

Ein zeitbeschränkter Test zur schnellen Erkennung von Schülerfehlern durch Junglehrkräfte – Qualitative Merkmale zur Schwierigkeitsbestimmung von Items

LENA PANKOW, HAMBURG; GABRIELE KAISER, HAMBURG & BRISBANE

Zusammenfassung: Zur Förderung von Lernchancen ist ein konstruktiver Umgang mit Schülerfehlern nötig, der das schnelle Erkennen von Schülerfehlern voraussetzt. Im Test zur schnellen Fehlererkennung, der im Rahmen der TEDS-FU-Studie durchgeführt wurde, waren Junglehrkräfte gefordert, typische Schülerfehler unter Zeitdruck zu erkennen. Zwei Merkmale – aus mathematikdidaktischer und aus wahrnehmungspsychologischer Perspektive – können in den beiden Tests für die Primar- und Sekundarstufe als schwierigkeitsgenerierend identifiziert und validiert werden. Die identifizierten Merkmale können als Analyseinstrument bei der Weiter- oder Neuentwicklung von Tests zur schnellen Fehlererkennung eingesetzt werden und bieten eine inhaltliche Grundlage zur Erklärung der Itemschwierigkeit.

Abstract: In order to promote learning opportunities, a fast recognition of students' errors is a necessary prerequisite for a constructive approach to students' errors. In the time-limited test for identifying students' errors from the TEDS-FU study, early career teachers were requested to recognize typical students' errors under time constraints. In the two tests for primary and secondary teachers, two characteristics could be identified as difficulty generating aspects, focusing the perspective of mathematics didactics and the psychology of perception. The identified characteristics can be used as an analysis tool for a new test for fast recognition of students' errors in the design process as well as the basis for the explanation of item difficulty.

1. Einleitung und Forschungsfrage

Lehrkräfte haben tagtäglich mit Schülerfehlern aller Art zu tun, die sich im Mathematikunterricht z. B. in systematische und unsystematische Fehler unterscheiden lassen (Radatz, 1980). Die schnelle Wahrnehmung dieser Fehler im Unterricht und der adäquate Umgang mit den Fehlern sind für Lehrkräfte bedeutsam. Die Fehler weisen auf den Lernprozess der Schüler*innen hin und „lassen erkennen, wie das Denken funktioniert“ (Beutelspacher, 2008, S. 87). Schoy-Lutz (2005) konnte zeigen, dass in 75 % der Schülerfehler im Mathematikunterricht echte Lerngelegenheiten verborgen sind. Voraussetzung zum Erkennen dieser Lerngelegenheiten ist das schnelle

Identifizieren der Schülerfehler. Die Lehrkräfte müssen darüber hinaus, um diese Fehler als Lerngelegenheiten nutzen zu können, entscheiden, ob ein Fehler für alle Lernenden von Bedeutung ist oder ob dieser individuell zu klären ist. Neben der benötigten Kompetenz der Klassenführung muss die Lehrkraft daher flexibel mit der Fehlersituation umgehen (vgl. Klug et al., 2013) und zudem über das mathematische und mathematikdidaktische Wissen verfügen, um die Fehler nachzuvollziehen und einordnen zu können. Im Test zur schnellen Fehlererkennung, der im Rahmen der nationalen Folgestudie der internationalen Vergleichsstudie „Teacher Education and Development Study in Mathematics – TEDS-M 2008“ (Blömeke et al., 2010) erhoben wurde, wurde die Kompetenz zur schnellen Fehlererkennung von Junglehrkräften im Fach Mathematik als ein Indikator für Lehrerexpertise erfasst, die im Folgenden vorgestellt wird (für Details zur TEDS-FU Studie siehe Blömeke et al., 2014). TEDS-FU setzte bei Lehrkräften der Primar- und der unteren Sekundarstufe zwei mathematisch unterschiedliche, aber strukturgleiche Tests zur schnellen Fehlererkennung ein. Es stellt sich a posteriori die Frage nach den schwierigkeitsgenerierenden Merkmalen der einzelnen Items, d. h. welche Merkmale eines Items machen dieses schwer bzw. leicht. Des Weiteren ist zu klären, ob diese Merkmale über die verschiedenen Lehramtstypen stabil sind, d. h. ob sich die gleichen Merkmale sowohl für Primarstufen- als auch für Sekundarstufenlehrkräfte zeigen.

Auf Basis einer mathematikdidaktischen Perspektive, aber auch einer wahrnehmungspsychologischen Perspektive aus der testbezogenen Diskussion über Fehlererkennungstests werden im Folgenden schwierigkeitsgenerierende Merkmale von Items theoriegeleitet entwickelt, die anhand der empirischen Ergebnisse überprüft werden. Die Analysen erlauben, zukünftig a priori die Itemschwierigkeit ähnlich konzipierter Tests zu bestimmen und bieten so die Möglichkeit, diese theoriegeleitet zu entwickeln. Zudem soll mit dem Artikel ein Beitrag zur Überprüfung der Güte des im Rahmen der TEDS-FU Studie entwickelten Testinstruments geleistet werden.

2. Theoretischer Rahmen

2.1 Lehrerprofessionswissen, Fehler und deren Erkennung

Die kognitiven Voraussetzungen, die Lehrkräfte für den Lehrberuf mitbringen müssen, werden in vielen Studien – wie in der COACTIV Studie (Kunter et al., 2011) oder in Studien der Michigan-Group (Hill, Ball & Schilling, 2008) – untersucht. In der TEDS-FU-Studie werden in Anlehnung an Shulman (1986) drei Wissensdimensionen des Lehrerprofessionswissens ausdifferenziert. Aufgrund ihrer empirisch gezeigten Relevanz für schnelle Fehlererkennung (Blömeke et al., 2014) werden im Folgenden zwei der drei Facetten genauer dargestellt. Die dritte Facette, General Pedagogical Knowledge (GPK), ist unter der Perspektive der Erkennung mathematischer Fehler kaum bedeutsam und wird deshalb im Folgenden nicht berücksichtigt.

Zum fachlichen Wissen, dem Content Knowledge (CK), gehört nach Shulman (1986) das Wissen, das den Inhalt umschließt und sich nicht auf reines Faktenwissen reduzieren lässt. Darüber hinaus soll die Lehrkraft auch die innere Struktur der Mathematik, die Grammatik und das Warum verstehen können (vgl. Shulman, 1986, S. 9). Für die Mathematik wird diese Wissenskomponente als *Mathematical Content Knowledge* (MCK) bezeichnet. Weiterhin unterscheidet Shulman (1986) das fachspezifisch-pädagogische Wissen, *Pedagogical Content Knowledge* (PCK), für die Mathematik *Mathematical Pedagogical Content Knowledge* (MPCK). Dieses Wissen beinhaltet

for the most regularly taught topics in one's subject area, the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations – in a word, the ways of representing and formulating the subject that make it comprehensible to others (Shulman, 1986, S. 9).

Des Weiteren hebt Shulman (1986, S. 9 f.) hervor, dass die Lehrkraft in Situationen, in denen sie mit Schülerfehlern konfrontiert wird, Strategien zur Behebung der Fehlvorstellungen haben muss, um das Wissen der Schüler*innen reorganisieren zu können.

Auch im Kompetenzmodell der COACTIV-Studie spielt das Erkennen von Schülerfehlern und der Umgang damit eine bedeutende Rolle, und zwar im fachdidaktischen Wissen, welches im Rahmen der Dimension Wissen über Schülervorstellungen ausdifferenziert wird in Wissen über „Fehlkonzeptionen, typische Fehler, Strategien“ (Kunter et al., 2011, S. 38).

Fehler und der Umgang mit diesen spielen insbesondere im Kontext des Lernens von Mathematik eine

große Rolle. Nach Oser, Hascher und Spychiger (1999) ist ein Fehler ein „von der Norm abweichender Sachverhalt oder ein von der Norm abweichender Prozess [...]“. Normen stellen also das Bezugssystem dar und ohne Normen bzw. Regeln wäre es nicht möglich, fehlerhafte und fehlerfreie Leistungen, das Richtige vom Falschen zu unterscheiden“ (S. 11). Heinze (2004) konkretisiert für den Mathematikunterricht Fehler als eine „Äußerung, die gegen die allgemeingültigen Aussagen und Definitionen der Mathematik sowie gegen allgemein akzeptiertes mathematisch-methodisches Vorgehen verstößt“ (S. 223). Aufgrund der hohen Relevanz für mathematisches Lernen wird im Folgenden auf den Fehler des regelhaften Missverständnisses fokussiert; sog. Flüchtigkeitsfehler, die ggf. aufgrund mangelnder Aufmerksamkeit oder einer andersartigen Störung der Ausführung der Prozedur auftreten (Malle, 1993), werden aufgrund ihrer fehlenden Regelmäßigkeit nicht berücksichtigt.

Eine Fehlersituation wird von Seifried und Wutke (2010) sowie von Heinze (2004) durch das Erkennen eines Fehlers definiert. Somit bildet die Identifikation des Fehlers bereits den ersten Schritt vor der vertieften Analyse des Fehlers (ähnlich Seidel & Prenzel, 2003). Die Wahrnehmung von Schülerfehlern ist darüber hinaus auch von dem erlernten Wissen über Schülerfehler beeinflusst (vgl. König et al., 2014). Aufgrund der Notwendigkeit, im Unterricht schnell zu reagieren, fokussieren wir im Folgenden auf schnelle Fehlererkennung als Voraussetzung für ein kompetentes Unterrichtshandeln und damit als eine zentrale Facette professioneller Kompetenz von Lehrkräften. Inhaltlich kann schnelle Fehlererkennung als eine Subkompetenz des umfassenderen Konstrukts der Diagnosekompetenz von Lehrkräften konzeptualisiert werden (Südkamp & Praetorius, 2017), die nach dem Ansatz von Blömeke, Gustafsson und Shavelson (2015) zur Kompetenz als Kontinuum als Teil der professionellen Unterrichtswahrnehmung, dem sog. Noticing, gezählt werden kann.

2.2 Itemschwierigkeit

Um die Kompetenzfacette ‚Schnelle Fehlererkennung‘ bei Lehrkräften empirisch erfassen zu können, ist es entweder möglich, die potentiell zu erkennenden Fehler induktiv aus vorliegendem Datenmaterial zu identifizieren oder deduktiv-systematisch aus der Literatur abzuleiten. Wir haben uns aufgrund fehlender Daten für die zweite Vorgehensweise entschieden. Eine wesentliche Komponente der Klassifikation der Itemschwierigkeit liegt in der Unterscheidung in eine formale und inhaltliche Komponente. Nach Kauertz (2008) bewirkt die formale Struktur in

Kompetenztests eine Schwierigkeit, die gezielt variiert oder konstant (wie in diesem Test: bei allen Items muss die falsche Schülerlösung identifiziert werden) gehalten werden solle. Besonders soll auf die inhaltlichen Aufgabenmerkmale der Items fokussiert werden, denen Schumann und Eberle (2011) ein bedeutendes Potenzial zur Aufklärung zuschreiben. Schumann und Eberle (2011) heben hervor, dass es möglich sein sollte, in einer nachträglichen Analyse der Items (post-Identifikation) eines Tests die beobachteten Itemschwierigkeiten in (fach-) didaktischer Hinsicht sinnvoll zu interpretieren. Auch Hartig (2007) betont dies als Ansatz zur Vorhersage schwierigkeitsbestimmender Aufgabenmerkmale, um etwas über die Validität eines Kompetenztests aussagen zu können. Angelehnt an Arbeiten von Cohors-Fresenborg et al. (2004) zu schwierigkeitgenerierenden Merkmalen zur Einschätzung der Aufgabenschwierigkeit bei der PISA-Studie werden im Folgenden Merkmale entwickelt, die auf die Items des Tests zur schnellen Fehlererkennung von TEDS-FU angewendet werden sollen.

Das erste Merkmal aus einer mathematikdidaktisch geprägten Perspektive untersucht die curriculare Einführung des Themas und Häufigkeit des curricularen Auftretens im nachfolgenden Mathematikunterricht (im Folgenden als *curriculare Relevanz* bezeichnet). Die Inhalte der Schulmathematik werden im deutschen Mathematikunterricht im Sinne des Spiralprinzips nach Bruner (1973) zu verschiedenen Zeitpunkten und verschiedenen Kontexten wieder aufgegriffen (Brinkmann, 2002). Fehler zu Themenfeldern, die in den unteren Klassenstufen unterrichtet werden, können auf diese Weise häufiger von Lehrkräften beobachtet werden, als jene, die erst in höheren Schulstufen eingeführt werden. Items, die typische Fehler zu Themenfeldern darstellen, die in unteren Klassenstufen unterrichtet werden, sollten somit zu leichteren Items gehören, als jene, die in höheren Klassenstufen zum ersten Mal unterrichtet werden. Im Sinne des Spiralprinzips sollten ebenfalls die Items leichter lösbar sein, die nach dessen Einführung häufig unterrichtet werden, da sie in anderen Themenfeldern involviert sind.

Um die Schwierigkeit hinsichtlich des ersten Auftretens des Themas bis hin zur Wiederholung des Themas im weiteren Verlauf des Mathematikunterrichts einordnen zu können, wurden die Bildungspläne der einzelnen Bundesländer analysiert. Die Themen, die in den Bildungsplänen der Bundesländer der Primarstufe sowie der unteren Sekundarstufe für das Fach Mathematik aufgeführt werden, unterscheiden sich nur wenig in der Reihenfolge der Themengebiete, ggf. aufgrund der starken Orientierung an den von der

KMK erlassenen Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss bzw. die Primarstufe, weshalb im Folgenden beispielhaft auf die Lehrpläne von Hamburg zurückgegriffen wird (vgl. Freie und Hansestadt Hamburg Behörde für Schule und Berufsbildung, 2012 a, b).

Das zweite Merkmal aus einer wahrnehmungspsychologischen Perspektive spiegelt den strukturellen Aufbau der Items wider. In dem genannten Merkmal (im Folgenden als *strukturelle Komplexität* bezeichnet) werden die Zeichen als Objekte – ihre Definition und ihre Grammatik (Rechenregeln/Gesetze) – näher betrachtet. Unter Bezug auf die in den Bildungsstandards formulierte allgemeine Kompetenz des Umgangs mit den symbolischen, formalen und technischen Elementen der Mathematik, gehen wir davon aus, dass die Zahlen und Symbole, aus denen die Items bestehen, seitens der Teilnehmenden wahrgenommen, interpretiert und im Gesamtzusammenhang beantwortet werden müssen. Dabei ist zu bemerken, dass die mathematischen Ausdrücke häufig in Gruppen oder Clustern zusammengefügt werden, die einer anderen Logik folgen als die der geschriebenen Sprache (vgl. Dyrvold, 2016, S. 11). Die Cluster, aus denen die Darstellungen der Items bestehen, sollen für die Analyse herangezogen werden, wobei hierarchische Beziehungen berücksichtigt werden (vgl. Drolling-Vetter, 2011). Bei dem Gebrauch der Zeichen muss jede*r Lernende die Regeln zum Gebrauch der Zeichen – in diesem Fall nach den Konventionen der Mathematik – erwerben, um sie entsprechend verwenden zu können, gleiches gilt für die Verwendung von Regeln (vgl. Brunner, 2013, S. 54). Um die Itemschwierigkeit mittels des zweiten Merkmals zu identifizieren, wird im ersten Schritt die Anzahl der verwendeten Zeichen, aus denen die Items bestehen, berücksichtigt. Die Zeichenlänge wird als komplexitätsgenerierend angesehen, da innerhalb einer kurzen Testsituation (4 Sekunden) alle Zeichen wahrzunehmen sind, was eine hohe Anforderung an die Reaktionsfähigkeit, die kognitive Aufnahme- und Verarbeitungsgeschwindigkeit darstellt. Dabei ist zu bemerken, dass Bedeutungseinheiten wie die Buchstabenfolge ‚sin‘ als ein Zeichen gewertet werden, während die Zahl 0,5 als zwei gesonderte Zeichen gezählt werden, da die Ziffern gesondert in das Stellenwertsystem eingeordnet werden müssen und damit zusätzliche kognitive Verarbeitungsschritte nötig sind. Im zweiten Schritt werden die zu verwendeten Rechenregeln, die zur Lösung des Items benötigt werden, festgestellt.

3. Methode

3.1 Studie und Stichprobe

TEDS-FU ist eine deutsche Folgestudie zu der 2008 durchgeführten internationalen Lehrerbildungsstudie TEDS-M (Teacher Education and Development Study in Mathematics). Die 2012 durchgeführte Studie TEDS-FU (vgl. Blömeke et al., 2014) erhob die Kompetenz von Junglehrkräften nach vier Jahren Berufserfahrung. TEDS-FU wurde in Form einer Onlinestudie mithilfe des vom DIPF (Deutsches Institut für Internationale Pädagogische Forschung) entwickelten CBA ItemBuilder (Rölke, 2012) durchgeführt. An zwei Tests, die für die Primarstufen- und Sekundarstufenlehrkräfte strukturgleich gestaltet, aber inhaltlich an das studierte Lehramt angepasst sind, nahmen 171 Sekundar- und 131 Primarstufenlehrkräfte teil. Neben den Instrumenten zur Erhebung des mathematischen Fachwissens, des fachdidaktischen Wissens, des allgemeinpädagogischen Wissens und ausgewählter Belief-Facetten aus TEDS-M, wurden Tests zur Erfassung der situierten Facetten von Lehrerprofessionalität neu entwickelt und eingesetzt (vgl. Blömeke et al., 2014; Kaiser et al., 2017). Neben diesen videobasierten Instrumenten (vgl. Kaiser et al., 2015) wurden wie in der COACTIV-Studie Tests zur Identifikation von Schülerfehlern entwickelt (Krauss et al., 2008; Krauss & Brunner, 2011), die im Gegensatz zur COACTIV-Studie die schnelle Erkennung von Schülerfehlern (mit Zeitbegrenzung) fokussierten.

3.2 Testinstrument und Analyse der Items

Der neu entwickelte Test zum schnellen Erkennen von Schülerfehlern intendierte eine möglichst unterrichtsnahe Erfassung von Schülerfehlern. Da die Lehrperson üblicherweise das von ihr im Unterricht behandelte Themengebiet kennt, wurde vorweg das Themengebiet, aus dem die gezeigten Fehler kommen würden, eingeblendet mit der Aufforderung, typische Fehler zu antizipieren (vgl. Pankow et al., 2016).

Der Sekundarstufentest besteht aus 16 und der Primarstufentest aus 15 typischen Schülerfehlern, die auf einschlägig bekannte Fehlkonzepte zurückzuführen und für mathematisches Lehren und Lernen bedeutsam sind.

Im Anschluss an die Antizipationsphase wurden den Probanden drei Schülerlösungen präsentiert, von der eine Lösung einen typischen Schülerfehler im Kontext des angekündigten Themengebietes enthält, während die anderen beiden Lösungen korrekt sind.

Die Buchstaben, zu den dargestellten Schülerlösungen werden über die Tastatur der Probanden eingegeben, und zwar über die Buchstaben S, D, F, auf denen die Finger der linken Hand während des Tests ruhen sollen. Im ersten Schritt wurde der typische Schülerfehler angekündigt, z. B. „Anwenden des Satzes des Pythagoras“ mit einem abgebildeten Dreieck. Es wurde maximal 5 Minuten Zeit gegeben, um sich mögliche Schülerfehler in diesem Bereich zu vergegenwärtigen.

Im zweiten Schritt, der Identifikation der falschen Schülerlösung, wurden drei Lösungen gezeigt und es musste die Taste mit der falschen Schülerlösung gedrückt werden. Um zu verhindern, dass die Aufgaben nachgerechnet werden konnten, wurde die Zeit zur Antwort auf vier Sekunden beschränkt. Die Zeit, die einer Expertenbefragung zufolge benötigt wird, um die Ergebnisse nachzurechnen, liegt bei ca. fünf bis sechs Sekunden. Es wurde bei der Testkonstruktion davon ausgegangen, dass erfahrene Lehrkräfte wissen, welche typischen Fehler in dem betreffenden Themengebiet zu erwarten sind, d. h. ihre Fehlerstrategie ist strukturiert, sodass sie den Fehler ohne Rechnung schnell, auf einen Blick erkennen. Innerhalb von vier Sekunden musste daher der zu dem Fehler zugehörige Buchstabe gedrückt werden, um die Antwort in das System einzugeben. Nach dieser Zeit war es nicht länger möglich, eine Antwort einzugeben, und das Item wurde als falsch bewertet. Als Leistungsindikator wurde die Anzahl der richtigen Antworten herangezogen.

Im Folgenden wird eine exemplarische Auswahl der Sekundarstufenitems bzgl. ihrer Schwierigkeit analysiert (Abb. 1). Das erste Item (Nr. 15) befasst sich mit der Addition zweier Brüche, da empirisch aufgezeigt wurde, dass 45 % aller Schülerfehler auf dieses Vorgehen zurückgehen, d. h. es werden Zähler und Nenner getrennt addiert (vgl. Wartha, 2007).

Weiterhin konnte Padberg (2009) zeigen, dass einige Schüler*innen ein eigenes Regelwerk (vgl. Allmendinger et al., 2013, 81 f.) zur Addition zweier Brüche entwickeln. Eichelmann, Narciss, Schnaubert und Melis (2012) weisen drauf hin, dass Schüler*innen konzeptuelles Wissen über Brüche besitzen müssen, unter anderem zu der Bedeutung des Zählers und des Nenners, aber auch zur adäquaten Bruchvorstellung selbst (vgl. Eichelmann et al. 2012, S. 30).

Das zweite Item (Nr. 14) zur Umformung einfacher Zahlterme mit Klammern beinhaltet zwei primäre Schwierigkeiten: Zum einen ist das Ausmultiplizieren eines Terms, in dem Klammern enthalten sind, komplexer als bei einem Term ohne Klammern, da zwei Rechenschritte sowie eine Verknüpfung (Distributivgesetz) zu beachten sind.

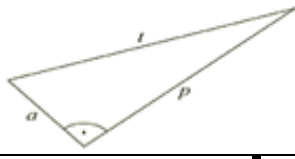
	Nr.	Ankündigung	„S“	„D“	„F“
Sekundarstufentest	15	Bruchrechnung: Addition zweier Brüche	$\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$	$\frac{2}{7} + \frac{3}{7} = \frac{5}{14}$	$\frac{1}{3} + \frac{2}{3} = 1$
	14	Umformung einfacher Zahlterme mit Klammern	$3 \cdot (5 + 2) = 3 \cdot 5 + 3 \cdot 2$	$3 + (5 + 2) = 3 + 5 + 2$	$3 - (5 + 2) = 3 - 5 + 2$
	7	Trigonometrie: Werte spezieller Winkel	$\sin 30^\circ = 0,5$	$\cos 0^\circ = 0$	$\tan 45^\circ = 1$
	8	Anwenden der <i>p-q</i> -Formel bei der quadratischen Gleichung $x^2 + 8x - 4 = 0$	$x^2 + 8x - 4 = 0$		
			$x_{1,2} = -4 \pm \sqrt{16 + 4}$	$x_{1,2} = -4 \pm \sqrt{16 - (-4)}$	$x_{1,2} = -4 \pm \sqrt{16 - 4}$
	5	Anwenden des Satzes des Pythagoras			
			$p^2 + a^2 = t^2$	$t^2 = a^2 + p^2$	$a^2 + t^2 = p^2$

Abb. 1: Auswahl (5 von 16) der Items zur schnellen Fehlererkennung aus dem TEDS-FU Sekundarstufentest

Zum anderen müssen Vorzeichen, wie in dem hier vorliegenden Fall, berücksichtigt werden, die nach Lüddecke (2015) auch unter anderem bei Problemlösungsprozessen Lösungsbarrieren beinhalten.

Das dritte Item (Nr. 7) aus der Trigonometrie kann unter zwei Sichtweisen interpretiert werden, was in der Literatur bisher kaum diskutierte Schwierigkeiten hervorrufen kann. So kann das Item einerseits funktional interpretiert werden als Frage nach den Funktionswerten von trigonometrischen Funktion zu speziellen Winkeln gegeben im Gradmaß. Andererseits ist auch eine geometrische, nichtfunktionale Auffassung möglich als Frage nach dem Verhältnis von Kathete und Hypothese im rechtwinkligen Dreieck zu speziellen Winkeln. Ein typischer Schülerfehler ist eine Verwechslung von $\cos 0^\circ$ mit $\sin 0^\circ$, sodass $\cos 0^\circ = 0$ angegeben wird.

Das vierte Item (Nr. 8) enthält einen typischen Schülerfehler, der bei der Lösung einer quadratischen Gleichung mittels der sog. *p-q*-Formel $x_{1,2} = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}$ entstehen kann. Neben einer möglichen fehlenden Normierung (vgl. Ritter & Voß 2015, 82 f.) spielen die Vorzeichen der Parameter „*p*“ und „*q*“ eine zentrale Rolle (Goebbels & Ritter 2013, S. 55), die häufig nicht beachtet werden.

Das fünfte Item (Nr. 5) zum Anwenden des Satzes des Pythagoras beinhaltet den wohlbekannten Fehler der Vertauschung von Katheten und Hypotenuse im

rechtwinkligen Dreieck, die durch die Nichtbeachtung der Definition der geometrischen Begriffe hervorgerufen sein kann oder durch die Fixierung auf übliche geometrische Darstellungen rechtwinkliger Dreiecke mit einer Kathete parallel zum unteren Rand der Darstellungsfläche (vgl. Draschoff, 2000).

3.3 Methodisches Vorgehen

Die Items beider Tests zur schnellen Fehlererkennung wurden basierend auf dem einparametrischen Raschmodell skaliert. Die EAP/PV-Reliabilität erreicht im Falle des Sekundarstufentests einen Wert von 0,62 und im Falle der Primarstufe von 0,64. Beide können als ausreichend angesehen werden (vgl. Magenheim et al., 2015; Robinson et al., 1991).

Zur Beantwortung der Forschungsfrage wird eine Analyse der Items durchgeführt, die bzgl. der Itemschwierigkeit extreme Ausprägungen aufweisen, da sie einerseits besonders einfach bzw. besonders schwer sind. Ein besonderer Fokus auf extreme Ausprägungen eines Merkmals oder auf Extremgruppen ist ein in der qualitativen Forschung weitverbreitetes Vorgehen. So werden z. B. im Rahmen der Typenbildung Teilgruppen gebildet, die hinsichtlich gewisser Merkmale besonders ähnlich sind, um Einsicht in allgemeinere Strukturen und Muster zu erhalten (vgl. Kelle & Kluge, 2010, S. 83 ff.). Auch in der quantitativen Forschung werden Items mit extremer Schwierigkeit bei der finalen Analyse ausgeschlossen, da sie aufgrund schwacher Diskrimination nicht

substantiell zur Messung des Konstrukts beitragen (vgl. Bond & Fox, 2007).

Um die Merkmale zur Itemschwierigkeit zu bestimmen, bewerteten drei Rater die Merkmale curriculare Einführung des Themas und Häufigkeit des curricularen Auftretens, Zeichen als Objekte – ihre Definition und ihre Grammatik (wie bspw. Rechengesetze). Diese Kodierung fand nach einer Schulung der drei Rater im Rahmen eines niedrig inferenten Verfahrens statt. Die Rater erhielten Angaben zum Rating der curricularen Einführung der Themen der aufgeführten Items sowie eine Anleitung, wie das spätere Auftreten zu raten ist. Die Angaben, wie Zeichen und deren Gesetzmäßigkeiten zu zählen seien, wurden anhand von Beispielen verdeutlicht (für Details siehe Kodierhinweise am Ende des Beitrages).

Die Interraterreliabilität für die Kodierung der Merkmale ergab einen substantiellen Übereinstimmungswert (vgl. Landis & Koch, 1977) des Fleiss' Kappas ($\kappa = 0,61$). Der dargestellte Wert erhöhte sich bezogen auf die Interraterreliabilität bei den leichten und schwierigen Items noch deutlich und erreichte einen Wert von $\kappa = 0,80$.

4. Ergebnisse

Die Analyse der Itemschwierigkeit im Test zur schnellen Fehlererkennung geschieht im Folgenden in zwei Schritten. Zunächst werden in Abschnitt 4.1. die Itemschwierigkeiten zum Primar- und Sekundarstufentest zur schnellen Fehlererkennung dargestellt, die mithilfe der Software ConQuest (vgl. Wu et al., 1998) geschätzt wurden. In Abschnitt 4.2. werden zunächst die qualitativen Merkmale, die zur Erklärung der Itemschwierigkeiten herangezogen werden, beschrieben und beim Sekundarstufentest exemplarisch an fünf Items dargestellt. Abschließend werden die Merkmale auf den Test zur schnellen Fehlererkennung der Primarstufe übertragen. Für dieses Vorhaben werden vier Items der Primarstufe zunächst kurz vorgestellt und im Anschluss ebenfalls hinsichtlich der schwierigkeitsgenerierenden Merkmale analysiert.

4.1 Schätzung der Itemschwierigkeiten

Die Schwierigkeit der einzelnen Items des Testinstruments wird mittels des Output des Programms ConQuest näher vorgestellt (Abb. 2 und 3).

Auf der linken Seite der Skala sind jeweils die teilnehmenden Personen repräsentiert, wobei jedes Kreuz einen Probanden repräsentiert.

Auf der linken Seite der Skala sind jeweils die teilnehmenden Personen repräsentiert, wobei jedes Kreuz einen Probanden repräsentiert.

Auf der rechten Seite kann anhand der Lage der Itemnummern die Schwierigkeit abgelesen werden. Die Schwierigkeitsgrade der Skala reichen von -3 bis +3. Items, die oberhalb von Null liegen, sind eher schwer, während Items, die unterhalb von Null liegen, eher leicht sind. Je weiter die Items von Null entfernt sind, desto seltener bzw. häufiger wurden sie richtig gelöst.

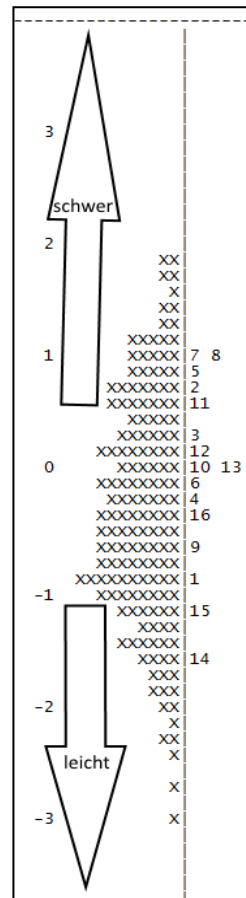


Abb. 2 (links): Itemschwierigkeiten des zeitbegrenzten Tests zur schnellen Fehlererkennung der Sekundarstufe

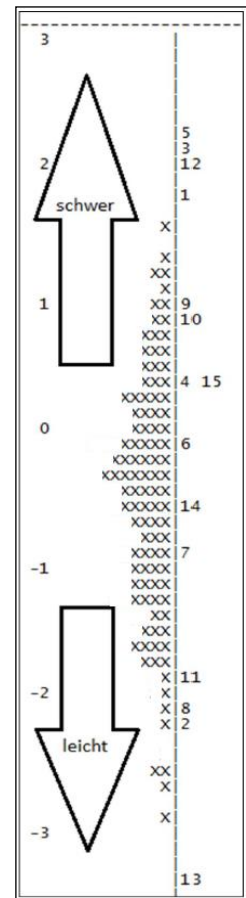


Abb. 3 (rechts): Itemschwierigkeiten des zeitbegrenzten Tests zur schnellen Fehlererkennung der Primarstufe

Die Outputs zeigen, dass die Itemstreuung für die Primarstufe (Abb. 3) gleichmäßiger über die Personenfähigkeiten verteilt ist als für die Sekundarstufe (Abb. 2). Dies bedeutet, dass für die Primarstufe Schätzungen in den Randbereichen besser möglich sind, während für die Sekundarstufe die Schätzungen im mittleren Schwierigkeitsbereich genauer sind.

Die Items Nr. 7 und Nr. 8 gehören im Sekundarstufentest (Abb. 2) zu den schwer zu lösenden Items, die Items Nr. 14 und Nr. 15 zu den leicht zu lösenden. Zur Beantwortung der Forschungsfrage soll in diesem Zusammenhang geklärt werden, wie die Schwierigkeit an beiden Enden der Skalen inhaltlich erklärt werden kann.

Im Vergleich zu dem Sekundarstufentest stellen die Items Nr. 2 und Nr. 8 im Primarstufentest (vgl. Abb. 3) Repräsentanten für leicht zu lösende Items dar, während die Items Nr. 1 und Nr. 12 als schwer zu lösende Items anzusehen sind. Hervorzuheben ist das Item Nr. 13, welches von allen Teilnehmer*innen richtig gelöst und daher als zu niedrig differenzierend ausgeschlossen wurde.

4.2 Qualitative Analysen zur Itemschwierigkeit

Im Folgenden werden die Merkmale zunächst auf die Items der Sekundarstufe bezogen, gefolgt von den Primarstufenitems.

Das Merkmal *curriculare Relevanz*, das die Einführung des Themas und Häufigkeit des curricularen Auftretens im weiteren Mathematikunterricht fokussiert, wird – wie in der Tab. 1 dargestellt – hinsichtlich zweier Aspekte untersucht: Zum einen wird die Klassenstufe des ersten Auftretens im Curriculum

analysiert, zum anderen wird die curriculare Häufigkeit der Verwendung in anderen Themengebieten nach der Einführung des Begriffs betrachtet. Dies wird im Folgenden für das Themengebiet „Addition zweier Brüche“ (Item Nr. 15 aus Abb. 1) exemplarisch ausgeführt. Das Thema ‚Addition zweier Brüche‘ tritt curricular in der Sekundarstufe insgesamt in 48 der 114 Themen, die unterrichtet werden, auf (ca. 40 %).

Um dies zu analysieren, wurde das Beispiel in einem schulinternen Curriculum für die Stadtteilschule (ein Zusammenschluss der ehemaligen Haupt- und Realschulen) Mathematik“ (vgl. Freie und Hansestadt Hamburg Behörde für Schule und Berufsbildung, 2012b) analysiert.

Die Unterrichtsvorhaben für die Jahrgänge 5-10 sind in diesem Beispielcurriculum in den Doppeljahrgängen 5/6 sowie 7/8 und 9/10 aufgeführt. Die Themen werden weiter in Inhalte differenziert. So wird das fünfte Thema in Jahrgang 5/6 „Einteilen – Verteilen“ in sieben Unterthemen eingeteilt wie bspw. Grundvorstellungen, Darstellungen, Größenvergleiche, Umwandlung von Bruch in Dezimalzahl und Grundrechenarten in Verbindung mit Brüchen und Dezimalzahlen.

Insgesamt entstehen so 114 Themen. Bei jedem der Fehler, die in beiden Tests aufgeführt wurden, war es so möglich, durch drei Rater zu bewerten, ob der Fehler in den Themengebieten des Unterrichts enthalten war oder nicht. Ähnliche curriculare Auswertungen wurden für die anderen Items ebenfalls vorgenommen, die Ergebnisse sind in Tab. 1 dargestellt.

Item		Merkmals	empirische Schwierigkeit (Rasch-Wert -3 (leicht) bis +3 (schwer) aus Abb. 2)	Einführung des Themas [in Klassenstufe(n)]	Häufigkeit curricularen Auftretens in anderen Themen ab Einführung
		Merkmals			
Sekundarstufen Items	Addition zweier Brüche (Item 15)		-1	5/6	ca. 40 %
	Umformen mit Klammern (Item 14)		-1,5	5	ca. 25 %
	Trigonometrie (Item 7)		1	9/10	ca. 3 %
	Quadratische Gleichung (p-q-Formel) (Item 8)		1	9/10	ca. 6 %
	Anwendung des Satzes des Pythagoras (Item 5)		1	9/10	ca. 4 %

Tab. 1: Analysen zum Merkmal *curriculare Relevanz*

Item		Merkmal	empirische Schwierigkeit (Rasch-Wert -3 (leicht) bis +3 (schwer) aus Abb. 2)	Merkmal		
				Zeichen		Regeln
				je Antwort	je Item	
Sekundarstufen Items	Addition zweier Brüche (Item 15)	-1	10 bis 11	32	3	
	Umformen mit Klammern (Item 14)	-1,5	13 bis 15	41	4	
	Trigonometrie (Item 7)	1	5 bis 6	16	9	
	Quadratische Gleichung (p - q -Formel) (Item 8)	1	9 bis 15	48	8	
	Anwendung des Satzes des Pythagoras (Item 5)	1	14	29	7	

Tab. 2: Analysen zum Merkmal *strukturelle Komplexität*

Zusammenfassend ist erkennbar, dass Items in dem zeitbeschränkten Test zur schnellen Fehlererkennung, die sich auf curriculare Inhalte in niedrigeren Klassenstufen beziehen (wie die Items Nr. 14 und Nr. 15), häufiger richtig beantwortet wurden als jene, die in höheren Klassenstufen unterrichtet wurden (Items Nr. 7, Nr. 8 und Nr. 5). Weiterhin zeigt sich, dass Items zu curricularen Themen, die häufig in nachfolgenden Klassenstufen wiederaufgenommen werden, häufiger richtig beantwortet wurden. Insgesamt tragen diese beiden Merkmale dazu bei, dass Items als schwer oder leicht angesehen werden können.

Die Auswertung des Merkmals *strukturelle Komplexität*, in dem Zeichen als Objekte – d. h. ihre Definition und ihre Grammatik (wie bspw. Rechengesetze) – analysiert werden, ist in der Tab. 2 wiedergegeben. Bezüglich des genannten Merkmals sind zwei unterschiedliche Aspekte zu analysieren. Zunächst werden die wahrzunehmenden Zeichen der Items beschrieben, während im Anschluss die zu verwendenden Regeln näher betrachtet werden, die zu einer Lösung als bekannt vorausgesetzt und korrekt angewandt werden müssen. Aus Platzgründen werden nur exemplarisch die Items ‚Addition zweier Brüche‘, ‚Quadratische Gleichungen (p - q -Formel)‘ sowie ‚Anwendung des Satzes des Pythagoras‘ detailliert dargestellt.

Im Item zur ‚Bruchrechnung‘ müssen die Brüche zunächst als solche identifiziert werden und im Anschluss hinsichtlich der Operation betrachtet werden. So bestehen die Bruchzahl $\frac{1}{2}$ wie auch die Bruchzahl $\frac{1}{4}$ und das Ergebnis jeweils aus insgesamt drei Zeichen. Es verbleiben der Operator und das Gleichheitszeichen. Das Gleichheitszeichen wird als ein Zeichen interpretiert. Insgesamt werden so 10–11 Zeichen je Antwortmöglichkeit erreicht. Wenn man die drei Antwortmöglichkeiten addiert, handelt es sich auf diese Weise um ca. 32 Zeichen, die im Rah-

men dieses Items wahrzunehmen sind. Die verwendeten Regeln, die in diesem Item Anwendung finden, sind die folgenden: Bildung und Umgang mit dem Hauptnenner sowie die Addition der Zähler und anschließendes Kürzen. Eine Analyse des Items zur ‚Lösung der quadratischen Gleichung mittels der p - q -Formel‘ zeigt, dass bei diesem Item die meisten Zeichen auftreten. Das Item besteht aus einer zusammengesetzten Antwort: Zunächst ist der übergeordnete Ausdruck, der aus 9 Zeichen besteht, wahrzunehmen, die zu den Antwortmöglichkeiten hinzuaddiert werden müssen, da der Ausdruck die Voraussetzung zur Lösung des Items darstellt. Zwei der Antwortmöglichkeiten umfassen 12 Zeichen, da das Symbol ‚ \pm ‘ als ein Zeichen gezählt wird. Die dritte Antwortmöglichkeit enthält 15 Zeichen, da neben zwei Klammern noch ein weiteres Vorzeichen hinzukommt. Neben der Zeichenanzahl werden ebenfalls die Rechengesetze überprüft: Die Indizes in dem Ausdruck ‚ $x_{1,2}$ ‘ benötigen Wissen über die Verwendung dieses Ausdrucks, nämlich, dass zwei Ergebnisse erwartet werden müssen. Ebenfalls sollte die Verwendung des ‚ \pm ‘-Zeichens bekannt sein. Die Verwendung des Rechenzeichens vor Klammern gehört ebenfalls in die Kategorie der als bekannt vorauszusetzenden Rechenregeln. Des Weiteren ist der Umgang mit Wurzeln von Bedeutung, da die Ergebnisse nicht einzeln berechnet werden dürfen, sondern aus dem gesamten Ausdruck die Wurzel zu ziehen ist. Bei dem Ziehen der Wurzel muss die Kenntnis, dass es im Bereich der reellen Zahlen nicht möglich ist, aus negativen Zahlen die Wurzel zu ziehen, vorhanden sein, um ein falsches Ergebnis einordnen zu können. Weiterhin ist Wissen zum Umgang mit Brüchen bzw. deren Umwandlung in Dezimalzahlen umgehen zu können.

Das Item zur ‚Anwendung des Satzes des Pythagoras‘ ist zusammengesetzt: Zunächst muss das Bild ei-

nes rechtwinkligen Dreiecks samt den Bezeichnungen wahrgenommen werden. Dieses besteht aus fünf Zeichen. Im Anschluss müssen die drei Antwortmöglichkeiten mit je acht Zeichen erfasst werden. Die Buchstaben je Antwortmöglichkeit sowie die Quadrate gelten als einzelne Zeichen, die wahrgenommen werden müssen. Darüber hinaus müssen Regeln zum Umgang mit den Zeichen inkorporiert sein, da neben dem Quadrieren ebenfalls die Überlegungen zu den Eigenschaften eines Dreiecks abgerufen werden müssen. Außerdem müssen die Voraussetzungen für die Anwendung und die Definition der Hypotenuse und der Katheten des Satzes des Pythagoras bekannt sein und angewendet werden. Abschließend müssen auch Kenntnisse zum Wurzelziehen und zur Gleichungsumformung angewendet werden.

Zusammenfassend ist erkennbar, dass das Merkmal *strukturelle Komplexität* geeignet ist, aus einer wahrnehmungspsychologischen Perspektive die Komplexität von Items vorherzusagen. Die Anzahl der Zeichen in einem Item beeinflusst die Schwierigkeit des Items, da die Zeit, in der die vorhandenen Zeichen wahrgenommen werden müssen, begrenzt ist. Allerdings ist nicht die reine Anzahl von Zeichen entscheidend, die die Schwierigkeit eines Items beeinflusst, sondern auch die Anzahl der anzuwendenden Regeln und somit die Art der Verknüpfungen zwischen den Zeichen. Die Subkategorien des Merkmals *strukturelle Komplexität* beeinflussen einander, so dass zu erkennen ist, dass die Items mit einer hohen Zeichenanzahl oder Anzahl von Regeln schwieriger sind als Items mit einer niedrigen Zeichenanzahl oder Anzahl von Regeln.

Abschließend lässt sich feststellen, dass sich bei den Analysen über alle Items der Sekundarstufe ein ähnliches, aber nicht so starkes Bild ergibt wie bei den Items aus der Extremgruppe der schwierigen bzw. einfachen Items. Ersichtlich spielen bei Items mittleren Schwierigkeitsgrads unterschiedliche Merkmale eine Rolle, so dass die zwei ausgewählten Merkmale an Bedeutung verlieren.

Allerdings weisen die Ergebnisse auch bei den mittelschweren Items in dieselbe Richtung. Insgesamt erscheinen damit weitere Analysen mit einem größeren Itempool nötig, um die oben vorgestellten Analysen zu erhärten. Aufgrund dieser Einschränkung wurden dieselben Analysen am Primarstufentest durchgeführt, um damit insgesamt zu einer größeren Breite in den Analysen zu gelangen und zusätzlich die Stabilität der Ergebnisse unabhängig von der unterrichteten Schulstufe der Testbeteiligten zu überprüfen.

5. Überprüfung der merkmalsbezogenen Analysen mithilfe der Items des Primarstufen-Tests

Die Merkmale zur Erklärung der Itemschwierigkeit wurden am Test zur schnellen Fehlererkennung aus der Sekundarstufe entwickelt und angewendet. Um zu abgesicherten Aussagen zu kommen, auch gerade bzgl. der Abhängigkeit von der unterrichteten Schulstufe, wird die Angemessenheit der schwierigkeitsgenerierenden Merkmale für Items anhand des Primarstufentests zur schnellen Fehlererkennung überprüft.

5.1 Vorstellung der Primarstufenitems

Die Items, die in dem Test zur schnellen Fehlererkennung verwendet wurden, beziehen sich auf Fehler, die im Grundschulunterricht typischerweise vorkommen. Zunächst sollen die Items mit niedriger Schwierigkeit dargestellt werden (Abb. 4).

Der Fehler mit der Null (Item Nr. 2) tritt zunächst bei der Bedeutungszuweisung der Null durch Grundschüler*innen auf. Das neutrale Element bei der Addition und der Subtraktion wird als ein solches auf die Multiplikation übertragen (Padberg & Benz, 2011). Der Fehler mit der Null ist einer der „häufigsten Einmaleins-Fehler“ (ebd., S. 147 f.). Eine weitere mögliche Ursache ist, dass die Schüler*innen eine falsche Vorstellung von der Null haben, sodass diese als „Nichts“ interpretiert wird und die 12 als Ergebnis angegeben wird, obwohl eine Multiplikation angegeben ist. Des Weiteren können Lernende die Vorstellung haben, dass das Ergebnis einer Multiplikation mindestens so groß sein muss wie der größte Faktor.

Beim Erkennen von Symmetrieachsen (Item Nr. 8) ist zu berücksichtigen, dass nach Franke und Reinhold (2016) die Lösungshäufigkeit sowohl von der Figur als auch der Lage der Achse abhängig ist. Sowohl bei Schulanfänger*innen als auch bei älteren Schüler*innen werden Schwierigkeiten sichtbar, wenn eine Spiegelachse schräg zur Bildkante verläuft. Eine Entwicklung des Symmetrieverständnisses ist daher auf vielfältige Weise durch eigene Erfahrungen zu vermitteln (vgl. Franke & Reinhold, 2016). In dem Test zur schnellen Fehlererkennung stellen die Lösungen „D“ und „F“ die Parallelität zu den Seitenkanten dar, sodass keine Schwierigkeit bei der Achsenerkennung zu erwarten ist.

Im Gegensatz dazu steht die Lösung „S“, in der die Spiegelachse die Diagonale darstellt. Die als schwierig identifizierten Items fokussieren andere Aspekte, z. B. Rechnen mit der Null.




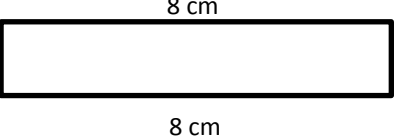
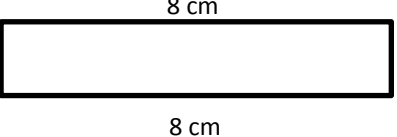
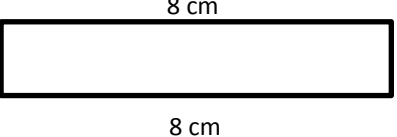
Nr.		Ankündigung	„S“	„D“	„F“
Primarstufenitems	1	Schriftliche Multiplikation	$\begin{array}{r} 31 \cdot 24 \\ 62 \\ \underline{124} \\ \dots \end{array}$	$\begin{array}{r} 31 \cdot 24 \\ 62 \\ \underline{124} \\ \dots \end{array}$	$\begin{array}{r} 31 \cdot 24 \\ 620 \\ \underline{124} \\ \dots \end{array}$
	2	Rechnen mit der Null	$12 - 0 = 12$	$12 \cdot 0 = 12$	$12 + 0 = 12$
	8	Eine Symmetrieachse einzeichnen			
	12	Umfang eines Rechtecks	 $U = 2 \cdot 8 \text{ cm}$	 $U = 2 \cdot 10 \text{ cm}$	 $U = 2\text{cm}+8\text{cm}+2\text{cm}+8\text{cm}$

Abb. 4: Auswahl (4 von 15) der Items zur schnellen Fehlererkennung aus TEDS-FU Primarstufentest

Sowohl die Arbeiten von Gerster (1982), wie später auch die von Padberg und Benz (2011), analysieren die Schwierigkeiten der schriftlichen Multiplikation (Item Nr. 1). Sie weisen neben Schwierigkeiten bei der Multiplikation mit der Null auf die Schwierigkeit des Umgangs mit dem Stellenwert hin, also auf die Möglichkeit der falschen Anordnung der errechneten Teilprodukte, sog. Stellenwertfehler (Padberg & Benz, 2011, S. 280). Der aufgeführte Fehler kann, so Padberg und Benz (2011), als systematischer und somit als typischer Fehler angesehen werden, da die ggf. nicht verstandene Bedeutung des Ausrückens der Ziffernreihe die Stellenwerte der einzelnen Rechnungen außer Acht lässt.

Weiterhin kann das frühe Weglassen der Endnull in der ersten Rechenzeile für die Fehlvorstellung verantwortlich sein.

Am Ende der Grundschulzeit werden Flächen und Umfänge von geometrischen Figuren berechnet, auf die in Item 12 Bezug genommen wird. Im Rahmen dieser Berechnungen spielt die Verwendung der korrekten Begriffe des Umfangs und des Flächeninhaltes und deren Unterscheidung eine Rolle. Franke und Reinhold (2016) weisen darauf hin, dass die Begriffe aufgrund der fehlenden Begriffsvorstellung häufig verwechselt werden (Franke & Reinhold, 2016). Auch wenn eine algebraische Betrachtung für die Berechnung des Flächeninhaltes und des Umfangs in der Grundschule noch nicht betrachtet werden, kann damit das Verständnis für Formeln zu den einschlägigen Berechnungen gefördert werden (Franke & Reinhold, 2016).

5.2 Überprüfung der merkmalsbezogenen Analyse der Aufgabenschwierigkeit mit Hilfe der Items des Primarstufen-Tests

Im Folgenden werden die beiden Merkmale zur Analyse der Aufgabenschwierigkeit, die für die Items des Sekundarstufentest entwickelt wurden, auf die Items der Primarstufe angewendet, um die Validität des Analyseinstruments zu überprüfen. Zunächst wird das Merkmal *curriculare Relevanz*, d. h. Einführung des Themas und Häufigkeit des curricularen Auftretens im weiteren Mathematikunterricht analysiert (Tab. 3).

Die schriftliche Multiplikation wird in Grundschulen in der 3./4. Klasse unterrichtet (vgl. Freie und Hansestadt Hamburg Behörde für Schule und Berufsbildung, 2012a, S. 3) und wird dann in weiteren acht von 42 möglichen Unterrichtsinhalten (vgl. Tab. 3) curricular aufgegriffen. Ebenso verhält es sich mit dem Themengebiet ‚Umfang eines Rechtecks‘ (Item 12).

Dieses Thema wird in der 4. Klasse unterrichtet und wird nach der Einführung in vier weiteren Themengebieten wiederaufgegriffen. Das Rechnen mit der Null, welches in Item Nr. 2 fokussiert wird, wird thematisch bei der Einführung der Addition in der 1. Klasse im Mathematikunterricht thematisiert und wird im Anschluss mit dem Bezug auf die weiteren Rechenarten sukzessive weitergeführt, sodass es insgesamt in 25 der 42 Unterrichtsthemen wiederaufgegriffen wird. Umgang mit Symmetrie (Item Nr. 8) wird zwar spielerisch in der 1.–2. Klassenstufe mit Hilfe von einfachen Spiegelungen an Flächen durchgeführt, lässt sich allerdings im weiteren Verlauf der Schullaufbahn nur in neun der 42 möglichen Unterrichtsinhalte finden.

Item		Merkmal	empirische Schwierigkeit (Rasch-Wert -3 (leicht) bis +3 (schwer) aus Abb. 3)	Merkmal Zeichen		Re- geln
				je Antwort	je Item	
Primarstufen Items	Schriftliche Multiplikation (Item 1)		1	7	21	6
	Fehler mit der Null (Item 2)		-2	5	15	2
	Symmetrie (Item 8)		-2	2	6	2
	Umfang eines Rechtecks (Item 12)		2	9 (Rechteck) +6-13 (abhängig von der Antwort- möglichkeit)	34	3

Tab. 3: Analysen zum Merkmal *curriculare Relevanz*

Item		Merkmal	empirische Schwierigkeit (Rasch-Wert -3 (leicht) bis +3 (schwer) aus Abb. 3)	Merkmal	
				Klassenstufe	Häufigkeit
Primarstufen Items	Schriftliche Multiplikation (Item 1)		1	3/4	ca. 25 %
	Fehler mit der Null (Item 2)		-2	1-3	ca. 60 %
	Einzeichnen der Symmetrieachse (Item 8)		-2	1/2	ca. 25 %
	Umfang eines Rechtecks (Item 12)		2	4	ca. 10 %

Tab. 4: Analysen zum Merkmal *strukturelle Komplexität*

Die Ergebnisse der Analysen zum Merkmal *strukturelle Komplexität*, d. h. die Auffassung von Zeichen als Objekte – ihre Definition und ihre Grammatik (wie bspw. Rechengesetze), werden zunächst übersichtsartig (Tab. 4) dargestellt und anschließend beispielhaft an Item 1 und 12 inhaltsbezogen analysiert.

Das Item zur „schriftlichen Multiplikation“ (Item Nr. 1) enthält fünf Zeichen bestehend aus zwei- bis dreistelligen Zahlen, die mit einer Multiplikation verbunden sind und im Anschluss untereinander angeordnet sind. Zusammengefasst handelt es damit auf diese Weise um 21 Zeichen, die im Rahmen dieses Items wahrzunehmen sind. Weiterhin müssen die Regeln zur schriftlichen Multiplikation von zweistelligen Zahlen berücksichtigt werden, ebenso wie folgende Regeln: Rechne von rechts nach links, verwende die Kenntnisse aus dem 1x1, berücksichtige eventuelle Überträge, beachte die Kommutativität der Multiplikation, beachte die stellenwertberücksichtigende Notation.

Das vierte Item zum ‚Umfang eines Rechtecks‘ (Item Nr. 12) besteht aus zwei Teilen, der Abbildung eines Rechtecks (ein Zeichen) sowie der jeweiligen Maßeinheiten. Das Rechteck kann als ein Zeichen gezählt werden, da dies als Gesamtfigur identifiziert werden muss.

Insgesamt, inkl. der zu identifizierenden Maßeinheiten an den Seiten des Rechtecks, handelt es sich um 9 Zeichen. Je Antwortmöglichkeit erhöhen sich die

wahrzunehmenden Zeichen auf 15 bis 22 Zeichen. Bei allen drei Antworten handelt es sich auf diese Weise um 34 Zeichen, die im Rahmen dieses Items zu erkennen sind. Für die Lösung des Items sind weiterhin Definitionen und Rechenregeln einzuhalten. So ist es notwendig, dass die Regeln des Einmaleins verinnerlicht sind, wie auch die Formel zur Berechnung des Umfangs eines Rechtecks, weiterhin ist die Verwendung der Einheiten zur korrekten Darstellung notwendig.

6. Diskussion der Ergebnisse, Grenzen der Studie und Schlussfolgerungen

Die quantitative Auswertung der Itemschwierigkeit (Abb. 2 und 3) konnte die unterschiedlichen Schwierigkeiten der Items, die in beiden Tests zur schnellen Fehlererkennung verwendet wurden, zeigen. Die Items, die einen hohen sowie einen niedrigen Schwierigkeitsgrad aufweisen, wurden genauer analysiert, da davon ausgegangen wurde, dass gerade bei stark oder schwach ausgeprägtem Schwierigkeitsgrad theoretisch entwickelte Erklärungsansätze greifen. Es zeigt sich, dass die literaturgeleiteten Merkmale zur Erklärung der qualitativen Schwierigkeit geeignet sind, die Itemschwierigkeit inhaltlich zu erklären (Tab. 5).

Das aus der mathematikdidaktischen Perspektive entwickelte Merkmal der *curricularen Relevanz* (d. h. curriculare Einführung des Themas und Häufigkeit

des curricularen Auftretens im weiteren Mathematikunterricht) kann die Itemschwierigkeit der analysierten Items des Sekundarstufentests inhaltlich aufklären. Es weisen solche Items eine geringe Itemschwierigkeit auf, die curricular früh eingeführt werden und anschließend im Rahmen eines spiralförmig geprägten Schulcurriculums häufig wieder aufgegriffen werden. Die Themen der Items, die später im Curriculum vorkommen und damit wohl auch später und damit seltener unterrichtet werden, konnten als schwieriger identifiziert werden. Dieses Merkmal konnte auf den strukturgleichen Primarstufentest übertragen werden, wo sich dieselben Ergebnisse zeigten. Dabei ist zu berücksichtigen, dass die Häufigkeit des Aufgreifens eines Themas in anderen curricularen Themenfeldern keinen exakten Wert darstellt, da die Werte auf Grundlage eines Ratings erfasst wurden. Die Daten ermöglichen aber eine recht gesicherte Einsicht in die Häufigkeit des curricularen Auftretens eines Themengebiets.

Das aus einer wahrnehmungspsychologischen Ebene entwickelte Merkmal *strukturelle Komplexität* (d. h. Auffassung von Zeichen als Objekte – ihre Definition und ihre Grammatik) ist ebenfalls gut geeignet, die Itemschwierigkeit aufzuklären. So ist die Anzahl der Zeichen in Kombination mit ihren Definitionen und Rechengesetzen zu sehen. Die Anzahl der Zeichen kann generell wie folgt ausgewertet werden: Je weniger Zeichen in der begrenzten Zeit wahrgenommen werden müssen, desto häufiger wird das Item richtig gelöst. Allerdings muss darüber hinaus auch die zweite Komponente des Merkmals beachtet werden: Je mehr Definitionen und Rechengesetze das Item zur Lösung benötigt werden, desto schwieriger ist das Item trotz weniger Zeichen, wie sich beispielsweise in dem Item zu ‚Trigonometrischen Funktionen bzw. Trigonometrie im rechtwinkligen Dreieck‘ zeigt. Bei diesem Item müssen insgesamt lediglich fünf Zeichen je Antwortkategorie wahrgenommen werden, allerdings ist es notwendig, die Werte der trigonometrischen Funktionen bzw. die entsprechenden Definitionen der Verhältnisse von Katheten und Hypotenuse im rechtwinkligen Dreieck zu kennen und die damit in Verbindung stehenden Voraussetzungen.

Die Ergebnisse zur Aufklärung der Itemschwierigkeit mittels der identifizierten Merkmale konnten von dem Test der schnellen Fehlererkennung der Sekundarstufe auf den Test der Primarstufe übertragen werden (Tab. 5).

Das Merkmal der *curricularen Relevanz* und damit der curricularen Verankerung und Vertrautheit mit dem Thema scheint am robustesten zu sein, da weder im Primar- noch im Sekundarstufentest Ausnahmen

zu erkennen sind. Die Items, die positive Werte in der ersten Spalte besitzen, gehören zu den schweren Items, während die Items mit negativen Ziffern zu den leichten Items gehören. Bei den Primarstufen-Items haben die Items ‚Schriftliche Multiplikation‘ und ‚Einzeichnen der Symmetrieachse‘ den gleichen Wert zum curricularen Auftreten in der weiteren Schullaufbahn, aber aus unterschiedlichen Gründen. Dies ist vermutlich dadurch bedingt, dass die ‚Schriftliche Multiplikation‘ erst am Ende der dritten bzw. zu Beginn der vierten Klasse erfolgen kann, da sie auf den anderen Rechenarten beruht, wohingegen Symmetrieachsen aufgrund ihrer relativ niedrigen Bedeutung im Grundschulcurriculum erst spät eingeführt werden. Des Weiteren können die zwei Teilkomponenten des Merkmals Zeichenanzahl und Anzahl der anzuwendenden Regeln nicht losgelöst voneinander betrachtet werden. So besteht das Item zum Umformen mit Klammern aus 41 Zeichen, benötigt allerdings vier Regeln zum Lösen.

Bezüglich der Zeichen würde es sich um ein schweres Item handeln, bzgl. der Regeln um ein leichtes Item. Umgekehrt würde es sich bei dem Item zur Trigonometrie um ein leichtes Item bzgl. der Zeichenanzahl handeln, da neun Regeln anzuwenden sind, um es zu lösen. Zusammenfassend können die Subfacetten des Merkmals die Itemschwierigkeit qualitativ erklären.

Die Streuung der Itemschwierigkeiten wird in den quantitativen Analysen deutlich, allerdings nicht inhaltlich erklärt. Die Analysemerkmale für Schüleraufgaben im Unterricht, die von Draxler (2006) entwickelt und durch Schumann und Eberle (2011) ergänzt wurden, konnten an die Gegebenheiten des vorliegenden Tests angepasst und analysiert werden.

Mithilfe der vorgestellten Merkmale, d. h. das aus einer mathematikdidaktisch geprägten Perspektive der *curricularen Relevanz* und das aus einer wahrnehmungspsychologisch geprägten Perspektive der *strukturellen Komplexität*, zur Aufklärung der Itemschwierigkeit in dem zeitbeschränkten Test war es möglich, die Unterschiede in der Itemschwierigkeit qualitativ aufzuklären. Obwohl die schwierigkeitsgenerierenden Merkmale für Items der Sekundarstufe entwickelt wurden, konnten damit auch die Itemschwierigkeiten des Primarstufentests erklärt werden, was als eine erste Validierung der Analysen anzusehen ist. Damit können die zwei aus unterschiedlichen Perspektiven identifizierten Merkmale als Basis zur Analyse der Schwierigkeit von Items dienen und einen Beitrag für die Weiterentwicklung von Tests zur schnellen Fehlererkennung leisten.

Items		Merkmale	empirische Schwierigkeit (Rasch-Wert -3 (leicht) bis +3 (schwer) aus Abb. 2 und Abb. 3)	1. Merkmal <i>curriculare Relevanz</i>		2. Merkmal <i>strukturelle Komplexität</i> Zeichen		Regeln
				Einführung des Themas [in Klassenstufe]	Häufigkeit des Auftretens	je Antwort	je Item	
Sekundarstufen Items	Trigonometrie (Item 7)		1	9/10	ca. 3 %	5 bis 6	16	9
	Quadratische Gleichung (p - q -Formel) (Item 8)		1	9/10	ca. 6 %	9+ 12 bis 15	48	8
	Anwendung Satz des Pythagoras (Item 5)		1	9/10	ca. 4 %	14	29	7
	Addition zweier Brüche (Item 15)		-1	5/6	ca. 40 %	10 bis 11	32	3
	Umformen mit Klammern (Item 14)		-1,5	5	ca. 25 %	13 bis 15	41	4
Primarstufen Items	Umfang eines Rechtecks (Item 12)		2	3/4	ca. 10 %	9 bis 13	34	3
	Schriftliche Multiplikation (Item 1)		1	3/4	ca. 25 %	7	21	6
	Einzeichnen der Symmetrieachse (Item 8)		-2	1/2	ca. 25 %	2	6	2
	Fehler mit der Null (Item 2)		-2	1–3	ca. 60 %	5	15	2

Tab. 5: Übersicht der Merkmale zur Schwierigkeitsgenerierung bei beiden zeitbeschränkten Tests

Dabei sollten in Weiterentwicklung solcher Tests aufgrund ihrer Einbindung in Studien zur Professionalität von Mathematiklehrkräften stärker mathematikdidaktisch-orientierte schwierigkeitsgenerierende Merkmale in den Vordergrund gerückt werden. Aber die Analysen machen deutlich, dass auch wahrnehmungspsychologische Aspekte eine nicht unwichtige Rolle spielen und berücksichtigt werden müssen.

Als eine Grenze der Studie muss festgestellt werden, dass die Ergebnisse nur für die einfachen und die schwierigen Items stabil sind, bei mittlerer Itemschwierigkeit scheinen noch andere Faktoren eine Rolle zu spielen, die über diese Analysen hinausgehen. Weitere Untersuchungen mit einem größeren Itempool erscheinen daher nötig, um einerseits die Relevanz der zwei rekonstruierten schwierigkeitsgenerierenden Aufgabenmerkmale zu erhärten und andererseits weitere Einflussfaktoren zu identifizieren. Dass die Ergebnisse sowohl für Primarstufenitems als auch für Sekundarstufenitems stabil sind, kann allerdings als ein Hinweis auf die Robustheit der Ergebnisse angesehen werden.

Abschließend kann festgestellt werden, dass das schnelle Erkennen von Schülerfehlern eine grundlegende Kompetenz von Lehrkräften darstellt und daher bereits in der Lehrerbildung gefördert werden sollte durch einschlägige Seminarsequenzen, wie von Heinrichs (2015) entwickelt. Als Konsequenz aus den Ergebnissen kann gefolgert werden, dass solche

Seminarsequenzen zwar mit Fehlern aus Themengebieten, die im jeweiligen Curriculum der unteren Klassenstufen auftreten, ansetzen sollten, da diese leichter zugänglich sind und die Lehramtsstudierenden damit weniger überfordern. Allerdings sollten aufgrund der hohen Relevanz für die Entwicklung der Kompetenz zur Fehlererkennung auch die schwerer zu identifizierenden Schülerfehler, deren Themen erst in den höheren Klassenstufen im Curriculum eingeführt werden, in den Seminarsequenzen behandelt werden, um damit Lehramtsstudierende für alle Schulstufen ihres Lehramts adäquat auszubilden.

Insgesamt weisen die Analysen auf die Bedeutung von Fehlern beim Umgang mit den in den Bildungsstandards geforderten Kompetenzen zur Verwendung mathematischer Darstellungen und zum Umgang mit symbolischen, formalen und technischen Elementen der Mathematik hin (Leiß & Blum, 2007), mit denen nicht nur Schülerinnen und Schüler, sondern auch die sie unterrichtenden Lehrkräfte vertraut sein sollten.

Literatur

- Allmendinger, H., Lengnink, K., Vohns, A. & Wickel, G. (2013). *Mathematik verständlich unterrichten. Perspektiven für Unterricht und Lehrerbildung*. Wiesbaden: Springer Fachmedien.
- Beutelspacher, A. (2008). Über die Unmöglichkeit und die Notwendigkeit von Fehlern in der Mathematik. In R. Caspary (Hrsg.), *Nur wer Fehler macht, kommt weiter. Wege zu einer neuen Lernkultur* (S. 86–96). Freiburg: Herder.
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond Dichotomies. Competence Viewed as a Continuum. *Zeitschrift für Psychologie*, 223 (1), 3–13.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.). (2010). *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.
- Blömeke, S., König, J., Busse, A., Suhl, U., Benthien, J., Döhrmann, M. et al. (2014). Von der Lehrerausbildung in den Beruf – Fachbezogenes Wissen als Voraussetzung für Wahrnehmung, Interpretation und Handeln im Unterricht. *Zeitschrift für Erziehungswissenschaft*, 17 (3), 509–542.
- Bond, T. & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: LEA.
- Brinkmann, A. (2002). *Über Vernetzungen im Mathematikunterricht - eine Untersuchung zu linearen Gleichungssystemen in der Sekundarstufe I*. Unveröffentlichte Dissertation. Duisburger Elektronische Texte.
- Bruner, J. S. (1973). *Der Prozeß der Erziehung*. Berlin: Berlin Verlag.
- Brunner, M. (2013). Didaktikrelevante Aspekte im Umfeld der Konzepte token und type. *Journal für Mathematik-Didaktik*, 34 (1), 53–72.
- 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*, 109-144. Wiesbaden: VS-Verlag für Sozialwissenschaften.
- Draschoff, S. (2000). *Lernen am Computer durch Konfliktinduzierung. Gestaltungsempfehlungen und Evaluationsstudie zum interaktiven computerunterstützten Lernen*. Münster: Waxmann.
- Draxler, D. (2006). *Aufgabendesign und basismodellorientierter Physikunterricht*. Unveröffentlichte Dissertation Duisburg-Essen (<https://d-nb.info/983890943/34>).
- Drollinger-Vetter, B. (2011). *Verstehenselemente und strukturelle Klarheit. Fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht*. Münster: Waxmann.
- Dyrvold, A. (2016). *Difficult to read or difficult to solve? The role of natural language and other semiotic resources in mathematics tasks*. Umeå: Print & Media.
- Eichmann, A., Narciss, S., Schnaubert, L. & Melis, E. (2012). Typische Fehler bei der Addition und Subtraktion von Brüchen – Ein Review zu empirischen Fehleranalysen. *Journal für Mathematik-Didaktik*, 33 (1), 29–57.
- Franke, M. & Reinhold, S. (2016). *Didaktik der Geometrie in der Grundschule* (3. Aufl.). Berlin: Springer Spektrum.
- Freie und Hansestadt Hamburg Behörde für Schule und Berufsbildung. (2012a). *Beispiel für ein schulinternes Fachcurriculum. Mathematik Grundschule*. Hamburg.
- Freie und Hansestadt Hamburg Behörde für Schule und Berufsbildung. (2012b). *Beispiel für ein schulinternes Fachcurriculum. Mathematik. Stadtteilschule Sekundarstufe I*. Hamburg.
- Gerster, H.-D. (1982). *Schülerfehler bei schriftlichen Rechenverfahren. Diagnose und Therapie*. Freiburg: Herder.
- Goebels, S. & Ritter, S. (2013). *Mathematik verstehen und anwenden – von den Grundlagen bis zu Fourier-Reihen und Laplace-Transformation* (2. Aufl.). Berlin: Springer.
- Hartig, J. (2007). Skalierung und Definition von Kompetenzniveaus. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)* (S. 83–99). Weinheim: Beltz Verlag.
- Heinrichs, H. (2015). *Diagnostische Kompetenz von Mathematik-Lehramtsstudierenden. Messung und Förderung*. Wiesbaden: Springer Fachmedien.
- Heinze, A. (2004). Zum Umgang mit Fehlern im Unterrichtsgespräch der Sekundarstufe I. *Journal für Mathematik-Didaktik*, 25 (3-4), 221–244.
- Hill, H.C., Ball, D.L. & Schilling, S.G. (2008) Unpacking Pedagogical Content Knowledge. *Journal for Research in Mathematics Education*, 39, 372-400.
- Kaiser, G., Blömeke, S., König, J., Busse, A., Döhrmann, M. & Hoth, J. (2017). Professional competencies of (prospective) mathematics teachers—cognitive versus situated approaches. *Educational Studies in Mathematics*, 94 (2), 161–182.
- Kaiser, G., Busse, A., Hoth, J., König, J., & Blömeke, S. (2015). About the Complexities of Video-Based Assessments. Theoretical and Methodological Approaches to Overcoming Shortcomings of Research on Teachers' Competence. *International Journal of Science and Mathematics Education*, 13 (2), 369–387.
- Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben*. Berlin: Logos Verlag.
- Kelle, U. & Kluge, S. (2010). *Vom Einzelfall zum Typus*. Wiesbaden: VS Verlag.
- Klug, J., Bruder, S., Kelava, A., Spiel, C. & Schmitz, B. (2013). Diagnostic competence of teachers. A process model that accounts for diagnosing learning behavior tested by means of a case scenario. *Teaching and Teacher Education*, 30, 38–46.
- König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A. & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education*, 38, 76–88.
- Krauss, S. & Brunner, M. (2011). Schnelles Beurteilen von Schülerantworten: Ein Reaktionszeittest für Mathematiklehrer/innen. *Journal für Mathematik-Didaktik*, 32(2), 233–251.
- Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M. et al. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie. *Journal für Mathematik-Didaktik*, 29 (3-4), 233–258.

- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse der Forschungsprogramms COACTIV*. Münster: Waxmann.
- Landis, J. R. & Koch G. G. (1977). The measurement of observer agreement for categorical data. In: *Biometrics*, 33, 1977, 159–174
- Leiß, D., & Blum, W. (2007). Beschreibung zentraler mathematischer Kompetenzen. In W. Blum, C. Drüke-Noe, R. Hartung & O. Köller. *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichts Anregungen, Fortbildungsideen* (3. Aufl.) (S. 33-50). Berlin: Cornelsen Scriptor.
- Lüddecke, J. (2015). *Fehler beim Problemlösen. Empirische Erkundungen zu Fehlern beim Bearbeiten mathematischer Probleme*. Hamburg: disserta Verlag.
- Magenheim, J., Schubert, S. & Schaper, N. (2015). Competencies in Computer Science Education. In T. Brinda, N. Reynolds, R. Romeike & A. Schwill (Hrsg.), *KEYCIT 2014: key competencies in informatics and ICT* (S. 33–56). Universitäts-Verlag Potsdam.
- Malle, G. (1993). *Didaktische Probleme der elementaren Algebra*. Braunschweig: Vieweg.
- Oser, F., Hascher, T. & Spychiger, M. (1999). Lernen aus Fehlern. Zur Psychologie des "negativen" Wissens. In W. Althof (Hrsg.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern* (S. 11–41). Wiesbaden: Verlag für Sozialwissenschaften.
- Padberg, F. & Benz, C. (2011). *Didaktik der Arithmetik. Für Lehrerbildung und Lehrerfortbildung*. Heidelberg: Spektrum.
- Padberg, F. (2009). *Didaktik der Bruchrechnung. Für Lehrerbildung und Lehrerfortbildung* (4., erw., Aufl.). Heidelberg: Springer Spektrum.
- Pankow, L., Kaiser, G., Busse, A., König, J., Blömeke, S., Hoth, J. et al. (2016). Early career teachers' ability to focus on typical students errors in relation to the complexity of a mathematical topic. *ZDM Mathematics Education*, 48 (1-2), 55–67.
- Radatz, H. (1980). *Fehleranalysen im Mathematikunterricht*. Wiesbaden: Vieweg+Teubner Verlag.
- Ritter, S. & Voß, U. (2015). *Erfolgreich Starten ins Ingenieurstudium. Grundlagen der Mathematik anwendungsorientiert erklärt*. Berlin: Springer Vieweg.
- Robinson, J. P., Shaver, P. R. & Wrightsman, L. S. (1991). Criteria for Scale Selection and Evaluation. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Hrsg.), *Measures of Personality and Social Psychological Attitudes* (S. 1-15). San Diego: Academic Press.
- Rölke, H. (2012). The ItemBuilder: A Graphical Authoring System for Complex Item Development. In T. Bastiaens & G. Marks (Hrsg.), *Proceedings of E-Learn. World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2012* (S. 344–353). Chesapeake: Aace. (<https://www.learn-technlib.org/p/41614>; letzter Zugriff am 15.02.2017).
- Schoy-Lutz, M. (2005). *Fehlerkultur im Mathematikunterricht. Theoretische Grundlegung und evaluierte unterrichtspraktische Erprobung anhand der Unterrichtseinheit "Einführung in die Satzgruppe des Pythagoras"*. Hildesheim: Franzbecker.
- Schumann, S. & Eberle, F. (2011). Bedeutung und Verwendung schwierigkeitsbestimmender Aufgabenmerkmale für die Erfassung ökonomischer und beruflicher Kompetenzen. In U. Faßhauer (Hrsg.), *Grundlagenforschung zum Dualen System und Kompetenzentwicklung in der Lehrerbildung* (S. 77–89). Leverkusen: Barbara Budrich.
- Seidel, T. & Prenzel, M. (2003). Mit Fehlern umgehen - Zum Lernen motivieren. *Praxis der Naturwissenschaften - Physik*, 51 (1), 30–34.
- Seifried, J. & Wuttke, E. (2010). Professionelle Fehlerkompetenz – Operationalisierung einer vernachlässigten Kompetenzfacette von (angehenden) Lehrkräften. *Wirtschaftspsychologie*, 12 (4), 17–28.
- Shulman, L. S. (1986). Those Who Understand. Knowledge Growth in Teaching. *Educational Researcher*, 15 (2), 4–14.
- Südkamp, A. & Praetorius, A.-K. (Hrsg.) (2017). *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen*. Münster: Waxmann.
- Wartha, S. (2007). *Längsschnittliche Untersuchungen zur Entwicklung des Bruchzahlbegriffs*. Hildesheim: Franzbecker.
- Wu, M. L., Adams, R. J. & Wilson, M. (1998). *ACER ConQuest. Generalised item reponse modelling software manual*. Melbourne: ACER Press.

Anschrift der Verfasserinnen

Lena Pankow
 Universität Hamburg,
 Fakultät für Erziehungswissenschaft
 Von-Melle-Park 8
 20146 Hamburg, Deutschland
lena.pankow@uni-hamburg.de
 ORCID ID: 0000-0001-8941-7133

Prof. Dr. Gabriele Kaiser
 Universität Hamburg
 Fakultät für Erziehungswissenschaft
 Von-Melle-Park 8
 20146 Hamburg, Deutschland
gabriele.kaiser@uni-hamburg.de
 Australian Catholic University
 Institute for Learning Sciences and Teacher Education
 ORCID ID: 0000-0002-6239-0169

Anhang

Kodierhinweise (Auszug aus dem Manual):

Rating des Merkmals: *curricularen Relevanz*

In einer Kreuztabelle wurden alle Themengebiete, die in dem Beispielcurriculum der Freien und Hansestadt Hamburg enthalten sind, auf der x-Achse aufgeführt. Die Items, aus denen der Test besteht, sind auf der y-Achse aufgeführt. In der so entstandenen Kreuztabelle wurde je Item geratet, ob das Thema des typischen Schülerfehlers in dem Themengebiet des Beispielcurriculums vorkommt (1) oder nicht (0).

Hinweis: Ein Thema des typischen Schülerfehlers kann erst dann mit 1 (enthalten) aufgeführt werden, wenn es zuvor eingeführt wurde. Bsp. Die Bruchrechnung wird erst in der 5/6 Klasse eingeführt, könnte allerdings schon in der Grundschule bei der Berechnung von Wahrscheinlichkeiten ihre Anwendung finden. Dies galt es beim Rating zu beachten.

Rating des Merkmals *strukturelle Komplexität*

Die *Anzahl der Zeichen*, aus denen die Items bestehen, soll bestimmt werden. Dafür gelten nach Literaturgrundlage die folgenden Regeln:

- Ein Bruch wird als 3 Zeichen gezählt, da zunächst erfasst werden muss, dass es sich um einen Bruch handelt. Des Weiteren muss der Wert des Bruchs bestimmt werden, hier müssen Zähler und Nenner in Beziehung gesetzt werden.
- Eine Zahl wie bspw. 30 wird hingegen als ein Zeichen gezählt, da die 3 und die 0 im Stellenwertsystem miteinander in Verbindung stehend gesehen werden muss.
- Zeichen, die eine Operation beschreiben, wie „+, −, ·, %, √“ oder auch „±, =, sin“, werden als ein Zeichen betrachtet.
- Enthaltene Einheiten wie „cm, ° oder m“ werden als ein Zeichen gezählt.
- Bei einigen Items ist über den Antwortmöglichkeiten eine Grafik zu sehen, die darin enthaltenen Zeichen müssen ebenfalls gezählt werden. Die Zeichen müssen allerdings nicht zu jeder Antwortmöglichkeit einzeln addiert werden, sondern nur einmal gesondert vermerkt werden.

In den Items enthaltene Regeln:

Die verwendeten Regeln, die zu einer Lösung als bekannt vorausgesetzt und korrekt angewandt werden müssen, sollen ebenfalls gezählt **und** aufgeführt werden.

Während bei einer einfachen Addition von zwei einstelligen Ziffern unter 5 nur eine Regel verwendet werden muss, ist es notwendig, bei einer Addition bei Ziffern über 5 neben der zuvor genannten Regel ebenfalls die des Stellenwerts einzuhalten. Wichtig beim Raten der Items ist das Aufführen der verwendeten Regeln, hier gilt es möglichst elementar vorzugehen. Man beginne mit den Regeln des Stellenwerts und „untersuche“ die einzelnen Operationen, die weiterhin notwendig sind.