des Erstautors in Prozent sowie die Orientierung im Text ausgewiesen.

Tabelle 2.4. *Eigenanteile und Orientierung im Text des Artikels: Überschätzen die Lehrmittelautor:innen den authentischen Lebensweltbezug von MINT-Aufgaben? Eine Studie zur Lernendenperspektive*

Eigenanteile	Orientierung im Text
Konzeption/Initiative = 90 % Theoriearbeit = 90 % Datenerhebung = 100 %[a] Analyse = 100 % Schreibarbeit = 90 %	Der Beitrag wurde vom Erstautor verfasst. Die Mitautor:innen haben durch Korrekturen, Literaturhinweise und Anmerkungen zur Qualitätssteigerung beigetragen.

3 Gesamtdiskussion der Publikationen und Schlussfolgerungen

Im vorliegenden Dissertationsprojekt wird auf Grundlage bestehender Kategoriensysteme ein Messinstrument zur Analyse von MINT-Aufgaben entwickelt und validiert (*Teilstudie I*). In der Folge dient das Aufgaben-Analyse-Instrument (AAI) dazu, Aufgaben aus dem *„MINT unterwegs“*-Projekt zu untersuchen. In einer ersten Anwendungsstudie (*Teilstudie II*) werden Aufgaben von unabhängigen Rater:innen beurteilt und nach den von den Autor:innen ausgewiesenen Aufgabentypen aus dem Lernprozessmodell gepoolt. Die indizierten Aufgabentypen und ihre AAI-Qualitätsmerkmale werden anschließend einem generell-theoretischen Expert:innen-Vorschlag gegenübergestellt. Das Vorhaben zeigt einerseits eine erste Anwendung des AAI auf und andererseits dienen die Analysen dazu, Schlussfolgerungen zur Aufgabenqualität des MINT-Projekts zu ziehen. In einer zweiten Anwendungsstudie (*Teilstudie III*) wird untersucht, ob sich die mit dem AAI beurteilten Aufgaben mittels Clusteranalysen in verschiedene Aufgabentypen des Lernprozessmodells gruppieren lassen. Diese Teilstudie weist auf eine weitere Anwendung des AAI hin und lässt die Schlussfolgerung zu, dass Aufgaben entsprechend ihren Qualitätsmerkmalen den verschiedenen Funktionen im Lernprozess zugeordnet werden können. In der letzten Anwendungsstudie dieses Dissertationsprojekts (*Teilstudie IV*) wird der Lebensweltbezug bei Aufgaben aus der Perspektive der Lernenden und der Lehrmittelautor:innen analysiert. Bei dieser Studie wird der Einsatzbereich des AAI erweitert und aufgezeigt, dass das Instrument in der Fachdidaktik nicht nur zur Potenzialanalyse von Aufgaben eingesetzt werden kann, sondern mit dem AAI-Instrument auch das Verhältnis von theoretischem Aufgabenpotenzial und tatsächlicher Wirkung – hier: Wie wirken die Aufgaben hinsichtlich des Lebensweltbezugs? – exploriert werden kann. Dabei zeigt sich, dass sich das Messen aus verschiedenen Perspektiven lohnt. So werden Merkmale wie der Lebensweltbezug in Aufgaben individuell wahrgenommen.

Aufgrund der unterschiedlichen Forschungszugänge, in Form von Entwicklungs-, Validierungs- und Anwendungsstudien, lässt sich zusammenfassend feststellen,

dass mit dem AAI MINT-Aufgaben gemessen und für weitere Analysen verwendet werden können.

3.1 Übergreifende methodische Diskussion der AAI-Skalen

Im Folgenden wird die übergeordnete Fragestellung: „Welches Potenzial lässt sich aus der Entwicklung, Validierung und Anwendung des AAI ableiten?“ in einem größeren Zusammenhang diskutiert. Hierzu werden Aspekte der AAI-Skalen ausgeführt, Erklärungsversuche dargelegt, Limitationen aufgegriffen und zukünftige Forschungsmöglichkeiten aufgezeigt.

Im Nachgang an die Befragungen im Rahmen des „*MINT unterwegs*“-Projekts berichteten Lehrpersonen, dass einige Lernende die sich ähnelnden Formulierungen im Fragebogen, trotz assortierter Itemtexte, erkannt hatten (Mummendey & Grau, 2014, S. 72–74). Wird beim Zielkonstrukt mit inhaltsgleichen Items gearbeitet, büßt das Instrument an Validität ein, da das Konstrukt nicht in seiner ganzen Breite abgedeckt wird (Döring & Bortz, 2016, S. 445; Streiner, 2003). Werden die statistischen Reliabilitätskennwerte der Skala *Lebensweltbezug* des hier vorgestellten Messinstruments betrachtet, scheinen zumindest die Kennwerte bei den Lernenden im mittleren Bereich von Cronbachs Alpha zu liegen ($M_\alpha = .78$, $SD_\alpha = 0.03$). Dies kann gemäß Streiner (2003) als moderate Inter-Item-Korrelation interpretiert werden. Werden hingegen die Reliabilitätskennwerte der fachdidaktischen Beurteilung betrachtet, fällt auf, dass die Cronbachs Alpha im oberen Bereich liegen ($M_\alpha = .93$, $SD_\alpha = 0.04$). Diese Werte können gemäß Streiner (2003) als exzellent interpretiert werden. Dass die Werte eine so hohe Reliabilität aufweisen, liegt wohl in der Natur der Sache, da die fachdidaktischen Rater:innen eine mehrtägige Schulung durchlaufen hatten. Dennoch könnte aus statistischer Sicht auf redundante Items geschlossen und somit die Validität geschmälert werden (Döring & Bortz, 2016, S. 445; Streiner, 2003). Deshalb sollten in zukünftigen Forschungsprojekten erneut die Cronbachs Alpha berechnet und analysiert werden. Damit die Items an inhaltlicher Breite gewinnen, wären sie gegebenenfalls zu überarbeiten.

Bereits Maier et al. (2010, 2014) haben auf Basis älterer Lernzieltaxonomien, fachdidaktischer Aufgabenklassifikationen und Befunde der Kognitionspsychologie ein einfaches Analyseraster mit 10 Dimensionen entwickelt. Aus diesem und

weiteren Systemen haben Wilhelm et al. (2014) und nachfolgend Luthiger et al. (2018) kompetenzfördernde Merkmale ausgewählt. Da das AAI auf Grundlage des LUKAS-Kategoriensystems (Luthiger et al., 2018) entwickelt wurde, kann die Qualität von MINT-Aufgaben mit neun Skalen erfasst werden. Theoretisch wären aus weiteren Kategoriensystemen andere Dimensionen denkbar (Blömeke et al., 2006; Heinle et al., 2022; Jordan et al., 2006; Schriebl et al., 2022). Die Anzahl Skalen könnten entsprechend mit dem modularen Aufbau des AAI erweitert werden. Werden dem AAI weitere Skalen hinzugefügt, würde dies einerseits der historischen Konzeptionsidee widerstreben und andererseits eine Handhabung in der Praxis, beispielsweise für Lehrpersonen oder Studierende in der fachdidaktischen Ausbildung, erschweren. Es gilt zu bedenken, dass bereits 26 Subskalen und 84 Items in der vorliegenden Form vorkommen (Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Daher sollte eher eine weitere Reduktion des AAI in Betracht gezogen werden, um das Instrument handhabbarer zu gestalten. Als mögliche Reduktionsindikatoren wären die nicht signifikanten Skalen aus der Clusteranalyse – *Teilstudie III* (Kapitel: 2.3 Empirische Unterscheidung von Aufgabentypen – eine explorative Studie) in Betracht zu ziehen. Dies hätte jedoch zur Folge, dass die Reduktion der Skalen auf Grundlage der untersuchten MINT-Aufgaben geschehen würde. Was wiederum die Frage aufwirft, inwiefern diese MINT-Aufgaben den Lernprozessmodell-Qualitätsstandards entsprechen. Allerdings gilt zu bedenken, dass bei der Entwicklung dieser Aufgaben im Jahre 2014 noch kein Messinstrument vorhanden war. Somit müssten die MINT-Aufgaben zuerst gemäß den theoretischen Expert:innen-Vorschlägen aus der *Teilstudie II* – Aufgabenqualität von MINT-Aufgaben (Kapitel: 2.2 Quality of task sets – An instrument for analysing science tasks with different functions along the learning process) iterativ überarbeitet werden. Eine erneute Beurteilung mit dem AAI und entsprechenden Aufgaben-Clusteranalysen sowie Signifikanztests der Cluster-Skalenmittelwerte könnte zielführend sein. Auf dieser Grundlage wäre für die Praxis ein Aufgabenprüfsystem mit prägnanten Fragen, wie dies bereits Wilhelm und Kalcsics (2017, S. 88–89) mit ihrem Prüfsystem kompetenzfördernder Aufgabensets oder Wildhirt und Luthiger (2020) mit ihrer vereinfachten Darstellung der LUKAS-Aufgabenprofile ausgearbeitet haben, abzuleiten.

Es ist zu erwähnen, dass in den AAI-Anwendungsstudien (*Teilstudie II* und *III*) die Daten aus der Validierungsstudie (*Teilstudie I*) als erweiterte Analyse verwendet wurden. Um das AAI breiter abzustützen, müsste, wie Messick (1995) vorschlägt, das Messinstrument an einer anderen und wenn möglich umfangreicheren Stichprobe eingesetzt werden. Dieser Schritt muss in zukünftigen Forschungsprojekten mitberücksichtigt werden (Stuppan, Wilhelm, & Bölsterli Bardy, 2022).

3.2 Zentrale Ergebnisse und Diskussion der einzelnen AAI-Skalen

Im Folgenden werden die einzelnen AAI-Skalen mit den Ergebnissen der vier *Teilstudien I* bis *IV* bezüglich der übergeordneten Fragestellung diskutiert. Das heißt, es werden zentrale Ergebnisse zusammengefasst, alternative Erklärungsversuche besprochen, Limitationen erörtert und zukünftige Forschungsmöglichkeiten aufgezeigt. Dabei wird die Diskussion, wenn immer möglich, chronologisch nach den aufgeführten *Teilstudien I* bis *IV* (Abbildung 1.1 und Kapitel: 2 Übersicht der Publikationen) geführt.

Kompetenzabbild

Die Ergebnisse der *Teilstudie I* zeigen, dass die geschulten Fachdidaktiker:innen gemäß Koo und Li (2016) moderate bis gute Interrater-Reliabilitäten in den Subskalen *singulär* (ICC(A,1) = .72)*, additiv* (ICC(A,1) = .85) sowie *integrativ* (ICC(A,1) = .73) der Skala *Kompetenzabbild* aufweisen. Es wird ersichtlich, dass die Subskala *additiv* die höchste Zustimmung in den untersuchten MINT-Aufgaben erhält (Stuppan, Wilhelm, & Bölsterli Bardy, 2022; Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Die beiden anderen Subskalen weisen eine rechtsschiefe Verteilung der Daten auf. Inwiefern das AAI Aufgaben mit einer hohen Ausprägung an singulärem oder integrativem Kompetenzabbild messen kann, ist Teil von zukünftigen Untersuchungen. Die negative Korrelation (r = -.54) der promaxrotierten Faktorenanalyse zwischen der Subskala *Kompetenzabbild – singulär* und *Kompetenzabbild – additiv* weist auf eine Beziehung der beiden Konstrukte hin. Analog zu den vertieften PISA-Analysen von Walter und Taskinen (2009) gilt die eigenständige Interpretierbarkeit eines Faktors bei Interfaktorkorre-

lationen bis $r = .85$ als gegeben (Thieroff et al., 2021). Aus testökonomischen und theoretischen Überlegungen wäre eine Umformulierung der Itemtexte denkbar (Lersch & Schreder, 2013; Luthiger et al., 2018, S. 60–61; Maier et al., 2014, S. 34–35; Nehring et al., 2016). Dies würde dazu führen, dass sich eine Skala mit einem Kontinuum von *singulär* bis *integrativ* bilden ließe. Eine Umsetzungsmöglichkeit besteht darin, in den Itemtexten das Wort „nur" zu streichen. Als Beispiele werden hier zwei Itemtexte genannt: *Kompetenzabbild – singulär*: „Bei dieser Aufgabe wird (nur) ein Aspekt einer (Sub-)Kompetenz anvisiert", oder *Kompetenzabbild – additiv*: „Diese Aufgabe erfasst (nur) mehrere Fähigkeiten oder Fertigkeiten einer oder mehrerer (Sub-)Kompetenzen."

In den untersuchten MINT-Aufgabentypen (*Teilstudie II*) zeigt sich, dass gemäß dem Expert:innen-Vorschlag in den Phasen der Konfrontation, Synthese und des Transfers das *Kompetenzabbild – integrativ* zu wenig Beachtung findet (Stuppan, Wilhelm, & Bölsterli Bardy, 2022). Diese Umstände müssten bei zukünftigen Aufgabenentwicklungsprojekten stärker von den Lehrmittelautor:innen mitberücksichtigt werden. Dadurch würden sich die Aufgabensets und die entsprechenden Aufgabentypen stärker an die theoretischen Expert:innen-Vorschläge annähern (Stuppan, Wilhelm, & Bölsterli Bardy, 2022). Ferner ist zu beachten, dass der Expert:innen-Vorschlag: *Kompetenzabbild – additiv* bei der Syntheseaufgabe keine essenzielle Übereinstimmung aufweist (Burke et al., 1999).

Obwohl sich die beiden Aufgabentypen Konfrontation und Übung in den Merkmalsausprägungen des Kompetenzabbildes theoretisch unterscheiden, werden in den Clusteranalysen und den anschließenden Signifikanztests der *Teilstudie III*, keine Skalenmittelwertunterschiede aufgedeckt. Dies lässt die Schlussfolgerung zu, dass die Lehrmittelautor:innen möglicherweise bei der Konstruktion der Lernaufgaben zu wenig auf die Kompetenzabbild-Merkmalsausprägungen geachtet haben (Stuppan et al., under review; Stuppan, Wilhelm, & Bölsterli Bardy, 2022). So wird gemäß dem Expert:innen-Vorschlag bei einer Konfrontation das *Kompetenzabbild – integrativ* und bei einer Übung das *Kompetenzabbild – singulär* als Qualitätsmerkmal gewählt (Luthiger et al., 2018; Stuppan et al., under review; Stuppan, Wilhelm, & Bölsterli Bardy, 2022).

Zusammenfassend lässt sich bezüglich der übergeordneten Fragestellung ableiten, dass mit dem AAI das *Kompetenzabbild* als Nominalskala erfasst werden kann,

jedoch eine Skala *Kompetenzabbild* als Kontinuum von *singulär* bis *integrativ* zu konstruieren und anschließend zu validieren wäre.

Lebensweltbezug

Die Untersuchungen (*Teilstudie I*) zu den Subskalen *konstruiert* (ICC(A,1) = .59), *authentisch* (ICC(A,1) = .55) sowie *real* (ICC(A,1) = .55) der Skala *Lebensweltbezug* machen deutlich, dass die geschulten Fachdidaktiker:innen gemäß Koo und Li (2016) moderate Interrater-Reliabilitäten aufweisen (Stuppan, Wilhelm, & Bölsterli Bardy, 2022; Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Dennoch zeigt sich im Vergleich zu den restlichen Subskalen des AAI, dass die beiden Subskalen *Lebensweltbezug – authentisch* und *Lebensweltbezug – real* tiefe Reliabilitätswerte haben. Werden die mittleren Fachdidaktiker:innen-Beurteilungen gemäß der promaxrotierten Faktorenanalyse betrachtet, zeigt sich, dass die beiden Subskalen *Lebensweltbezug – konstruiert* und *Lebensweltbezug – authentisch* über alle untersuchten MINT-Aufgaben hinweg eine mittlere Zustimmung verbuchen. Gemäß der Faktorenanalyse liegt eine negative Korrelation zwischen der Subskala *Lebensweltbezug – konstruiert* als auch *Lebensweltbezug – authentisch* sowie zwischen *Lebensweltbezug – konstruiert* und *Lebensweltbezug – real* vor (mit je r = -.71). Dies deutet auf eine Beziehung der drei untersuchten Konstrukte hin (Bohl et al., 2015; Brandl, 2016). Dass die Subskala *Lebensweltbezug – authentisch* und *Lebensweltbezug – real* positiv korrelieren (r = .67) ist ein weiteres Indiz, dass die untersuchten Konstrukte miteinander in Verbindung stehen. Aus empirischer und theoretischer Perspektive ist es folgerichtig, dass die Subskalen nicht als Kontinuum abgebildet werden (Bohl et al., 2015; Brandl, 2016; Stuppan et al., 2023; Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). So können Aufgaben zugleich einen konstruierten und einen authentischen Lebensweltbezug beinhalten. Weiter zeigen die Ergebnisse von *Teilstudie IV*, dass dies sowohl bei den fachdidaktischen Lehrmittelautor:innen als auch in der Einschätzung der Lernenden vorkommen kann. Allerdings haben andere Autor:innengruppen wie Maier et al. (2010, S. 67) oder Stäudel et al. (2012) in ihren Kategoriensystemen theoretisch angenommen, dass der Lebensweltbezug aus einer hierarchischen Stufung besteht (Dimension: Lebensweltbezug beinhaltet die Ausprägungen: keine, konstruiert, authentisch und real). Die Annahme kann mit den vorliegenden Daten in

diesem Dissertationsprojekt nicht bestätigt werden (Stuppan et al., 2023; Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022).

Auffallend ist die tiefe fachdidaktische Potenzialeinschätzung der analysierten Aufgaben bei der Subskala *Lebensweltbezug – real* (Skalenausprägung: $M = 1.28$, $SD = 1.12$). Diese Befunde decken sich wiederum mit weiteren Autor:innengruppen wie Maier et al. (2010) oder Heinle et al. (2022). Wie präzise das AAI reale Aufgaben messen kann, ist aus empirischer Sicht nicht eindeutig zu beantworten, da die Daten eine rechtsschiefe Verteilung haben (Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Um dies zu prüfen, müssten in zukünftigen Projekten Aufgaben mit hohem realem Lebensweltbezug beurteilt werden.

Weiter zeigt sich in den Untersuchungen (*Teilstudie II*), dass die Erarbeitungsaufgaben gemäß dem theoretischen Expert:innen-Vorschlag einen zu geringen konstruierten Lebensweltbezug aufweisen (Stuppan, Wilhelm, & Bölsterli Bardy, 2022). Die MINT-Autor:innengruppe hätte demzufolge bei der Entwicklung der Erarbeitungsaufgaben ein Augenmerk auf „konstruierte Aufgaben“ zu setzen. Ähnliche Ergebnisse zeigen sich in den Clusteranalysen (*Teilstudie III*): Die untersuchten Aufgabentypen (Konfrontation und Übung) unterscheiden sich in keiner Subskala von Lebensweltbezug signifikant (Stuppan et al., under review).

Die dritte Anwendungsstudie (*Teilstudie IV*) macht deutlich, wie heterogen der Lebensweltbezug von den Lernenden und den Lehrmittelautor:innen wahrgenommen wird. In der Anwendungsstudie wird hervorgehoben, dass es somit für die Lehrmittelautor:innen bei der Konzeption von Lernaufgaben wie auch bei der Beurteilung von Lernaufgaben anspruchsvoll ist, den tatsächlich von den Lernenden wahrgenommenen Lebensweltbezug von Aufgaben zu antizipieren (Stuppan et al., 2023).

Sind die Gelingensbedingungen für eine reale Aufgabe dennoch gegeben, führt dies unter anderem zur Diskussion, inwiefern dieser reale Lebensweltbezug in der Schule offengelegt werden soll. So geben beispielsweise Akbaba et al. (2018) oder Fuhrmann (2020) zu bedenken, dass ein realer Lebensweltbezug und dessen Offenlegung in der Klasse Nebenwirkungen hat. Dies, weil die Privatsphäre der Lernenden vor der Klasse offengelegt werde und so unter anderem familiäre Traditionen oder Gegebenheiten zu einem Ausgrenzen führen könnten. Die ausge-

führten Aspekte bezüglich des Lebensweltbezugs sollten in zukünftigen Aufgabenentwicklungsprojekten mitberücksichtigt werden.

Insgesamt deuten die Befunde der Studien darauf hin, dass der Lebensweltbezug von Aufgaben mit dem AAI gemessen werden kann. Die erhobenen Messwerte zeigen in Anbetracht der verschiedenen Perspektiven auf, dass eine differenzierte Sichtweise bezüglich der Lebensweltwahrnehmung bei Lernaufgaben weiter untersucht werden muss (Schriebl et al., 2022; Stuppan et al., 2023).

Lernendenvorstellung

In den Untersuchungen (*Teilstudie I*) zeigt sich, dass die Interrater-Reliabilität der unabhängigen Rater:innen in den drei Subskalen *erheben* (ICC(A,1) = .57), *abgleichen* (ICC(A,1) = .73) sowie *reflektieren* (ICC(A,1) = .64) der Skala *Lernendenvorstellung* nach Koo und Li (2016) als moderate Übereinstimmungswerte zu interpretieren sind (Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Während die beiden Subskalen *Lernendenvorstellung – erheben* und *Lernendenvorstellung – abgleichen* mittelmäßige fachdidaktische Beurteilungen aufweisen, sind die Daten bei der Subskala *reflektieren* rechtsschief verteilt (M = 1.30, SD = 0.95). Dies deutet darauf hin, dass in den 146 untersuchten MINT-Aufgaben das Reflektieren von Lernendenvorstellungen kaum vorkommt (Stuppan et al., under review; Stuppan, Wilhelm, & Bölsterli Bardy, 2022). Wegen der Rechtsschiefe kann empirisch nicht abschließend beurteilt werden, inwiefern das AAI das Reflektieren von Lernendenvorstellungen messen kann.

Weiter muss dem Umstand Rechnung getragen werden, dass die Subskala *Lernendenvorstellung – abgleichen* im validierten AAI aktuell nur zwei Items aufweist. Hier wäre zu empfehlen, die Anzahl Items auf drei bis fünf zu erhöhen (Döring & Bortz, 2016; Mummendey & Grau, 2014).

Ferner kommt hinzu, dass die Skalenwertvorschläge der Expert:innen bezüglich der Aufgabentypen Konfrontation, Synthese und Transfer keine essenzielle Übereinstimmung in der Subskala *Lernendenvorstellung – reflektieren* aufweisen (Stuppan, Wilhelm, & Bölsterli Bardy, 2022).

Die untersuchten Erarbeitungsaufgaben weisen gegenüber den Expert:innen-Vorschlägen tiefere Skalenwerte in den Subskalen *abgleichen* und *reflektieren* der Skala *Lernendenvorstellung* auf (*Teilstudie II*).

In zukünftigen Entwicklungsprojekten von Lernaufgaben wäre es ratsam, bei der Konstruktion von Erarbeitungsaufgaben ein Augenmerk auf die Subskala *Lernendenvorstellung – abgleichen* und *Lernendenvorstellung – reflektieren* zu legen. Morrison und Lederman (2003) konnten in ihrer Studie zur Bedeutung und Berücksichtigung von Lernendenvorstellung aufzeigen, dass Lehrpersonen keinerlei Diagnosen und Reflexionen von Lernendenvorstellung im Unterricht durchführten. Dies könnte unter anderem damit zu tun haben, dass, wie die hier dargelegten Studien verdeutlichen, bereits bei der Entwicklung von Lernaufgaben keine solche Voraussetzung angelegt wird (Heeg et al., 2021). Dennoch gilt anzumerken, dass in den naturwissenschaftlichen Didaktiken, der Lernendenvorstellung ein hoher Stellenwert beigemessen wird und eine konsequente Berücksichtigung zu einem größeren Lernerfolg führen kann (Baalmann et al., 2004; Chen et al., 2020; Heeg et al., 2021; Taber, 2019).

Zusammenfassend ist festzuhalten, dass mit dem AAI das Potenzial der Lernendenvorstellung in Lernaufgaben erfasst werden kann. Nach der getätigten Clusteranalyse von *Teilstudie III* weisen alle drei Subskalen *erheben*, *abgleichen* sowie *reflektieren* der Skala *Lernendenvorstellung* in den gefundenen Zwei-Clusterlösungen signifikante Skalenwerte auf (Stuppan et al., under review). Die Ergebnisse der Analyse machen deutlich, dass im Konfrontationsaufgaben-Cluster das Aufgreifen von Lernendenvorstellungen einen hohen Stellenwert hat. Es empfiehlt sich, in der Subskala *Lernendenvorstellung – abgleichen* die Anzahl Items zu erhöhen und die Expert:innen erneut zu der Subskala *Lernendenvorstellung – reflektieren* zu befragen.

Wissensart

Die Interrater-Reliabilität der *Teilstudie I* zeigt, dass die drei Subskalen *Fakten* (ICC(A,1) = .58), *Konzepte* (ICC(A,1) = .56) und *Metakognition* (ICC(A,1) = .73) der Skala *Wissensart* nach Koo und Li (2016) moderate und die Subskala *Prozeduren* (ICC(A,1) = .79) gute fachdidaktische Rater:innen-Übereinstimmung auf-

weisen (Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Bis auf die Subskala *Wissensart – Metakognition* sind die Beurteilungen der 146 Aufgaben in den Subskalen *Wissensart – Fakten* und *Wissensart – Konzepte* im mittleren Ratingskalenbereich. *Wissensart – Metakognition* weist mit einem Skalenmittelwert von 0.60 ($SD = 0.84$) den tiefsten Zustimmungswert aller AAI-Subskalen aus, somit kommt die Metakognition in den untersuchten Aufgaben kaum vor (Stuppan, Wilhelm, & Bölsterli Bardy, 2022; Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Aufgrund der geringen Anzahl an Aufgaben, in denen die Metakognition gemessen wurde, ist eine Schlussfolgerung, inwiefern mit dem AAI das Potenzial der Metakognition in Aufgaben gemessen werden kann, nur bedingt möglich.

Die tiefen Skalenmittelwerte der metakognitiven Beurteilungen stehen in Diskrepanz zu den Expert:innen-Vorschlägen (*Teilstudie II* und *III*) und den untersuchten MINT-Aufgabentypen: Konfrontation, Erarbeitung, Synthese und Transfer, weil gemäß den Expert:innen-Vorschlägen bei den genannten Typen die Metakognition in den Aufgaben vorkommen sollte (Sjuts, 2022; Stuppan, Wilhelm, & Bölsterli Bardy, 2022; Wilhelm et al., 2014). Ähnliche Befunde haben Kleinknecht et al. (2011), Maier et al. (2010) und Sjuts (2022) in ihren Studien berichtet: Es wurde kaum eine Aufgabe ermittelt, in der explizit auf eine Aktivierung von metakognitivem Wissen hingearbeitet wird. Zudem ist anzumerken, dass die Metakognition theoretisch in weitere Strategien eingeteilt werden könnte. So hat beispielsweise bereits Flavell (1979) die Metakognition in metakognitives Wissen, metakognitive Überwachung und Selbstregulierung differenziert. In Bezug auf die Expert:innen-Vorschläge (*Teilstudie II*) der Subskala *Wissensart – Konzepte* in den Syntheseaufgaben fällt auf, dass sie keine essenzielle Übereinstimmung aufweisen.

Weiter legen die Ergebnisse der Clusteranalysen (*Teilstudie III*) dar, dass sich die untersuchten Aufgabentypen (Konfrontation und Übung) in keiner Subskala von *Wissensart* signifikant unterscheiden (Stuppan et al., under review).

Im Rahmen einer vom Autor noch nicht veröffentlichten Studie zur Konvergenz wurden vier Aufgaben aus dem Forschungsprojekt von Maier et al. (2011) mit dem AAI beurteilt. Dabei stellte sich heraus, dass unter anderem die Forschungsgruppe rund um Maier und Jordan in ihrem Kategoriensystem Prozeduren eng mit

Algorithmen aus der Mathematik gleichsetzen. Das AAI erfasst Prozeduren vielfältiger, indem abgesehen von dem Problemlösen auch Fertigkeiten, Arbeits- und Handlungsweisen berücksichtigt werden.

Insgesamt deuten die Befunde der Teilstudien darauf hin, dass mit dem AAI die *Wissensart* in Aufgaben von geschulten Fachdidaktiker:innen gemessen werden kann. In zukünftigen Forschungsprojekten sollte beachtet werden, dass das AAI bei der Subskala *Wissensart – Metakognition* in bisherigen Untersuchungen erst tiefe Skalenwerte gemessen hat. Somit weisen die analysierten Aufgaben deutlich tiefere Beurteilungen aus, als von den Expert:innen vorgeschlagen (Stuppan et al., under review; Stuppan, Wilhelm, & Bölsterli Bardy, 2022). Diese Befunde decken sich mit weiteren Studien zur Aufgabenanalyse. So haben Autor:innen wie Den Elzen-Rump et al. (2008), Kleinert et al. (2021), Maier et al. (2010) oder Sjuts (2022) in ihren untersuchten Aufgaben tiefe Zustimmungswerte bezüglich der Metakognition berichtet.

Wissensaktivität

In den Untersuchungen der *Teilstudie I* zeigen die erhobenen Subskalen *reproduzieren* (ICC(A,1) = .66), *transferieren* (ICC(A,1) = .57) sowie *kreieren* (ICC(A,1) = .65) der Skala *Wissensaktivität* nach Koo und Li (2016) jeweils moderate Beurteilerübereinstimmung der Fachdidaktiker:innen auf (Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Die Skalenmittelwerte der promaxrotierten Faktorenanalyse liegen knapp unter dem mittleren Ratingskalenbereich und haben somit eine leicht rechtsschiefe Verteilung.

Werden die Expert:innen-Vorschläge (*Teilstudie II*) betrachtet, zeigt sich, dass bei den Konfrontationsaufgaben eine geringe Übereinstimmung bei der Subskala *Wissensaktivität – transferieren* vorliegt. Dies könnte damit zu tun haben, dass ein Transfer, wie es Barnett und Ceci (2002) sowie Gysin et al. (2022) vorschlagen, weiter in Inhalt: Was wird transferiert? (Content: What is transferred?) sowie Kontext: Wann und wo wird Lernen transferiert, von und bis wohin (Context: When and where learning is transferred from and to?) differenziert werden könnte und sich die Expert:innen daher uneins waren.

Bei der getätigten Clusteranalyse (*Teilstudie III*) wird deutlich, dass die Subskalen *Wissensaktivität – reproduzieren* und *Wissensaktivität – kreieren* sich in der gefundenen Zwei-Clusterlösung signifikant unterscheiden. Die Subskala *Wissensaktivität – transferieren* unterscheidet sich in den untersuchten Aufgaben gemäß den Clusteranalysen hingegen nicht signifikant (Stuppan et al., under review).

Zusammenfassend kann auf der Grundlage der empirischen Befunde geschlussfolgert werden, dass mit dem AAI die *Wissensaktivität* in den drei Subskalen *reproduzieren*, *transferieren* sowie *kreieren* gemessen werden kann (Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Weitere Subdimensionen sind nicht auszuschließen, diese müssten aufgrund einer intendierten Sparsamkeit und Handlichkeit des AAI geprüft werden (Krüger & Upmeier zu Belzen, 2021; Maier et al., 2010, S. 27; Töpfer et al., 2022).

Repräsentationsform

Die Ergebnisse der Untersuchungen (*Teilstudie I*) deuten darauf hin, dass das Kontinuum *Transformation* bis *keine Transformation* (ICC(A,1) = .77) der Skala *Repräsentationsform* bei den geschulten Fachdidaktiker:innen gemäß Koo und Li (2016) gute Interrater-Reliabilitäten aufweist (Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Ähnliche Befunde haben Maier et al. (2010, S. 69) mit ihrem hochinferenten Ratingverfahren berichtet. Da der Ratingskalenbereich der Repräsentationsform mit der promaxrotierten Faktorenanalyse im mittleren Bereich liegt (M = 2.39, SD = 1.13), kann aufgrund der empirischen Validierungsstudien davon ausgegangen werden, dass das AAI die Skala *Repräsentationsform – Transformation* bis *keine Transformation* als Kontinuum in den untersuchten MINT-Aufgaben messen kann.

Weiter weisen die Expert:innen-Vorschläge bei allen Aufgabentypen essenzielle Übereinstimmungen bei den Skalenausprägungen auf (*Teilstudie II*). Ferner zeigen sich bei den Vergleichen zwischen den beurteilten MINT-Aufgaben und den Expert:innen-Vorschlägen keine nennenswerten Abweichungen.

Zudem zeigt sich in den getätigten Clusteranalysen (*Teilstudie III*), dass der Konfrontationsaufgaben-Cluster gegenüber dem Übungsaufgaben-Cluster keine signifikanten Skalenmittelwerte hat.

Zusammenfassend kann geschlussfolgert werden, dass mit der Skala *Repräsentationsform* des AAI in den getätigten Validierungs- und Anwendungsstudien die Repräsentationsform der untersuchten MINT-Aufgaben auf einem Kontinuum von *Transformation* bis *keine Transformation* erfasst werden kann.

Offenheit

Teilstudie I zeigt als Ergebnis gemäß Koo und Li (2016) moderate bis gute Inter-rater-Reliabilitäten in den Subskalen *Lösungsvorgehen frei* bis *erklärt* (ICC(A,1) = .77), *Resultat vielfältig* bis *einheitlich* (ICC(A,1) = .81) sowie *Bearbeitung vorstrukturiert* bis *unstrukturiert* (ICC(A,1) = .67) der Skala *Offenheit* (Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Alle drei Subskalen: *Offenheit – Lösungsvorgehen frei*, *Offenheit – Resultat vielfältig* und *Offenheit – Bearbeitung vorstrukturiert*, weisen nach der promaxrotierten Faktorenanalyse eine mittelmäßige fachdidaktische Beurteilung auf (Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Dies deutet darauf hin, dass in den untersuchten MINT-Aufgaben die *Offenheit* als Konstrukt gemessen werden kann. Die durchgeführte, explorative Faktorenanalyse zeigt eine positive Korrelation (r = .54) zwischen der Subskala *Offenheit – Lösungsvorgehen frei* und der Subskala *Offenheit – Resultat vielfältig* auf. Des Weiteren ist eine negative Korrelation (r = -.50) zwischen der Subskala *Offenheit – Lösungsvorgehen frei* und der Subskala *Offenheit – Bearbeitung vorstrukturiert* zu berichten. Dies deutet auf eine Beziehung der drei untersuchten Konstrukte hin. Aus empirischer und theoretischer Perspektive ist es sinnvoll, dass die Subskalen nicht als Kontinuum abgebildet werden (Arnold et al., 2017; Kleinert et al., 2021; Schukajlow & Krawitz, 2020; Stuppan et al., under review; Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Eine mögliche Erklärung der negativen Korrelation zwischen *Offenheit – Lösungsvorgehen frei* und *Offenheit – Bearbeitung vorstrukturiert* ist, dass die beiden theoretischen Enden der jeweiligen Kontinuen gegensätzlich sind. So wäre gegebenenfalls in Betracht zu ziehen, die Subskala *Offenheit – Bearbeitung* als Kontinuum von *unstrukturiert* bis *vorstrukturiert* auszurichten. Dieses Vorgehen würde die beiden Enden der Subskala *Offenheit – Bearbeitung* tauschen und die jetzige ausgewiesene negative Korrelation wäre positiv. Da in den meisten Aufgaben des Lernprozessmodells eine Strukturierung der Aufgabe vorgeschlagen wird, wird aus theoretischer Sicht von

einem Tausch der Skalen-Enden abgeraten (Stuppan, Wilhelm, & Bölsterli Bardy, 2022).

Werden die Skalenausprägungen der Expert:innen-Vorschläge (*Teilstudie II*) bezüglich der Skala *Offenheit* betrachtet, sind bei allen Aufgabentypen essenzielle Übereinstimmungen festzustellen (Burke et al., 1999; Stuppan, Wilhelm, & Bölsterli Bardy, 2022). Weiter zeigt sich bei den Vergleichen zwischen den beurteilten MINT-Aufgaben und den Expert:innen-Vorschlägen keine nennenswerten Abweichungen (Stuppan, Wilhelm, & Bölsterli Bardy, 2022).

Die beiden Subskalen *Offenheit – Lösungsvorgehen frei* und *Offenheit – Resultat vielfältig* sind bei der getätigten Clusteranalyse (*Teilstudie III*) bedeutsam. Mit deren Ratingskalenwerten sind unter anderem die Aufgabentypen, Konfrontationsaufgaben und Übungsaufgaben, empirisch identifiziert worden (Stuppan et al., under review).

Abschließend kann festgehalten werden, dass mit dem AAI in den iterativen Validierungs- und Anwendungsstudien die Skala *Offenheit* mit ihren Subskalen *Lösungsvorgehen frei* bis *erklärt*, *Resultat vielfältig* bis *einheitlich* sowie *Bearbeitung vorstrukturiert* bis *unstrukturiert*, MINT-Aufgaben messen und entsprechend abbilden kann.

Lernunterstützung

Die erhobenen AAI-Subskalen (*Teilstudie I*) *Hilfestellung* (ICC(A,1) = .65), *Austauschmöglichkeit* (ICC(A,1) = .75) sowie *Lehrpersonenfeedback* (ICC(A,1) = .78) der Skala *Lernunterstützung* weisen nach Koo und Li (2016) in den Teilstudien moderate bis gute Urteilerübereinstimmungen auf (Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Die Skalenmittelwerte der promaxrotierten Faktorenanalyse liegen bei der Subskala *Lernunterstützung – Hilfestellung* und *Lernunterstützung – Lehrpersonenfeedback* unter dem mittleren Ratingskalenbereich. Dies deutet auf eine leicht rechtsschiefe Verteilung hin. Hingegen weist die Subskala *Lernunterstützung – Austauschmöglichkeit* eine leichte linksschiefe Verteilung auf (M = 2.87, SD = 1.11).

Werden die Vorschläge zu den Ratingskalenwerten der Expert:innen bezüglich der einzelnen Aufgabentypen aus dem Lernprozessmodell begutachtet (*Teilstudie*

II), fällt auf, dass bei der Syntheseaufgabe die Subskala *Lernunterstützung – Hilfestellung* und bei der Übungsaufgabe die Subskala *Lernunterstützung – Lehrpersonenfeedback* eine nicht essenzielle Übereinstimmung aufweist (Burke et al., 1999; Stuppan, Wilhelm, & Bölsterli Bardy, 2022). Weiter haben die untersuchten Erarbeitungsaufgaben und Übungsaufgaben gegenüber den Expert:innen-Vorschlägen eine tiefere Skalenwertausprägung in der Subskala *Lernunterstützung – Lehrpersonenfeedback* (Stuppan, Wilhelm, & Bölsterli Bardy, 2022). Diesen Umständen müsste in zukünftigen Entwicklungsprojekten zu Lernaufgaben Rechnung getragen werden, zum Beispiel indem einerseits die Expert:innen erneut zu den ausgeführten Subskalen befragt und andererseits die Lehrmittel-Autor:innen explizit auf die Relevanz des Lehrpersonenfeedbacks in den Aufgabentypen Erarbeitung und Übung hingewiesen werden, wie das unter anderem Buholzer et al. (2020), Chin (2006), Hattie und Wollenschläger (2014) sowie Jurik et al. (2014) vorschlagen.

In der Clusteranalyse (*Teilstudie III*) zu den Aufgabentypen zeigt sich, dass sich die Konfrontationsaufgaben und Übungsaufgaben nur in der Subskala *Lernunterstützung – Hilfestellung* signifikant unterscheiden (Stuppan et al., under review). Gemäß den Vorschlägen der Expert:innen sollten jedoch Konfrontationsaufgaben und Übungsaufgaben in allen drei Subskalen der AAI-Skala *Lernunterstützung* voneinander abweichen.

Weiter ist zu erwähnen, dass Hattie und Yates (2015) Feedbacks in weitere Subdimensionen wie Feedback zur Aufgabe, Feedback zum Lernprozess, Feedback zur Selbstregulation oder Feedback zur Person differenzierten. Ergänzend könnten digital aufbereitete Lernprogramme den Lernenden ebenso ein Feedback geben (Erhel & Jamet, 2013; Goldin et al., 2017; Zhu et al., 2020). Hier stellt sich die Frage, inwiefern digitale Lernprogramme als Lehrpersonenfeedback betrachtet werden können (Narciss, 2018). Dies müsste gegebenenfalls in zukünftigen Studien untersucht werden. In Anbetracht eines sparsamen und einfach handhabbaren Messinstruments ist die erwähnte Ausdifferenzierung des Feedbacks hinsichtlich Erweiterungen des AAI zu prüfen und abzuwägen (Luthiger et al., 2018; Maier et al., 2010; Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022).

Auf der Grundlage der empirischen Befunde kann geschlussfolgert werden, dass mit dem AAI die *Lernunterstützung* in den drei Subskalen *Hilfestellung*, *Aus-*

tauschmöglichkeit sowie *Lehrpersonenfeedback* gemessen werden kann. Da sich die untersuchten MINT-Aufgaben (Konfrontation und Übung) in den Merkmalsausprägungen der *Lernunterstützung* wenig unterscheiden, die Expert:innen diese jedoch vorschlagen, sind Lehrmittelautor:innen diesbezüglich in zukünftigen Entwicklungsarbeiten darauf hinzuweisen. So wäre beispielsweise bei der Entwicklung von Konfrontationsaufgaben auf die *Lernunterstützung – Austauschmöglichkeit* zu achten (Stuppan et al., under review; Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022).

Lernweg

Die Ergebnisse der *Teilstudie I* weisen gemäß Koo und Li (2016) moderate bis gute Interrater-Reliabilitäten in den Subskalen *kompensierend* (ICC(A,1) = .73), *profilbildend* (ICC(A,1) = .79) sowie *selbstdifferenzierend* (ICC(A,1) = .54) der Skala *Lernweg* auf (Stuppan, Wilhelm, Bölsterli Bardy, et al., 2022). Im Vergleich zu den übrigen Subskalen des AAI hat die Subskala *Lernweg – selbstdifferenzierend* den tiefsten Reliabilitätswert. Mit Blick auf die mittleren Fachdidaktiker:innen-Beurteilungen gemäß der promaxrotierten Faktorenanalyse zeigt sich, dass die beiden Subskalen *Lernweg – kompensierend* und *Lernweg – selbstdifferenzierend* eine mittlere Zustimmung erhalten (Affeldt et al., 2018; Baumann et al., 2018; Brauns & Abels, 2021; Stebler et al., 2022). Hingegen zeigt die Subskala *Lernweg – profilbildend* eine leicht rechtsschiefe Verteilung der Daten (M = 1.46, SD = 1.24). Daraus lässt sich ableiten, dass in den untersuchten Aufgaben ein profilbildender Charakter im Sinne einer Wahlmöglichkeit nach individuellen Interessen nicht in einer hohen Ausprägung gemessen werden konnte (Götz & Nett, 2011; Mötteli et al., 2022). Inwiefern Aufgaben mit einer hohen Ausprägung an profilbildendem Charakter mit dem AAI gemessen werden können, ist Teil von zukünftigen Untersuchungen.

Es zeigt sich hingegen bei den Expert:innen-Vorschlägen (*Teilstudie II*), dass alle *Lernweg*-Subskalen nach Burke et al. (1999) eine substanzielle Übereinstimmung haben. Was für eine theoretische Plausibilität spricht. Bei der Vergleichsstudie zwischen den theoretischen Expert:innen-Vorschläge und den erhobenen MINT-Aufgaben zeigt sich, dass der ganzzahlige Index, in Anlehnung an Altman (1999),

gute bis perfekte Übereinstimmungen aufweist. Es wird illustriert, dass mit der gewählten ganzzahligen Abstufung keine Differenzen zwischen den theoretischen Vorschlägen und den untersuchten MINT-Aufgaben bestehen (Stuppan, Wilhelm, & Bölsterli Bardy, 2022).

Ebenso verdeutlicht die Clusteranalyse aus der *Teilstudie III*, dass die beiden Subskalen *Lernweg – kompensierend* und *Lernweg – selbstdifferenzierend* einen wichtigen Faktor für die empirische Trennung der Aufgabentypen darstellen. Beide Faktoren weisen im Anschluss an die Clusteranalyse signifikante Skalenmittelwertunterschiede auf (Stuppan et al., under review).

Insgesamt zeigen die Befunde der Teilstudien, dass die beiden *Lernweg*-Subskalen *kompensierend* und *selbstdifferenzierend* einen substanziellen Beitrag zur Klassifizierung von Aufgaben leisten. In zukünftigen Forschungsprojekten muss beachtet werden, dass die AAI-Subskala *Lernweg – profilbildend* in bisherigen Untersuchungen erst tiefe Skalenwerte gemessen hat.

3.3 Ausblick

In den vorgestellten vier Studien wurde auf die Entwicklung, Validierung sowie erste AAI-Anwendungen eingegangen. Abgeleitet aus weiteren Überlegungen könnten sich zukünftige Forschungen mit folgenden hier einzeln umschriebenen Fragestellungen beschäftigen. Dabei werden die Fragestellungen in zwei Stränge aufgeteilt: 1) Aufgaben in Unterrichtssituationen und 2) Implikationen im Bereich Lehre, Weiter- und Hochschulbildung.

1) Aufgaben in Unterrichtssituationen

Hierzu stellen sich zurzeit folgende Fragen:

- Welche weiteren Skalen aus dem AAI müssten Lehrpersonen sowie Lernende beurteilen, um die Aufgabenbearbeitung in den Unterrichtssituationen abbilden zu können?
- Welche Merkmalsausprägungen des AAI werden von den Lehrkräften während des Unterrichtens verändert und wie kommen diese Veränderungen zu-

stande (Lehrpersonen-Kommentar, Unterrichtssituationen, Professionskompetenz der Lehrperson etc.)?

- Gibt es einen Unterschied in der Lernhaltung bei den Schüler:innen in Abhängigkeit davon, ob sie das Lernprozessmodell kennen oder nicht?
- Welche AAI-Merkmalsausprägungen der Konfrontationsaufgabe führen bei den Lernenden zu einer hohen epistemischen Neugier?

Um diesen und weiteren Fragen bezüglich Aufgaben in Unterrichtssituationen nachzugehen, wären nebst quantitativen Befragungen mittels Fragebogen auch Videoanalysen oder leitfadengestützte Interviews einzusetzen.

2) Implikationen im Bereich Lehre, Weiter- und Hochschulbildung

Zum aktuellen Zeitpunkt stellen sich die folgenden Fragen:

- Welche Ressourcen (Testmanual, Rating-Tabellen, Beispielaufgaben, Zeitaufwände) müssen für qualitativ hohe Ratings eingesetzt werden?
- Welche Professionskompetenzen unterstützen eine qualitätsvolle Aufgabenanalyse?
- Wie kann das AAI in der Aus- und Weiterbildung von Lehrpersonen genutzt werden (beispielsweise zur diagnostischen Expertise von (angehenden) Lehrkräften)?
- Wie kann das AAI von Lehrmittelautor:innen zur Verbesserung der Aufgabenqualität genutzt werden?
- Wie kann das AAI von Fachdidaktiker:innen zur Verbesserung der Dozierendentätigkeit genutzt werden?
- Wie muss das AAI jeweils angepasst werden, um den unterschiedlichen Anwendungsbereichen gerecht zu werden (Erstellung / Optimierung / Analyse von Aufgaben)?
- Wie muss das AAI jeweils angepasst werden, um den unterschiedlichen Anwendern gerecht zu werden (Forschende, Autor:innen, Lehrkräfte, Studierende, Schüler:innen)?

Wie bereits in Kapitel 3.1 ausgeführt, ist eine Vereinfachung des AAI anzustreben, um es in der Praxis handhabbarer zu machen. Offen ist zurzeit, welche Merkmale mit welcher Ausprägung zu dieser Vereinfachung beitragen, um mit wenigen Fragen die Qualität einer Aufgabe erfassen zu können. Dies führt zu folgender Fragestellung:

- Welche Merkmale aus dem AAI beinhaltet ein in der Praxis einsetzbares Prüfsystem, um die Qualität von Aufgaben beurteilen zu können?

3.4 Fazit

Insgesamt zeigen die Befunde der vier Studien, dass das entwickelte und validierte AAI MINT-Aufgaben in neun Qualitätsmerkmalen messen kann. Die Resultate können in weiteren Analysen dazu genutzt werden, Aufgaben verschiedenen Typen nach dem Lernprozessmodell zuzuordnen. Diese Einteilung dient dazu, die Aufgaben qualitativ zu überarbeiten und somit näher an eine theoretische Setzung von Expert:innen heranzuführen. Werden die vorgeschlagenen Skalen-Modifikationen am AAI vorgenommen, ermöglicht dies im Sinne des Design-Based-Research-Ansatzes eine zyklische Optimierung. Es zeigt sich in den Analysen, dass die Perspektiven der Beurteilungen (Fachdidaktiker:innen, Lernende oder Lehrpersonen) und die jeweilige Situation (Potenzialanalyse, Umsetzung im Unterricht) entscheidende Einflussgrößen sind. Das anwendungsbezogene Messinstrument kann weiteren interessierten Personen zur Verfügung gestellt werden. Die Modularität des AAI lässt eine Adaption, wie ein Hinzufügen von weiteren Qualitätsmerkmalen, zu.

Literatur

Adamina, M. (2010). Mit Lernaufgaben grundlegende Kompetenzen fördern. In P. Labudde (Hrsg.), *Fachdidaktik Naturwissenschaft: 1.-9. Schuljahr* (1. Aufl.). Haupt.

Affeldt, F., Siol, A., Markic, S., & Eilks, I. (2018). Neue Ansätze zur Differenzierung im Schülerlabor: Differentiation in experimental work in non-formal learning environments. *CHEMKON*, *25*(7), 255–262. https://doi.org/10.1002/ckon.201700502

Ainsworth, S. E. (2006). DeFT: A Conceptual Framework for Considering Learning with Multiple Representations. *Learning and Instruction*, *16*(3), 183–198.

Akbaba, Y., Bräu, K., & Fuhrmann, L. (2018). Schulische Aufgaben mit Lebensweltbezug. Nebenwirkungen jenseits didaktischer Absichten. In M. Martens, K. Rabenstein, K. Bräu, M. Fetzer, H. Gresch, I. Hardy, & C. Schelle (Hrsg.), *Konstruktionen von Fachlichkeit: Ansätze, Erträge und Diskussionen in der empirischen Unterrichtsforschung* (S. 247–259). Verlag Julius Klinkhardt.

Altman, D. G. (1999). *Practical statistics for medical research*. Chapman & Hall/CRC.

Anderson, L. W., & Krathwohl, D. R. (Hrsg.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete ed). Longman.

Arnold, J., Kremer, K., & Mayer, J. (2017). Scaffolding beim Forschenden Lernen: Eine empirische Untersuchung zur Wirkung von Lernunterstützungen. *Zeitschrift für Didaktik der Naturwissenschaften*, *23*(1), 21–37. https://doi.org/10.1007/s40573-016-0053-0

Artelt, C., & Wirth, J. (2014). Kognition und Metakognition. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie: Mit Online-Materialien* (6., vollständig überarbeitete Auflage, S. 167–192). Beltz.

Astleitner, H. (2006). *Aufgaben-Sets und Lernen: Instruktionspsychologische Grundlagen und Anwendungen*. Peter Lang.

Baalmann, W., Frerichs, V., Weitzel, H., Gropengießer, H., & Kattmann, U. (2004). Schülervorstellungen zu Prozessen der Anpassung—Ergebnisse einer Interviewstudie im Rahmen der Didaktischen Rekonstruktion. *Zeitschrift für Didaktik der Naturwissenschaften*, *10*, 7–28.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Baumann, T., Kieserling, M., Struckholt, S., & Melle, I. (2018). Verbrennungen—Eine Unterrichtseinheit für inklusiven Unterricht. *CHEMKON*, *25*(4), 160–170. https://doi.org/10.1002/ckon.201800016

Beerenwinkel, A., Parchmann, I., & Gräsel, C. (2007). Chemieschulbücher in der Unterrichtsplanung – Welche Bedeutung haben Schülervorstellungen? *Chemkon*, *14*(1), 7–14.

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.

Bohl, T., Drüke-Noe, C., Hoppe, H., Kleinknecht, M., Maier, U., & Metz, K. (2015). Was bringt diese Aufgabe? Lehrerinnen und Lehrer analysieren das kognitive Potenzial von Aufgaben. *Pädagogik*, *67*(5), 28–31.

Bölsterli Bardy, K., & Wilhelm, M. (2018). Von kompetenzorientierten zu kompetenzfördernden Aufgaben im Schulbuch. *Erziehung & Unterricht*, *168*(1/2), 121–129.

Brandl, W. (2016). Lernstand erheben, Lernwirksamkeit feststellen, Lernleistung beurteilen – eine Skizze des Bezugssystems. *Haushalt in Bildung & Forschung*, *5*(4), 3–24. https://doi.org/10.3224/hibifo.v5i4.25691

Braund, M., & Reiss, M. (2006). Towards a More Authentic Science Curriculum: The contribution of out-of-school learning. *International Journal of Science Education*, *28*(12), 1373–1388. https://doi.org/10.1080/09500690500498419

Brauns, S., & Abels, S. (2021). Die Anwendung naturwissenschaftlicher Untersuchungsmethoden inklusiv gestalten – Naturwissenschaftsdidaktische Theorie und Empirie erweitern mit dem Kategoriensystem inklusiver naturwissenschaftlicher Unterricht (KinU). *Zeitschrift für Didaktik der Naturwissenschaften*, *27*(1), 231–249. https://doi.org/10.1007/s40573-021-00135-0

Bremerich-Vos, A. (2008). Benjamin S. Bloom (und andere) revisited. In A. Bremerich-Vos, D. Granzer, & O. Köller (Hrsg.), *Lernstandsbestimmung im Fach Deutsch. Gute Aufgaben für den Unterricht.* (S. 29–49). Beltz.

Brod, G. (2021). How Can We Make Active Learning Work in K–12 Education? Considering Prerequisites for a Successful Construction of Understanding. *Psychological Science in the Public Interest*, *22*(1), 1–7. https://doi.org/10.1177/1529100621997376

Brophy, J., & Alleman, J. (1991). Activities as Instructional Tools: A Framework for Analysis and Evaluation. *Educational Researcher*, *20*(4), 9–23. https://doi.org/10.3102/0013189X020004009

Brühwiler, C., Hollenstein, L., Affolter, B., Biedermann, H., & Oser, F. (2017). Welches Wissen ist unterrichtsrelevant?: Prädiktive Validität dreier Messinstrumente zur Erfassung des pädagogisch-psychologischen Wissens von Lehrpersonen. *Zeitschrift für Bildungsforschung*, *7*(3), 209–228. https://doi.org/10.1007/s35834-017-0196-1

Büchter, A., & Leuders, T. (2016). *Mathematikaufgaben selbst entwickeln: Lernen fördern – Leistung überprüfen* (7., überarbeitete Neuauflage). Cornelsen.

Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2., aktualisierte und erw. Aufl., [Nachdr.]). Pearson Studium.

Buholzer, A., Baer, M., Zulliger, S., Torchetti, L., Ruelmann, M., Häfliger, A., & Lötscher, H. (2020). Formatives Assessment im alltäglichen Mathematikunterricht von Primarlehrpersonen: Häufigkeit, Dauer und Qualität. *Unterrichtswissenschaft*, *48*(4), 629–661. https://doi.org/10.1007/s42010-020-00083-7

Burke, M. J., Finkelstein, L. M., & Dusig, M. S. (1999). On Average Deviation Indices for Estimating Interrater Agreement. *Organizational Research Methods*, *2*(1), 49–68. https://doi.org/10.1177/109442819921004

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.

Chen, C., Sonnert, G., Sadler, P. M., & Sunbury, S. (2020). The Impact of High School Life Science Teachers' Subject Matter Knowledge and Knowledge of Student Misconceptions on Students' Learning. *CBE—Life Sciences Education*, *19*(1), ar9. https://doi.org/10.1187/cbe.19-08-0164

Chin, C. (2006). Classroom Interaction in Science: Teacher questioning and feedback to students' responses. *International Journal of Science Education*, *28*(11), 1315–1346. https://doi.org/10.1080/09500690600621100

Den Elzen-Rump, V., Leutner, D., & Wirth, J. (2008). Lernstrategien im Unterrichtsalltag. In S. Kliemann (Hrsg.), *Schülergerechtes Arbeiten in der Sekundarstufe I: Diagnostizieren und Fördern in der Sekundarstufe I* (S. 101–111). Cornelsen.

Dori, Y., & Sasson, I. (2013). A three-attribute transfer skills framework-part I: Establishing the model and its relation to chemical education. *Chem. Educ. Res. Pract.*, *14*. https://doi.org/10.1039/C3RP20093K

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5., vollständig überarbeitete, aktualisierte und erweiterte Auflage). Springer.

Duit, R. (1995). Vorstellungen und Lernen von Physik und Chemie. Zu den Ursachen vieler Lernschwierigkeiten. *Plus lucis*, *2*, 11–18.

Dunekacke, S., Wittmann, G., & Jenßen, L. (2022). Differenzierte Beschreibung der Handlungsplanung von angehenden Erzieher*innen mit Blick auf die mathematische Lernunterstützung – Eine explorative Studie. *Unterrichtswissenschaft*. https://doi.org/10.1007/s42010-022-00158-7

DVS. (2020). *MINT unterwegs Konzept*. https://volksschulbildung.lu.ch/-/media/Volksschulbildung/Dokumente/unterricht_organisation/faecher_wost_lehrmittel/faecher/Primar/Mint/mint_konzept_unterwegs.pdf?la=de-CH

EDK. (2011). *Grundkompetenzen für die Naturwissenschaften. Nationale Bildungsstands*. Bern: Schweizerische Konferenz der kantonalen Erziehungsdirektoren.

Ellis, H. C. (1965). *The transfer of learning*. Macmillan.

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, *67*, 156–167. https://doi.org/10.1016/j.compedu.2013.02.019

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Flechsig, K.-H. (2008). Komplexe Aufgaben in der beruflichen Aus- und Weiterbildung. 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. 241–258). Waxmann Verlag.

Friedrich, H. F., & Mandl, H. (1992). Lern- und Denkstrategien—Ein Problemaufriss. In H. Mandl & H. F. Friedrich (Hrsg.), *Lern- und Denkstrategien: Analyse und Intervention* (S. 3–54). Hogrefe.

Fuhrmann, L. (2020). Pizza und Liebe im Unterricht. Die Konstruktion von Wissen über Aufgaben mit Lebensweltbezug. *Zeitschrift für interpretative Schul- und Unterrichtsforschung*, *9*(2020), 38–51. https://doi.org/10.25656/01:24557

Gagné, R. M., Briggs, L. J., & Wager, W. W. (1992). *Principles of instructional design* (4th ed). Harcourt Brace Jovanovich College Publishers.

Goldin, I., Narciss, S., Foltz, P., & Bauer, M. (2017). New Directions in Formative Feedback in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education*, *27*(3), 385–392. https://doi.org/10.1007/s40593-016-0135-7

Götz, T., & Nett, U. E. (2011). Selbstreguliertes Lernen. In T. Götz (Hrsg.), *Emotion, Motivation und selbstreguliertes Lernen* (S. 143–184). Schöningh.

Gropengießer, H., & Marohn, A. (2018). Schülervorstellungen und Conceptual Change. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 49–67). Springer.

Gysin, D., Müller, A., & Brovelli, D. (2022). Nutzung von metakognitiven Strategien beim Transfer von physikalischen Konzepten – Entwicklung und Anwendung eines Messinstruments für Transferstrategien. *Progress in Science Education (PriSE)*, 20-42 Pages. https://doi.org/10.25321/PRISE.2023.1325

Hattie, J., & Wollenschläger, M. (2014). A conceptualization of feedback. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen: Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 135–149). Waxmann.

Hattie, J., & Yates, G. C. R. (2015). *Lernen sichtbar machen aus psychologischer Perspektive* (W. Beywl & K. Zierer, Hrsg.; 1. Auflage, überarbeitete deutschsprachige Ausgabe von "Visible learning and the science of how we learn"). Schneider Verlag Hohengehren.

Heeg, J., Bittorf, R. M., & Schanze, S. (2021). Erforschung potenzieller Entwicklungsverläufe diagnostischer Fähigkeiten angehender Chemielehrkräfte hin-

sichtlich Lernendenvorstellungen – Die Bedeutung individueller Vorstellungen über Lernendenvorstellungen. *Zeitschrift für Didaktik der Naturwissenschaften*, *27*(1), 17–44. https://doi.org/10.1007/s40573-021-00124-3

Heinle, A., Schiepe-Tiska, A., Reinhold, F., Heine, J.-H., & Holzberger, D. (2022). Supporting student motivation in class: The motivational potential of tasks. *Zeitschrift für Erziehungswissenschaft*. https://doi.org/10.1007/s11618-022-01090-3

Heins, J. (2017). *Lenkungsgrade im Literaturunterricht: Zum Einfluss stark und gering lenkender Aufgabensets auf das Textverstehen*. Springer-Verlag.

Herrington, A., & Herrington, J. (Hrsg.). (2006). *Authentic learning environments in higher education*. Information Science Pub.

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In *Conceptual and procedural knowledge: The case of mathematics* (S. 1–27). Lawrence Erlbaum Associates, Inc.

Hild, P., Buff, A., Gut, C., & Parchmann, I. (2020). Adaptives kompetenzbezogenes Feedback beim selbstständigen praktisch-naturwissenschaftlichen Arbeiten: Eine empirische Untersuchung zur Wirksamkeit unterschiedlicher Feedbackformen. *Zeitschrift für Didaktik der Naturwissenschaften*, *26*(1), 19–35. https://doi.org/10.1007/s40573-020-00109-8

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99–107. https://doi.org/10.1080/00461520701263368

Johnson, D. W., & Johnson, R. T. (2009). An Educational Psychology Success Story: Social Interdependence Theory and Cooperative Learning. *Educational Researcher*, *38*(5), 365–379. https://doi.org/10.3102/0013189X09339057

Johnson, T. R., & Krems, J. F. (2001). Use of current explanations in multicausal abductive reasoning. *Cognitive Science*, *25*(6), 903–939. https://doi.org/10.1207/s15516709cog2506_2

Jonas, K. (2020). *Zum Verhältnis von Interesse und Lebensweltbezug bei Lernaufgaben im Unterrichtsfach Geschichte*. https://opus4.kobv.de/opus4-unipassau/frontdoor/index/index/docId/774

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.

Jurik, V., Gröschner, A., & Seidel, T. (2014). Predicting students' cognitive learning activity and intrinsic learning motivation: How powerful are teacher statements, student profiles, and gender? *Learning and Individual Differences*, *32*, 132–139. https://doi.org/10.1016/j.lindif.2014.01.005

Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E., & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften*, *16*, 135–153.

Kingston, N., & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice*, *30*(4), 28–37. https://doi.org/10.1111/j.1745-3992.2011.00220.x

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Klein, P., Kuhn, J., & Müller, A. (2018). Förderung von Repräsentationskompetenz und Experimentbezug in den vorlesungsbegleitenden Übungen zur Experimentalphysik: Empirische Untersuchung eines videobasierten Aufgabenformates. *Zeitschrift für Didaktik der Naturwissenschaften*, *24*(1), 17–34. https://doi.org/10.1007/s40573-018-0070-2

Kleinert, S. I., Isaak, R. C., Textor, A., & Wilde, M. (2021). Die Nutzung gestufter Lernhilfen zur Unterstützung des Experimentierprozesses im Biologieunterricht – eine qualitative Studie. *Zeitschrift für Didaktik der Naturwissenschaften*, *27*(1), 59–71. https://doi.org/10.1007/s40573-021-00126-1

Kleinknecht, M., Bohl, T., Maier, U., & Metz, K. (Hrsg.). (2013). *Lern- und Leistungsaufgaben im Unterricht: Fächerübergreifende Kriterien zur Auswahl und Analyse*. Klinkhardt.

Kleinknecht, M., Maier, U., Metz, K., & Bohl, T. (2011). Analyse des kognitiven Aufgabenpotenzials. Entwicklung und Erprobung eines allgemeindidaktischen Auswertungsmanuals. *Unterrichtswissenschaft*, *39*(4), 328–344.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E., & Vollmer, H. (2003). *Zur Entwicklung nationaler Bildungsstandards: Eine Expertise. Stand Juni 2003*.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Krüger, D., & Upmeier zu Belzen, A. (2021). Kompetenzmodell der Modellierkompetenz – Die Rolle abduktiven Schließens beim Modellieren. *Zeitschrift für Didaktik der Naturwissenschaften*, *27*(1), 127–137. https://doi.org/10.1007/s40573-021-00129-y

Lee, D., Huh, Y., Lin, C.-Y., & Reigeluth, C. M. (2018). Technology functions for personalized learning in learner-centered schools. *Educational Technology Research and Development*, *66*(5), 1269–1302. https://doi.org/10.1007/s11423-018-9615-9

Leisen, J. (2006). Aufgabenkultur im mathematisch-naturwissenschaftlichen Unterricht. *Der mathematische und naturwissenschaftliche Unterricht*, *59*(5), 260–266.

Leisen, J., Kruczinna, R., & Bennung, R. (Hrsg.). (1999). *Methoden-Handbuch: Deutschsprachiger Fachunterricht (DFU)*. Pagina.

Lenske, G., Thillmann, H., Wirth, J., Dicke, T., & Leutner, D. (2015). Pädagogisch-psychologisches Professionswissen von Lehrkräften: Evaluation des ProwiN-Tests. *Zeitschrift für Erziehungswissenschaft*, *18*(2), 225–245. https://doi.org/10.1007/s11618-015-0627-5

Lersch, R., & Schreder, G. (2013). *Grundlagen kompetenzorientierten Unterrichtens: Von den Bildungsstandards zum Schulcurriculum*. Budrich.

Lombardi, D., Shipley, T. F., & Astronomy Team, Biology Team, Chemistry Team, Engineering Team, Geography Team, Geoscience Team, and Physics Team. (2021). The Curious Construct of Active Learning. *Psychological Science in the Public Interest*, *22*(1), 8–43. https://doi.org/10.1177/1529100620973974

Luthiger, H., Wilhelm, M., Wespi, C., & Wildhirt, S. (Hrsg.). (2018). *Kompetenzförderung mit Aufgabensets: Theorie – Konzept – Praxis* (1. Auflage). hep.

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). Klinkhardt.

Maier, U., Kleinknecht, M., Metz, K., Schymala, M., & Bohl, T. (2010). *Entwicklung und Erprobung eines Kategoriensystems für die fächerübergreifende Aufgabenanalyse*. Universität (Erlangen, Nürnberg). Lehrstuhl für Schulpädagogik.

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Mietzel, G. (2007). *Pädagogische Psychologie des Lernens und Lehrens*. Hogrefe.

Morrison, J. A., & Lederman, N. G. (2003). Science teachers' diagnosis and understanding of students' preconceptions. *Science Education*, *87*(6), 849–867. https://doi.org/10.1002/sce.10092

Mötteli, C., Grob, U., Pauli, C., Reusser, K., & Stebler, R. (2022). „Choice and voice" in Schulen mit personalisierten Lernkonzepten aus Sicht der Schülerinnen und Schüler. *Unterrichtswissenschaft*, *50*(2), 287–308. https://doi.org/10.1007/s42010-021-00122-x

Müller, A., & Helmke, A. (2008). Qualität von Aufgaben als Merkmale der Unterrichtsqualität – verdeutlicht am Fach Physik. In J. Thonhauser (Hrsg.), *Aufga-*

ben 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* (6., korrigierte Auflage). Hogrefe.

Nachtigall, V., Rummel, N., & Serova, K. (2018). Authentisch ist nicht gleich authentisch – Wie Schülerinnen und Schüler die Authentizität von Lernaktivitäten im Schülerlabor einschätzen. *Unterrichtswissenschaft*, *46*(3), 299–319. https://doi.org/10.1007/s42010-018-0020-1

Narciss, S. (2018). Feedbackstrategien für interaktive Lernaufgaben. In S. Kracht, A. Niedostadek, & P. Sensburg (Hrsg.), *Praxishandbuch Professionelle Mediation* (S. 1–24). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-54373-3_35-1

Nehring, A., Stiller, J., Nowak, K. H., Upmeier zu Belzen, A., & Tiemann, R. (2016). Naturwissenschaftliche Denk- und Arbeitsweisen im Chemieunterricht – eine modellbasierte Videostudie zu Lerngelegenheiten für den Kompetenzbereich der Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften*, *22*(1), 77–96. https://doi.org/10.1007/s40573-016-0043-2

Neubrand, J. (2002). *Eine Klassifikation mathematischer Aufgaben zur Analyse von Unterrichtssituationen. Selbsttätiges Arbeiten in Schülerarbeitsphasen in den Stunden der TIMSS-Video-Studie.* Franzbecker.

Niggli, A. (2013). *Didaktische Inszenierung binnendifferenzierter Lernumgebungen: Theorie – Empirie – Konzepte – Praxis*. Klinkhardt.

Niss, M., Blum, W., & Galbraith, P. (2007). Introduction. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Hrsg.), *Modelling and Applications in Mathematics Education* (Bd. 10, S. 3–32). Springer US. https://doi.org/10.1007/978-0-387-29822-1_1

Plath, J. (2020). Verstehensprozesse bei der Bearbeitung realitätsbezogener Mathematikaufgaben: Klassische Textaufgaben vs. Zeitungstexte. *Journal für Mathematik-Didaktik*, *41*(2), 237–266. https://doi.org/10.1007/s13138-019-00148-w

Reinfried, S. (2016). Kompetenzorientierte Lernaufgaben – mehr als alter Wein in neuen Schläuchen? *Geographie aktuell & Schule*, *38*(223), 4–14.

Reusser, K. (2014). *Aufgaben – Träger von Lerngelegenheiten und Lernprozesse im kompetenzorientierten Unterricht*. Seminar, 4, 77-101.

Rieck, K. (2005). *Gute Aufgaben* (D. Hoffmann, G. Friege, M. Prenzel, R. Demuth, & K. Rieck, Hrsg.). IPN Leibniz-Institut f. d. Pädagogik d. Naturwissenschaften an d. Universität Kiel.

Riemeier, T. (2007). Moderater Konstruktivismus. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung: Ein Handbuch für*

Lehramtsstudenten und Doktoranden (S. 69–80). Springer-Verlag Berlin Heidelberg.

Ruelmann, M., Torchetti, L., Zulliger, S., Buholzer, A., & Praetorius, A.-K. (2021). Kognitiv-motivationale Schüler*innenprofile und ihre Bedeutung für die Schüler*innenwahrnehmung der Lernunterstützung durch die Lehrperson. *Unterrichtswissenschaft*, *49*(3), 395–422. https://doi.org/10.1007/s42010-021-00100-3

Schecker, H., & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, *12*, 45–66.

Schriebl, D., Müller, A., & Robin, N. (2022). Modelling Authenticity in Science Education. *Science & Education*. https://doi.org/10.1007/s11191-022-00355-x

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T.-Y., & Lee, Y.-H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching*, *44*(10), 1436–1460. https://doi.org/10.1002/tea.20212

Schukajlow, S., & Krawitz, J. (2020). Ist Lösungsvielfalt lernförderlich? Multiple Lösungen beim Mathematischen Modellieren. *MNU Journal*, *73*, 182–187.

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Siemon, J., Scholkmann, A., & Paulsen, T. (2018). Beschreibung von Formen lehrerseitigen Unterstützungsverhaltens im offenen Unterricht: Eine Analyse anhand von Videodaten. *Zeitschrift für Bildungsforschung*, *8*(1), 19–41. https://doi.org/10.1007/s35834-018-0206-y

Sjuts, J. (2022). Metakognition innerhalb von Aufgaben und Aufgabenbearbeitungen in Mathematik. *Zeitschrift für Mathematikdidaktik in Forschung und Praxis*, *3*. https://doi.org/10.48648/R2R9-6D11

Stäudel, L., Bohl, T., Merk, S., & Rehm, M. (2012). Aufgaben im naturwissenschaftlichen Unterricht. Allgemeindidaktische, fachdidaktische und fachliche Expertise. *Journal für Lehrerinnen und Lehrerbildung*, *12*, 26–32.

Stebler, R., Pauli, C., & Reusser, K. (2022). Personalisiertes Lernen als schulisches Bildungskonzept. Erscheinungsformen, Qualitätsmerkmale und Forschungsbefunde. In G. Brägger & H.-G. Rolff (Hrsg.), *Handbuch Lernen mit digitalen Medien* (2., korrigierte Auflage, S. 402–430). Beltz.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building Student Capacity for Mathematical Thinking and Reasoning: An Analysis of Mathematical Tasks Used in Reform Classrooms. *American Educational Research Journal*, *33*(2), 455–488. JSTOR. https://doi.org/10.2307/1163292

Streiner, D. L. (2003). Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*, *80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18

Stuppan, S., Bölsterli Bardy, K., Schmid, A. M., & Wilhelm, M. (2023). Überschätzen die Lehrmittelautor:innen den authentischen Lebensweltbezug von

MINT-Aufgaben? Eine Studie zur Lernendenperspektive. *Zeitschrift für Didaktik der Naturwissenschaften*, *29*(1), 9. https://doi.org/10.1007/s40573-023-00158-9

Stuppan, S., Wilhelm, M., & Bölsterli Bardy, K. (2022). Quality of task sets – An instrument for analysing science tasks with different functions along the learning process. *Progress in Science Education (PriSE)*, *Vol. 5 No. 1 (2022): Special Issue: Tasks in Science Education*, 76–93. https://doi.org/10.25321/prise.2022.1330

Stuppan, S., Wilhelm, M., & Bölsterli Bardy, K. (under review). Empirische Unterscheidung von Aufgabentypen – eine explorative Studie. *Swiss Journal of Educational Research*.

Stuppan, S., Wilhelm, M., Bölsterli Bardy, K., & Künzle, R. (2022). Messinstrument zur Analyse und Kategorisierung von MINT-Aufgaben – Konstruktion und Validierung. *Tagungsband der 5. Tagung Fachdidaktiken*, 533–539. https://doi.org/10.33683/dida.22.05.86

Sumfleth, E., & Nakoinz, S. (2019). Chemie verstehen – beobachtbare makroskopische Phänomene auf submikroskopischer Ebene modellbasiert interpretieren. *Zeitschrift für Didaktik der Naturwissenschaften*, *25*(1), 231–243. https://doi.org/10.1007/s40573-019-00101-x

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Taber, K. S. (2019). Alternative Conceptions and the Learning of Chemistry. *Israel Journal of Chemistry*, *59*(6–7), 450–469. https://doi.org/10.1002/ijch.201800046

Thieroff, B., Schubert, J. C., & Gölitz, D. (2021). Entwicklung und empirische Validierung eines kontextorientierten Skalenmodells zur Erfassung des Interesses von Schüler*innen am Klimawandel. *Zeitschrift für Didaktik der Naturwissenschaften*, *27*(1), 45–57. https://doi.org/10.1007/s40573-021-00125-2

Töpfer, C., Hapke, J., Liebl, S., & Sygusch, R. (2022). Kompetenzorientierung im Sport: Eine Taxonomie für den Sportunterricht. *German Journal of Exercise and Sport Research*, *52*(4), 570–583. https://doi.org/10.1007/s12662-022-00831-0

Ulrich, N. (2019). *Interaktive Lernaufgaben in dem digitalen Schulbuch eChemBook: Einfluss des Interaktivitätsgrads der Lernaufgaben und des Vorwissens der Lernenden auf den Lernerfolg*. Logos Verlag.

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in Teacher–Student Interaction: A Decade of Research. *Educational Psychology Review*, *22*(3), 271–296. https://doi.org/10.1007/s10648-010-9127-6

von Aufschnaiter, C., & Alonzo, A. C. (2018). Foundations of formative assessment: Introducing a learning progression to guide preservice physics teachers'

video-based interpretation of student thinking. *Applied Measurement in Education*, *31*(2), 113–127. https://doi.org/10.1080/08957347.2017.1408629

Vorholzer, A., & von Aufschnaiter, C. (2020). Dimensionen und Ausprägungen fachinhaltlicher Kompetenz in den Naturwissenschaften – ein Systematisierungsversuch. *Zeitschrift für Didaktik der Naturwissenschaften*, *26*(1), 1–18. https://doi.org/10.1007/s40573-020-00107-w

Walter, O., & Taskinen, P. (2009). Naturwissenschaftsbezogene Motivationen und Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund in Deutschland: Der Einfluss der Generation, der Herkunft und des Elternhauses. In M. Prenzel & J. Baumert (Hrsg.), *Vertiefende Analysen zu PISA 2006: Zeitschrift für Erziehungswissenschaft* (S. 185–203). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91815-0_10

Weiss, L., & Müller, A. (2015). The notion of authenticity in the PISA units in physical science: An empirical analysis. *Zeitschrift für Didaktik der Naturwissenschaften*, *21*(1), 87–97. https://doi.org/10.1007/s40573-015-0025-9

Wildhirt, S., & Luthiger, H. (2020). Das LUKAS-Modell: Ein Weg der interdisziplinären Zusammenarbeit zwischen Allgemeiner Didaktik und den Fachdidaktiken in der Lehrerinnen- und Lehrerbildung der Pädagogischen Hochschule Luzern. In D. Scholl, S. Wernke, D. Behrens, & K. Zierer (Hrsg.), *Thementeil Allgemeine Didaktik und Fachdidaktik* (S. 57–74). Schneider Verlag Hohengehren.

Wilhelm, M., & Kalcsics, K. (2017). *Lernwelten Natur – Mensch – Gesellschaft: Fachdidaktische Grundlagen. Ausbildung. Studienbuch 3. Zyklus* (1. Auflage). Schulverlag.

Wilhelm, M., Luthiger, H., & Wespi, C. (2014). *Kategoriensystem für ein kompetenzorientiertes Aufgabenset*. Entwicklungsschwerpunkt Kompetenzorientierter Unterricht, Pädagogische Hochschule Luzern.

Winkler, I. (2011). *Aufgabenpräferenzen für den Literaturunterricht: Eine Erhebung unter Deutschlehrkräften*. VS Verlag für Sozialwissenschaften / Springer Fachmedien Wiesbaden, Wiesbaden. https://doi.org/10.1007/978-3-531-92698-8

Zhu, M., Liu, O. L., & Lee, H.-S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education*, *143*, 103668. https://doi.org/10.1016/j.compedu.2019.103668

Anhang A: Publikation Teilstudie I

Messinstrument zur Analyse und Kategorisierung von MINT-Aufgaben – Konstruktion und Validierung

Sebastian Stuppan[a,b], Markus Wilhelm[a,b], Katrin Bölsterli Bardy[a] & Roland Künzle[a]

[a] Pädagogische Hochschule Luzern

[b] Pädagogische Hochschule Heidelberg, Deutschland

Eingereicht: Juli 2021 / Akzeptiert: Januar 2022

DOI: https://doi.org/10.33683/dida.22.05.86

Abstract

In einem kompetenzorientierten Unterricht haben Aufgaben eine entscheidende Bedeutung. Mit Hilfe von Aufgabensets werden Gelegenheiten geschaffen, damit Lernende ihre Kompetenzen individuell entwickeln und darin gefördert werden. Im Forschungsvorhaben wird ein Messinstrument zur Analyse und Kategorisierung von MINT-Aufgaben erarbeitet. Dieser Artikel beschreibt und diskutiert die Konstruktion und Validierung des Aufgaben-Analyse-Instruments (AAI).

Keywords

Validierungsstrategie, Messinstrument; MINT – Aufgaben, Kategorisierung, Skalenentwicklung

Detail Ergebnisse Teilstudie I

Konstrukt (Grundlage: Literatur und Theorien)	**1.1 Konzeptspezifikation: Anzahl Unteraspekt eines Konstrukts** (Grundlage: Literatur und Theorien)	**2.1 Operationalisierung** (Item-Konstruktion)	**2.2 Operationalisierung** (Qualitative Gespräche: Fachdidaktiker:innen)	**3.1 Einzelne Experten:innen-Befragung** (Qualität der Items)	**3.2 Experten:innen-Befragung** (Evaluation)	**3.3 Experten:innen-Befragung** (Qualität der Items und Konsens-Rating von MINT-Aufgaben)	**4.1 Experten:innen-Pretest und Revision** (Evaluation und Konsens-Rating von MINT-Aufgabe 6110)	**4.2 Experten:innen-Pretest und Revision** (Rating von MINT-Aufgabe 4110 1a)	**4.3 Expert:innen-Pretest und Revision** (Rating von MINT-Aufgabe 4110 1b)	**Theoriebasierter und empirisch geklärter Itempool**
Kompetenzabbild	**3**	9 Items	Item Änderung			9 Items umformulieren	ICC(C,1): .97** 2 Items umformulieren	ICC(C,1): - Testmanual anpassen	ICC(C,1): .71**	**9 Items**
Lebensweltbezug	**3**	15 Items		3 Items umformulieren - 1 Item	- 2 Items	3 Items umformulieren	ICC(C,1): .58**	ICC(C,1): - Testmanual anpassen	ICC(C,1): .67**	**12 Items**
Lernendenvorstellung	**3**	17 Items	Item Änderung	- 4 Items	+ 3 Items	- 4 Items	ICC(C,1): .71**	ICC(C,1): .91**	ICC(C,1): .83**	**12 Items**
Wissensart	**4**	16 Items		- 3 Item		3 Items umformulieren	ICC(C,1): .84**	ICC(C,1): .50**	ICC(C,1): .78**	**13 Items**
Wissensaktivität	**4**	19 Items	Item Änderung	+ 1 Item 3 Items umformulieren	- 2 Items	- 4 Items 2 Items umformulieren	ICC(C,1): .76** 3 Items umformulieren	ICC(C,1): - Testmanual anpassen	ICC(C,1): .69**	**14 Items**
Repräsentationsform	**3**	12 Items	Item Änderung	+ 2 Items	- 3 Items		ICC(C,1): .59**	ICC(C,1): - 7 Items leicht umformulieren	ICC(C,1): .68**	**11 Items**
Offenheit	**7**	20 Items	Item Änderung	+ 2 Items	- 4 Items	+ 4 Items	ICC(C,1): .90** 3 Items umformulieren	ICC(C,1): .87** 1 Item leicht umformulieren	ICC(C,1): .87**	**22 Items**
Lernunterstützung	**3**	14 Items	Item Änderung		- 2 Items		ICC(C,1): .93** - 2 Items	ICC(C,1): .74**	ICC(C,1): .91**	**10 Items**
Lernweg	**3**	14 Items	Item Änderung		+ 1 Item - 4 Items	1 Item umformulieren	ICC(C,1): .53**	ICC(C,1): .57**	ICC(C,1): .63**	**11 Items**

Signifikanzniveaus: ** $p < .001$; * $p < .01$; . $p < .05$

Abbildung A.1. Darstellung der Konstruktion des theoriebasierten und empirisch geklärten Itempools

Anhang B: Publikation Teilstudie II

Quality of task sets – An instrument for analysing science tasks with different functions along the learning process

Sebastian Stuppan[a,b], Markus Wilhelm[a,b] & Katrin Bölsterli Bardy[a]

[a] Pädagogische Hochschule Luzern

[b] Pädagogische Hochschule Heidelberg, Deutschland

Received: October 2021 / Accepted: April 2022

DOI: https://doi.org/10.25321/prise.2022.1330

Structured Abstract

Background: In competency-oriented education, tasks in science subjects have increased in importance in recent years. In addition, it is suggested that to develop a competency, a set of tasks with different functional embeddings is required. An option for the arrangement of tasks is explained in the Learning Process Model and starts with the so-called "confrontation" task at the beginning of a new topic, challenging learners with a new problem. Next, learners build up the required concepts and skills with the help of development tasks to be able to solve the initial problem, followed by exercises to train and expand the competency. To sum up, learners solve the initial problem in a synthesis task and may then need to apply the competency in a transfer task. Each task type (e.g., confrontation task) is described by the weighting of the following nine scales: 1) chart of competencies, 2) relationship to daily life, 3) learners' conceptions, 4) knowledge, 5) knowledge activities, 6) forms of representation, 7) task openness, 8) learning supports, and 9) learning paths. Each scale contains between one and four subscales that descri-

be the task types by their weighting. The description in the form of scale values is taken up empirically.

Purpose: This study combines existing task scales from research with the different functions of tasks along the learning process model. An expert panel worked in a general-theoretical manner, and trained lecturers rated tasks with the instrument to analyse tasks.

Sample/setting, Design and Methods: In our study, we used the existing scales from the *Instrument to Analyse Tasks* (IAT). To develop the experts' proposed scale values, we consulted four experts. We calculated the AD_M index (average absolute deviation) as a quality-control measure for the experts' level of agreement. According to the tasks' scale values as rated by trained lecturers ($N = 2$), we selected 25 of the 146 science education tasks from the project *MINT*[1] *unterwegs* ("STEM on the move"). In the comparison, we calculated the score differences between the experts' scale values and the rated task.

Results: The results show that it is possible to describe different task types of a task set with the IAT when the scale values are obtained from expert proposals. Moreover, the IAT scale values obtained from the expert proposals are quite similar to those from task ratings by trained lecturers.

Conclusions: This study indicates that experts can distinguish and characterize the Learning Process Model's various task types by weighting IAT scale proposals. Furthermore, it has been shown that tasks can be analysed with the IAT. When the results of the task analysis by trained lecturers are compared to the experts' proposals, the tasks can be revised reasonably and optimized for the learning process.

Keywords: *competency, science, task sets, learning, model, instrument, analysis, task quality*

[1] "MINT" is the German acronym for "Mathematik, Informatik, Naturwissenschaften und Technik", the English version being "STEM" (science, technology, engineering and mathematics). Sample tasks from the project *MINT unterwegs* are available on the website: https://mint-erleben.lu.ch (retrieved 20 January 2022). The project has been running since 2016.

Anhang C: Publikation Teilstudie III

Empirische Unterscheidung von Aufgabentypen – eine explorative Studie

Sebastian Stuppan[a,b], Markus Wilhelm[a,b] & Katrin Bölsterli Bardy[a]

[a] Pädagogische Hochschule Luzern

[b] Pädagogische Hochschule Heidelberg, Deutschland

Eingereicht: Februar 2023

Zusammenfassung

Mit dem wechselnden Fokus von der Input- zur Outcomeorientierung und von der Schul- zur Unterrichtsqualität rückt die Aufgabenqualität in den Forschungsfokus. Damit der in den Lehrplänen anvisierte kumulative Kompetenzaufbau der Lernenden unterstützt werden kann, gewinnen aufeinander aufbauende Aufgaben an Bedeutung. Der in dieser Studie vorgestellte Lernprozessmodellansatz unterscheidet dazu fünf Aufgabentypen zur Konfrontation, Erarbeitung, Übung, Synthese und zum Transfer beim Lernen. Inwiefern sich diese Typen in Aufgabensets tatsächlich empirisch unterscheiden, wird in dieser Studie untersucht. Dazu werden 58 MINT-Aufgaben mittels Clusteranalysen zur Eruierung der Aufgabentypen in Gruppen eingeteilt. Die Befunde der Studie zeigen, dass eine Typisierung grundsätzlich möglich ist.

Schlüsselwörter

Clusteranalyse, Lernaufgaben, Lernprozess, Aufgabenqualität, Aufgabenanalyse

Empirical differentiation of task types - an explorative study

Abstract

As attention has shifted over the last decade from input to outcome orientation and from school to teaching quality, researchers have increasingly focused on task quality. To support learners' cumulative competency development as envisaged by curricula, sets of tasks that build on each other are critical. Using the Learning Process Model presented in this study, tasks can be divided into five types: confrontation, development, exercise, synthesis, and transfer, to promote competency-enhancing education. The extent to which this theoretical assumption of task types is reflected in the task sets is not yet clear. In this study, 58 STEM tasks were grouped using cluster analysis to identify task types. The results of the study show that a typology of tasks can in principle be derived.

Keywords

Cluster analysis, learning tasks, learning process, task quality, task analysis

Detail Ergebnisse Teilstudie III

Tabelle C.1. *AAI: Skalen, Subskalen, Cronbachs α, Beispielitems und Anzahl Items (Stuppan et al., 2022)*

Skala	Subskala	*Cronbachs α*	Beispielitem (Anzahl Items)
Kompetenzabbild	singulär	.99	… wird lediglich ein Aspekt einer (Sub-)Kompetenz bearbeitet (3).
	additiv	.98	… werden nur einige Aspekte einer oder mehrerer (Sub-)Kompetenzen bearbeitet (3).
	integrativ	.98	… werden möglichst viele Aspekte einer oder mehrerer (Sub-)Kompetenzen bearbeitet (3).
Lebensweltbezug	konstruiert	.96	... ist ausgedacht (3).
	authentisch	.93	… könnte mit dem echten Leben der Lernenden etwas zu tun haben (3).
	real	.96	… müssen die Lernenden tatsächlich in ihrem Leben lösen (3).
Lernendenvorstellung	erheben	.91	Die Lernenden müssen ihre Vermutung aufschreiben oder erzählen (3).
	abgleichen	.90	Die Lernenden wissen nach dem Bearbeiten …, ob ihre Lösungsidee stimmt (2).
	reflektieren	.82	Die Lernenden müssen reflektieren, warum ihre Lösung mit dem wissenschaftlichen Konzept übereinstimmt oder nicht (3).
Wissensart	Fakten	.91	Das Lösen … erfordert einen Umgang mit Fakten (3).
	Prozeduren	.96	Die Aufgabenbearbeitung beinhaltet den Umgang mit Fertigkeiten (3).
	Konzepte	.87	Bei der Aufgabenbearbeitung müssen die Lernenden konzeptionelles Wissen anwenden (3).
	Metakognition	.89	… muss überlegt werden, was beim gewählten Vorgehen zielführend oder unangemessen ist (3).
Wissensaktivität	reproduzieren	.88	Eine ähnliche Aufgabe haben die Lernenden schon gelöst (4).
	transferieren	.94	... müssen die Lernenden ihr Wissen aus dem Unterricht in einer neuen Situation anwenden (4).
	kreieren	.96	Die Lernenden müssen … neue Lösungsideen entwickeln oder erproben (4).
Repräsentationsform	Transformation	.93	Die Aufgabenbearbeitung ... erfordert eine andere Darstellungsform (4).
Offenheit	Lösungsvorgehen frei	.95	... lässt offen, wie bei der Lösung vorzugehen ist (4).
	Resultat vielfältig	.96	… lässt mehrere Resultate zu (4).
	Bearbeitung vorstrukturiert	.92	… ist für die Lösungsfindung vorstrukturiert (4).
Lernunterstützung	Hilfestellung	.97	… hat es Hinweise, sodass die Lernenden diese Aufgabe einfacher lösen können (3).
	Austauschmöglichkeit	.88	Beim Lösen … können sich die Lernenden mit anderen austauschen (3).
	Lehrpersonenfeedback	.98	Die Lösungsvorschläge der Lernenden werden … von der Lehrperson kommentiert (3).
Lernweg	kompensierend	.93	... ermöglicht, fachliche Lücken zu schließen (3).
	profilbildend	.92	… bietet ein Wahlangebot nach Interesse an (3).
	selbstdifferenzierend	.92	… bietet den Lernenden die Gelegenheit, verschiedene Lösungstiefen zu wählen (3).

Anmerkungen. 5-stufige Ratingskala (von 0 = stimmt nicht bis 4 = stimmt sehr); drei Punkte „…“ bezeichnen „(in) diese/r Aufgabe“

Tabelle C.2. *Zusammenfassung der Rating-Ergebnisse zur Einschätzung der MINT-Aufgaben anhand der AAI-Skalen*

Skala	Subskala	Relative Häufigkeit der Ratingskala (in %)				
		0	1	2	3	4
Kompetenzabbild	singulär	59	24	7	10	0
	additiv	17	0	7	43	33
	integrativ	41	28	17	14	0
Lebensweltbezug	konstruiert	3	45	10	26	16
	authentisch	7	14	22	52	5
	real	29	36	17	17	0
Lernendenvorstellung	erheben	0	0	3	53	43
	abgleichen	17	31	29	10	12
	reflektieren	21	55	21	3	0
Wissensart	Fakten	2	7	19	71	2
	Prozeduren	21	14	16	48	2
	Konzepte	0	3	17	66	14
	Metakognition	59	29	12	0	0
Wissensaktivität	reproduzieren	16	36	31	10	7
	transferieren	7	17	31	36	9
	kreieren	19	22	40	12	7
Repräsentationsform	Transformation	3	24	28	33	12
Offenheit	Lösungsvorgehen frei	5	17	16	45	17
	Resultat vielfältig	10	45	5	22	17
	Bearbeitung vorstrukturiert	0	7	33	31	29
Lernunterstützung	Hilfestellung	22	52	16	10	0
	Austauschmöglichkeit	3	3	14	38	41
	Lehrpersonenfeedback	22	43	17	16	2
Lernweg	kompensierend	17	12	22	48	0
	profilbildend	31	24	22	22	0
	selbstdifferenzierend	9	12	26	45	9

Anmerkung. 5-stufige Ratingskala (von 0 = stimmt nicht bis 4 = stimmt sehr)

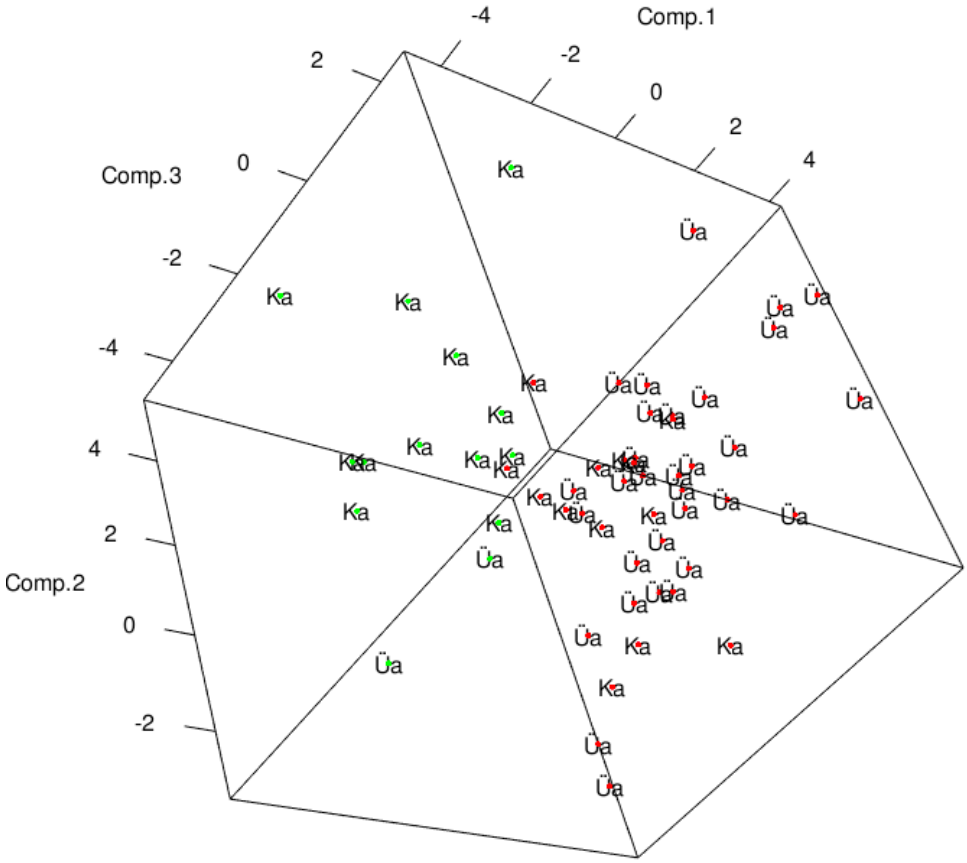

Abbildung C.1. 3D Hauptkomponentenanalyse der PAM-Clusterlösung mit 46.6% Varianzaufklärung, rote Punkte = Cluster 1 ($n = 44$ Aufgaben), grüne Punkte = Cluster 2 ($n = 14$ Aufgaben), Ka = Gemäß MINT-Autor:innen Konfrontationsaufgaben, Üa = Gemäß MINT-Autor:innen Übungsaufgaben

Tabellarisches Verzeichnis der untersuchten „*MINT unterwegs*“-Lernaufgaben.

Tabelle C.3. *Untersuchte MINT-Aufgaben*

Aufgabentitel	Webadresse der „*MINT unterwegs*“-Aufgabe (abgerufen am 20.8.2022)	Aufgabentyp gemäß MINT-Autor:innen	PAM-Cluster-Zuordnung
01 Welche Sportarten können wir wählen?	https://mint-erleben.lu.ch/Zyklus2/show/Körper/laufen	Ka	1
04 Welcher Muskel ist beim Armdrücken angespannt?	https://mint-erleben.lu.ch/Zyklus2/show/Körper/laufen	Üa	1
01 Der verletzte Arm	https://mint-erleben.lu.ch/Zyklus2/show/Körper/Knochenmara	Ka	1
03 Was macht deinen Knochen stabil?	https://mint-erleben.lu.ch/Zyklus2/show/Körper/Knochenmara	Üa	1
01 Die verschwundenen Muskeln	https://mint-erleben.lu.ch/Zyklus2/show/Körper/Gips	Ka	2
04 Warum Gelenke nicht quietschen…	https://mint-erleben.lu.ch/Zyklus2/show/Körper/Gips	Üa	1

Tabelle C.3. (Fortsetzung)

Aufgabentitel	Webadresse der „*MINT unterwegs*“-Aufgabe (abgerufen am 20.8.2022)	Aufgabentyp gemäß MINT-Autor:innen	PAM-Cluster-Zuordnung
05 Beugen und strecken: Hier spielen zwei Muskeln gegeneinander!	https://mint-erleben.lu.ch/Zyklus2/show/Körper/Gips	Üa	1
01 Wenn das Herz kräftig pocht	https://mint-erleben.lu.ch/Zyklus2/show/Körper/fit	Ka	1
01 Vorbeugen ist besser als heilen	https://mint-erleben.lu.ch/Zyklus2/show/Körper/Belastung	Ka	1
03 Knochen müssen Belastungen aushalten	https://mint-erleben.lu.ch/Zyklus2/show/Körper/Belastung	Üa	1
01 Luca kommt zu spät zur Schule	https://mint-erleben.lu.ch/Zyklus2/show/energie/Gesichter	Ka	2
03 Auf der Suche nach Energieformen	https://mint-erleben.lu.ch/Zyklus2/show/energie/Gesichter	Üa	1
01 Lucas Hamster produziert Strom	https://mint-erleben.lu.ch/Zyklus2/show/energie/Energiespur	Ka	1
01 Ein Belüftungssystem für Luca ausdenken	https://mint-erleben.lu.ch/Zyklus2/show/energie/Energiequelle	Ka	2
03 Mache dein eigenes Elektroquiz	https://mint-erleben.lu.ch/Zyklus2/show/energie/Energiequelle	Üa	1
04 Luca möchte die Geschwindigkeit des Motors verändern	https://mint-erleben.lu.ch/Zyklus2/show/energie/Energiequelle	Üa	1
04 Kann ein Auto mit einer «Essig-Batterie» fahren?	https://mint-erleben.lu.ch/Zyklus2/show/energie/Energiequelle	Üa	1
01 Temperatursturz im Schulzimmer	https://mint-erleben.lu.ch/Zyklus2/show/energie/Tierewinter	Ka	1
02 Unsere Forschungskonferenz - 1. Teil	https://mint-erleben.lu.ch/Zyklus2/show/energie/Tierewinter	Üa	1
02 Unsere Forschungskonferenz - 2. Teil	https://mint-erleben.lu.ch/Zyklus2/show/energie/Tierewinter	Üa	1
02 Abschluss Forschungskonferenz	https://mint-erleben.lu.ch/Zyklus2/show/energie/Tierewinter	Üa	1
Woraus besteht eine Wasserglace?	https://mint-erleben.lu.ch/Zyklus2/show/chemie > NMG3b_Stoffe_Lehrpersonenmaterial_V40 S. 9 [a]	Ka	2
Unterscheidung der Begriffe «Stoff», «Material» und «Gegenstand»	https://mint-erleben.lu.ch/Zyklus2/show/chemie > NMG3b_Stoffe_Lehrpersonenmaterial_V40 S. 11 [a]	Üa	1
Stoffe beschreiben und unterscheiden: Mit welchen Eigenschaften?	https://mint-erleben.lu.ch/Zyklus2/show/chemie > NMG3b_Stoffe_Lehrpersonenmaterial_V40 S. 13 [a]	Üa	1
Wer ist der Täter? Teil 1: Wer hat den Drohbrief geschrieben?	https://mint-erleben.lu.ch/Zyklus2/show/chemie > NMG3b_Stoffe_Lehrpersonenmaterial_V40 S. 16 [a]	Ka	2
Sind in Wasser gelöster Zucker und Wasserdampf noch vorhanden, obwohl sie unsichtbar sind?	https://mint-erleben.lu.ch/Zyklus2/show/chemie > NMG3b_Stoffe_Lehrpersonenmaterial_V40 S. 22 [a]	Ka	2
Denkmodell oder Modell von der sichtbaren Welt?	https://mint-erleben.lu.ch/Zyklus2/show/chemie > NMG3b_Stoffe_Lehrpersonenmaterial_V40 S. 25 [a]	Üa	1
Kannst du Haushaltszucker in einen neuen Stoff umwandeln?	https://mint-erleben.lu.ch/Zyklus2/show/chemie > NMG3b_Stoffe_Lehrpersonenmaterial_V40 S. 27 [a]	Ka	2
Auf welchen Bildern entstehen neue Stoffe?	https://mint-erleben.lu.ch/Zyklus2/show/chemie > NMG3b_Stoffe_Lehrpersonenmaterial_V40 S. 29 [a]	Üa	1
01 Hell und dunkel > Aufgabe A	https://mint-erleben.lu.ch/Zyklus2/show/Optik/lichtschatten	Ka	2
04 Schattentheater	https://mint-erleben.lu.ch/Zyklus2/show/Optik/lichtschatten	Üa	1
01 Der tote Winkel	https://mint-erleben.lu.ch/Zyklus2/show/Optik/Rückspiegel	Ka	2
04 Laser-Ziel-Spiel	https://mint-erleben.lu.ch/Zyklus2/show/Optik/Rückspiegel	Üa	1
03 Löffelspiegel und Kosmetikspiegel	https://mint-erleben.lu.ch/Zyklus2/show/Optik/Rückspiegel	Üa	1

Tabelle C.3. (Fortsetzung)

Aufgabentitel	Webadresse der „*MINT unterwegs*“-Aufgabe (abgerufen am 20.8.2022)	Aufgabentyp gemäß MINT-Autor:innen	PAM-Cluster-Zuordnung
01 Der vergessliche Detektiv	https://mint-erleben.lu.ch/Zyklus2/show/Optik/Lupen	Ka	1
03 Wasserlupe bauen und untersuchen	https://mint-erleben.lu.ch/Zyklus2/show/Optik/Lupen	Üa	1
01 Sonnenbrand trotz T-Shirt?	https://mint-erleben.lu.ch/Zyklus2/show/Optik/Sonnenlicht	Ka	2
04 Farbige Schatten	https://mint-erleben.lu.ch/Zyklus2/show/Optik/Sonnenlicht	Üa	1
01 Indianer fangen Fische	https://mint-erleben.lu.ch/Zyklus2/show/Optik/Lichtbrechung	Ka	1
01 Was macht einen elektrischen Gegenstand aus?	https://mint-erleben.lu.ch/Zyklus2/show/Elektrizität/Elektrostatik	Ka	2
01 Das Abenteuer der Familie Da Silva	https://mint-erleben.lu.ch/Zyklus2/show/Elektrizität/Stromkreis	Ka	2
02 Was ist ein Kurzschluss?	https://mint-erleben.lu.ch/Zyklus2/show/Elektrizität/Schalt	Üa	1
01 Wie ist eine Lichterkette aufgebaut?	https://mint-erleben.lu.ch/Zyklus2/show/Elektrizität/Lichterkette	Ka	1
03 Stimmen die folgenden Überlegungen?	https://mint-erleben.lu.ch/Zyklus2/show/Elektrizität/verstanden	Ka	1
01 Wie funktionieren die Schaltungen?	https://mint-erleben.lu.ch/Zyklus2/show/Elektrizität/verstanden	Üa	1
01 Der Rasenmähroboter im Garten	https://mint-erleben.lu.ch/Zyklus2/show/Robotik/Roboter	Ka	1
03 Was haben der Mensch und ein Roboter gemeinsam?	https://mint-erleben.lu.ch/Zyklus2/show/Robotik/Roboter	Üa	1
04 Von Sensoren und Aktoren	https://mint-erleben.lu.ch/Zyklus2/show/Robotik/Roboter	Üa	1
06 Das Roboterspiel "Ich, der Blue-Bot"	https://mint-erleben.lu.ch/Zyklus2/show/Robotik/Roboter	Üa	1
01 Der Bodenroboter entdeckt die Schatzinsel	https://mint-erleben.lu.ch/Zyklus2/show/Robotik/Problemlösung	Ka	1
03 Die Bodenroboter tauschen ihre Plätze	https://mint-erleben.lu.ch/Zyklus2/show/Robotik/Problemlösung	Üa	1
04 Mensch-Thymio-Roboter	https://mint-erleben.lu.ch/Zyklus2/show/Robotik/Problemlösung	Üa	1
05 Mit dem Thymio-Roboter verschiedene Aufgaben lösen: Level 2	https://mint-erleben.lu.ch/Zyklus2/show/Robotik/Problemlösung > AUFGABE C Level 2 [a]	Üa	1
02 Wir verlassen die Schatzinsel	https://mint-erleben.lu.ch/Zyklus2/show/Robotik/Problemlösung	Üa	1
05 Mit dem Thymio-Roboter verschiedene Aufgaben lösen: Level 1	https://mint-erleben.lu.ch/Zyklus2/show/Robotik/Problemlösung > AUFGABE B Level 1 [a]	Üa	1
01 Die technische Entwicklung der Roboter	https://mint-erleben.lu.ch/Zyklus2/show/Robotik/Gesellschaft	Ka	1
03 Roboterkonstruktionen > Aufgabe A und B	https://mint-erleben.lu.ch/Zyklus2/show/Robotik/Gesellschaft	Üa	2
03 Roboterkonstruktionen > Aufgabe C	https://mint-erleben.lu.ch/Zyklus2/show/Robotik/Gesellschaft	Üa	2

Anhang D: Publikation Teilstudie IV

Überschätzen die Lehrmittelautor:innen den authentischen Lebensweltbezug von MINT-Aufgaben? Eine Studie zur Lernendenperspektive

Sebastian Stuppan[a,b], Katrin Bölsterli Bardy[a], Andrea Maria Schmid[a,b] & Markus Wilhelm[a,b]

[a] Pädagogische Hochschule Luzern

[b] Pädagogische Hochschule Heidelberg, Deutschland

Eingereicht: Juni 2022 / Akzeptiert: März 2023 / Publiziert: 29 April 2023

DOI: https://doi.org/10.1007/s40573-023-00158-9

Zusammenfassung

In der Scientific Literacy und der Context-Based Science Education wird dem Lebensweltbezug eine hohe Bedeutung zugeschrieben. In vielen Forschungsprojekten zum Lebensweltbezug, ein Aspekt der Aufgabenqualität, werden schriftliche Aufgabenstellungen beurteilt und ausgewertet. Eine Schwierigkeit beim Analysieren von Aufgaben ist die Tatsache, dass die Bewertung von geschulten Beurteilenden anders ausfallen kann als durch die Zielgruppe im Unterricht. Die Dokumentenanalyse bildet das objektive Potenzial der Aufgaben aus Sicht der jeweiligen Beurteiler:innen ab. Die Einbettung der Aufgaben im Unterricht, das Niveau der Lösungen sowie die Aufgabeneinschätzung der Lernenden werden weggelassen. Studien, die sich mit der Analyse der schriftlichen Aufgaben und der Aufgabenbeurteilungen durch Lernende im Unterricht auseinandersetzen, existieren derzeit kaum. Um diese Forschungslücke zu schmälern, wird in diesem Vorhaben verglichen, wie fachdidaktische Lehrmittelautor:innen ihre entwickelten Aufgaben einschätzen und Lernende dieselben Aufgaben im Unterricht nach deren Be-

arbeitung wahrnehmen. Hierzu beurteilten Lernende ($N = 805$) ausgewählte MINT-Aufgaben ($N = 16$) auf deren Lebensweltbezug (min. 66, max. 182 Lernende pro Aufgabe) mithilfe des Aufgaben-Analyse-Instruments. Die Lernenden-Einschätzungen werden anschließend inferenzstatistisch der fachdidaktischen Perspektive der Lehrmittelautor:innen gegenübergestellt und diskutiert. Die Befunde dieser Studie zeigen, dass die Lernenden in den untersuchten MINT-Aufgaben die Lebenswelt divers wahrnehmen und es keine geschlechtsspezifischen Unterschiede gibt. Die Gegenüberstellung der Lernenden-Einschätzungen und der Setzung der Lehrmittelautor:innen macht deutlich, dass die Autor:innen in der Regel die Aufgaben authentischer beurteilen, als die Lernenden diese wahrnehmen. Aufgrund der doch deutlichen Diskrepanz zwischen den zwei Perspektiven wird diskutiert, welchen Mehrwert der Einbezug der Praxis in einem iterativen Verfahren beim Erstellen von Lernaufgaben haben könnte, um den Lebensweltbezug aus der Perspektive der Lernenden stringenter aufnehmen zu können.

Schlüsselwörter

Lebensweltbezug, Lernaufgaben, Einschätzungen, Aufgabenanalysen, Perspektiven, MINT-Lernmaterialien

Do authors of teaching materials overestimate the authenticity of STEM tasks' relationship to daily life? A study from the learner perspective

Abstract

In scientific literacy and context-based science education, the relationship to daily life is of upmost importance. As an aspect of task quality, many research projects assess and evaluate the relationship to daily life of written tasks. A major difficulty in task analysis is that trained lecturers' analyses can be different from those of the target group in the classroom, as document analysis shows only the objective potential of tasks from the assessors' perspective. The embedding of tasks in lessons, the level of solutions, and task evaluation by learners have been overlooked.

At present, hardly any studies exist dealing with the analysis of written tasks and task assessments by learners in the classroom. To reduce this research gap, this project compares how subject-specific education authors assess the tasks they have developed and how learners perceive the same tasks in class after they have been completed. For this purpose, learners ($N = 805$) assessed selected STEM tasks ($N = 16$) for their relationship to daily life (min. 66, max. 182 learners per task) using the Instrument to Analyse Tasks. The learners' assessments are analyzed and discussed using inferential statistics from the perspective of teaching-material authors. This research shows that learners have diverse perceptions of the relationship to daily life in the investigated STEM tasks. Moreover, learner gender does not have a significant influence here. Comparing the learners' and authors' assessments of the teaching materials makes it clear that the authors generally assess the tasks to be more authentic than do the learners. Because of the obvious discrepancy between the two perspectives on tasks' relationship to daily life, the potential benefit of iteratively including the learner's perspective when creating tasks is discussed.

Keywords

relationship to daily life, learning task, perceptions, Task analysis, perspectives, STEM Learning materials

Detail Ergebnisse Teilstudie VI

Tabelle D.1. *Eigenschaften der Items der Skala Lebensweltbezug – konstruiert (Fragebogen-Pretest, N = 220 Lernende)*

Item	*M*	*SD*	pi	$\alpha_{\text{item del}}$	Itemtext der Lernenden
S_LK102	52	35	.52	.69	Diese Aufgabe wurde ausgedacht.
S_LK303	48	37	.48	.61	Diese Aufgabe wurde erfunden.
S_LK701	35	34	.35	.71	Diese Aufgabe wurde zusammengedichtet.

M_{rit} = .51; Cronbachs α = .76

Anmerkungen. Die Werte sind in % als Anteil an der Maximalpunktzahl; pi = Itemschwierigkeiten; $\alpha_{\text{item del}}$ = Cronbachs Alpha falls Item entfernt; rit = korrigierte item-total-correlation (Trennschärfe)

Tabelle D.2. *Eigenschaften der Items der Skala Lebensweltbezug – authentisch (Fragebogen-Pretest, N = 220 Lernende)*

Item	*M*	*SD*	pi	$\alpha_{\text{item del}}$	Itemtext der Lernenden
S_LA102	52	33	.52	.78	Diese Aufgabe könnte mit meinem echten Leben etwas zu tun haben.
S_LA201	48	33	.47	.68	Diese Aufgabe könnte in meinem Alltag vorkommen.
S_LA202	50	34	.50	.70	Diese Aufgabe könnte auch außerhalb meines Schulalltags vorkommen.

M_{rit} = .56; Cronbachs α = .79

Anmerkungen. Die Werte sind in % als Anteil an der Maximalpunktzahl; pi = Itemschwierigkeiten; $\alpha_{\text{item del}}$ = Cronbachs Alpha falls Item entfernt; rit = korrigierte item-total-correlation (Trennschärfe)

Tabelle D.3. *Eigenschaften der Items der Skala Lebensweltbezug – real (Fragebogen-Pretest, N = 220 Lernende)*

Item	*M*	*SD*	pi	$\alpha_{\text{item del}}$	Itemtext der Lernenden
S_LR102	46	35	.46	.78	Diese Aufgabe muss ich tatsächlich in meinem Leben lösen.
S_LR303	35	33	.35	.62	Diese Aufgabe muss ich außerhalb der Schule auch bearbeiten.
S_LR601	30	33	.30	.65	Diese Aufgabe muss ich außerhalb der Schule auch lösen.

M_{rit} = 0.52; Cronbachs α = .76

Anmerkungen. Die Werte sind in % als Anteil an der Maximalpunktzahl; pi = Itemschwierigkeiten; $\alpha_{\text{item del}}$ = Cronbachs Alpha falls Item entfernt; rit = korrigierte item-total-correlation (Trennschärfe)

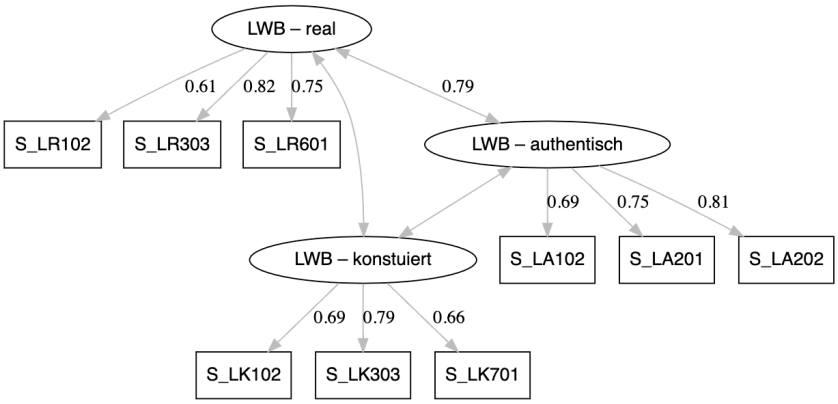

Abbildung D.1. Drei-Faktoren Modell der Subskalen: *konstruiert*, *authentisch* und *real* von *Lebensweltbezug* (LWB) (Fragebogen-Pretest, $N = 220$ Lernende). Model fit: χ^2 (24) = 43.325, p = .009, CFI = .965, SRMR = .043, GFI = .950, RMSEA = .066

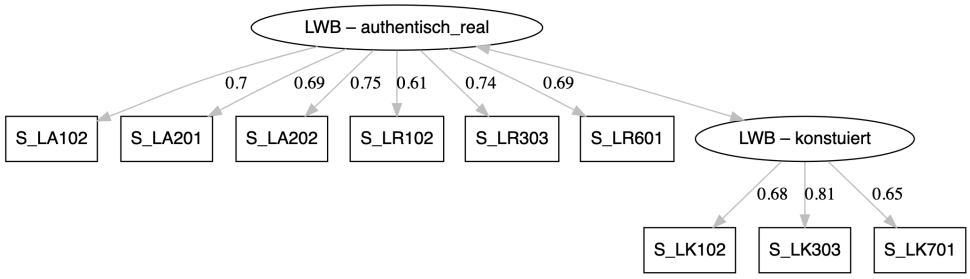

Abbildung D.2. Zwei-Faktoren Modell der Subskalen: *konstruiert*, *authentisch_real* von *Lebensweltbezug* (LWB) (Fragebogen-Pretest, $N = 220$ Lernende). Model fit: χ^2 (26) = 69.933, $p < .001$, CFI = .917, SRMR = .057, GFI = .916, RMSEA = .098

Eidesstattliche Erklärung

Ich erkläre, dass ich die vorgelegte Arbeit selbständig angefertigt, dabei keine anderen Hilfsmittel als die im Quellen- und Literaturverzeichnis genannten benutzt, alle aus Quellen und Literatur, einschließlich des Internets, wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht und auch die Fundstellen einzeln nachgewiesen habe.

Diese Arbeit ist in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

31.3.2023

Sebastian Stuppan

Danksagung

Diese Arbeit wäre ohne die Unterstützung, das Mitwirken und die Beratung vieler Menschen nie geschrieben worden. Ihnen allen gilt mein herzlicher Dank.

Besonders danke ich Prof. Dr. Markus Wilhelm für die Betreuung der Dissertation, die fachlich kompetenten, wertvollen Rückmeldungen und das große Vertrauen in meine Arbeit. Ich danke für die familienfreundlichen Arbeitsbedingungen, die mir als Vater von drei Mädchen entgegengebracht wurden, und die wissenschaftliche Förderung, von der ich in den vergangenen fünf Jahren reichlich profitieren durfte.

Ein weiterer Dank geht an meine beiden Co-Betreuer:innen Prof. Dr. Katrin Bölsterli Bardy und Prof. Dr. Markus Rehm für die zahlreichen konstruktiven Anregungen, im Speziellen in den Vor- und in den Nachbereitungen von Tagungen, die wir gemeinsam bestritten haben.

Danken möchte ich dem Forschungsfonds der PH Luzern, dem Institut für Fachdidaktik Natur, Mensch, Gesellschaft der PH Luzern und der Aebli-Näf-Stiftung für die finanzielle Unterstützung sowie dem binationalen Promotionskolleg der PH Heidelberg und PH Luzern für die geschaffenen Rahmenbedingungen, die diese Dissertation erst ermöglichten.

Ein weiterer Dank geht an die Dienststelle Volksschulbildung Luzern sowie den *„MINT unterwegs“*-Autor:innen für deren Bereitschaft, mir die Aufgabensets zur Verfügung zu stellen.

Danken möchte ich dem LUKAS (Luzerner Modell zur Entwicklung Kompetenzfördernder Aufgabensets)-Team für die Bereitschaft, sich meinen vielen Fragen zu stellen und sich mit der Expertise einzubringen. – Vielen lieben Dank: Prof. Dr. Herbert Luthiger, Prof. Dr. Markus Wilhelm, Claudia Wespi und Prof. Dr. Susanne Wildhirt.

Ich danke Sophia und Leonie Aregger sowie Roland Künzle für das Mitwirken beim Digitalisieren der Daten und beim Beurteilen der MINT-Aufgaben. Die solide und pflichtbewusste Hintergrundarbeit war ein bedeutendes Rückgrat dieser Forschungsarbeit.

Prof. Dr. Marianne Müller, Dr. Stella Bollmann und Prof. Dr. Jürg Schwarz danke ich für die Beratung bei statistischen Fragestellungen und die Hilfe bei den Herausforderungen der Statistikprogrammumgebung R.

Ein großer Dank geht an alle Lernenden, Lehrpersonen und Fachdidaktiker:innen, die an den Befragungen teilgenommen haben. Die vielen Begegnungen und das engagierte Mitdenken habe ich in besonderem Maße geschätzt.

Danken möchte ich auch meinen (ehemaligen) Mitdoktorierenden, im Besonderen Daniel Gysin und Dr. Andrea Maria Schmid. Der Austausch sowie das gegenseitige Verständnis für diverse Situationen haben dazu beigetragen, dieses Dissertationsvorhaben voranzutreiben.

Ein weiteres Dankeschön geht an Natalie Ehrenzweig, Dr. Arthur Vayloyan, Michael Lück, meine Lehrerkolleg:innen aus Alpnach und meinen Eltern Donat und Ursula Stuppan. Sie haben sich immer die Zeit genommen, meine Anliegen anzuhören, Textbausteine kritisch gegengelesen und mit wertvollen Anmerkungen kommentiert.

Ein herzlicher Dank geht an meine Schwiegermutter Giuseppa Beyeler-Caduff für die unzähligen abgesprochenen und spontanen Kinderbetreuungsstunden. Als Folge dessen hatte ich viele Freiräume für Sitzungen, Beratungen, Konferenzbesuche und Zeit zum Verfassen der Manuskripte.

Mein innigster Dank geht an meine Frau Sabine Stuppan. Für die stetige Unterstützung, die Begleitung und den Rückhalt. Ohne die Bereitschaft, mir mit zusätzlicher Familienarbeit den Rücken freizuhalten, wäre diese Promotion nicht möglich gewesen.