

# Beurteilungen der Schwierigkeit von Mathematikaufgaben durch Schüler und Schülerinnen

ALEXANDER SALLE, OSNABRÜCK; MIRKO GETZIN, OSNABRÜCK; RUDOLF VOM HOFE, BIELEFELD

**Zusammenfassung:** Die angemessene Beurteilung von Aufgaben stellt eine zentrale Fähigkeit für das Gelingen selbstgesteuerter Lernprozesse bei Lernenden dar. Um herauszufinden, auf welche Weise Schwierigkeitsbeurteilungen von Lernenden ausfallen und welche Aufgabenmerkmale sie für ihre Urteile heranziehen, wurden Bearbeitungen von Blütenaufgaben mit nicht-hierarchisch gekennzeichneten Teilaufgaben analysiert. 1169 Schülerinnen und Schüler der Klassenstufen 5–7 wählten die für sie einfachste und schwierigste Teilaufgabe und begründeten ihre jeweilige Auswahl schriftlich. Das große Spektrum auftretender Beurteilungen zeigt eine hohe Heterogenität der Schwierigkeitsurteile. Die formulierten Begründungen hingegen nehmen aufgabenübergreifend vor allem auf eigene Tätigkeiten (leichteste Teilaufgaben) und mangelndes Lese- und Aufgabenverständnis (schwierigste Teilaufgabe) Bezug, während konkrete Aufgabenmerkmale selten als Begründung genannt werden.

**Abstract:** The appropriate assessment of tasks is a central skill for the success of self-directed learning processes by learners. In order to find out how learners decide about the difficulty of a task and which reasons are relevant for their assessments, written assessments and explanations of 1169 pupils in grades 5–7 were analyzed. “Blütenaufgaben” with their non-hierarchically marked subtasks are a suitable task format for an investigation of such assessments. The students chose the easiest and most difficult subtask and explained their choices. The results show a great variety of different assessments. However, the explanations refer primarily to the learners’ own activities (easiest subtasks) and a lack of reading and task comprehension (most difficult subtask). Specific characteristics of subtasks were only mentioned seldomly.

## 1. Einleitung

Die Berücksichtigung der Heterogenität von Schulklassen bzw. Lerngruppen ist eine zentrale Forderung, welche die Gestaltung des Mathematikunterrichts in den letzten Jahren bestimmt (Gemeinsame Kommission Lehrerbildung der GDM, DMV & MNU, 2017). Als grundlegend hierfür gilt, dass mathematische Lernprozesse individuell erfolgen und auf den individuellen Bedürfnissen der Lernenden fußen (ebd.; Prediger, 2008; Prediger & Scherres, 2012). Im Sinne eines umfassenden Inklusions-

verständnisses für den Mathematikunterricht (Benölken, Dixel & Berlinger, 2018, S. 43 ff.) birgt diese Verschiedenheit der Individuen, insbesondere ihrer Lernausgangslagen mitsamt aller individuellen Bedürfnisse, ein Potenzial, das es durch die konkrete Ausrichtung von Lerngelegenheiten gewinnbringend auszuschöpfen gilt. Durch individuelle Diagnose und Förderung mathematischer Lehr-Lern-Prozesse im gemeinsamen Unterricht können individuelle Potenziale erfasst und fortentwickelt werden sowie Gemeinsamkeit und Zugehörigkeit innerhalb einer Lerngruppe durch den sozialen Austausch der Lernenden geschaffen werden (ebd.; Häsel-Weide & Prediger, 2017). Der Zweck dieses pädagogisch bewussten Umgangs mit Vielfalt im Unterricht ist neben der potenzialorientierten Ausschöpfung individueller Lernausgangslagen auch die fächerübergreifende Lebensvorbereitung durch Demokratisierung und gesellschaftliche Teilhabe (Vock & Gronostaj, 2017, S. 5 ff.), indem Schülerinnen und Schüler zunehmend Verantwortung für ihr eigenes Lernen übernehmen und eigene Entscheidungen im Lernprozess, wie z. B. Aufgabenauswahlen, treffen.

Eine viel genutzte und effektive Möglichkeit zur konkreten Ausgestaltung solch individuellen Lernens in Schulklassen stellt der Einsatz differenzierender Lernumgebungen dar (Leuders & Prediger, 2016). Im Sinne der inneren Differenzierung bzw. Binnendifferenzierung können die Bedürfnisse aller Lernenden innerhalb der Lernumgebungen insbesondere durch eine adressatengerechte Aufgabenauswahl berücksichtigt werden: Schülerinnen und Schüler lösen keine feststehende Menge von Aufgaben, vielmehr bearbeiten sie je nach Befähigung und Bedürfnis individuell unterschiedliche und unterschiedlich viele Aufgaben auf unterschiedliche Weisen. Vor allem binnendifferenzierende Aufgaben ermöglichen Schülerinnen und Schülern, selbständig eine Auswahl von Teilaufgaben zu treffen (ebd., S. 9 und S. 15).

Welche Mathematikaufgaben dabei von welchen Lernenden bearbeitet werden, lässt sich im Wesentlichen auf zwei verschiedene Arten entscheiden (Heymann, 1991, S. 63 ff.): Entweder stellt die Lehrperson Aufgaben zusammen und weist sie den einzelnen Lernenden individuell zu („Paralleldifferenzierung“; Prediger & Scherres, 2012) oder die Lernenden wählen entsprechende Aufgaben aus einer – häufig von der Lehrperson kuratierten – Aufgabensammlung aus (ebd.; Scherres, 2013; Drüke-Noe, 2018).<sup>1</sup> In beiden

Fällen ist die Aufgabenschwierigkeit ein zentrales Merkmal für die Auswahl von Aufgaben zur Bearbeitung durch Schülerinnen und Schüler (Drüke-Noe, 2018; Prediger, 2008).

Die Aufgabenauswahl in der ersten, geschlosseneren Variante beruht als „passgenaue Ausrichtung jeder einzelnen Aufgabe auf einen Schüler“ (Büchter & Leuders, 2016, S. 104) auf der diagnostischen Kompetenz von Lehrerinnen und Lehrern. Wie Lehrkräfte Mathematikaufgaben insbesondere hinsichtlich der Aufgabenschwierigkeit beurteilen, (zu-)ordnen und auswählen, ist seit Jahren wichtiger Bestandteil der Professionalisierung von Mathematiklehrkräften (Büchter & Leuders, 2005) und in den letzten Jahren zunehmend untersucht worden (u. a. Brophy, 2000; Helmke, 1987; Hoffmann & Böhme, 2014; Karst, 2012; Mudiappa & Artelt, 2014; Ostermann, Leuders & Nückles, 2015; Prediger, 2008; Schrader, 2011; Schrader & Artelt, 2009).

In der zweiten, offeneren Variante liegt die Verantwortung für die Angemessenheit der Aufgabenauswahl im Wesentlichen bei den Lernenden. Im Rahmen selbstgesteuerten Lernens<sup>2</sup> ist dabei die realistische Beurteilung von Aufgabenschwierigkeiten und anderen Differenzierungsaspekten, wie Anspruchsniveau, Lerntempo, Zugangsweisen oder sprachliche Anforderungen, insbesondere für eine adäquate Aufgabenauswahl und Organisation der eigenen Lernhandlungen sowie die Beurteilung des persönlichen Lernstands und das Erkennen etwaiger Lücken im notwendigen Vorwissen relevant (Efklides & Vlachopoulos, 2012; Hasselhorn & Gold, 2013, S. 95 ff.; Klein-Landeck, Karau & Landeck, 2016, S. 26). Wie Schülerinnen und Schüler hierbei konkret ein Urteil über die Schwierigkeit von Mathematikaufgaben vor bzw. nach ihrer Bearbeitung fällen, liegt bisher kaum im Fokus der fachdidaktischen Forschung. Auch ist bislang unklar, inwiefern Schülerinnen und Schüler in der Lage sind, Mathematikaufgaben sachgerecht auszuwählen, um dadurch „das [zu] lernen, was für sie relevant ist“ (Leuders & Prediger, 2016, S. 143).

Dieses Wissen ist jedoch mit Blick auf die konkrete Umsetzung eines potenzialorientierten, differenzierenden Fachunterrichts hochrelevant: Unter der Prämisse, dass die individuellen Potenziale gewinnbringend in die jeweiligen Lernprozesse integriert werden sollen, muss gewährleistet sein, dass Lernende in selbstgesteuerten Arbeitsphasen diese Potenziale auch erkennen und Aufgaben dahingehend beurteilen können, welches Potenzial diese Aufgaben zur Erreichung ihrer individuellen Lernziele bergen. Dabei obliegt es den Lernenden, Urteile über Aufgaben zu fällen, welche sie zum Lernen verwenden möchten. Diese Urteile umfassen möglichst realistische

Einschätzungen der Art der Aufgabenanforderungen, der Bearbeitungsdauer bzw. des -umfangs, des notwendigen Vorwissens sowie insbesondere die Beurteilung der subjektiv empfundenen Aufgabenschwierigkeit und Einschätzungen personenspezifischer Merkmale (Getzin, i. V.). Dabei greifen die Schülerinnen und Schüler auf Wissen zurück, welches sie bei der Reflexion von vergangenen Aufgaben, also *nach* deren Bearbeitung, gewonnen haben. Vor allem die subjektiv empfundene Schwierigkeit wird dabei von ihnen besonders häufig zur Beschreibung und Beurteilung von Mathematikaufgaben hinzugezogen (ebd.).<sup>3</sup>

In der vorliegenden Arbeit wird daher untersucht, inwieweit Schülerinnen und Schüler die *Schwierigkeit* von Mathematikaufgaben als bedeutsamen Differenzierungsaspekt (Leuders & Prediger, 2016) subjektiv beurteilen und wie sie ihre Urteile begründen. Dabei soll die Reihung und lexikographische Bezeichnung der Teilaufgaben als beeinflussender Faktor ausgeschaltet werden (Astleitner, 2008, S. 70 ff.; Krippner, 1992; Leuders & Prediger, 2016, S. 132 ff.; Leutner, 2004, S. 289 ff.). Auf diese Weise soll ein Beitrag zu einem wichtigen, bisher wenig betrachteten Baustein selbstgesteuerten Mathematiklernens geleistet werden, der ein tieferes Verständnis von individueller Förderung ermöglicht und praxisrelevante Implikationen für den Mathematikunterricht sowie für das Lernen mit binnendifferenzierenden Aufgaben generiert.

Nach einer Aufarbeitung des theoretischen Rahmens und des Forschungsstandes werden die Forschungsfragen präzisiert und die darauf basierende Anlage der durchgeführten empirischen Studie dargestellt. Im Anschluss daran werden Ergebnisse zu den individuellen Schwierigkeitsurteilen und deren Begründungen vorgestellt sowie abschließend diskutiert.

## 2. Theoretischer Rahmen und Forschungsstand

Aufgaben bilden beim Lernen und Lehren von Mathematik ein zentrales Vermittlungsinstrument. Bruder (2008) versteht unter ihnen eine „Aufforderung zum Lern-Handeln im Mathematikunterricht“ (ebd., S. 19). Aufgaben eröffnen damit einen Zugang, Mathematik differenziert zu unterrichten (u. a. „innere Differenzierung durch Aufgaben“; Prediger, 2008).

Im Sinne eines differenziert gestalteten Unterrichts arbeiten Schülerinnen und Schüler gegebenenfalls an denselben Aufgaben in unterschiedlicher Weise oder aber an verschiedenen Aufgaben, die sie aus einem vorgefertigten Aufgabenpool auswählen können. In beiden Fällen sind es die Schülerinnen und Schüler, welche aufgabenbezogene Entscheidungen treffen.

## 2.1 Aufgabenschwierigkeit und schwierigkeitsbeeinflussende Aufgabenmerkmale

Die Bestimmung einer universellen Aufgabenschwierigkeit ist im Allgemeinen nicht möglich, da hierfür stets auch subjektive Faktoren, wie z. B. Vorkenntnisse zu Inhalt und Format, eine Rolle spielen (Aebli 1980, nach Schukajlow, 2011, S. 46 f.).

Die Schwierigkeit einer Aufgabe kann zum einen *nach der Bearbeitung* durch eine konkrete Personengruppe als empirischer Parameter bestimmt werden (a posteriori), zum anderen durch Analysen *vor einer Bearbeitung* der Aufgabe (a priori). Hier wird nochmals unterschieden, ob eine konkrete Lerngruppe in den Blick genommen oder die Aufgabe ohne Berücksichtigung einer Lerngruppe und allein auf Basis schwierigkeitsbeeinflussender Aufgabenmerkmale beurteilt wird (Drücke-Noe & Siller, 2018; Ostermann, Leuders & Nückles, 2015). Des Weiteren kann von einer Person nach Bearbeitung einer Aufgabe ein subjektiv-naiver Schätzwert für deren Schwierigkeit bestimmt werden (ebd.). Solche Schwierigkeitsurteile erfolgen häufig auf einer erdachten Schwierigkeitsskala „leicht – mittel – schwer“ (Drücke-Noe, 2018) oder auf dem Kontinuum „einfachste – schwierigste“ und erfolgen im Sinne einer Bearbeitungsreflexion (Landmann et al., 2009). Sie bieten damit individuell Anhaltspunkte zur eigenen Lernorganisation und zur Einschätzung zukünftiger Aufgaben (Drücke-Noe, 2018).

Auf Basis konkreter Bearbeitungen durch Lernende (a posteriori) lässt sich die Schwierigkeit einer Aufgabe durch die relative Lösungshäufigkeit, also den Anteil korrekter Lösungen an allen Bearbeitungen, bestimmen (u.a. Astleitner, 2008). Je häufiger die Aufgabe gelöst wird, desto einfacher scheint die Aufgabe für die betrachtete Stichprobe zu sein (Moosbrugger & Kelava, 2012, S. 77 ff.). Liegen Bearbeitungen hinreichend vieler Aufgaben vor, so kann die Aufgabenschwierigkeit mit geeigneten statistischen Modellen, wie z. B. dem Rasch-Modell aus der Item-Response-Theorie (ebd., S. 236 ff.), unter Einbezug von Personenfähigkeiten noch elaborierter festgelegt werden. Die Güte der Vorhersagekraft solcher Schwierigkeitsurteile profitiert daher von einem großen Stichprobenumfang (ebd.).

Vor einer Bearbeitung hingegen (a priori) lässt sich die Schwierigkeit einer Aufgabe mit Hilfe von Aufgabenanalysen beurteilen. Dazu wird bestimmt, welche Ausprägungen schwierigkeitsbeeinflussender Merkmale eine Mathematikaufgabe aufweist (u. a. Prenzel et al., 2002) und zu welchem Einfluss diese Ausprägungen auf die Aufgabenbearbeitungen durch eine (fiktive oder konkret bekannte) Personengruppe führen können. Solche Merkmale können sich zum einen auf die oberflächliche Beschaffenheit der

Aufgabe beziehen, wie zum Beispiel auf das Antwortformat oder die Formulierung des Aufgabentextes (ebd.). Zum anderen können schwierigkeitsbeeinflussende Merkmale inhaltlicher bzw. tiefenstruktureller Art sein; dazu zählen beispielsweise die logische Komplexität intendierter Lösungswege oder deren Anforderungen, die erforderlichen kognitiven Aktivitäten zur Lösung einer Aufgabe, die Offenheit von Ausgangslage, Lösungsweg und Ergebnis u.v.m. (u. a. Büchter & Leuders, 2016; Leuders & Prediger, 2016). Über empirisch ermittelte Effektgrößen der einzelnen Merkmale auf die empirische Aufgabenschwierigkeit kann so durch die Kombination der Merkmale und ihrer Ausprägungen die Schwierigkeit einer Aufgabe für eine Personengruppe geschätzt und reguliert werden (s. Tab. 1).

Aufgabenmerkmale	Referenz
Enthält Grafik	Prenzel et al. (2002)
Enthält bildliche Information	Sekundarstufen
Antwortformat (frei, lang bzw. kurz)	
Enthält langen Aufgabentext	
Verarbeitung von Textinformation	
Logisches Verknüpfen	
Räumliches Modell aufbauen	
Erfordert etwas auszurechnen	
Erfordert divergentes Denken	
Erfordert Terminologisches Wissen	
Faktenwissen notwendig	
Kontraintuitives Wissen notwendig	
Funktionale Zusammenhänge	
Internationale Kompetenzklasse	Neubrand et al. (2002)
Curriculare Wissensstufe	Sekundarstufe
Außermathematischer Kontext	
Umfang der Verarbeitung	
Internationale Kompetenzklasse	
Rein innermathematische Deduktion	
Multiple Lösbarkeit	
Grundvorstellungsintensität	Blum et al. (2004)
	Sekundarstufen
Kognitive Komplexität	Cohors-Fresenborg et al. (2004)
Formelhandhabung	
Formalisierung von Wissen	
	Sekundarstufen
Inhalt	Leuders & Prediger (2016)
Notwendiges Vorwissen	Sekundarstufen
Aktivierte Grundvorstellungen	
Kognitive Aktivitäten	
Sprachliche Komplexität	
Zugangsvielfalt	
Offenheit & Geschlossenheit	
<i>Im Kontext der Arithmetik:</i>	Rathgeb-Schnierer & Rechtsteiner (2018)
Spezifische Zahlenmerkmale	Primarstufe
Spezifische Zahlbeziehungen	
Durchzuführende Rechenoperation	
Anzahl von Rechenschritten	

Tab. 1: Ausgewählte Aufgabenmerkmale mit dem Potenzial zur Schwierigkeitsbeeinflussung.

Erfolgt die a-priori-Schwierigkeitsbeurteilung einer Aufgabe durch eine Lehrperson für eine konkrete Lerngruppe, so können darüber hinaus Personenmerkmale der Lernenden (z. B. Vorwissensstand, Bekanntheitsgrad des Aufgabenformats oder Einstellung zur Mathematik) sowie Gruppenmerkmale (z. B. Zusammenarbeitspräferenzen oder Leistungshierarchien) von der beurteilenden Lehrkraft hinzugezogen werden (Karst, 2012).

Schwierigkeitsbeeinflussende Aufgabenmerkmale können demnach als „Stellschrauben“ angesehen werden, um die Aufgabenschwierigkeit im Sinne der Differenzierung zu variieren (Prediger & Scherres, 2012); dass ein Aufgabenmerkmal die empirische Schwierigkeit einer Aufgabe beeinflusst, ist im Allgemeinen jedoch nicht immer der Fall (Neubrand et al., 2002; Prenzel et al., 2002).

Die Kenntnis solcher Aufgabenmerkmale mit dem Potenzial, die Aufgabenschwierigkeit beeinflussen zu können, spielt für das Finden von Schwierigkeitsurteilen eine große Rolle. In den folgenden Abschnitten wird das für solche Urteile notwendige Wissen und die verschiedenen Perspektiven bei einer Schwierigkeitsbeurteilung beschrieben.

## 2.2 Wissen über Aufgaben

Das Wissen über Aufgaben und ihre Schwierigkeit wird als Bestandteil des deklarativen metakognitiven Wissens verstanden (Flavell, 1984; Kaiser & Kaiser, 2006) und umfasst unter anderem das Wissen um inhalts- sowie strukturspezifische Merkmale einer Mathematikaufgabe, das Wissen um Aufgabentypen sowie das Wissen um Anforderungsniveaus und individuelle Schwierigkeiten einer Aufgabe für Lernende (De Corte et al., 2011; Kaiser & Kaiser, 2006). Des Weiteren umfasst es Wissen über den Kontext, in dem eine Aufgabe gestellt wird, die äußere Erscheinung (Oberflächenstruktur) und die Syntax einer Mathematikaufgabe sowie speziell bei Lehrpersonen jenes Wissen darüber, inwiefern diese Komponenten spezielle Differenzierungsaspekte, wie zum Beispiel ihre Schwierigkeit, beeinflussen (Ostermann, Leuders & Nückles, 2015; Schneider & Artelt, 2010).

Dieses Wissen ist sowohl für Lehrende als auch für Lernende relevant. Für Lehrerinnen und Lehrer ermöglicht es die Antizipation intendierter Bearbeitungsprozesse einer Mathematikaufgabe durch Lernende (ebd.), indem die mathematische Komplexität der Aufgabe, ihre Oberflächenstruktur sowie Wissen über typische Schülerschwierigkeiten, Lern- und Verstehenshürden, die bei der Aufgabebearbeitung auftreten können, und potentielle Alltagseinflüsse auf die Lernenden bei der Aufgabenstellung berücksichtigt werden (ebd., S. 51; Hadjidemetriou & Williams, 2002; Lengnink, 2009).

Auch für Lernende spielen Teile dieses Aufgabewissens eine entscheidende Rolle im Verlauf ihres individuell gestalteten, selbstregulierten Lernprozesses: Bei der *Planung* des Lernprozesses, wenn vor dem Hintergrund gesteckter Ziele vom Lernenden beispielsweise entschieden wird, zuerst eine leichte Aufgabe zu bearbeiten, oder die Entscheidung darüber, Textaufgaben in den Hausaufgaben erst nach allen anderen Aufgaben zu bewältigen; bei der *Suche nach Bearbeitungsstrategien*, indem eine Aufgabe mit bekannten Aufgabenformaten verglichen wird (Aßmus & Förster, 2013); bei der *Reflexion* einer Aufgabebearbeitung, wenn Aufgabentypen anhand ihrer auftretenden Merkmale kategorisiert werden und entsprechende Strategien für zukünftige Aufgabebearbeitungen bereitgestellt werden (Getzin, i. V.; Landmann et al., 2009).

Das Wissen über Mathematikaufgaben ist somit zentral für die Beurteilung von Aufgabenschwierigkeiten. Auf welche Weise Schwierigkeiten von Aufgaben aus der Perspektive von Lehrenden oder Lernenden beurteilt werden und welche Urteile im Detail dabei zu fällen sind, wird im Folgenden näher betrachtet.

## 2.3 Beurteilung von Aufgabenschwierigkeiten durch Lehrerinnen und Lehrer

Um gestuft differenzierte Aufgabenangebote oder binnendifferenzierende Materialien entlang der Aufgabenschwierigkeit zu entwickeln oder auszuwählen, ist neben der Analyse der Lerngruppe eine adäquate Beurteilung von Aufgabenschwierigkeiten zentral (Karst, 2012, S. 87 ff.; Ostermann, Leuders & Nückles, 2015, S. 48). Für eine angemessene Schwierigkeitsbeurteilung von Aufgaben ist es aus der Perspektive von Lehrenden unabhängig von der Schulstufe relevant, eine Vielzahl schwierigkeitsgenerierender Aufgabenmerkmale zu kennen und diese variieren zu können (Prediger, 2008; Rathgeb-Schnierer & Rechtsteiner, 2018). Hierbei ist nicht nur die Kenntnis der Aufgabenmerkmale an sich bedeutsam (deklarativer Aspekt des Wissens), sondern auch die Kenntnis über verschiedene Ausprägungen eines Merkmals und deren Einflüsse auf bspw. die Schwierigkeit (prozeduraler Aspekt des Aufgabewissens). Die Beurteilung von Aufgaben und ihren spezifischen Anforderungen stellt damit neben der Beurteilung von Personenmerkmalen eine Facette fachbezogener diagnostischer Kompetenz von Lehrenden dar (Schrader, 2009; Südkamp, Möller & Pohlmann, 2008) und bedarf komplexer Entscheidungen zur Gestaltung von Lehr-Lernprozessen (Ostermann, Leuders & Nückles, 2015, S. 49).

In empirischen Untersuchungen beurteilten Lehrpersonen die Schwierigkeit von Aufgaben für Lerngruppen oftmals ungenau (a priori), d. h. abweichend von

einer empirisch ermittelten a posteriori Aufgabenschwierigkeit (ebd.; Nathan & Koedinger 2000; Spinath, 2005; Thonhauser et al., 2003). So zeigt Spinath (2005) auf, dass diagnostische Fehlurteile von Seiten der Grundschullehrkräfte auf einer falschen Einschätzung von leistungsrelevanten Schülermerkmalen fußen können. Neben der Rolle von solchen Personenmerkmalen, dem Fachwissen<sup>4</sup> und Unterrichtserfahrungen<sup>5</sup> beeinflussen aber auch speziell Aufgabenmerkmale diese ungenauen Urteilsfindungen: Für den Inhaltsbereich Funktionen zeigen Ostermann, Leuders und Nückles (2015) im Sekundarstufenkontext, dass Aufgaben mit bildlichen Informationen (Funktionsgraphen) in ihrer Schwierigkeit stärker unterschätzt werden als solche Aufgaben mit tabellarischer Darstellung.

Auch Thonhauser et al. (2003) stellen sich die Frage, wie Lehrerinnen und Lehrer aufgabenbezogen diagnostische Urteile bilden, und stellen fest, dass Lehrende der Klassenstufe 8 vorrangig oberflächliche Aufgabenmerkmale zur Urteilsbegründung hinzuziehen, wie zum Beispiel die Textlänge, die Verwendung von Grafiken, spezielle Zahlenmerkmale in Rechenaufgaben und den generellen Inhaltsbereich, in dem die Aufgabe zu verorten ist. Die Tiefenstruktur einer Aufgabe, die maßgeblich ihre kognitive Komplexität bestimmt, bleibt dabei weitestgehend außer Acht (ebd.). Diese Merkmale spielen lediglich dann eine Rolle, wenn die Schwierigkeitsbeurteilung anhand eines exemplarischen Lösungsweges erfolgt. McElvany et al. (2009) bestätigen diese Befunde in ihrer empirischen Studie zur aufgabenspezifischen Schwierigkeitsbeurteilung von Lehrerinnen und Lehrern, wobei sie untersuchten, wie Lehrende vor allem die oberflächliche Beschaffenheit des Lernmediums (Text-Bild-Verbindungen) zur Schwierigkeitsbeurteilung hinzuziehen.

Nathan und Koedinger (2000) zeigen darüber hinaus, dass sich insbesondere Eigenschaften der Aufgabentextformulierung (formal-abstrakte vs. alltägliche Sprache) auf die ungenauen Schwierigkeitsurteile von Lehrpersonen auswirken. Die Genauigkeit von Schwierigkeitsurteilen ist jedoch trainierbar: Die Sensibilisierung von Lehrenden für den Einfluss schwierigkeitsgenerierender Aufgabenmerkmale führt im Inhaltsbereich Funktionen insgesamt zu einer passgenaueren Beurteilung von Aufgabenschwierigkeiten (Ostermann, Leuders & Nückles, 2015). Während die Schwierigkeitsurteile von Lehrpersonen in den letzten Jahren intensiver beforscht wurden, liegen für Urteile von Schülerinnen und Schülern nur wenige belastbare Ergebnisse vor.

## 2.4 Beurteilung von Aufgabenschwierigkeiten durch Schülerinnen und Schüler

Welche Aufgabenmerkmale Lernende wahrnehmen und welche aus ihrer Perspektive tatsächlich schwierigkeitsbeeinflussend sind, wurde vor allem im Rahmen des *flexiblen Rechnens*, also dem situationsbedingten, individuellen Reagieren von Lernenden auf Aufgabenmerkmale in der Arithmetik der Primarstufe untersucht. Dabei wurden mathematisch-inhaltliche Merkmale nachgewiesen, die Schülerinnen und Schüler der Primarstufe bei der Schwierigkeitsbeurteilung verwenden (Rathgeb-Schnierer & Rechtsteiner, 2018; Selter, 2000): Eigenschaften der Zahlen und deren Beziehung untereinander, Art und Anzahl der durchzuführenden Rechenoperationen, Zahlwahrnehmung im Lösungskontext und Bekanntheit des Aufgabentyps. Über die Arithmetik hinaus wird in Bezug auf den Begriff des *algebraischen Struktur-sinns* (u. a. Janßen, 2016) im Sekundarstufenkontext beforscht, welche Aufgabenmerkmale von Lernenden bei Term- und Gleichungsumformungen wahrgenommen und zielorientiert genutzt werden. Inwieweit diese und andere mathematisch-inhaltliche Aufgabenmerkmale zu Schwierigkeitsbeurteilungen beitragen, ist bisher nur wenig erforscht (*flexibles algebraisches Handeln*; Block, 2014, S. 199). Auch ist weitestgehend unklar, welche inhaltlichen Merkmale Einfluss auf die subjektive Schwierigkeitsbeurteilung von Aufgaben anderer Inhaltsbereiche in der Sekundarstufe nehmen.

Für die subjektive Beurteilung von Aufgabenschwierigkeiten spielt die Vertrautheit bzw. Bekanntheit von Aufgabenformaten eine große Rolle. Aßmus, Fritzlär und Förster (2018) beforschen hierzu, inwiefern Schülerinnen und Schüler der Primarstufe Analogien zwischen Aufgaben erkennen und diese als Strategie der Schwierigkeitsverminderung verwenden. Erste Ergebnisse zeigen, dass mathematisch begabte Grundschulkindern Analogien zwischen Problemen erkennen und bei Problembearbeitungen davon profitieren, indem sie Analogien als heuristische Strategie verwenden (ebd.). Inwieweit dies in der Sekundarstufe trägt, ist bislang unklar.

Einen ähnlichen Einbezug bekannter Zusammenhänge beim Verstehen einer (vorrangig textbasierten) Aufgabenstellung fasst Reusser (1997) theoretisch im Situationsmodell zusammen. Demnach vergegenwärtigen sich Lernende beim Lesen einer Aufgabe deren Sachstruktur und erzeugen dabei eine mentale Repräsentation der entsprechenden Situation – das so genannte (mentale) Situationsmodell. Die hierdurch erzeugte Interpretation der Aufgabengestalt basiert insbesondere auf den individuellen Erfahrungen der aufgabenbearbeitenden Person sowie ihrem Vorwissen und offenbart schließlich die in der Aufgabe zu

erarbeitende Informationslücke. Zur Beurteilung der Schwierigkeit von Aufgaben *vor ihrer Bearbeitung* wird somit also auf die Erfahrung der Schwierigkeitsbeurteilung vergangener Aufgaben *nach ihren Bearbeitungen* zurückgegriffen, indem sie zur Interpretation der Aufgabengestalt hinzugezogen werden. Beim Identifizieren der Informationslücke im Situationsmodell offenbaren sich den Lernenden zudem die aufgabenspezifischen Anforderungen, die zum Schließen dieser Lücke durch die Aufgabenbearbeitung erforderlich sind (ebd.). Diese aufgedeckten Anforderungen bieten u. a. Anhaltspunkte zur konkreten Schwierigkeitsbeurteilung (Getzin, i. V.).

Die subjektive Schwierigkeitsbeurteilung von Aufgaben kann auch von anderen dispositionalen Einflüssen im Aufgabenbearbeitungsprozess beeinflusst werden (ebd.), wie z. B. von dem mathematikspezifischen Selbstkonzept sowie der Selbstwirksamkeitserwartung. Als Teil des akademischen Selbstkonzeptes bildet das mathematische Selbstkonzept die manifestierte Vorstellung einer Person über sich selbst (Pekrun & Zirngibl, 2004, S. 192) bzgl. mathematikbezogener Überzeugungen (affektiv) und der eigenen mathematischen Kompetenz (kognitiv-evaluativ) (Bong & Skaalvik, 2003; Möller & Trautwein, 2015). Für Prozesse der Schwierigkeitsbeurteilung von Mathematikaufgaben bedeutet dies, dass sich das mathematische Selbstkonzept in der generalisierten selbstbezogenen Wahrnehmung der beurteilenden Person und den daraus resultierenden personenspezifischen Annahmen äußert (Schwanzer et al., 2005, S. 183), z. B. indem Lernende mit hohem kognitiv-evaluativen mathematischen Selbstkonzept tendenziell dazu neigen könnten, Aufgaben als leicht zu empfinden, weil sie über das notwendige mathematisch-inhaltliche Vorwissen verfügen und reflektieren können. Belastbare empirische Befunde über solche Zusammenhänge sind nicht bekannt. Getzin (i. V.) konnte feststellen, dass die subjektiv empfundene Aufgabenschwierigkeit von Schülerinnen und Schülern vor der Aufgabenbearbeitung häufig im Zuge eines vorgestellten Bearbeitungsprozesses festgemacht wird, wobei sie sich die Frage stellen, ob sie generell in der Lage dazu wären, eine korrekte Lösung der Mathematikaufgabe anzufertigen. Dies deckt sich mit dem theoretischen Konstrukt der Selbstwirksamkeitserwartung als „subjektive Gewissheit, eine neue oder schwierige Aufgabe auch [unter Widerständen] erfolgreich zu bearbeiten“ (Schmitz & Schwarzer, 2000, S. 12), welche im Gegensatz zum Selbstkonzept an konkrete Handlungen, Probleme bzw. Aufgaben gekoppelt ist und über die reine Selbstwahrnehmung hinaus geht (Pekrun & Zirngibl, 2004, S. 193). Die Erfassung dieser Widerstände innerhalb einer Mathematikaufgabe erfolgt also individuell, so dass verschiedene Schülerinnen und Schüler bei derselben

Aufgabe nicht nur einen unterschiedlichen Schwierigkeitsgrad empfinden können, sondern auch ganz konkret unterschiedliche Merkmale der Aufgabe als „Widerstand“ wahrnehmen können.

Insgesamt zeigt sich, dass Faktoren, die den Schwierigkeitsgrad einer Mathematikaufgabe bestimmen, von Schülerinnen und Schülern subjektiv und unterschiedlich wahrgenommen werden (Krauthausen & Scherer, 2007). Insbesondere kann diese subjektive Beurteilung von jener Schwierigkeitsbeurteilung der Lehrenden abweichen (ebd.).

Ein zusätzlich wichtiger Aspekt der Aufgabenpräsentation, welcher die Schwierigkeitsbeurteilung durch Lernende beeinflusst, ist die hierarchische Auszeichnung einer Aufgabe und deren Position innerhalb einer Aufgabenserie (Astleitner, 2008; Leutner, 2004).

## 2.5 Reihung und Indizierung von Aufgaben

Die Reihung von Aufgaben kann auf Basis verschiedener Differenzierungsaspekte, wie bspw. den sprachlichen Anforderungen oder dem Einsatz von Darstellungen, vorgenommen werden (Leuders & Prediger, 2016). Die a priori anhand von Aufgaben- und Personenmerkmalen beurteilte Aufgabenschwierigkeit stellt bei Mathematikaufgaben den am häufigsten eingesetzten Differenzierungsaspekt dar, um Aufgabenreihungen vorzunehmen, zu variieren und diese dem Lernstand von Schülerinnen und Schülern anzupassen (Büchter & Leuders, 2016; Hußmann & Prediger, 2007; Klein-Landeck, Karau & Landeck, 2016; Prediger, 2008). Die Aufgabenreihung mit stetig ansteigendem Schwierigkeitsgrad gilt dabei in der disziplinübergreifenden Literatur generell als lernförderlich (Astleitner, 2008; Leutner, 2004; Nußbaum & Leutner, 1986a; 1986b; Thonhauser, 2008). Dies wurde für unterschiedliche Unterrichtsbedingungen empirisch bestätigt (ebd.). Darüber hinaus zeigen Aufgabensequenzen mit stark wechselnder Veränderung der Schwierigkeit kognitiv entlastende Wirkungen auf Lernende (*Task-Switching-Effect*; Burghardt & Swan, 2013; Gilbert & Shallice, 2002).

Die Reihung von Aufgaben nach bestimmten Differenzierungsaspekten gibt Schülerinnen und Schülern einerseits Orientierungsmöglichkeiten, um Aufgaben im Rahmen des selbstregulierten Lernprozesses einschätzen und auswählen zu können (Landmann et al., 2009; Leuders & Prediger, 2016). Andererseits wird durch das häufige Auftreten von Aufgabenreihungen nach aufsteigendem Schwierigkeitsgrad in der Unterrichtspraxis schließlich auch die Wahrnehmung von Aufgabengruppen durch Lernende beeinflusst: So nehmen Lernende letzt dargestellte Aufgaben häufig automatisch als schwierigste Aufgabe wahr (*Sequential-Difficulty-Effect*; Goldstein, 2008; Lannie & Martens, 2004). Auch gibt die Studie von Lannie und

Martens (2004) Hinweise darauf, dass sich die Wahrnehmung einer Aufgabe als besonders schwierig negativ auf die Bearbeitungsbereitschaft und die Lernwirksamkeit von Schülerinnen und Schülern bzw. Studierenden auswirkt.

Während die beschriebenen Studien sich auf klassische Kennzeichnungen wie a), b), c) etc. und der damit implizierten Reihenfolge von Mathematikaufgaben beziehen, werden für den Mathematikunterricht weitere alternative Varianten vorgeschlagen, die jeweils bestimmte Aspekte der Organisation des Lernens aufgreifen. Im Rahmen differenzierender Lernumgebungen werden beispielsweise Schwierigkeitsgrade von Aufgaben ausgezeichnet. Dies kann durch Sternchen (\* = leicht, \*\* = mittel, \*\*\* = schwer) bzw. Punkte für visualisierte Schwierigkeitsstufen geschehen (Klein-Landeck, Karau & Landeck, 2016) oder durch eine Einteilung in Wahl- und Pflichtaufgaben, in denen Mindestanforderungsniveaus und erhöhte Anforderungsniveaus abgesteckt werden (ebd.; Bruder, 2008; Leuders & Prediger, 2016). Auch ist eine Kennzeichnung von Mathematikaufgaben durch konkrete Anforderungen, Heuristiken oder Tätigkeitsprofile denkbar (Bruder & Collet, 2011). Dies soll Schülerinnen und Schülern ermöglichen, Aufgaben bezüglich der Passung ihrer individuellen Bedürfnisse im Rahmen selbstregulierter Lernprozesse einschätzen zu können (Landmann et al., 2009; Leuders & Prediger, 2016).

Eine noch freiere Kennzeichnung ohne erkennbare Schwierigkeitsbewertung und Reihenfolge wird in den Blütenaufgaben des Projekts SINUS.NRW vorgenommen (Salle, 2015; 2016): Aufbauend auf dem im Projekt MABIKOM („Mathematische binnendifferenzierende Kompetenzentwicklung in einem mit neuen Technologien unterstützten Mathematikunterricht“) entwickelten Format der Blütenaufgabe (Bruder, Reibold & Wehrse, 2013) werden bei den Sinus-Blütenaufgaben ausgehend von einem Sachkontext vier unabhängige Teilaufgaben formuliert. Diese Teilaufgaben weisen keine Kennzeichnungen von Schwierigkeiten o.ä. auf, sondern werden mit Spielkartensymbolen referenziert und in einer unspezifischen Reihenfolge dargestellt. Damit sind die Orientierungsmöglichkeiten verringert und eigenständige aufgabenbezogene Urteile von den bearbeitenden Lernenden notwendig. Erste Ergebnisse geben Hinweise auf die Vielfalt der Schwierigkeitsurteile und die Unterschiedlichkeit der jeweiligen Begründungen dieser Urteile (Salle, vom Hofe & Pallack, 2014).

### 3. Forschungsfragen

Das vorige Kapitel zeigt, dass im Rahmen der Professionsforschung von Lehrerinnen und Lehrern bereits diverse Studien zu *Schwierigkeitsbeurteilungen*

*durch Lehrende* (für Lerngruppen) vorliegen. Da die Schwierigkeit von Aufgaben ein wesentlicher Differenzierungsaspekt für den Mathematikunterricht ist, ist es jedoch auch für Schülerinnen und Schüler relevant, Aufgabenschwierigkeiten subjektiv und im Rahmen ihres Lernprozesses adäquat beurteilen zu können. Von Seiten der *Lernenden* ist jedoch nur wenig über aufgabenbezogene Urteilsprozesse bekannt, insbesondere im Bereich der Sekundarstufen.

Bei den Studien verschiedener Disziplinen mit einem Fokus auf die Lernenden wurde zudem oftmals der Einfluss der Aufgabenreihung auf aufgabenbezogene Urteile bzw. Lernprozesse untersucht (vgl. Abschnitt 2.5), welcher – wie zuvor berichtet – die subjektive Beurteilung von Aufgabenschwierigkeiten beeinflusst. Im Fokus dieser Studien stand zumeist ein natives Schwierigkeitsverständnis von Aufgaben wie in Abschnitt 2.1 beschrieben. Wie uniform oder unterschiedlich Lernende Mathematikaufgaben in Bezug auf ihre Schwierigkeit beurteilen, falls keine Orientierung durch eine Reihung oder Indizierung vorgegeben ist, ist jedoch unklar. Welche Gründe Lernende für diese subjektiven Schwierigkeitsurteile anführen und welche Eigenschaften einer Aufgabe sie als leicht oder schwierig wahrnehmen bzw. bezeichnen, ist ebenfalls weitestgehend ungeklärt.

Die Kenntnis von solchen Prozessen der Urteilsfindung und deren Begründungen kann einen wichtigen Beitrag zum Wissen über selbstgesteuerte Lernprozesse im Mathematikunterricht leisten. Zu wissen, wie und mit welchem Fokus Schülerinnen und Schüler den Schwierigkeitsgrad von Mathematikaufgaben beurteilen, ist zudem zentral für die Gestaltung differenzierender Lernumgebungen. Die Berücksichtigung des Wissens über Aufgaben und Gründe für deren Schwierigkeit würde die Auswahl des Aufgabenpools in sachgerechter Weise ermöglichen.

In diesem Beitrag wird daher die schülerinnen- und schülerseitige Schwierigkeitsbeurteilung von Mathematikaufgaben untersucht, bei denen die Teilaufgaben weder lexikographisch noch anderweitig hierarchisch gekennzeichnet sind. Im Fokus steht dabei das in Abschnitt 2.1 umschriebene subjektiv-naive Verständnis der Aufgabenschwierigkeit und keine quantitative Schätzung des empirischen Parameters der Aufgabenschwierigkeit. Die subjektive Einschätzung der Aufgabenschwierigkeiten soll zudem *nach* der Bearbeitung der Aufgaben vorgenommen werden. Aufgrund der Kenntnis aller Teilaufgaben und möglicher Bearbeitungsweisen können die Lernenden einerseits eine möglichst hohe Zahl dieser Aufgaben vergleichend beurteilen und haben darüber hinaus andererseits detailliertere Kenntnisse über die einzelnen Anforderungen.

Ziel ist es, auf der Basis einer breiten Stichprobe explorativ erste Kenntnisse über das Beurteilungsverhalten von Schülerinnen und Schülern bei einer ausgewählten Sammlung nicht-gereihter Aufgaben in einer typischen Unterrichtsumgebung zu erhalten. Zudem sollen die Begründungen für die Schwierigkeitsbeurteilungen näher untersucht werden, um Erkenntnisse darüber zu gewinnen, weshalb Schülerinnen und Schüler Aufgaben als schwierig oder leicht empfinden. Auf Basis der empirischen Daten sollen dann Muster beschrieben werden, um so explorativ Kenntnisse über das Beurteilungs- und Begründungsverhalten von Lernenden zu generieren, die als Grundlage für Folgestudien dienen können.

Liegt einem Schüler oder einer Schülerin eine Mathematikaufgabe mit mehreren Teilaufgaben vor, so wird in der Folge das Paar „Leichteste Teilaufgabe“ und „Schwierigste Teilaufgabe“ als *individuelles Schwierigkeitsurteil* eines Schülers bzw. einer Schülerin bezeichnet.

Der Beitrag geht somit zwei Forschungsfragen nach:

- 1) Wie vielfältig sind die individuellen subjektiven Schwierigkeitsurteile von Schülerinnen und Schülern in Bezug auf unabhängige, nicht-hierarchisch gekennzeichnete Teilaufgaben?
- 2) Wie begründen die Schülerinnen und Schüler ihre individuellen subjektiven Schwierigkeitsurteile einer Teilaufgabe als leichteste oder schwierigste?

## 4. Anlage der Studie

### 4.1 Design

Da nur wenige Kenntnisse zu schülerinnen- und schülerseitigen Beurteilungen von Mathematikaufgaben vorliegen, ist ein *induktives, hypothesengenerierendes Vorgehen* angezeigt; ein quantitativ-schließendes Vorgehen liefe hier Gefahr, die Vielfalt der Beurteilungs- und Begründungsprozesse zu wenig zu berücksichtigen (Lamnek & Krell, 2016). Um eine größtmögliche Offenheit gegenüber dem Forschungsfeld zu bewahren, wird ein großer und vielfältiger Datensatz ausgewertet, in dem Schülerinnen und Schüler verschiedener Schulformen, verschiedener Schulstufen und Bearbeitungen unterschiedlicher Aufgaben einbezogen sind (ebd.).

Für einen möglichst umfassenden Einblick in die Beurteilungsprozesse wurde zudem ein Erhebungsdesign mit Einzelarbeits- und Gruppenarbeitsphase gewählt. In der Einzelarbeit haben Schülerinnen und Schüler die Möglichkeit, sich mit den Teilaufgaben allein auseinanderzusetzen (eigenständige Teilaufgabenauswahl). In der darauffolgenden Arbeitsphase sollten die Schülerinnen und Schüler in einer gemeinsamen Besprechung auch mit jenen Teilaufgaben in

Kontakt kommen, die sie in der Einzelarbeitsphase nicht bearbeiteten. Anschließend füllten alle Probandinnen und Probanden einen Fragebogen (s. Abschnitt 4.2) aus. Um möglichst vielfältige Informationen über die Beurteilungs- und Begründungsprozesse zu erhalten, werden offene Frageformate eingesetzt; insbesondere werden keine Antwortmöglichkeiten vorgeben, um die Informationsbereitschaft der Probandinnen und Probanden nicht negativ zu beeinflussen (Lamnek & Krell, 2016).

Der Tatsache, dass die Schwierigkeitsurteile *nach* der Bearbeitung erhoben wurden, wird in der Auswertung Rechnung getragen, indem u. a. berücksichtigt wird, ob Schülerinnen und Schüler Teilaufgaben bearbeitet haben oder nicht.

### 4.2 Fragebogen

Die Auswertungen des vorliegenden Beitrags basieren auf den Items 2 und 3 des eingesetzten Fragebogens, in denen die Schülerinnen und Schüler die leichteste bzw. schwierigste Teilaufgabe angeben und eine Begründung anfügen (s. Abb. 1). In die Lücken wird dabei das jeweilige Aufgabensymbol eingetragen (subjektiv-naives Schwierigkeitsverständnis; geschlossenes Format; Moosbrugger & Kelava, 2012, S. 43 ff.). Im Freitextfeld werden die Begründungen angefügt (offenes Freitextfeld; ebd., S. 40 f.). Aufgrund des Einsatzes in verschiedenen Schulstufen wurde berücksichtigt, das Antwortformat möglichst wenig zu variieren, um vor allem jüngeren Schülerinnen und Schülern den Zugang zum Testinstrument zu erleichtern. Nach der Gruppenarbeitsphase wurde eine orientierende Instruktion zur Bearbeitung des Fragebogens durchgeführt.

Da die Studie Teil eines größeren Forschungsvorhabens zum unterrichtlichen Einsatz von Blütenaufgaben ist, enthielt der eingesetzte Fragebogen weitere Items, die zur Hälfte im Freitextformat und zur Hälfte zum Ankreuzen gestaltet wurden. Die Items beziehen sich nicht auf die hier thematisierten Forschungsfragen, sondern fokussieren u. a. die Zustimmung zum Format der Aufgaben, die Bewertung der Durchführung sowie der Gruppenarbeitsphase, die eigene Bearbeitungsreihenfolge und die Bewertung des generellen Aufgabenformats. Die Fragebogenkonstruktion ist im Sinne der intuitiven Konstruktionsstrategie (Moosbrugger & Kelava, 2012, S. 36) erfolgt.

2. Die Teilaufgabe \_\_\_\_\_ war für mich am einfachsten, weil: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

3. Die Teilaufgabe \_\_\_\_\_ war für mich am schwierigsten, weil: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Abb. 1: Eingesetzte Items.

### 4.3 Stichprobe

Insgesamt nahmen 1528 Schülerinnen und Schüler von fünf Schulen in Nordrhein-Westfalen (zwei Gymnasien, zwei Gesamtschulen und eine Realschule) an der Studie teil. Alle letztendlich ausgewählten Schulen partizipierten an einem SINUS.NRW-Projekt. Die Erhebung wurde im regulären Schulunterricht der fünf Schulen realisiert, um eine hohe Authentizität und Vielfalt der Daten für das induktive Vorgehen bei der Hypothesengenerierung zu gewährleisten (Lamnek & Krell, 2016).

Ein Datensatz gilt als vollständig, wenn sowohl eine eindeutige *Festlegung für die einfachste* als auch für *die schwierigste Teilaufgabe* getroffen wurde und damit ein gültiges individuelles Schwierigkeitsurteil vorliegt. Dementsprechend werden alle Analysen des Beitrags auf Basis von 1169 Schülerinnen- und Schülerbearbeitungen vorgenommen. Diese umfassen Schülerinnen und Schüler aus der Jahrgangsstufe 5 ( $n_5 = 335$ ), Jahrgangsstufe 6 ( $n_6 = 452$ ) und Jahrgangsstufe 7 ( $n_7 = 382$ ). Insgesamt wurden somit 359 Bearbeitungen aufgrund unvollständiger Datensätze ausgeschlossen. Eine Auswertung der unvollständigen Datensätze ergab keinerlei Hinweise darauf, dass die Verteilung der Ausschlüsse auf die Jahrgangsstufen noch auf die Aufgaben zu systematischen Verzerrungen der Ergebnisse führen.

### 4.4 Eingesetzte Aufgaben

Alle eingesetzten Aufgaben haben dasselbe Format mit einem zentralen Sachkontext und je vier voneinander unabhängigen Teilaufgaben (Salle, 2016; s. auch Abb. 2). Jede Teilaufgabe orientiert sich dabei an einer von vier unterschiedlichen Charakteristiken: Vorwärtsarbeiten, Rückwärtsarbeiten, komplexe Teilaufgabe und offene Teilaufgabe. Komplexe Teilaufgaben können beispielsweise Begründungsaufgaben, Aufgaben zu Darstellungswechseln, etc. sein. Die Teilaufgaben sind nicht mit Buchstaben oder Zahlen gekennzeichnet, sondern durch die Spielkartensymbole Kreuz, Pik, Herz und Karo. Die Zuordnung der Symbole zu bestimmten Charakteristiken sowie die Anordnung der jeweiligen Teilaufgaben sind dabei in jeder Blütenaufgabe willkürlich erfolgt.

Da die Blütenaufgaben thematisch in den jeweils aktuellen Schulunterricht eingebunden werden sollten, wurden bei den Fachlehrerinnen und Fachlehrern für jeden Jahrgang die aktuellen Themen des Halbjahres erfragt. Pro Jahrgang wurden drei passende thematische Aufgaben formuliert, aus denen jede Lehrkraft

genau eine Aufgabe für den Einsatz in der eigenen Klasse auswählen konnte. Insgesamt wurden sieben von neun möglichen Aufgaben eingesetzt (s. Tab. 2).

Die eingesetzten Aufgaben behandeln folgende Themen: 5.1 „Fahrradüberprüfung“ (s. Abschnitt 4.5); 5.2 „Nordseeurlaub“: Bestimmung von Kosten und Kostenteilung für einen Urlaub von zwei Familien – Grundrechenarten mit ganzzahligen Geldbeträgen; 5.3 „Wiegeaufgabe“: Bestimmung von Zielgewichten und fehlenden Gewichten auf Basis eines vorhandenen Wägesatzes – Grundrechenarten mit ganzzahligen Gewichten; 6.1 „Erdbeermilchshake“: Bestimmung von Zutatenmengen sowie deren Anteile an einem Liter, Bestimmung aufgeteilter Flüssigkeitsmengen auf Basis eines gegebenen Rezeptes sowie Formulierung eines eigenen Rezeptes – Elementare Bruchrechnung mit Flüssigkeitsmengen in Litern (s. Anhang); 6.2 „Schatzsuche“ Eintragen, Ablesen und Beschreiben von Routen und Orten auf einer Inselkarte – Grundlegende Orientierung im Koordinatensystem; 7.1 „Klassenausflug“: Berechnung von Eintrittspreis und Aufwendungen für Zubehör für den Kegelbahnbesuch einer ganzen Klasse – Einfache Proportionalität und Grundrechnen mit Dezimalbrüchen bei Kostenbestimmung; 7.2 „Geländespiel“: Weg-Zeit-Protokolle konkurrierender Gruppen in einem Geländespiel – Eigenschaften funktionaler Zusammenhänge zwischen Weg und Zeit sowie Nutzung grafischer und sprachlicher Darstellungen solcher Zusammenhänge.

Aufgrund der Unterschiedlichkeit der eingesetzten Blütenaufgaben und ihrer Teilaufgaben wird kein aufgabenübergreifender Vergleich anhand ihrer Charakteristiken vorgenommen. Die Darstellung der Teilaufgaben im Beitrag wird ebenfalls durch Spielkartensymbole vorgenommen.

### 4.5 Analyse der eingesetzten Blütenaufgabe „Fahrradüberprüfung“

Im Folgenden wird exemplarisch die Blütenaufgabe 5.1 „Fahrradüberprüfung“ dargestellt, um den eingesetzten Aufgabentyp in seinen Details abzubilden (s. Abb. 2). Die Aufgabe hat als zentralen Kontext die Ergebnisse einer Sicherheitsüberprüfung von Fahrrädern. Die Fahrräder gehören den Schülerinnen und Schülern einer Klasse. Alle beanstandeten Mängel sind in einem Balkendiagramm aufgeführt. Inhaltlich fokussiert diese Blütenaufgabe damit auf das Ablesen von Häufigkeiten aus einem Balkendiagramm.

Aufgabe (Jahrgangsstufe.Index)	5.1	5.2	5.3	6.1	6.2	7.1	7.2
Anzahl Klassen	3	6	6	8	14	4	20
Anzahl Schülerinnen und Schüler	55	134	146	176	276	77	305

Tab. 2: Zuordnung der untersuchten Klassen zu den eingesetzten Aufgaben sowie Anzahl von in die Studie einbezogenen individuellen Schwierigkeitsurteilen und vollständigen Begründungen.

Der gemeinsame Sachkontext ist durch die Thematisierung von Fahrrädern alltagsnah für Fünftklässlerinnen und Fünftklässler, was den Lernenden einen leichteren Zugang zur Aufgabe ermöglichen könnte. Ausgehend von dem zentralen Sachkontext weist die Aufgabe vier unabhängige Teilaufgaben verschiedener Charakteristiken auf. Aufgrund ihrer thematischen Verwandtschaft ist die curriculare Wissensstufe aller Teilaufgaben auf demselben Niveau zu verorten.

Als Teilaufgabe zur Charakteristik „Vorwärtsarbeiten“ wurde Herz formuliert. Nach der Auswahl des richtigen Balkens „Beleuchtung“ muss die Anzahl der Beanstandungen abgelesen werden. Hier wird von der Beanstandungskategorie auf die Ausprägung geschlossen. Den Kern des zu leistenden Transfers stellt hier der Zusammenhang der Achsenbeschriftungen „Beanstandungen“ und „Anzahl“ dar. Insgesamt ist der Aufgabentext eher kurz und die Aufgabe ist im Frageformat formuliert. Die verwendeten Wörter sind weitestgehend alltagsgebräuchlich, lediglich „beanstandet“ könnte in der Zielgruppe ein noch unbekanntes, jedoch im Kontext erschließbares Wort sein. Zudem werden keine alternativen Darstellungsformen für Daten sowie keine Zusatzinformationen im Aufgabentext dargeboten. Dies weist zusätzlich auf eine geringe kognitive Komplexität hin. Insbesondere sind nur wenige Bearbeitungsschritte bis zur Lösung nötig.

Für die Charakteristik „Rückwärtsarbeiten“ wurde die Teilaufgabe Pik formuliert, in der andersherum

von der maximalen sowie minimalen Ausprägung im Diagramm auf die Beanstandung (die Kategorie) geschlossen werden muss. Gegenüber der Teilaufgabe Herz sind damit die Rollen von Gegebenheiten und Ziel der Teilaufgabe vertauscht. Im Vergleich zu Herz sind bei dieser Teilaufgabe außerdem mehrere Ablesehandlungen vorzunehmen, was die kognitive Komplexität erhöht. Durch den implizit geforderten Vergleich aller abzulesenden Zahlen, um das Minimum bzw. das Maximum zu ermitteln, erscheint die kognitive Komplexität im Vergleich zu Herz höher. Zwar ist auch bei dieser Teilaufgabe der Aufgabentext eher kurz und die Aufgabenstellung nicht operationalisiert, sondern im Frageformat gestellt. Die sprachlogische Komplexität ist hier jedoch höher als in Herz, denn eine Schwierigkeit stellt hier die Verknüpfung der synonym gebrauchten Wörter „Mangel“ und „Beanstandungen“ dar. Auch ist die Satzstruktur komplexer aufgrund der Verwendung einer Aufzählung von zwei Fragen.

In der „komplexen Teilaufgabe“ Karo wird eine Behauptung formuliert, welche die Differenz der Summe von Beanstandungen (31) und der Fahrradanzahl (27) als Fehler bei der Diagrammerstellung bezeichnet. Mögliche Bearbeitungsschritte bei dieser Aufgabe sind die Bestimmung der Summe von tatsächlich gezählten Beanstandungen sowie die Erläuterung der Differenz im Sachzusammenhang. Durch multiple Lösungsmöglichkeiten hierfür wird die Schwierigkeit der Aufgabe vermindert.

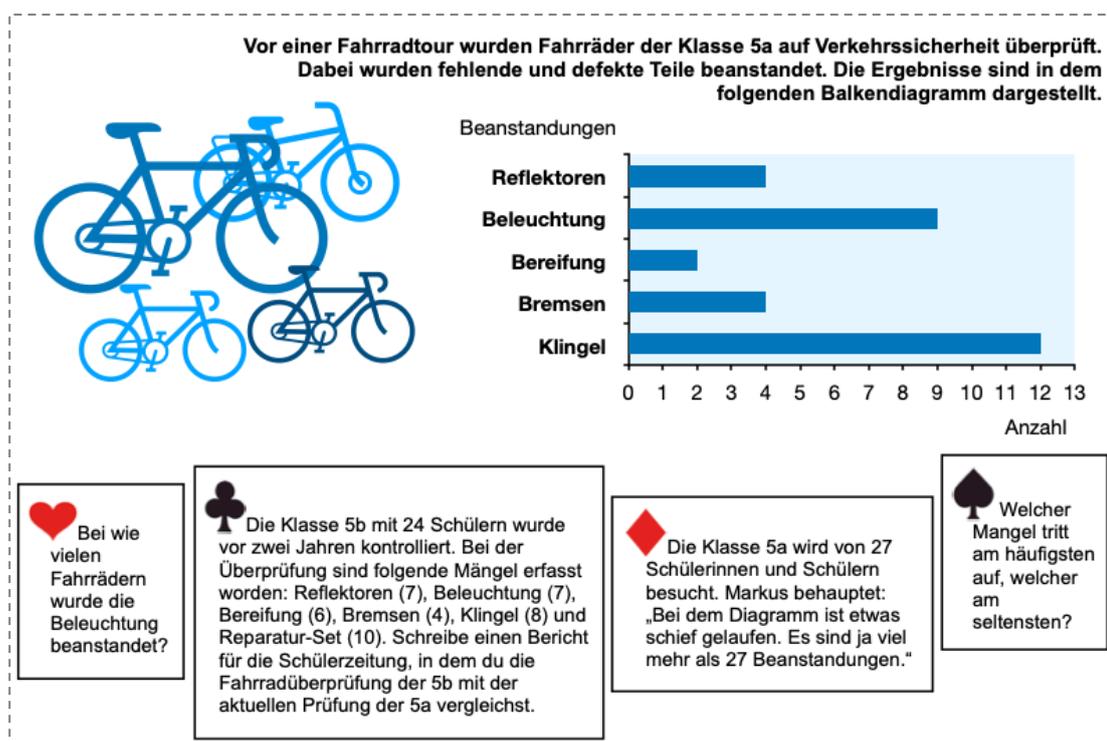


Abb. 2: Blütenaufgabe 5.1 „Fahrradüberprüfung“.

Dies kann durch selbst erdachte konkrete Beispiele geschehen („das Fahrrad von Markus war total kaputt und hatte keine Reflektoren, kein Licht, kaputte Bremsen, kaputte Reifen und keine Klingel“) oder allgemeinere Beschreibungen („Bei einem Fahrrad kann mehr kaputt sein, es muss nicht immer nur eins kaputt sein.“). Außerdem muss die Zahl 27 als zusätzliche Information im Aufgabentext verarbeitet und im Sachkontext interpretiert werden. Die kognitive Komplexität dieser Teilaufgabe ist somit aufgrund der dargestellten multiplen Anforderungen besonders hoch. Zudem ist der Umfang des Aufgabentextes im Vergleich zu Herz oder Pik höher. Das enthaltene Vokabular und die Satzstruktur sind dabei jedoch vergleichsweise einfach gehalten. Der durch die im Aufgabentext formulierte Behauptung ausgelöste kognitive Konflikt gibt jedoch hinreichenden Anlass, um die Aufgabenanforderung transparent zu machen.

Die Teilaufgabe Kreuz wurde für die Charakteristik „offene Teilaufgabe“ formuliert. Diese Teilaufgabe fordert dazu auf, einen Schülerbericht zu schreiben, in dem ein zusätzlicher Datensatz von Beanstandungen mit dem im Aufgabenkontext gegebenen Datensatz zu vergleichen ist. Zwar ist die Aufgabenstellung durch die Verwendung des Operators „Schreibe“ eindeutig formuliert, die konkrete Anforderung wird aufgrund der Offenheit des tatsächlichen Antwortformats jedoch variabel gehalten. Durch diese Öffnung können Schülerinnen und Schüler eigenständig entscheiden, ob der Bericht weitere Darstellungsformen (z. B. ein weiteres Balkendiagramm) enthält. Je nach Bearbeitung variiert die kognitive Komplexität dieser Aufgabe. Die sprachlogische Komplexität hingegen erweist sich bei dieser Teilaufgabe als besonders groß, da eine Aufzählung verwendet wird, weitere Textinformationen verarbeitet werden müssen, erneut das Wort „Mängel“ synonym zu „Beanstandungen“ verwendet wird und weil die Textlänge in dieser Teilaufgabe am umfangreichsten ist.

Insgesamt zeigt sich, dass die verschiedenen Teilaufgaben der Blütenaufgabe 5.1 „Fahrradüberprüfung“ trotz ihrer thematischen Verwandtschaft unterschiedliche Anforderungsprofile aufweisen. Damit geben die Teilaufgaben Anlass zur eigenständigen Aufgabenauswahl sowie damit einhergehend zur Beurteilung der subjektiv empfundenen Schwierigkeit durch Schülerinnen und Schüler.

#### 4.6 Durchführung

Die Bearbeitung der Blütenaufgaben und der Fragebögen fand in einer Doppelstunde (90 Minuten) statt, die in allen Schulen innerhalb eines einmonatigen Zeitfensters realisiert wurde und sich in die Unterrichtsreihe der jeweiligen Klasse einfügte. Jeder Schülerin und jedem Schüler lag ein vierseitiges

Testheft vor, welches die eingesetzten Mathematikaufgaben und die weiteren Materialien enthielt. Nach einer kurzen Begrüßung (5 Minuten) lasen die Schülerinnen und Schüler den Sachkontext und alle Teilaufgabenstellungen ihrer jeweiligen Blütenaufgabe und klärten gemeinsam mit der Lehrperson unklare Begriffe. Zudem wurde das Vorgehen der Stunde erklärt (5-10 Minuten). Die Bearbeitung der Aufgabe begann mit einer Einzelarbeitsphase, in der die Schülerinnen und Schüler beliebige und beliebig viele Teilaufgaben bearbeiten konnten (15-20 Minuten). Darauf folgte eine Gruppenarbeitsphase, in der die Lernenden in zufälligen Gruppen von zwei bis vier Personen weiterarbeiteten.<sup>6</sup> Sie sollten dabei vorhandene Lösungen vergleichen und gemeinsam besprechen sowie noch nicht bearbeitete Teilaufgaben gemeinsam bearbeiten (20-25 Minuten). Alle Schülerinnen und Schüler arbeiteten in der Gruppenphase mit einer anderen Stiftfarbe als zuvor. Dann wurden die Fragebögen ausgefüllt (5-10 Minuten) und abschließend die Lösungen der Aufgaben besprochen (10-15 Minuten).

Der Fragebogen wurde nach der Bearbeitung der Aufgaben ausgefüllt, damit explizit die Erfahrungen der Bearbeitung und die Gruppenphase mit in die Beurteilung der Aufgaben einfließen konnten. Anhand der erfragten Urteilsbegründungen sollte überprüft werden, inwiefern die getroffenen Urteile anhand von Aspekten der Aufgabebearbeitung oder beispielsweise durch die Sozialform begründet werden.

Durch die drei Phasen i) Lesen der Blütenaufgabe und Klärung unklarer Begriffe, ii) Einzelarbeit und iii) Gruppenarbeit sollte erreicht werden, dass die Schülerinnen und Schüler alle Teilaufgaben gesichtet haben, um sich unabhängig von Klassenkameradinnen und -kameraden über die Teilaufgaben Gedanken machen und diese nach eigener Präferenz bearbeiten zu können. In der Gruppenarbeitsphase mussten sie sich (erneut) mit Teilaufgaben auseinandersetzen, welche sie während der Einzelarbeitsphase ggf. noch nicht bearbeitet hatten.

Die Entscheidung, Schwierigkeitsbeurteilungen nach der Bearbeitung treffen zu lassen, hatte folgende Gründe (s. Abschnitt 2.3): Erstens sollten in der Studie möglichst vielfältige Begründungen herausgearbeitet werden; eine genauere Kenntnis der Aufgabe fördert diese Vielfalt (vgl. Thonhauser et al., 2003). Zweitens zeigt sich in diesen Studien, dass – wenn Aufgabenmerkmale im Fokus stehen – vor allem tiefenstrukturelle Merkmale seltener berücksichtigt werden (ebd.; vgl. auch Leuders & Nückles, 2015). Bleibt eine Bearbeitung aus, ist der Einbezug tiefenstruktureller Merkmale eher als unwahrscheinlich einzustufen. Diese sind jedoch in Bezug auf die

mathematische Kompetenzentwicklung in differenzierendem Unterricht besonders relevant.

#### 4.7 Auswertung

Die Antworten der geschlossenen Items des Fragebogens werden hinsichtlich der Häufigkeit ihres Auftretens quantitativ ausgewertet, um aufgabenbezogen die Anzahlen der jeweiligen individuellen Schwierigkeitsurteile bestimmen zu können. Diese Häufigkeiten werden anschließend miteinander verglichen und deskriptiv-statistisch ausgewertet, um eventuelle Muster im Beurteilungsverhalten nachzuweisen.

Zudem wird für jede Gruppe und jede Klasse (nach Ausschluss von Urteilen, in denen dieselbe Teilaufgabe als einfachste und schwierigste ausgewählt wurde) ein *Unterschiedsquotient* als Verhältniszahl berechnet (Bourier, 2011, S. 119 ff.). Für die Kennzahl der *Gruppen* wird der Quotient aus der Anzahl der in der Gruppe gegebenen unterschiedlichen individuellen Antworten und der Anzahl der Gruppenmitglieder (= maximale Zahl unterschiedlicher Antworten) gebildet. Dieses Verhältnis ergibt 1, wenn innerhalb einer Gruppe jedes Gruppenmitglied eine andere Schwierigkeitsbegründung angibt. Für die Kennzahl der *Klassen* wird der Quotient aus der Anzahl der in der Klasse gegebenen unterschiedlichen individuellen Antworten und der Anzahl maximal möglicher unterschiedlicher Antworten (zwölf Kombinationen aus einfachster und schwierigster Teilaufgabe) gebildet. Auch hier ist die Unterschiedlichkeit der Antworten umso höher, je größer das Verhältnis ist (maximal 1). Aufgrund der unterschiedlichen Bezugsgrößen sind die beiden Unterschiedsquotienten als Verhältniszahlen nicht vergleichbar, sondern zeigen lediglich auf Gruppen- bzw. Klassenebene an, wie individuell unterschiedlich innerhalb einzelner Gruppen bzw. Klassen geantwortet wurde. Die Prüfung dieser Maße ermöglicht es damit, Beeinflussungen der individuellen Beurteilungsverhalten durch etwaige Gruppen- bzw. Klassenzugehörigkeiten zu untersuchen.

Die Kategorisierung der offenen Begründungen wird anhand einer qualitativen Inhaltsanalyse vorgenommen (Mayring, 2015). In einer Pilotstudie, die ebenfalls der oben beschriebenen Prozedur folgte, wurde ein theorie- und empiriebasiertes Kategoriensystem formuliert, welches in der vorliegenden Studie angewandt und durch Ergänzungen und Abgrenzungen weiterentwickelt wurde. Im ersten Schritt wurden dazu auf Basis der gesichteten Literatur mögliche Kategorien für Begründungen deduktiv hergeleitet und formuliert (s. Abschnitte 2.3–2.5). Dabei wurden Studien zur Schwierigkeitsbeurteilung sowohl aus der Perspektive Lernender wie auch Lehrender herangezogen, um die Kategorien möglichst breit

anzulegen. Die Ergebnisse der Literaturrecherche wurden geordnet und in einem vorläufigen Kategoriensystem zusammengefasst. Anschließend wurden die empirischen Daten der Vorstudie mit diesem System codiert. Die Zuordnungen der Antworten zu den Kategorien wurden anschließend detaillierter analysiert, um gegebenenfalls Kategorien zusammenzufassen oder neue Kategorien zu formulieren (induktive Fortentwicklung des Kategoriensystems). Antworten in den Kategorien „Sonstige“ und „nicht-codierte“ Antworten wurden gesondert hinsichtlich der Formulierung neuer Kategorien analysiert. Das final verwendete Kategorienschema umfasst zwölf Kategorien und gliedert sich wie folgt auf:

- 1) *Kontextbezogene Begründung* – Begründungen, die den spezifischen Sachkontext der Aufgabe thematisieren. (*Beispiel: „... einfachste, weil ich die Himmelsrichtungen kenne.“*) Diese Kategorie umfasst keine Begründungen, die äußere Gegebenheiten der Aufgabenbearbeitung (Unterrichtszeit, Einzelarbeit, ...) benennen.
- 2) *Mathematisch-inhaltliche Begründung* – Antworten, die begriffliches oder technisches Wissen sowie mathematische Objekte oder deren Eigenschaften benennen. (*Beispiel: „... schwierigste, weil es da um Funktionen ging.“*)
- 3) *Prozessbezogene Begründung* – Begründungen, in denen explizit der Aufgabenoperator oder zur Bearbeitung notwendige bzw. nicht notwendige Tätigkeiten genannt werden. (*Beispiel: „... einfachste, weil man da nur rechnen musste.“*)
- 4) *Strategisch-organisatorische Begründung* – Begründungen mit Bezug zur Regulation des eigenen Lernprozesses bzw. der Aufgabenbearbeitung, insbesondere Begründungen, welche (metakognitive) Strategien zur optimalen Ressourcennutzung (Werkzeuge, Schreibmaterial, Zeit, kognitive Ressourcen...) oder Aufwandsoptimierung enthalten. (*Beispiel: „... schwierigste, weil sie so lang gedauert hat.“*)
- 5) *Formal-gestalterische Begründung* – Antworten, welche auf die Oberflächenstruktur von Aufgaben fokussieren, wie Repräsentationsformen mathematischer Objekte innerhalb der Aufgabe, Antwortformat, bereitgestellte Informationen, Textlänge, etc. (*Beispiel: „... einfachste, weil da so wenig Text stand.“*)
- 6) *Begründung durch Lese- & Aufgabenverständnis* – Begründungen mit konkretem Bezug zur Formulierung des Aufgabentextes oder zur Gestaltung gegebener Informationen. (*Beispiel: „... schwierigste, weil ich die Aufgabenstellung nicht verstanden habe.“*)

- 7) *Redundante Begründung* – Antworten, in denen die Wahl der einfachsten/ schwierigsten Teilaufgabe durch ihre Einfachheit/ Schwierigkeit selbst begründet wird. (Beispiel: „... einfachste, weil sie so leicht war.“)
- 8) *Vorwissensbasierte Begründung* – Begründungen auf Grundlage von inhaltlichem Vorwissen mit Mathematikbezug, auch Aussagen über eigene Kompetenzen. (Beispiel: „... schwierigste, weil ich nicht so gut Graphen zeichnen kann.“)
- 9) *Begründung anhand persönlicher Präferenz* – Begründungen, die auf dem individuellen Interesse, der Motivation oder auf emotionalen Aspekten basieren, z. B. Aussagen über die individuelle Bearbeitungsbereitschaft und über persönliche Vorlieben. (Beispiel: „... einfachste, weil ich Spaß an der Aufgabe hatte.“)
- 10) *Begründung anhand Lösungserfolg oder -strategie* – Antworten, in denen über den (Miss-) Erfolg konkreter Bearbeitungsschritte oder über das (Nicht-) Erreichen eines (Teil-) Ergebnisses berichtet wird. (Beispiel: „... schwierigste, weil ich nicht zum Ergebnis gekommen bin.“)
- 11) *Begründung durch Aufgabenvergleich* – Begründungen durch Vergleiche zwischen zwei oder mehr Teilaufgaben oder Antworten mit konkretem Bezug zu anderen Teilaufgaben. (Beispiel: „... einfachste, weil ich die anderen Aufgaben nicht lösen konnte.“)
- 12) *Sonstige Begründung* – Begründungen, bei denen keine Zuordnung zu anderen Kategorien möglich ist.

Jede Antwort wird zunächst mit genau einer Kategorie versehen, da aufgrund der Kürze der Antworten nur mit vereinzelt Mehrfachcodierungen zu rechnen ist. Liegen zu einem individuellen Schwierigkeitsurteil multiple Begründungen unterschiedlicher Kategorien vor, so werden nur der Kategoriencode der erstgenannten Begründung sowie der zusätzliche Code „Mehrfachantwort“ vergeben. Insgesamt befinden sich in 71 Begründungen der einfachsten Teilaufgabe und 50 Begründungen der schwierigsten Teilaufgabe mehrfache Zuordnungen, was insgesamt ca. 5 % aller Urteile entspricht. Eine Detailanalyse zeigt, dass durch die als Mehrfachantwort codierten Begründungskategorien keine Kategorie systematisch benachteiligt ist. Daher werden Mehrfachantworten in Folge nicht mehr durchgängig berücksichtigt.

Die Codierung der Urteilsbegründungen anhand des dargelegten Codierschemas wurde von zwei Codierern vorgenommen. Etwa 15 % des verwendeten Datenmaterials wurden von beiden Codierern zur Überprüfung codiert. Aus diesen unabhängigen

Vorgängen resultierte eine prozentuale Beurteiler-Übereinstimmung (Wirtz & Kutschmann, 2007) von ca. 85 % für die Zuordnung der Kategorien zu den Urteilsbegründungen. Bei Annahme einer zufälligen Übereinstimmung von 50 % ergibt sich damit Cohens Kappa (ebd.; Cohen, 1960) als  $\kappa = 0.7$ , was nach Landis und Koch (1977) einer „beachtlichen Übereinstimmung“ entspricht. Aufgetretene Diskrepanzen wurden durch eine Diskussion aufgelöst und in den Leitfaden als Ankerbeispiele aufgenommen.

Alle Schülerinnen und Schüler erhielten zudem für jede der vier bearbeitbaren Teilaufgaben eine Bewertung der jeweiligen Bearbeitung: 0 – nicht bearbeitete Teilaufgabe; 1 – Ansätze vorhanden; 2 – größtenteils richtig gelöst; 3 – vollständig richtig gelöst. Diese Bewertung wurde von acht Codierern und Codierern vorgenommen. Alle Codierenden erhielten ein Training, indem Daten aus der Vorstudie mit dem Kategoriensystem codiert und im Nachhinein diskutiert wurden. Es handelte sich um zwei Doktoranden der Mathematikdidaktik und sechs Masterstudierende. Für die Klassifizierung der vorliegenden Daten wurden dann 48 Bearbeitungen von allen unabhängig bewertet. Die Übereinstimmung lag hierfür ebenfalls bei 85 %. Um zudem mögliche Zusammenhänge zwischen der Güte der Aufgabenbearbeitung und den vorliegenden Begründungskategorien zu untersuchen, wird für die kategorialen Merkmale der Kontingenzkoeffizient *Cramér's V* berechnet.

## 5. Ergebnisse

Die Ergebnisse zu den Teilaufgaben sind aufgabenbezogen und nicht -übergreifend zu verstehen, da die Themenbereiche und die Gestaltung der eingesetzten Blütenaufgaben sehr unterschiedlich ausfallen.

### 5.1 Individuelle Schwierigkeitsurteile

Von den 16 möglichen individuellen Schwierigkeitsurteilen (z. B. Pik als einfachste und Herz als schwierigste Teilaufgabe) werden im Folgenden nur diejenigen zwölf berücksichtigt, bei denen unterschiedliche Teilaufgaben als leichteste und schwierigste benannt wurden (insgesamt gab es zwölf Fälle identischer Nennungen, s. Abb. 3).

#### 5.1.1 Beurteilungsergebnisse der Aufgaben

Die Verteilung der individuellen Schwierigkeitsurteile (kurz: Beurteilungsergebnisse) für alle eingesetzten Blütenaufgaben zeigt drei unterschiedliche Muster (s. Abb. 3a-g):

- 1) Relativ einheitliche individuelle Schwierigkeitsurteile treten bei den Aufgaben 5.1 und 5.3 auf. In Aufgabe 5.1 wird das Beurteilungsergebnis Herz (einfachste) und Kreuz (schwierigste) von etwa der Hälfte aller Schülerinnen und Schüler getroffen.

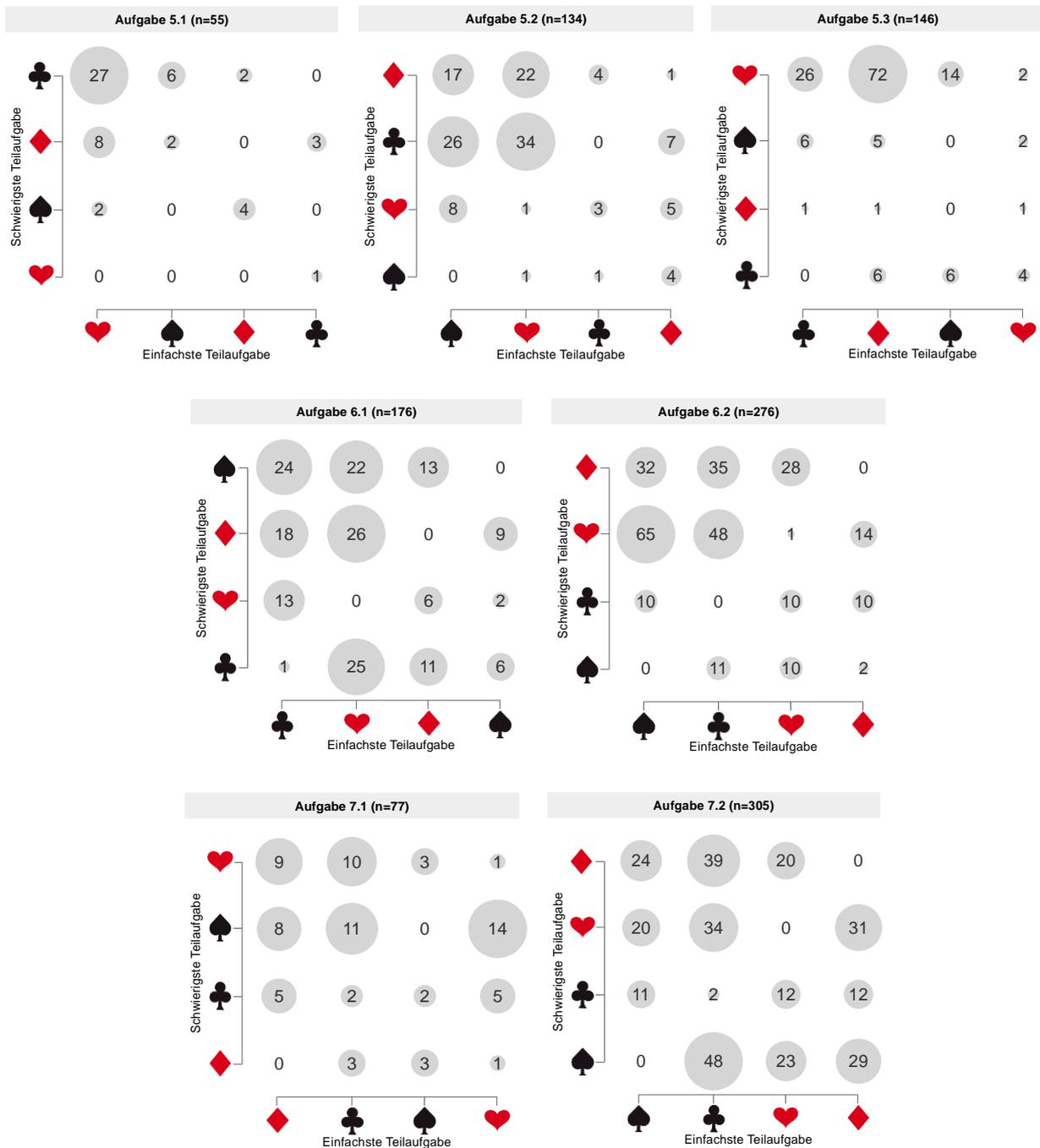


Abb. 3a-g: Darstellung der individuellen Schwierigkeitsurteile für alle eingesetzten Blütenaufgaben.

In Aufgabe 5.3 beurteilen knapp die Hälfte der Lernenden Karo als einfachste und Herz als schwierigste Teilaufgabe. Während das zweit- (ca. 18 %) und dritthäufigste (ca. 10 %) Beurteilungsergebnis die Teilaufgabe Herz als schwierigste Teilaufgabe nennt, variiert die Angabe zur einfachsten.

2) Bei Aufgabe 5.2 treten vier Beurteilungsergebnisse mit einer Häufigkeit von 13 % und mehr auf. Hier dominiert demnach nicht ein Schwierigkeitsurteil, sondern Kombinationen aus Pik und Herz (einfachste) sowie Kreuz und Karo (schwierigste).

Diese beiden Muster (Blütenaufgaben 5.1, 5.2 und 5.3) weisen folglich auch eine geringere Vielfalt der getroffenen individuellen Schwierigkeitsurteile auf (s. Abb. 4).

3) Auch bei den Blütenaufgaben 6.1, 6.2, 7.1 und 7.2 zeigen sich Unterschiede in den Beurteilungsergebnissen, jedoch treten die Ergebnisse jeweils mit einer substanziell höheren Häufigkeit auf. Dies äußert sich zum einen in den absoluten Anzahlen (s. Abb. 3), zum anderen in der Vielfalt der jeweiligen Schwierigkeitsurteile (s. Abb. 4).

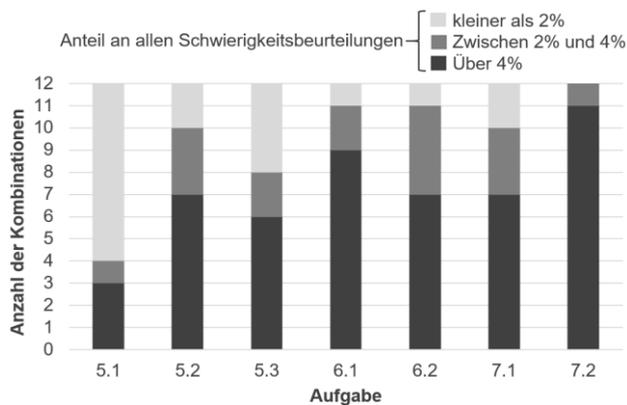


Abb. 4: Anzahlen von Beurteilungsergebnissen mit Häufigkeiten. Für jede Aufgabe ist aufgeführt, wie viele der zwölf möglichen Beurteilungsergebnisse (ohne identische Nennungen für die einfachste und schwierigste Teilaufgabe) mit einer entsprechenden Häufigkeit genannt wurden. Schwarze Säulen zeigen an, wie viele Schwierigkeitsbeurteilungen mit einer Häufigkeit von mehr als 4 % aller Datensätze gewählt wurden, mindestens mussten dies fünf Nennungen sein. Dunkelgraue Säulen zeigen an, wie viele Schwierigkeitsurteile mit einer Häufigkeit zwischen 4 % und 2 % genannt wurden, mindestens mussten hier jedoch drei Nennungen auftreten. Hellgraue Säulen zeigen an, wie viele Schwierigkeitsbeurteilungen mit einer Häufigkeit von 2 % oder weniger gewählt werden.

Die vier Blütenaufgaben der Jahrgangsstufen 6 und 7 weisen mehr individuelle Schwierigkeitsurteile mit höheren Anzahlen auf (s. Abb. 4). Die Schwierigkeitsbeurteilungen können unabhängig davon abgegeben werden, ob die jeweilige Teilaufgabe zur Bearbeitung von Lernenden ausgewählt wurde oder nicht sowie ob diese korrekt bearbeitet wurden oder nicht. Dies wird in der Folge betrachtet.

### 5.1.2 Schwierigkeitsbeurteilungen und Bearbeitungen der Teilaufgaben

Für jedes individuelle Schwierigkeitsurteil wurde überprüft, ob die Schülerinnen und Schüler die jeweils als leichteste oder schwierigste beurteilte Teilaufgabe in der Einzelarbeitsphase oder in der Gruppenarbeitsphase bearbeiteten. Eine Teilaufgabe gilt als bearbeitet, wenn die jeweilige Bearbeitung mindestens als „1 – Ansätze vorhanden“ codiert wurde.

Die subjektiv als einfachste beurteilte Teilaufgabe wurde in ca. 89 % aller Fälle bereits in der Einzelarbeit und in ca. 95 % der Fälle in Einzel- oder Gruppenarbeit bearbeitet. Etwa 5 % der Schülerinnen und Schüler beurteilten demnach eine Teilaufgabe als leichteste, die sie weder in der Einzel- noch in der Gruppenarbeitsphase bearbeitet hatten. Es gibt insgesamt nur 432 Einzelbearbeitungen (etwa 37 % aller Fälle), die eine vollständig korrekte Lösung der jeweils subjektiv einfachsten Aufgabe angefertigt haben. 766 Einzelbearbeitungen (65,5 %) haben zumindest eine Lösung der subjektiv einfachsten Aufgabe,

die in Teilen korrekt ist (vollständig korrekte Lösungen inbegriffen). In der Gruppenarbeit sind die Anteile vollständig korrekter Lösungen bzw. in Teilen korrekter Lösungen nur geringfügig gestiegen (etwa 2 % bzw. 4 %).

Die subjektiv schwierigste Teilaufgabe wurde in ca. 63 % aller Fälle bereits in der Einzelarbeit und in ca. 80 % der Fälle in Einzel- oder Gruppenarbeit bearbeitet. Etwa 20 % der Schülerinnen und Schüler beurteilten demnach eine Teilaufgabe als schwierigste, die sie weder in Einzel- noch in der Gruppenarbeitsphase bearbeitet hatten. Außerdem bewirkte die Gruppenarbeit in 17 % der Fälle eine Auseinandersetzung mit der subjektiv am schwierigsten empfundenen Teilaufgabe, obwohl eine Bearbeitung in der Einzelarbeitsphase nicht erfolgte. Es werden typischer Weise insgesamt also deutlich häufiger Teilaufgaben als schwierigste beurteilt, die von den Schülerinnen und Schüler gar nicht bearbeitet wurden, im Vergleich zur Beurteilung der einfachsten Teilaufgabe. Unter den Einzelbearbeitungen der jeweils subjektiv schwierigsten Teilaufgabe befinden sich lediglich 194 Lösungen (etwa 16,5 % aller Fälle), die vollständig korrekt sind. Auch hier ist etwa die doppelte Anzahl an Lösungen zumindest in Teilen korrekt (403 Bearbeitungen, vollständig korrekte Lösungen inbegriffen). Der Anstieg korrekter Lösungen in Gruppenarbeit ist auch für die subjektiv schwierigste Teilaufgabe gering, fällt jedoch im relativen Vergleich mit dem Anstieg bei der subjektiv einfachsten Teilaufgabe größer aus (etwa 4 % bzw. 6 %).

### 5.1.3 Verschiedenheit von individuellen Beurteilungsergebnissen

Neben den in Abschnitt 4 beschriebenen beabsichtigten Effekten einer Gruppenarbeitsphase besteht die Möglichkeit, dass durch die Diskussion in den Gruppen oder durch die herrschende Unterrichtskultur die individuellen Schwierigkeitsurteile vereinheitlicht wurden. Um diese Effekte zu untersuchen, werden die individuellen Schwierigkeitsurteile aller Gruppen ( $n = 386$ ) und Klassen auf Gemeinsamkeiten mit Hilfe der in Abschnitt 4.7 beschriebenen Unterschiedsquotienten analysiert.

In nahezu der Hälfte aller Gruppen werden keine identischen Schwierigkeitsbeurteilungen abgegeben (Unterschiedsquotient Gruppen = 1, s. Abb. 5a).

Neben ca. 50 Gruppen mit einem Maß von 0,5 (entweder zwei identische Schwierigkeitsbeurteilungen in einer Zweiergruppe oder zwei unterschiedliche Schwierigkeitsbeurteilungen in einer Vierergruppe) gilt für die meisten Gruppen, dass mehr als die Hälfte der Gruppenmitglieder paarweise unterschiedliche Schwierigkeitsbeurteilungen abgegeben haben.

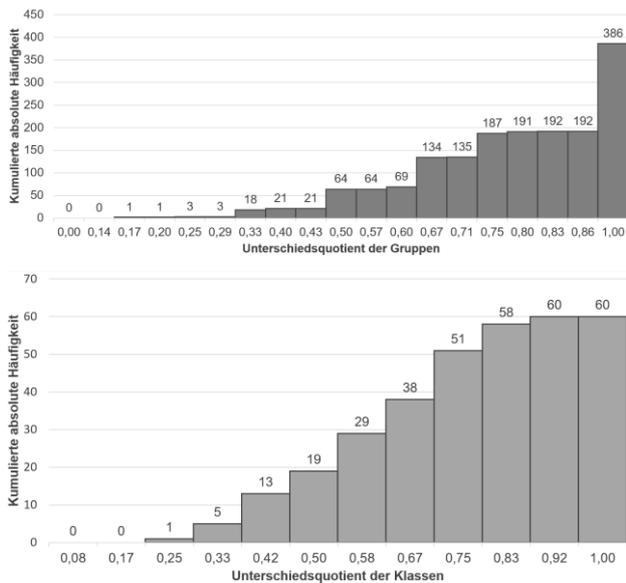


Abb. 5a und 5b: Kumulierte Häufigkeitsverteilung der Unterschiedsquotienten. Oben: Heterogenität innerhalb von Gruppen der Zusammenarbeit (variable Klassenbreite aufgrund möglicher Gruppengrößen). Unten: Heterogenität innerhalb von Schulklassen (äquidistante Klasseneinteilung).

Auch für die Klassen weist die Verteilung des Unterschiedsquotienten eine deutliche Verlagerung nach rechts auf (s. Abb. 5b). Insgesamt liegt die deutliche Mehrheit der Klassen oberhalb von 0,5, was auf vielfältige individuelle Schwierigkeitsurteile hinweist.

## 5.2 Begründungen der individuellen Schwierigkeitsurteile

Im Folgenden werden die Begründungen für die in 5.1 dargestellten Beurteilungen näher untersucht.

### 5.2.1 Begründungen für die einfachste und schwierigste Teilaufgabe

Über alle Blütenaufgaben hinweg dominiert sowohl für die Begründung der einfachsten als auch für die

Begründung der schwierigsten Aufgabe jeweils eine Begründungskategorie: Bei der Beurteilung der einfachsten Teilaufgabe die Kategorie „prozessbezogene Begründung“, bei der schwierigsten Teilaufgabe die Kategorie „Lese- und Aufgabenverständnis“ (s. Abb. 6). Dieses typische Muster zeigt sich ebenfalls, wenn man jede der Blütenaufgaben einzeln betrachtet (s. Abb. 7). Schülerinnen und Schüler beziehen sich bei ihren *Begründungen der einfachsten Teilaufgabe* sehr oft auf konkrete Tätigkeiten, die sie bei der Bearbeitung der Aufgabe ausgeführt haben.

Darüber hinaus finden sich unter den jeweils fünf meistcodierten Begründungskategorien pro Blütenaufgabe bestimmte Kategorien wieder, die aufgabenübergreifend als typisch bezeichnet werden können.

In allen sieben Balken (s. Abb. 7) zeigen sich „strategisch-organisatorische“ Begründungen unter den fünf meistgenannten Kategorien. Hier beziehen sich Schülerinnen und Schüler häufig unspezifisch auf die kurze Dauer des Bearbeitungsprozesses: „... weil ich es am schnellsten lösen konnte“. Die Antworten referenzieren keine Details der Teilaufgabe und greifen den Lösungsprozess höchstens als „Rechnen“ auf.

Bei fünf Aufgaben wird auf das „Lese- und Aufgabenverständnis“ Bezug genommen, dass also die Aufgabe beispielsweise „gut zu verstehen war“. Bei vier Aufgaben wird vorhandenes „Vorwissen“ als Begründung angeführt. Hier wird ebenfalls selten auf konkrete Details eingegangen, sondern Vorwissen häufig als „das“ oder „es“ in die Begründung eingebunden. Eine Aufgabe wurde demnach als einfachste beurteilt, „... weil da eigentlich das gemacht werden musste, was ich kann“ oder „... weil ich das konnte“. In wenigen Fällen werden konkrete Inhalte genannt, wie z. B. „... weil ich sehr gut Brüche dividieren kann“ oder Bezug zu vorcurricularen Erfahrungen genommen.

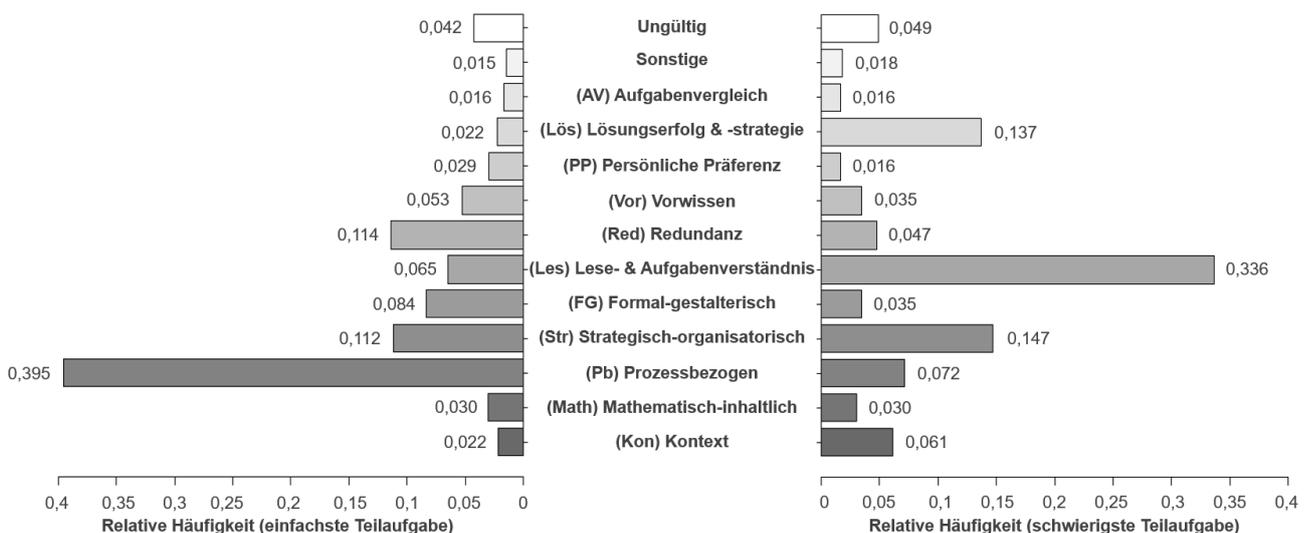


Abb. 6: Relative Häufigkeitsverteilung der Begründungskategorien über alle eingesetzten Blütenaufgaben (links Begründungen für einfachste Teilaufgabe; rechts Begründungen für die schwierigste Teilaufgabe; n = 1169).

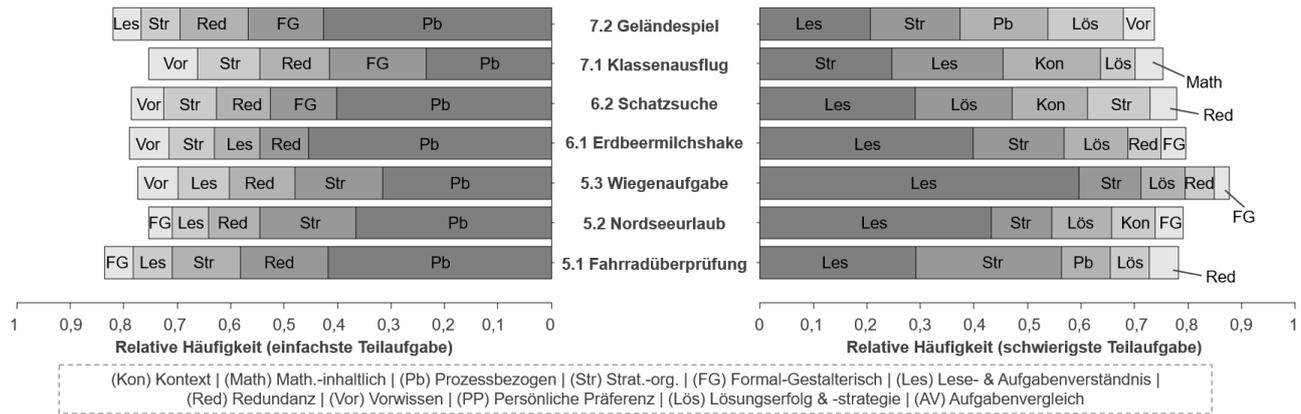


Abb. 7: Auffächerung der fünf häufigsten Begründungskategorien je Blütenaufgabe für die Beurteilung der einfachsten Teilaufgabe (links) und der schwierigsten Teilaufgabe (rechts). Absolute Häufigkeiten je Aufgabe sind Tabelle 2 zu entnehmen.

Ebenfalls bei allen sieben Blütenaufgaben vertreten ist die Kategorie „Redundanz“, in der Schülerinnen und Schüler beispielsweise in tautologischer Weise als Grund aufführen, dass die Teilaufgabe die leichteste sei, weil „die Aufgabe einfach ist“.

Bei den Begründungen der schwierigsten Teilaufgabe dominiert die Kategorie „Lese- und Aufgabenverständnis“: Bei sechs Aufgaben wird sie teilweise deutlich als häufigste Kategorie angeführt (s. Abb. 7). Eine weitere wichtige Kategorie bilden wie bei den einfachsten Teilaufgaben „strategisch-organisatorische“ Begründungen. Diese nehmen typischerweise sehr häufig Bezug zum zeitlichen Aufwand, wie zum Beispiel „... weil die Aufgabe so lang dauerte“ oder „... weil ich so lange darüber nachgedacht habe, was ich schreiben könnte“.

Bei den weiteren Kategorien zeigt sich kein so deutlich übergreifendes Muster wie bei den vorigen beiden Kategorien, dennoch finden sich gleiche Kategorien an unterschiedlicher Stelle unter den fünf meistgenannten Kategorien wieder. Die Kategorie „Lösungserfolg und -strategie“ bezieht sich auf die eigene Aufgabenlösung und begründet die Schwierigkeit häufig dadurch, dass das eigene Ergebnis nicht korrekt war, z. B. „... weil ich das nicht so gut hinbekommen habe und nicht wirklich weiß, ob die Aufgabe richtig ist“, oder dass eine Lösungsstrategie fehlte, z. B. „... weil ich nicht wusste, wie man rechnen musste“. Weiterhin ist bei drei Aufgaben die Begründung der Aufgabenschwierigkeit anhand der Schwierigkeit oder dem Nicht-Mögen des „Kontextes“ unter den fünf meistgenannten Kategorien. Die Kategorie „Redundanz“ tritt bei vier Aufgaben auf und enthält analog zur einfachsten Aufgabe Begründungen wie „... weil sie schwer war“.

In den individuellen Schwierigkeitsbegründungen (Kombination aus „Begründung für einfachste“ und „Begründung für schwierigste“) einer Schülerin bzw. eines Schülers wird überwiegend auf

unterschiedliche Begründungskategorien zurückgegriffen. Dieselbe Begründungskategorie bei der Begründung der einfachsten und der schwierigsten Teilaufgabe liegt lediglich bei insgesamt 166 Schülerinnen und Schülern vor (14,2 % aller Fälle) (s. Abb. 8).

Damit haben etwa 85,8 % der Schülerinnen und Schüler unterschiedliche Begründungskategorien bei der einfachsten Teilaufgabe und der schwierigsten Teilaufgabe gegeben. Aufgrund der Vielzahl der Kategorien lassen sich im Rahmen der Auswertungen keine „typischen“ Konstellationen bestimmen.

In einer weiteren Analyse mittels des Kontingenzkoeffizienten Cramér’s V konnten keine Zusammenhänge zwischen der Korrektheit der Bearbeitungen (vgl. Abschnitt 5.1.2) und der Begründung anhand spezifischer Begründungskategorien (vgl. Kategoriensystem in Abschnitt 4.7) gefunden werden.

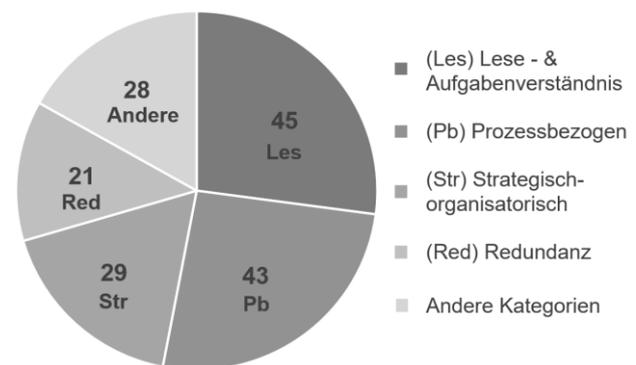


Abb. 8: Absolute Häufigkeiten von Lernenden mit identischen Kategorien bei der Begründung der einfachsten und der schwierigsten Teilaufgabe.

Bei sechs Blütenaufgaben findet sich die Kategorie „formal-gestalterisch“ ebenfalls unter den fünf meistgenannten Begründungskategorien. Hier begründen Lernende die Einfachheit der Teilaufgaben mit einer geringen Textlänge oder dem Vorhandensein eines Bildes.

### 5.2.3 Verschiedenheit der Begründungen in Gruppen und Klassen

Analog zu Abschnitt 5.1.3 wurden auch für die Urteilsbegründungen Unterschiedsquotienten für Gruppen und Klassen ermittelt. Für die Berechnung wurden nur Schülerinnen und Schüler einbezogen, die sowohl für die Wahl der einfachsten als auch der schwierigsten Aufgabe eine Begründung angegeben haben. In einer Gruppe wurden keine gültigen Begründungen getroffen, daher ist die Gesamtzahl der Gruppen hier 385. Die Anzahl berücksichtigter Klassen beläuft sich erneut auf 60.

Von 385 Gruppen weisen 13 einen Unterschiedsquotienten von exakt 0,5 auf, während nur eine Gruppe unter einem Wert von 0,5 liegt. Die übrigen Gruppen weisen allesamt Quotienten von über 0,5 auf. In 328 Gruppen wurden sogar vollständig unterschiedliche Begründungen angegeben (Unterschiedsquotient Gruppen = 1), wobei 45 dieser Gruppen nur aus einer Person bestehen. Insgesamt weisen die individuellen Schwierigkeitsbegründungen ungeachtet der dominierenden Kategorien „Prozessbezogen“ und „Lese- und Aufgabenverständnis“ eine hohe Heterogenität innerhalb der Gruppen auf.

Ähnlich heterogen sind die individuellen Schwierigkeitsbegründungen innerhalb der Klassen. Die maximale Anzahl möglicher Kombinationen von Begründungen für die leichteste und schwierigste Teilaufgabe ist die Anzahl von Schülerinnen und Schülern der Klasse. Von 60 Klassen finden sich lediglich 3, in denen der Quotient unter 0,5 liegt. Das bedeutet, dass die Hälfte der Schülerinnen und Schüler dieser Klassen Begründungskombinationen wählt, die auch ein Klassenkamerad oder eine Klassenkameradin auswählt. Die übrigen Klassen verteilen sich zwischen den Maßzahlen 0,5 und 0,9, wobei in 29 Klassen mindestens 70 % der individuellen Schwierigkeitsbeurteilungen paarweise unterschiedlich waren.

Sowohl die Analyse der Klassen als auch der Gruppen geben in der Breite keine Hinweise auf

uniformierende Einflüsse der jeweiligen Lerngruppen bezüglich der Schwierigkeitsbegründungen.

### 5.3 Detailauswertung der Aufgabe Erdbeermilchshake

In der Folge wird exemplarisch für die Aufgabe „Erdbeermilchshake“ (s. Anhang) dargestellt, wie sich die Begründungen der Schwierigkeitsurteile auf die einzelnen Teilaufgaben verteilen.

Auch bei den einzelnen Teilaufgaben finden sich die beiden häufigsten Begründungskategorien in ähnlicher Ausprägung wieder (s. Abb. 9), wie sie sich auch kumuliert für die gesamte Blütenaufgabe zeigen (s. Abb. 7). Lediglich bei der Teilaufgabe Herz tritt die Begründungskategorie „strategisch-organisatorisch“ für die Beurteilung der schwierigsten Aufgabe am häufigsten auf. Jedoch wurde die Teilaufgabe Herz von den Schülerinnen und Schülern lediglich in 21 Fällen als schwierigste Teilaufgabe eingestuft, wodurch einzelne Antworten einen höheren Einfluss auf die Begründungskategorien haben als bei häufiger beurteilten Teilaufgaben (vgl. Zeilensummen im Bubbleplot der Aufgabe 6.1 „Erdbeermilchshake“, Abb. 3).

Insgesamt werden Teilaufgaben, welche eine hohe sprachlogische Komplexität sowie eine vergleichsweise große Textlänge aufweisen (Kreuz und Karo, s. Anhang), vorrangig aufgrund der Kategorie „Text- und Aufgabenverständnis“ als besonders schwierig beurteilt wurden. Für die anderen Teilaufgaben ist diese Begründungskategorie weniger stark ausgeprägt.

Die Kategorie „Formal-Gestalterisch“ tritt bei Begründungen der schwierigsten Teilaufgabe nicht unter den fünf meistgenannten auf, dafür jedoch bei der Begründung der einfachsten Teilaufgabe bei der offenen Aufgabe (Pik) und der Aufgabe Kreuz. Dies deckt sich mit der Verwendung weiterer Darstellungsformen in diesen Teilaufgaben.

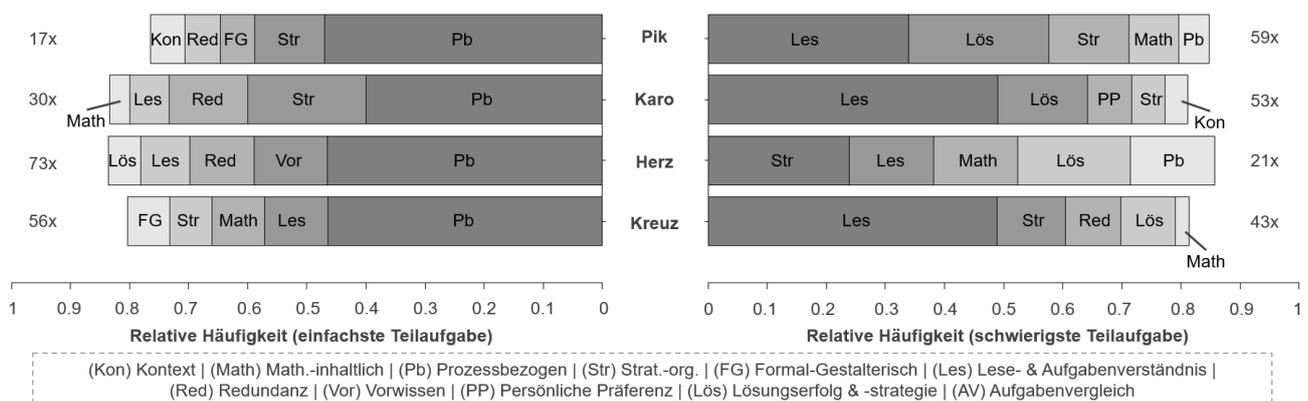


Abb. 9: Darstellung der fünf häufigsten Urteilsbegründungen nach Teilaufgaben für die Aufgabe 6.1 Erdbeermilchshake. Am Ende der Balken sind die jeweiligen absoluten Häufigkeiten aufgelistet.

Analysiert man die konkreten Begründungen in den Kategorien, so zeigen sich innerhalb der Kategorien weitere Muster. Für die Begründungen der Kategorie „Prozessbezogen“ werden unterschiedliche Bezüge zur Aufgabenstellung sichtbar. Etwa ein Drittel der Antworten bezieht sich auf einen in der Teilaufgabe genannten Operator bzw. eine Tätigkeit, beispielsweise „... weil man sich etwas eigenes ausdenken konnte“. Die übrigen zwei Drittel der Schülerinnen und Schüler begründen mit Tätigkeiten, die im Aufgabentext nicht genannt werden, sondern bspw. während der Bearbeitung auftreten, z. B. „... weil ich nur erweitern, kürzen und dividieren musste“. Ein weiteres Charakteristikum dieser Antworten ist das Wort „nur“, welches in ca. 80 % der Begründungen dieser Kategorie genannt wird, z. B. „... weil man da nur ein bisschen plus und mal rechnen muss“ (s. Abb. 10).

2. Die Teilaufgabe 8 war für mich am einfachsten, weil: man da nur ein bisschen Plus und Mal rechnen musste.

Abb. 10: Beurteilung der einfachsten Teilaufgabe mit prozessbezogener Begründung.

Zum einen könnte eine ausgeführte Tätigkeit so als einfach gekennzeichnet werden – evtl. auch im Gegensatz zu anderen, nicht genannten oder bei schwierigeren Aufgaben auftretenden Tätigkeiten. Zum anderen können derartige Formulierungen auch strategisch-organisatorische Aspekte transportieren, nämlich dass die Tätigkeit zügig durchzuführen ist.

Parallel zeigt sich bei Begründungen der Kategorie „Lese- und Aufgabenverständnis“ als schwierigste Teilaufgabe häufiger das Wort „erst“, beispielsweise „... weil die habe ich erst nicht verstanden“. Damit könnten die Schülerinnen und Schüler anzeigen wollen, dass sie zwar zu Beginn ihrer Beschäftigung mit der Aufgabe Schwierigkeiten hatten (Einzelbearbeitungsphase), im Anschluss jedoch in der Lage waren, sie zu bearbeiten (spätere Einzelbearbeitung oder ggf. Gruppenbearbeitungsphase). Dies wird expliziert in der Begründung „... weil ich sie nicht sofort verstanden habe. Nachher aber schon“. Wird eine solche zeitliche Spezifizierung nicht vorgenommen, wird absoluter formuliert: „... weil ich habe sie nicht gecheckt“.

2. Die Teilaufgabe 8 war für mich am einfachsten, weil: man sich etwas eigenes ausdenken konnte.

3. Die Teilaufgabe 8 war für mich am schwierigsten, weil: man sich da selbst ein Rezept ausdenken muss.

Abb. 11a und 11b: Begründung der Beurteilung von Teilaufgabe Pik als einfachste (oben) und schwierigste (unten).

In Bezug auf die offene Aufgabe (Pik) finden sich – bei unterschiedlichen Schülerinnen und Schülern – ebenfalls z. T. identische Antworten für die Begründung der einfachsten bzw. schwierigsten Teilaufgabe, beispielsweise „... weil man sich da selbst ein Rezept ausdenken muss“ (s. Abb. 11a und 11b).

Solche nahezu identischen Lösungen zeigen sich auch bei den offenen Aufgaben anderer eingesetzter Blütenaufgaben.

## 6. Diskussion

### 6.1 Zusammenfassende Interpretation

*Frage 1: Wie vielfältig sind die individuellen subjektiven Schwierigkeitsurteile von Schülerinnen und Schülern in Bezug auf unabhängige, nicht-hierarchisch gekennzeichnete Teilaufgaben?*

Bei den Schwierigkeitsurteilen zeigt sich eine aufgabenspezifisch unterschiedlich ausgeprägte Vielfalt. Über alle Blütenaufgaben hinweg treten mindestens drei verschiedene individuelle Schwierigkeitsbeurteilungen in substantieller Menge auf, bei einigen Aufgaben sogar bis zu 11 (vgl. Abb. 4). Schwierigkeitsbeurteilungen sind demnach nicht einheitlich durch die beurteilte Aufgabe festgelegt, sondern im Sinne des selbstregulierten Lernens und der Bindendifferenzierung hochgradig individuell beeinflusst (Landmann et al., 2009; Ostermann et al., 2015; Schneider & Artelt, 2010;). Inwieweit eine Aufgabenreihung mit stetig ansteigendem (empirischen) Schwierigkeitsgrad auch vor dem Hintergrund der subjektiven Schwierigkeitseinschätzung am lernförderlichsten ist, bleibt eine zu klärende Frage (Astleitner, 2008; Leutner, 2004).

In den Aufgaben der fünften Klassenstufe zeigt sich eine deutlich geringer ausgeprägte Vielfalt der individuellen Schwierigkeitsbeurteilungen als bei den Aufgaben der sechsten und siebten Klassenstufe. Aufgrund der in Inhalt und Formulierung völlig unterschiedlichen Aufgaben sind Rückschlüsse auf einen Einfluss der Klassenstufe jedoch nur vorsichtig zu treffen. Würde sich dieses Muster in weiteren Studien bestätigen, so könnte eine mögliche Erklärung hierfür das in der Schullaufbahn wachsende individuelle Wissen von Schülerinnen und Schülern über Mathematikaufgaben sein, welches zu differenzierteren und demnach diverseren Beurteilungen der Schwierigkeit von Mathematikaufgaben führen könnte (De Corte et al., 2011; Flavell, 1984; Kaiser & Kaiser, 2006).

Weiterhin zeigt sich, dass die Formulierung und die Gestaltung einer Aufgabe (insbesondere bzgl. schwierigkeitsbeeinflussender Merkmale) über die Bedeutung für die Bearbeitung hinaus (Neubrand et

al., 2002; Prenzel et al., 2002) auch Einfluss auf die Schwierigkeitsbeurteilungen und deren Vielfalt haben kann. Dies wird auch durch die einige konkrete Begründungen bestätigt, wenn beispielsweise Bezüge zu den geforderten Rechenoperationen oder zu konkreten Eigenschaften der verwendeten Zahlen hergestellt, wenn andere bekannte Aufgabenformate benannt oder der Umfang einer Aufgabenbearbeitung angeführt werden.

Es kann weitestgehend ausgeschlossen werden, dass die Gruppenarbeit sowie die Klassenzugehörigkeit zu einer Vereinheitlichung der Urteile führen. Vielmehr scheint hier die Gruppenarbeit die Auseinandersetzung mit schwierigen Aufgaben zu fördern, da die Bearbeitungen der als am schwierigsten beurteilten Aufgabe während dieser Phase deutlich ansteigen. Dies stärkt die Perspektive eines *gemeinschaftlich durchgeführten* differenzierenden Unterrichts (Benölken, Dixel & Berlinger, 2018; Salle et al., 2014). Wie genau sich dieser Einfluss manifestiert und welche Diskussionen in den einzelnen Gruppen geführt werden, ist an dieser Stelle unklar und stellt eine lohnende Perspektive für weitere Studien dar.

*Frage 2: Wie begründen die Schülerinnen und Schüler ihre individuellen subjektiven Schwierigkeitsurteile einer Teilaufgabe als leichteste oder schwierigste?*

Bei allen eingesetzten Blütenaufgaben sind zwei Kategorien mit Abstand am häufigsten codiert worden: Ungeachtet der jeweiligen Blütenaufgabe werden „prozessbezogene“ Begründungen bei der Beurteilung der subjektiv einfachsten und Begründungen der Kategorie „Lese- und Aufgabenverständnis“ bei der Beurteilung der subjektiv schwierigsten Teilaufgabe deutlich häufiger formuliert als andere Begründungen. Die in der Studie erhobenen Begründungen scheinen in der Breite demnach nur bedingt aufgabenspezifisch zu sein. Die beobachteten Begründungen sind zudem bis auf wenige Ausnahmen auf Basis von Aufgabenmerkmalen, personenbezogenen Merkmalen oder Kombinationen dieser Merkmale getroffen worden. Personengruppenspezifische Merkmale (Karst, 2012) wurden nur selten als Begründungen genannt (z. B. „..., weil wir in der Gruppe alle nicht wussten, was zu tun ist“).

Begreift man den Aufbau eines tragfähigen Situationsmodells als Kernvoraussetzung einer Aufgabenbearbeitung, so ist die Dominanz der beiden Begründungskategorien und deren Zusammenhang zu den Bearbeitungen plausibel (Reusser, 1997): Kann ein solches mentales Modell (subjektiv) erfolgreich gebildet werden, liegt insbesondere auf der Hand, welche Tätigkeiten vermeintlich auszuführen sind bzw. ausgeführt wurden. Die Aufgabe wird dementsprechend bei vorliegender Aufgabenbearbeitung als

leichter beurteilt. Gelingt ein solcher Aufbau nicht, ist unklar, „was zu tun ist“. In der Konsequenz bleibt die Bearbeitung aus, da die „Aufgabe nicht verstanden“ wurde. Im Sinne der Selbstwirksamkeitserwartung würde die Aufgabe also als schwierig gelten, da eine Hürde durch die Aufgabe offengelegt wurde, die nach Selbsteinschätzung der Schülerinnen und Schüler nicht oder nur mit sehr großem Aufwand bewältigt werden kann. Die Lernenden nehmen sich in solchen Situationen nicht selbstwirksam wahr.

Nur ca. 14,2 % der Schülerinnen und Schüler orientieren sich bei beiden Begründungen an einem festen Merkmal der Aufgabe oder einem selbst festgelegten Kriterium; die Mehrheit der Lernenden wählt unterschiedliche Kategorien bei den Begründungen für die subjektive Einfachheit bzw. Schwierigkeit verschiedener Teilaufgaben. Lernende sind es gewohnt, Schwierigkeiten von Aufgaben auf einer gedachten Schwierigkeitsskala von „leicht – mittel – schwer“ einzuordnen (Drüke-Noe, 2018). Die Gründe für eine solche Einordnung sind nach den vorliegenden Ergebnissen jedoch häufig vielfältig und nicht auf nur einem Merkmal basierend.

In den Ergebnissen zeigt sich zudem nochmals der Zusammenhang zwischen subjektiven a priori und a posteriori Schwierigkeitsurteilen. Demnach werden bei der subjektiven Schwierigkeitsbeurteilung der einfachsten Aufgabe *nach der Bearbeitung* vielmehr die eigenen Tätigkeiten bewertet, anstatt konkrete Merkmale der Aufgabe in den Blick zu nehmen (prozessbezogene Begründungen). Das Lese- und Aufgabenverständnis hingegen bildet eine Barriere *vor der Aufgabenbearbeitung*. Aspekte, die ebenfalls während einer Aufgabenbearbeitung relevant würden und andere Kategorien betreffen (z. B. mathematischer Gegenstand, Vorwissen), werden häufig daher nicht näher erfahren und auch nicht thematisiert. Dies deckt sich mit dem Beurteilungsverhalten von Lehrkräften: Während Lehrende vor der Bearbeitung (a priori) häufig vorrangig Oberflächenmerkmale von Aufgaben zur Begründung ihrer Schwierigkeitsurteile verwenden (McElvany et al., 2009; Ostermann, Leuders & Nückles, 2015; Thonhauser et al., 2003), nutzen sie a posteriori auch eigene Aufgabenbearbeitungen mit tiefenstrukturellen Aufgabenmerkmalen zur subjektiven Urteilsfindung und -begründung (Thonhauser et al., 2003), wie z. B. mathematisch-inhaltliche oder tätigkeitsbezogene Begründungen.

Bezüglich a priori Beurteilungen kann demnach vermutet werden, dass sich „prozessbezogene“ Begründungen hier sehr viel seltener und „Lese- und Aufgabenverständnis“-Begründungen in ähnlicher Zahl zeigen werden als bzw. wie bei a posteriori Urteilen.

Aus den Detailbetrachtungen der einzelnen Begründungen lassen sich zudem weitere Muster bezüglich der individuellen Schwierigkeitsurteile feststellen:

- 1) Es tritt außerhalb der Kategorie „Prozessbezogen“ eine Vielzahl *unspezifischer Begründungen* auf. Diese nehmen in diesem Fall keine konkreten Details der Aufgabe auf, sondern referenzieren die Teilaufgabe als Ganzes durch Pronomen wie „es“, „das“ oder „sie“ und verbleiben somit auf einer allgemeinen Ebene (insbesondere „Lese- und Aufgabenverständnis“, „Vorwissen“, „Lösungserfolg und -strategie“ und „Strategisch-organisatorisch“). Empirisch wirksame schwierigkeitsbeeinflussende Merkmale wie Grundvorstellungsintensität, sprachlogische Komplexität, oberflächliche Gestaltungsmerkmale etc. (Neubrand et al., 2002; Prenzel et al., 2002) finden sich nur selten in den Begründungen wieder.
- 2) *Begründungen mit nahezu demselben Wortlaut* treten sowohl bei der Beurteilung der einfachsten als auch der schwierigsten Teilaufgabe (insbesondere bei offenen Aufgaben) auf. Dieselbe Tätigkeit kann demnach als subjektiv schwierigkeitsgenerierend wie auch schwierigkeitsmildernd wahrgenommen werden.

Insbesondere bei den offenen Teilaufgaben ist hier zu berücksichtigen, dass Aufgabenlösungen unterschiedlich ausführlich, komplex oder anspruchsvoll ausfallen können und demnach auch die Tätigkeiten als entsprechend eher einfach oder eher schwierig beurteilt werden können. Die Vertrautheit mit der Bearbeitung solch offener Aufgaben kann zudem ein wichtiger Faktor dafür sein, inwiefern beispielsweise das „Selbstaufdenken“ als einfach oder schwierig beurteilt wird.

## 6.2 Methodische Reflexion und Desiderate

Die vorgestellten Ergebnisse zu Schwierigkeitsbeurteilungen und -begründungen sind in einer größeren Projektgruppe entstanden und immer wieder in gemeinsamen Sitzungen diskutiert worden. Dies führt zu einer hohen intersubjektiven Nachvollziehbarkeit der Kategorisierungen und der resultierenden Folgerungen. Die Daten erlauben zudem eine Identifikation von typischen Einschätzungen und Begründungen auf einer breiten Grundlage. Obgleich die Ergebnisse natürlich im Rahmen der Behandlung spezifischer Aufgaben in spezifischen Lerngruppen entstanden sind, erlaubt die hohe Anzahl von Probandinnen und Probanden das Formulieren weitergehender Hypothesen, da aufgrund der verschiedenen Schulformen und Klassenstufen hinreichend viele verschiedene Fälle in die Analysen einbezogen wurden.

Die subjektiven Schwierigkeitsurteile wurden in der vorliegenden Studie unterrichtsintegriert erhoben, was die inhaltliche Validität erhöht, da die Schwierigkeitsurteile in gewohnter bzw. „natürlicher“ Unterrichtsumgebung getroffen wurden.

Aufgrund dieser Tatsache wurden mehrere Blütauftgaben verwendet (s. Abschnitt 4.2). Um die Unterrichtsbindung zu gewährleisten und die gewünschten Breite der Datenbasis zu erreichen, wurden keine personenbezogenen Daten über den Leistungsstand der Schülerinnen und Schüler, gleichwohl jedoch Daten zur Leistung bei der Bearbeitung der Blütauftgaben ermittelt. Vergleiche der subjektiven Beurteilungen mit solchen Leistungsdaten eröffnen jedoch zusätzliche Perspektiven auf die Erforschung von Aufgabenbeurteilungsprozessen und bedürfen einer Berücksichtigung in zukünftigen Studien in diesem Forschungsfeld (Getzin, i. V.).

Die Urteile und Begründungen wurden *nach* der Bearbeitung der Aufgaben in Einzel- und Gruppenarbeit erhoben. Auf diese Weise sollte gewährleistet werden, dass sich die Schülerinnen und Schüler mit möglichst allen Aufgaben auseinandersetzen und dementsprechend vielfältige Begründungen formulieren. Diese Erkenntnisse können wiederum genutzt werden, um Erkenntnisse über a priori Einschätzungen zu gewinnen. Analog zu obigen Betrachtungen führt die vertiefte Auseinandersetzung der Schülerinnen und Schüler mit den Aufgaben vermutlich zu einem breiteren sowie an tiefenstrukturellen Merkmalen und Tätigkeiten orientierten Beurteilungs- und Begründungsverhalten, während eine a priori Beurteilung vermehrt oberflächliche Merkmale heranzieht (vgl. McElvany et al., 2009; Ostermann, Leuders & Nückles, 2015; Thonhauser et al., 2003). Die jeweiligen Unterschiede im Prozess der Beurteilungen a posteriori und a priori bieten Potenzial für detailliertere, insbesondere qualitative Studien. Über die Variation bestimmter Merkmale von einzelnen Aufgaben könnte zudem geklärt werden, inwiefern schwierigkeitsbeeinflussende Merkmale einen Einfluss auf subjektive Beurteilungen und Begründungen haben.

Die Ergebnisse der Studie betrachten schülerseitige Begründungen inhaltlich übergreifend und fassen daher beispielsweise Tätigkeiten und Merkmale aus Inhaltsbereichen zusammen (bspw. „Berechnen“, „Zeichnen“ und „Ablesen von Koordinaten“ als prozessbezogene Begründungen). Untersuchungen, die gezielt einzelne Inhaltsgebiete und daher auch feinere Nuancen von Aufgabenstellungen für die Schwierigkeiten und Schwierigkeitsbeurteilungen von Schülerinnen und Schülern in den Blick nehmen, liegen bisher fast ausschließlich für die Primarstufe vor (Block, 2014; Rathgeb-Schnierer & Rechtsteiner, 2018; Selter, 2000). Eine entsprechende inhaltliche Vertiefung

in der Sekundarstufe könnte hier weiterführende spezifischere Erkenntnisse liefern.

Eine weitere Studienlimitation liegt darin, dass für die Begründungen nur kleine Antwortfelder zur Verfügung standen, was die Schülerinnen und Schüler vermutlich zu eher kürzeren Antworten veranlasste. Inwieweit eine Aufforderung zu längeren Begründungen differenziertere Antworten zur Folge gehabt hätte, kann an dieser Stelle nicht festgestellt werden.

Ein vereinheitlichender Einfluss der Gruppenarbeit auf die Schwierigkeitsbeurteilungen und -begründungen konnte statistisch weitestgehend ausgeschlossen werden. Inwieweit sich jedoch die Gruppenarbeit darüber hinaus auf die individuellen Urteile und Begründungen auswirkt, ob beispielsweise das Urteil vor der Gruppenarbeit ein anderes war als danach, bleibt ebenfalls offen. Um dies näher in den Blick zu nehmen, können Erhebungen mit Hilfe des lauten Denkens (u. a. Sandmann, 2014) oder mittels einer Interaktionsanalyse (u. a. Krummheuer, 2011) Einflüsse der Interaktionen diesbezüglich klären.

Offen bleibt auch die Frage, inwiefern die Klassenstufe einen systematischen Einfluss auf Schwierigkeitsbeurteilungen hat. Dies könnte beispielsweise mittels stufenübergreifender Aufgaben in den Blick genommen werden. Zudem könnten längsschnittliche Betrachtungen bei denselben Lernenden weitere Einsichten in Entwicklungen bei Beurteilungs- und Begründungsprozessen bringen.

Die vorliegende Studie fokussiert auf die *Beurteilung* von Aufgaben hinsichtlich ihrer subjektiv empfundenen *Schwierigkeit*. Diese stellt im Rahmen selbst- und binnendifferenzierender Lernumgebungen ein wichtiges Kriterium zur Auswahl von Aufgaben dar. In der vorliegenden Studie werden jedoch weder Aussagen über das tatsächliche Aufgabenauswahlverhalten getroffen noch über weitere Faktoren, die für diese Auswahl relevant sein können (diese traten lediglich teilweise in den Begründungen der Schülerinnen und Schüler auf). Folgestudien könnten auf der Basis der vorliegenden Ergebnisse zum einen klären, wie das Verhältnis zwischen Auswahlverhalten und beurteilter subjektiver Aufgabenschwierigkeit ist. Zum anderen würde sich so die Möglichkeit ergeben, die subjektiv empfundene Aufgabenschwierigkeit als Einflussgröße zu weiteren möglichen Merkmalen wie dem personenspezifischen Interesse an einer Mathematikaufgabe, der geschätzten Bearbeitungsdauer, dem Erkennen notwendigen Vorwissens und nötiger Ressourcen zur Bearbeitung einer Aufgabe oder auch dem Erfassen von Analogien zwischen unterschiedlichen Aufgaben in Beziehung zu setzen.

### 6.3 Implikationen für die Schulpraxis

Wie bereits beschrieben, sind Lehrende wie auch Lernende es gewohnt, Mathematikaufgaben auf einer gedachten Schwierigkeitsskala von „leicht - mittel - schwer“ zu beurteilen (Drüke-Noe, 2018). Dies liegt vor allem an der in Schulbüchern üblichen Charakterisierung von Mathematikaufgaben entlang ihrer Schwierigkeit. Eine solche Zuordnung ermöglicht es Lernenden zum Beispiel, anschließend selbstständig entscheiden zu können, welche Aufgabentypen ihnen noch Schwierigkeiten bereiten und intensiver gelernt werden müssen (ebd., Drüke-Noe & Siller, 2018). Ferner ermöglicht erst diese Fähigkeit das niveaueingemessene Lernen in differenzierenden Lernumgebungen (Prediger & Scherres, 2012), welches vor allem im Sinne des selbstregulierten Lernens (z. B. in Freiarbeit, Wochenplanarbeit und ähnlichen Unterrichtskonzepten; u. a. Klein-Landeck, Karau & Landeck, 2016) erstrebenswert ist.

Die vorliegende Studie zeigt, dass die Schwierigkeitsurteile über Mathematikaufgaben sowie deren Begründungen von Schülerinnen und Schüler hochgradig individuell erfolgen. Lernende weisen die Teilaufgaben zwar auf einer Skala von „einfachste Teilaufgabe“ bis „schwierigste Teilaufgabe“ ein; die individuell verwendeten und in den Begründungen thematisierten subjektiv schwierigkeitsbeeinflussenden Faktoren, die der Reduktion zugrunde liegen, sind jedoch sehr vielfältig. Wie Krauthausen und Scherer (2007) berichten, nehmen Schülerinnen und Schüler Aufgabenmerkmale unterschiedlich wahr, so dass zwei Lernende aufgrund desselben wahrgenommenen Merkmals zu zwei unterschiedlichen Schwierigkeitsurteilen kommen können. Dies wurde auch in dieser Studie deutlich und stellt damit eine empirische Stütze für die Notwendigkeit differenzierenden Mathematikunterrichts dar (Benölken, Dixel & Berlinger, 2018). Die vorliegenden Ergebnisse zeigen, dass insbesondere die (durchgeführten oder antizipierten) Tätigkeiten der Aufgabenbearbeitung eine wichtige Rolle bei der Beurteilung spielen. Dies könnte z. B. im Unterricht durch die konsequente Nutzung und Reflexion von Operatoren und den damit einhergehenden Aufgabenanforderungen berücksichtigt werden.

Kennzeichnungen von besonders schwierigen Aufgaben (z. B. durch Sternchen; u. a. Büchter & Leuders, 2016, S. 105) können zwar eine Orientierung für Schülerinnen und Schüler darstellen, vor dem Hintergrund der empirischen Ergebnisse ist „schwierig“ jedoch relativ zu sehen. Dies liegt vor allem daran, dass Schülerinnen und Schüler aufgrund ihrer Kompetenzen auf unterschiedliche Weise mit schwierigkeitsbeeinflussenden Aufgabenmerkmalen

umgehen können (Krauthausen & Scherer, 2013; Leuders & Prediger, 2016).

Begegnen könnte man dem beispielsweise durch geeignete Vorwissensaktivierungen, um Schülerinnen und Schülern transparent zu machen, welche Inhalte und Tätigkeiten in den anstehenden Aufgaben gefordert werden oder mit welchen Darstellungsformen in einer Aufgabe operiert wird. Auf diese Weise erhalten die Schülerinnen und Schüler Hinweise und Anker, wie sie eine Aufgabenstellung, die sie „nicht verstanden“ haben, dennoch bearbeiten können. Weiterhin ist eine differenziertere Kennzeichnung von schwierigen Aufgaben denkbar, z. B. als Aufgaben für Experten „..., die mit besonders komplizierten Brüchen rechnen können“ oder „..., die keine Probleme mit Funktionsgraphen haben“ (Differenzierung der etablierten „Expertenaufgaben“, vgl. für Aufgaben zum Problemlösen Bruder & Collet, 2008, S. 13 f.). Ausprägungen solcher Parameter könnten – in einer altersgerechten Formulierung – Schülerinnen und Schülern als differenziertere Orientierungshilfe gegenüber der gröberen Sternchen-Kennzeichnung dienen.

Beurteilungsprozesse zu üben und Beurteilungskompetenzen auszubilden ist zentral für das Gelingen eines differenzierenden Mathematikunterrichts, der den Schülerinnen und Schülern Verantwortung für ihren eigenen Lernprozess überträgt. Aufgabenreflexionen (z. B. durch Checklisten und Selbstdiagnoseaufgaben; u. a. Leuders & Prediger, 2016, S. 180 ff.), Vergleiche von Vorher-Nachher-Schwierigkeitsurteilen (z. B. in Planung und Rückschau von Problemlöseprozessen), die Analyse konkreter Anforderungsprofile (u. a. Bruder & Collet, 2008) und die Verdeutlichung von Analogien zwischen Mathematikaufgaben (Aßmus & Förster, 2013; Aßmus, Fritzlar & Förster, 2018) stellen vielversprechende Aktivitäten für die Förderung von Beurteilungskompetenzen dar. Der gezielte Einsatz des in der Studie verwendeten Aufgabenformats der Blütenaufgaben kann dabei genutzt werden, Schülerinnen und Schüler zur Auseinandersetzung mit und Reflexion von Mathematikaufgaben und ihren spezifischen Unterschieden sowie Anforderungen anzuregen.

## Resümee

Die Ergebnisse zeigen sowohl die Vielfalt, die mit subjektiven Schwierigkeitsbeurteilungen und -begründungen einhergeht, als auch Aspekte, die sich in der Breite als weniger vielfältig erweisen. Vor allem aber zeigt die Studie, dass viele Fragen in Bezug auf die Beurteilung von Mathematikaufgaben durch Schülerinnen und Schüler noch offen und damit vielversprechende Gegenstände zukünftiger Forschungsvorhaben sind. Die Erkenntnisse der vorliegenden

Analysen können als Grundlage und Hypothesen für Folgeuntersuchungen dienen. Aufgabeneinschätzungen durch Schülerinnen und Schüler sind hinsichtlich der zunehmenden Relevanz von Differenzierung und selbstgesteuertem Lernen in der Unterrichtspraxis von besonderer Bedeutung und demnach ein gewinnbringendes Forschungsfeld für zukünftige Studien.

## Anmerkungen

<sup>1</sup> Hier sind Mischformen denkbar, die im Anteil des Entscheidungsspielraums der Lernenden variieren können.

<sup>2</sup> Selbststeuerung ist hier im Sinne des selbstgesteuerten Lernens nach Weinert (1982, S. 102) als eine Lernform zu verstehen, „bei der die Lernenden in hohem Maße beeinflussen können, ob, was, wann, wie und woraufhin sie lernen“.

<sup>3</sup> Untersuchungen weiterer aufgabenbezogener Urteile, neben dem hier fokussierten Schwierigkeitsurteil, erfolgen im Zuge der Dissertationsstudie von Getzin (i. V.). Zwischenergebnisse sind dokumentiert in Tagungsbeiträgen, s. Getzin (2017; 2018).

<sup>4</sup> Ostermann, Leuders und Nückles (2015) betrachten den so genannten Expert-Blind-Spot-Effekt, welcher die abnehmende Urteilsgenauigkeit der Schwierigkeitseinschätzung mit zunehmendem fachmathematischem Wissen beschreibt (vgl. auch Nathan & Koedinger, 2000).

<sup>5</sup> In der Untersuchung von Ostermann, Leuders und Nückles (2015) nimmt die „Verschätzung“ zwischen den Gruppen Studierende, Referendarinnen und Referendare sowie Lehrkräfte mit zunehmender Praxiserfahrung ab.

<sup>6</sup> Vereinzelt traten Gruppen bestehend aus fünf bis sieben Personen auf.

## Danksagung

Die vorgestellte Studie wurde im Rahmen des SINUS.NRW-Projektes vom Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen gefördert. An dieser Stelle sei allen Beteiligten des Projektes gedankt, insbesondere den Kolleginnen und Kollegen der teilnehmenden und mitwirkenden Schulen. Zudem danken wir den Codiererinnen und Codierern, die zu der Auswertung des umfangreichen Datenmaterials beigetragen haben.

## Literatur

- Artelt, C. (2009). Diagnostische Urteile von Lehrkräften im Bereich der Lesekompetenz. In A. Bertschi-Kaufmann & C. Rosebrock (Hrsg.), *Literalität: Bildungsaufgabe und Forschungsfeld* (S. 125–136). Weinheim: Juventa.
- Astleitner, H. (2008). Die lernrelevante Ordnung von Aufgaben nach der Aufgabenschwierigkeit. In J. Thonhauser, (Hrsg.), *Aufgaben als Katalysatoren von Lernprozessen* (S. 65–80). Münster u. a.: Waxmann.
- Aßmus, D. & Förster, F. (2013). ViStAD – Erste Ergebnisse einer Video-Studie zum analogen Denken bei

- mathematisch begabten Grundschulkindern. *mathematica didactica*, 36, 45–65.
- Aßmus, D., Förster, F. & Fritzlär, T. (2018). Ähnlichkeiten zwischen mathematischen Problemen aus Sicht von Grundschulkindern. In B. Rott, A. Kuzle & R. Bruder (Hrsg.), *Problemlösen unterrichten und untersuchen, Tagungsband der Herbsttagung des GDM-Arbeitskreises Problemlösen in Darmstadt 2017* (S. 21–35). Münster: WTM-Verlag.
- Benölken, R., Dixel, T. & Berlinger, N. (2018). Mathematikunterricht und Potentialorientierung. In M. Veber, R. Benölken & M. Pfitzner (Hrsg.), *Potentialorientierte Förderung in den Fachdidaktiken* (S. 43–59). Münster: Waxmann.
- Block, J. (2014). Eine didaktische Landkarte quadratischer Gleichungen als Konzeptualisierung für flexibles algebraisches Handeln. In J. Roth & J. Ames (Hrsg.), *Beiträge zum Mathematikunterricht 2014* (Bd. 1, S. 197–200). Münster: WTM-Verlag.
- Bong, M. & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15(1), 1–40.
- Bourier, G. (2011). *Beschreibende Statistik* (9. aktualisierte Auflage). Wiesbaden: Springer Fachmedien.
- Brophy, J. (2000). *Teaching*. Educational Practices Series, Vol. 1. Brüssel: International Academy of Education.
- Bruder, R. (2008): Vielseitig mit Aufgaben arbeiten. In R. Bruder; T. Leuders & A. Büchter (Hrsg.), *Mathematikunterricht entwickeln. Bausteine für kompetenzorientiertes Unterrichten* (S. 18–52). Berlin: Cornelsen Scriptor.
- Bruder, R. & Collet, C. (2011). *Problemlösen lernen im Mathematikunterricht*. Berlin: Cornelsen Scriptor.
- Bruder, R., Reibold, J. & Wehrse, T. (2013). *Binnendifferenziertes Aufgabenmaterial für den Mathematikunterricht der Sek. I*. Braunschweig: Schroedel.
- Büchter, A. & Leuders, T. (2005). Appropriate Problems for Learning and for Performing – an Issue for Teacher Training. *Zentralblatt für Didaktik der Mathematik*, 37(5), 343–350.
- Büchter, A. & Leuders, T. (2016). *Mathematikaufgaben selbst entwickeln, Lernen fördern – Leistung überprüfen*. Berlin: Cornelsen.
- Burghardt, H. & Swan, M. (2013). Task design for systemic improvement: principles and frameworks. In C. Margolinas (Hrsg.), *Task Design in Mathematics Education, Proceedings of ICMI Study 22* (S. 431–439). Oxford: HAL Open Access.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Cohors-Fresenborg, E., Sjuts, J. & Sommer, N. (2004). Komplexität von Denkvorgängen und Formalisierung von Wissen. In M. Neubrand (Hrsg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland, Vertiefende Analysen im Rahmen von PISA 2000* (S. 109–144). Wiesbaden: VS Verlag für Sozialwissenschaften.
- De Corte, E., Mason, L., Depaeppe, F. & Verschaffel, L. (2011). Self-regulation of mathematical knowledge and skills. In B. J. Zimmerman & D. H. Schunk (Hrsg.), *Handbook of self-regulation of learning and performance* (S. 155–172). New York/London: Routledge.
- Drücke-Noe, C. (2018). Einfach – mittel – schwierig ..., Wenn das so einfach wäre: Aufgaben unterschiedlichen Schwierigkeitsgrades entwickeln. *mathematik lehren*, 209, 9–12.
- Drücke-Noe, C. & Siller, H.-S. (2018). Aufgaben als Aufgabe. *mathematik lehren*, 209, 2–8.
- Efklides, A. & Vlachopoulos, S. P. (2012). Measurement of metacognitive knowledge of self, task, and strategies in mathematics. *European Journal of Psychological Assessment*, 28(3), 227–239.
- Flavell, J. H. (1986). Annahmen zum Begriff Metakognition sowie zur Entwicklung von Metakognition. In F. E. Weinert & R. H. Kluwe (Hrsg.), *Metakognition, Motivation und Lernen* (S. 23–31). Stuttgart: Kohlhammer.
- Gemeinsame Kommission Lehrerbildung der GDM, DMV & MNU (2017). Fachdidaktik für den inklusiven Mathematikunterricht. *GDM-Mitteilungen*, 103, 42–46.
- Getzin, M. (i. V.). *Mathematikaufgaben einschätzen – Entwicklung einer Grounded Theory über aufgabenbezogene Entscheidungen von Schülerinnen und Schülern*. Dissertation, Universität Osnabrück.
- Getzin, M. (2017). Die Einschätzung von Aufgaben und ihren Schwierigkeiten durch Lernende der Mathematik. In U. Kortenkamp & A. Kuzle (Hrsg.), *Beiträge zum Mathematikunterricht 2017* (S. 309–312). Münster: WTM-Verlag.
- Getzin, M. (2018). Mathematikaufgaben einschätzen – Wie blicken Schülerinnen und Schüler auf Aufgaben der Inhaltsbereiche Funktionen und Geometrie? In Fachgruppe Didaktik der Mathematik der Universität Paderborn (Hrsg.), *Beiträge zum Mathematikunterricht 2018* (S. 599–602). Münster: WTM-Verlag.
- Gilbert, S. J. & Shallice, T. (2002). Task switching: a PDP model. *Cognitive Psychology*, 44(3), 297–337.
- Goldstein, H. (2008). The Effect of Perceived Difficulty and Locus of Control on Task Performance. *The Huron University College Journal of Learning and Motivation*, 46(1), 42–56.
- Hadjidemetriou, C. & Williams, J. S. (2002). Teachers' pedagogical content knowledge: graphs, from a cognitivist to a situated perspective. In A. D. Cockburn & E. Nardi (Hrsg.), *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, S. 57–64). Norwich: University of East Anglia.
- Häsel-Weide, U. & Prediger, S. (2017). Förderung und Diagnose im Mathematikunterricht – Begriffe, Planungsfragen und Ansätze. In M. Abshagen, B. Barzel, J. Kramer, T. Riecke-Baulecke, B. Rösken-Winter & C. Selter (Hrsg.), *Basiswissen Lehrerbildung: Mathematik unterrichtet mit Beiträgen für den Primar- und Sekundarstufenbereich* (S. 167–181). Seelze: Friedrich/Klett Kallmeyer.
- Hasselhorn, M. & Gold, A. (2013). *Pädagogische Psychologie, Erfolgreiches Lernen und Lehren* (3., vollst. überarbeitete und erw. Auflage). Stuttgart: Kohlhammer.
- Heymann, H. W. (1991). Innere Differenzierung im Mathematikunterricht. *mathematik lehren*, 49, 63–66.
- Hoffmann, L. & Böhme, K. (2014). Wie gut können Grundschullehrkräfte die Schwierigkeit von Deutsch- und Mathematikaufgaben beurteilen? *Psychologie in Erziehung und Unterricht*, 1, 42–55.
- Hußmann, S. & Prediger, S. (2007). Mit Unterschieden rechnen – Differenzieren und Individualisieren. *Praxis der Mathematik in der Schule*, 49(17), 1–8.
- Janßen, T. (2016). *Ausbildung algebraischen Struktursinns im Klassenunterricht, Lernbezogene Neudeutung eines*

- mathematikdidaktischen Begriffs*. Bremen: Dokumentenserver Staats- und Universitätsbibliothek. Abgerufen von <https://media.suub.uni-bremen.de/bitstream/elib/1084/1/00105386-1.pdf>.
- Kaiser, A. & Kaiser, R. (2006). *Metakognition*. München: Luchterhand.
- Karst, K. (2012). *Kompetenzmodellierung des diagnostischen Urteils von Grundschullehrern*. Empirische Erziehungswissenschaft, Band 35. Münster: Waxmann.
- Klein-Landeck, M., Karau, C. & Landeck, I. (2016). *Mit Freiarbeit zu neuer Lernkultur* (2., überarbeitete Neuauflage). Berlin: Cornelsen.
- Krauthausen, G. & Scherer, P. (2007). *Einführung in die Mathematikdidaktik* (3. Auflage). Heidelberg: Spektrum.
- Krippner, W. (1992). *Mathematik differenziert unterrichten*. Schroedel: Hannover.
- Krummheuer, G. (2011). *Interaktionsanalyse*. Abgerufen von [http://www.fallarchiv.uni-kassel.de/wp-content/uploads/2010/07/krummheuer\\_inhaltsanalyse.pdf](http://www.fallarchiv.uni-kassel.de/wp-content/uploads/2010/07/krummheuer_inhaltsanalyse.pdf).
- Lamnek, S. & Krell, C. (2016). *Qualitative Sozialforschung*. Weinheim, Basel: Beltz Verlag.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Landmann, M., Perels, F., Otto, B. & Schmitz, B. (2009). Selbstregulation. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 49–72). Heidelberg: Springer.
- Lannie, A. L. & Martens, B. K. (2004). Effects of task difficulty and type of contingency on students' allocation of responding to math worksheets. *Journal of Applied Behaviour Analysis*, 37(1), 53–65.
- Lengnink, K. (2009). Vorstellungen bilden: Zwischen Lebenswelt und Mathematik. In T. Leuders, L. Hefendehl-Hebeker & H.-G. Weigand (Hrsg.), *Mathemagische Momente* (S. 120–129). Berlin: Cornelsen.
- Leuders, T. & Prediger, S. (2016). *Flexibel differenzieren und fokussiert fördern im Mathematikunterricht*. Berlin: Cornelsen.
- Leutner, D. (2004). Instructional-design principles for adaptivity in open learning environments. In N. M. Seel & S. Dijkstra (Hrsg.), *Curriculum, plans and processes in instructional design: international perspectives* (S. 289–307). Mahwah: Erlbaum.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse, Grundlagen und Techniken* (12., überarbeitete Auflage). Weinheim/Basel: Beltz.
- McElvany, N., Schroeder, S., Richter, T., Hachfeld, A., Baumert, J., Schnotz, W., Horz, H. & Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für Pädagogische Psychologie*, 23, 223–235.
- Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Auflage). Berlin/Heidelberg: Springer.
- Möller, J. & Trautwein, U. (2015). Selbstkonzept. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 177–199). Berlin/Heidelberg: Springer.
- Mudiappa, M. & Artelt, C. (Hrsg.) (2014). *BiKS – Ergebnisse aus den Längsschnittstudien, Praxisrelevante Befunde aus dem Primar- und Sekundarschulbereich*. Bamberg: University of Bamberg Press.
- Nathan, M. J. & Koedinger, K. R. (2000). An investigation of teachers' beliefs of students' algebra development. *Cognition and Instruction*, 18(2), 209–237.
- Neubrand, M., Klieme, E., Lüdtke, O. & Neubrand, J. (2002). Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung. *Unterrichtswissenschaft*, 30, 100–119.
- Nußbaum, A. & Leutner, D. (1986a). Die Auswirkung der Schwierigkeit textbegleitender Fragen auf die Lernleistung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 18(3), 230–244.
- Nußbaum, A. & Leutner, D. (1986b). "Learning by discovery" using rules of varying difficulty. *The German Journal of Psychology*, 10, 309.
- Ostermann, A., Leuders, T. & Nückles, M. (2015). Wissen, was Schülerinnen und Schülern schwer fällt, Welche Faktoren beeinflussen die Schwierigkeitseinschätzung von Mathematikaufgaben? *Journal für Mathematik-Didaktik*, 36, 45–76.
- Pekrun, R. & Zirngibl, A. (2004). Schülermerkmale im Fach Mathematik. In M. Prenzel (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland* (S. 191–210). Münster: Waxmann.
- Prediger, S. (2008). Mit der Vielfalt rechnen – Aufgaben, Methoden und Strukturen für den Umgang mit Heterogenität im Mathematikunterricht. In S. Hußmann, A. B. Liegmann, E. Nyssen, K. Racherbäumer & C. Walzebug (Hrsg.), *Individualisieren, differenzieren, vernetzen* (S. 129–139). Hildesheim: Franzbecker.
- Prediger, S. & Scherres, C. (2012). Niveaugemessenheit von Arbeitsprozessen in selbstdifferenzierenden Lernumgebungen – Qualitative Fallstudie am Beispiel der Suche aller Würfelnetze. *Journal für Mathematik-Didaktik*, 33(1), 143–173.
- Prenzel, M., Häußler, P., Rost, J. & Senkbeil, M. (2002). Der PISA-Naturwissenschaftstest: Lassen sich die Aufgabenschwierigkeiten vorhersagen? *Unterrichtswissenschaft*, 30(2), 120–135.
- Rathgeb-Schnierer, E. & Rechtsteiner, C. (2018). *Rechnen lernen und Flexibilität entwickeln, Grundlagen – Förderung – Beispiele*. Berlin/Heidelberg: Springer Spektrum.
- Salle, A., vom Hofe, R. & Pallack, A. (2014). Differenzierter Unterricht mit Blütenaufgaben. In J. Roth & J. Ames (Hrsg.), *Beiträge zum Mathematikunterricht 2014* (Bd. 2, S. 1047–1050). Münster: WTM-Verlag.
- Salle, A. (2015). Blütenbildung bei Minusgraden. *Mathematik 5–10*, 33, 18–21.
- Salle, A. (2016). Mathematisch kreativ werden – Back to the roots: Aufgabenvariation in Blütenaufgaben. *mathematik lehren*, 195, 29–33.
- Sandmann, A. (2014). Lautes Denken – die Analyse von Denk-, Lern- und Problemlöseprozessen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 179–188). Berlin/Heidelberg: Springer Spektrum.
- Scherres, C. (2013). *Niveaugemessenes Arbeiten in selbstdifferenzierenden Lernumgebungen, Eine qualitative Fallstudie am Beispiel einer Würfelnetz-Lernumgebung*. Berlin/Heidelberg: Springer Spektrum.
- Schmitz, G. S. & Schwarzer, R. (2000). Selbstwirksamkeitserwartung von Lehrern. *Zeitschrift für Pädagogische Psychologie*, 14(1), 12–25.
- Schneider, W. & Artelt, C. (2010). Metacognition and mathematics education. *ZDM – The International Journal on Mathematics Education*, 42, 149–161.

- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1, 27–52.
- Schrader, F.-W. (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(3), 237–245.
- Schrader, F.-W. (2011). Lehrer als Diagnostiker. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 683–698). Münster: Waxmann.
- Schukajlow, S. (2011). *Mathematisches Modellieren, Schwierigkeiten und Strategien von Lernenden als Bausteine einer lernprozessorientierten Didaktik der neuen Aufgabenkultur*. Münster u. a.: Waxmann.
- Schwanzer, A. D., Trautwein, U., Lüdtke, O. & Sydow, H. (2005). Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener. *Diagnostica*, 51(4), 183–194.
- Selter, C. (2000). Vorgehensweisen von Grundschüler(inne)n bei Aufgaben zur Addition und Subtraktion im Zahlenraum bis 1000. *Journal für Mathematik-Didaktik*, 21, 227–258.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19(1–2), 85–95.
- Südkamp, A., Möller, J. & Pohlmann, B. (2008). Der Simulierte Klassenraum, Eine experimentelle Untersuchung zur diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 22(34), 261–276.
- Thonhauser, J. (2008). Warum (neues) Interesse am Thema „Aufgaben“? In J. Thonhauser (Hrsg.), *Aufgaben als Katalysatoren von Lernprozessen* (S. 13–27). Münster u. a.: Waxmann.
- Thonhauser, J., Buschmann, I., Gastager, A. U.; Moser, G. & Schmich, J. (2003). Benchmarks von innen – ein Projekt in progress. *Lernende Schule*, 6(24), 57–59.
- Vock, M. & Gronostaj, A. (2017). *Umgang mit Heterogenität in Schule und Unterricht*. Berlin: Friedrich-Ebert-Stiftung.
- Weinert, F. E. (1982). Selbstgesteuertes Lernen als Voraussetzung, Methode und Ziel des Unterrichts. *Unterrichtswissenschaft*, 10(2), 99–110.
- Wirtz, M. A. & Kutschmann, M. (2007). Analyse der Beurteilerübereinstimmung für kategoriale Daten mittels Cohens Kappa und alternativer Maße. *Die Rehabilitation*, 46(6), 370–377.

Anschrift der Verfasser

Alexander Salle  
Universität Osnabrück  
Institut für Mathematik  
Albrechtstr. 28a  
49076 Osnabrück  
alexander.salle@uni-osnabrueck.de

Mirko Getzin  
Universität Osnabrück  
Institut für Mathematik  
Albrechtstr. 28a  
49076 Osnabrück  
mirko.getzin@uni-osnabrueck.de

Rudolf vom Hofe  
Universität Bielefeld  
Fakultät für Mathematik  
Institut für Didaktik der Mathematik  
Universitätsstr. 25  
33615 Bielefeld

## Anhang

**Melanie, Peter, Vladislav und Irem möchten einen Erdbeermilchshake mixen. Sie finden folgendes Rezept:**

**Erdbeermilchshake**

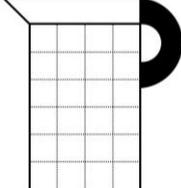


**Rezept (für vier große Shakes)**

Zutaten

- $\frac{2}{3}$  Liter Milch
- $\frac{1}{6}$  Liter süße Sahne
- $\frac{1}{6}$  Liter Erdbeermark (aus 250 g Erdbeeren)
- 3 Esslöffel Zucker

♣ Im Rezeptbuch sollen die Anteile der flüssigen Zutaten (einschließlich dem Erdbeermark) als Füllung einer 1-Liter-Karaffe verdeutlicht werden. Zeichne die Karaffe in dein Heft und trage im Bereich der Rechteckfläche die entsprechenden Füllhöhen der einzelnen Anteile übereinander ein.



♥ Melanie bereitet für sich und ihre beste Freundin Erdbeermilchshakes zu. Dazu verwendet sie das obige Rezept und schreibt es für zwei Personen um. Welche neuen Mengen erhält sie? Notiere deine Ergebnisse als Rezept für zwei Personen.

♠ Denke dir ein eigenes Milchshake-Rezept für 6 Personen aus und gib deine Zutaten mit Bruchzahlen an. Wie viel bekommt jede Person?

♦ Ein übliches Trinkglas fasst 0,2 l. Wie viele Gläser erhält man aus dem Rezept für vier große Shakes?

Darstellung der Blütenaufgabe „Erdbeermilchshake“ (Aufgabe 6.1)