




Einschätzung mathematischer Fähigkeiten 4- bis 6-jähriger Kinder im Bereich Mengen und Zahlen durch frühpädagogische Fachkräfte

Lena Aumann  · Hedwig Gasteiger · Laura Tabeling · Rosa Maria Puca

Eingegangen: 4. November 2020 / Angenommen: 7. Dezember 2021 / Online publiziert: 12. Januar 2022
© Der/die Autor(en) 2022

Zusammenfassung Frühpädagogische Fachkräfte in Kindertageseinrichtungen sind durch entsprechende Bestimmungen auf Bundes- und Landesebene angehalten, Kinder auf Grundlage ihres aktuellen Lernstandes zum fachlich tragfähigen Weiterlernen anzuregen und individuelle Entwicklungsverläufe zu dokumentieren. Dafür müssen die Fachkräfte in der Lage sein, den Lernstand der Kinder adäquat einzuschätzen. Die hier berichtete Studie analysiert die Genauigkeit von Einschätzungen frühpädagogischer Fachkräfte ($N=49$) von Leistungen vier- bis sechsjähriger Kinder ($N=160$) in 17 Items im Bereich Mengen und Zahlen. Zur Aufklärung von Varianz in der Einschätzungsgenauigkeit wurden das Geschlecht, der sozioökonomische Status und die mathematische Leistung der Kinder im Bereich Mengen und Zahlen sowie das mathematikdidaktische Wissen der Fachkräfte als mögliche Merkmale erhoben. Es zeigte sich im Schnitt eine mäßige Einschätzungsgenauigkeit mit einer leichten Tendenz zur Überschätzung der Leistungen. Als einziges einflussnehmendes Merkmal auf die Einschätzungsgenauigkeit stellte sich in dieser Untersuchung die Leistung der Kinder im Bereich Mengen und Zahlen heraus: Die Einschätzung der frühpädagogischen Fachkräfte war genauer, wenn die Kinder bessere mathematische Fähigkeiten zeigten. Der sozioökonomische Status und das Geschlecht der Kinder sowie das mathematikdidaktische Wissen der Fachkräfte zeigten keinen signifikanten Einfluss auf die Einschätzungsgenauigkeit.

Schlüsselwörter Frühpädagogische Fachkräfte · Diagnostische Fähigkeiten · Einschätzungsgenauigkeit · Frühes mathematisches Lernen · Kindertagesstätte

Lena Aumann (✉) · Hedwig Gasteiger · Laura Tabeling
Institut für Mathematik (Mathematikdidaktik), Universität Osnabrück, Albrechtstraße
28A, 49076 Osnabrück, Deutschland
E-Mail: lena.aumann@uni-osnabrueck.de

Rosa Maria Puca
Institut für Psychologie (Pädagogische Psychologie), Universität Osnabrück, Osnabrück, Deutschland

MESC-Codes C601 · D601

Assessments of 4 to 6 Year old Children's Number Skills by Early Childhood Teachers

Abstract In Germany, regulations at federal and state levels require early childhood teachers to encourage children's learning based on their current level of education and to document individual development. For this, teachers have to be able to adequately assess children's level of education. The current study analyzes early childhood teachers' ($N=49$) ability to accurately assess number skills of four to six year old children ($N=160$) on 17 selected items. As possible influences on assessment accuracy, we also collected data on children's sex, socio-economic status and their actual number skills as well as on teachers' mathematical pedagogical content knowledge. We found, on average, that teachers were able to fairly assess the children's number skills, with a slight tendency to overestimate their performance. It turned out that in this study children's number skills are the only influencing factor on early childhood teachers' assessment accuracy. The assessment was more accurate when the children showed better math skills. Socio-economic status and sex of the child as well as the teachers' mathematical pedagogical content knowledge was not predictive for early childhood teachers' assessment accuracy.

Keywords Early childhood teachers · Diagnostic skills · Assessment accuracy · Early mathematics education · Kindergarten

1 Einleitung

Allgemeine und bereichsspezifische Fähigkeiten von Kindern angemessen einschätzen zu können, ist für adäquates pädagogisches Handeln, mit dem Ziel, eine anregende Lernumgebung zu schaffen, unumstritten (Fisher et al. 1981; Schrader 1989; Rogalla und Vogt 2008; Anders et al. 2010). Im Schulkontext ist es bspw. eine zentrale Aufgabe von Lehrkräften, genau einzuschätzen, welche Lernvoraussetzungen ihre Schülerinnen und Schüler mitbringen, welche individuellen Kenntnisse und Fertigkeiten sie haben und welche Aufgaben sie bewältigen können. Nur dann können sie den Lernenden passgenaue Lernmöglichkeiten bieten, die an deren Vorwissen ansetzen und so ein Weiterlernen ermöglichen (Anders et al. 2010; Artelt und Gräsel 2009; Rogalla und Vogt 2008; Schrader und Praetorius 2018; Spinath 2005). Mit Blick auf internationale Schulleistungsuntersuchungen wird jedoch deutlich, dass gerade die Einschätzung des Vorwissens von Schülerinnen und Schülern und insbesondere das Erkennen von spezifischen Schwierigkeiten offensichtlich keine leichte Aufgabe ist: Nur ca. 11 % der in PISA 2000 diagnostizierten Risikoschülerinnen und -schüler, denen erhebliche Schwierigkeiten bei einem Einstieg ins Berufsleben prognostiziert werden konnten, wurden von den Lehrkräften korrekt als besonders gefährdet eingeschätzt (Baumert et al. 2001). Es scheint insofern begründet, dass die KMK seit 2001 methodische und diagnostische Kompetenz als eines von sieben Handlungsfeldern für die Professionalisierung von Lehrkräften aufführt und Maß-

nahmen zum Erwerb und zur Fortbildung dieser Kompetenzen als Teilbereich des professionellen Handelns fordert (Avenarius et al. 2003).

Im Gegensatz zum Schulkontext ist nur wenig untersucht, wie genau frühpädagogische Fachkräfte kindliche Kompetenzen einschätzen. Ein Blick in die Forderungen der Kultusminister- und Jugendministerkonferenz des Bundes und der für den vorschulischen Bereich zuständigen Ministerien einzelner Länder zeigt jedoch, dass sich die Grundaufgabe der frühpädagogischen Fachkräfte nicht wesentlich von denen der Lehrkräfte unterscheidet: Kinder sollen auf Grundlage ihres Lernstandes und ihrer individuellen Lernmöglichkeiten mit adaptiven Angeboten und durch individuelle Förderung zum fachlich tragfähigen Weiterlernen angeregt werden (Benz et al. 2017; JMK/KMK 2004; NMW 2018; MKFFI NRW 2018; StMAS 2016). Um Kinder also angemessen begleiten und unterstützen zu können, müssen sich Fachkräfte einen guten Überblick darüber verschaffen, welche Lernvoraussetzungen bei den Kindern bereits vorliegen (Benz et al. 2015; Dollinger 2013). Dies geschieht v. a. durch Beobachtung der Kinder in Alltagssituationen und einer Einschätzung der individuellen Fähigkeiten (Benz et al. 2017). Basierend auf den Voraussetzungen der Kinder kann die frühpädagogische Fachkraft dann den täglichen Gruppenalltag im Hinblick auf Anregungen zum Weiterlernen auf unterschiedlichem Niveau planen und reflektieren.

Besondere Bedeutung kommt der Einschätzung kindlicher Fähigkeiten zu, wenn beobachtete Fähigkeiten und frühe Entwicklungsprozesse nicht nur der Charakterisierung des Lernstandes dienen: Bund und Länder fordern in ihren Vorgaben für Kindertageseinrichtungen das Erkennen von Entwicklungsrisiken oder besonderen Begabungen (JMK/KMK 2004; NMW 2018; MKFFI NRW 2018; StMAS 2016). Bezogen auf Entwicklungsrisiken erwiesen sich insbesondere frühe mathematische Kompetenzen zu Mengen und Zahlen als maßgeblich prädiktiv für spätere (mathematische) Schulleistungen der Kinder (Duncan et al. 2007; Nguyen et al. 2016; Jordan et al. 2010; Dornheim 2013; Krajewski und Schneider 2006; Weißhaupt et al. 2006). Deshalb ist insbesondere eine genaue Einschätzung der kindlichen Kompetenzen im Bereich Mengen und Zahlen von Bedeutung. Sie ist sowohl Voraussetzung für eine passgenaue Förderung als auch von Relevanz, wenn es darum geht, besondere Begabungen oder Entwicklungsrückstände frühzeitig zu erkennen, die unmittelbare Auswirkungen auf Lern- und Entwicklungschancen von Kindern haben können (Gasteiger und Benz 2016).

2 Informelle Einschätzungen im Kontext von diagnostischer Kompetenz

Während schriftliche und mündliche Lernstandserfassungen und das Fällen diagnostischer Urteile im Schulkontext zum Aufgabenspektrum der Lehrkräfte gehören, sind es im frühpädagogischen Kontext eher Beobachtungen und informelle Einschätzungen im Alltag, die der Fachkraft einen Überblick über Fähigkeiten des Kindes geben und die „eine wichtige Rolle für die [...] Gestaltung von Spiel- und Lernumgebungen und schließlich für den Kompetenzerwerb der Kinder“ spielen (Benz et al. 2017). Diese im Alltag der Kindertagesstätte vorgenommenen *informellen Einschätzungen*

können der *Diagnostischen Kompetenz* zugeordnet werden und sind von *formalen Diagnosen* abzugrenzen.

Diagnostische Kompetenz ist nach Schrader und Praetorius (2018, S. 92) die „Fähigkeit, die [...] anfallenden Diagnoseleistungen so zu erbringen, dass sie bestimmten Güte- oder Qualitätskriterien genügen“. Diagnoseleistungen in pädagogischen Kontexten können formal und methodisch abgesichert sein, aber auch auf informellen, subjektiven Einschätzungen beruhen (Schrader und Praetorius 2018). *Formale Diagnosen* bedienen sich dabei standardisierter Beobachtungen oder Tests, aus denen systematisch und reflektiert Ergebnisse gewonnen werden (Schrader und Praetorius 2018; Dollinger 2013). Diese Form der Diagnose erlaubt es, Leistungen und Entwicklungsstände von Kindern methodisch abgesichert zu beurteilen. So können mit formalen Diagnoseverfahren bspw. Risiken für Schwierigkeiten beim Mathematiklernen oder für Rechenschwäche festgestellt werden (Benz et al. 2015). Die *alltägliche* Aufgabe von frühpädagogischen Fachkräften ist jedoch nicht die Beurteilung der kindlichen Leistung, sondern die Persönlichkeitsstärkung und die Förderung des natürlichen Wissensdrangs und der Lernfreude der Kinder (§ 22 Abs. 1 SGB VIII; § 2 Abs. 1 NKiTaG). Da in Kindertagesstätten daher in der Regel keine standardisierten Diagnosen mit dem Ziel der Beurteilung der frühen mathematischen Fähigkeiten von Kindern gefordert sind (anders als für sprachliche Fähigkeiten in einigen Bundesländern, s. bspw. § 3 Abs. 1 KiTaG; § 12 Abs. 2 BayKiBiG; § 19 Abs. 3 KiBiz), nimmt die formale Diagnostik mathematischer Fähigkeiten im alltäglichen Handeln der Kindertagesstätte eine eher untergeordnete Rolle ein. Im Elementarbereich sind daher *informelle Einschätzungen*, in der Regel in Form von sensiblen (Alltags-)Beobachtungen über einen längeren Zeitraum, von großer Bedeutung (Benz et al. 2017; Kilday et al. 2012; Kowalski et al. 2018). Sie sind eher implizit, treten also in alltäglichen Situationen der Kindertagesstätte auf und haben immer auch einen subjektiven Charakter. Gleichzeitig sind sie oft Grundlage für Überlegungen zur adaptiven Gestaltung des Gruppenalltags oder zur Auswahl von Angeboten und individuellen Fördermaßnahmen (Dollinger 2013).

In Untersuchungen zur Qualität von Einschätzungen und zur Operationalisierung dieser informellen Form der diagnostischen Kompetenz wird häufig die *Einschätzungsgenauigkeit* herangezogen. Die Einschätzungsgenauigkeit ist dabei in der Regel die Übereinstimmung zwischen den getroffenen Einschätzungen (z. B. zur Intelligenz, Mathematikleistung) und den in der Realität tatsächlich vorhandenen Merkmalsausprägungen (z. B. gemessen durch standardisierte Tests) (Helmke 2021; Schrader und Praetorius 2018). Dabei kann die Einschätzungsgenauigkeit davon abhängen, ob direkte oder indirekte Einschätzungen getroffen werden. *Indirekt* oder *uninformiert* sind Einschätzungen genereller Merkmale oder Fähigkeiten. Eine indirekte Einschätzung wäre also bspw., ob das Kind über eine Vorstellung von Mengen mit bis zu zehn Elementen verfügt. *Direkt* oder *informiert* hingegen sind Einschätzungen, bei denen die einschätzende Person eine ganz konkrete Leistung der einzuschätzenden Personen in einer bestimmten Aufgabe (z. B. in einem Test) einschätzt (Kowalski et al. 2018), bspw., ob das Kind die richtige Anzahl nennen kann, wenn eine Menge von sechs Plättchen gezeigt wird. Erwartungsgemäß sind direkte Einschätzungen genauer als indirekte, da die einschätzende Person Informationen darüber erhält, welche konkrete Fähigkeit eingeschätzt werden soll, während sich die

Fachkräfte bei indirekten Einschätzungen eher auf allgemeine, ganzheitliche Kenntnisse des Kindes stützen müssen (Hoge und Coladarci 1989; Südkamp et al. 2012; Dollinger 2013; Ostermann et al. 2019).

Während formale Diagnosen bei korrekter Durchführung in der Regel zielgerichtet methodisch gesicherte Ergebnisse zum Leistungsstand der Kinder liefern, fordern informelle Einschätzungen von den frühpädagogischen Fachkräften einiges an Fachkompetenz: das Wissen, welche Alltagssituationen das Potenzial bieten, frühe mathematische Kompetenzen beobachten zu können und die Kompetenz, Informationen aus Beobachtungen und Äußerungen im Alltag so zu bündeln, dass der Lernstand des Kindes genau eingeschätzt werden kann (Gasteiger und Benz 2016).

3 Empirische Erkenntnisse zur Einschätzungsgenauigkeit im pädagogischen Kontext

Trotz der Relevanz informeller Einschätzungen für die tägliche Arbeit in Kindertagesstätten gibt es zur Einschätzungsgenauigkeit frühpädagogischer Fachkräfte bislang nur wenige empirische Erkenntnisse. Aus diesem Grund werden zunächst einige Ergebnisse zur Einschätzungsgenauigkeit von Lehrkräften herangezogen, um Erkenntnisse über mögliche Schwierigkeiten bei der Einschätzung mathematischer Kompetenzen oder über mögliche Zusammenhänge zu gewinnen. Daraufhin werden Ergebnisse zur Einschätzungsgenauigkeit mathematischer Kompetenzen von frühpädagogischen Fachkräften betrachtet.

3.1 Einschätzungsgenauigkeit von Lehrkräften

Unterschiedliche Studien kommen in der Untersuchung der Einschätzungsgenauigkeit von Lehrkräften zu heterogenen Ergebnissen. Studien, in denen Lehrkräfte allgemeine Merkmale von Schülerinnen und Schülern einschätzten (bspw. Intelligenz, schulische Motivation, Leistungsängstlichkeit oder methodische Kompetenzen), zeigen eine durchschnittlich eher geringe Genauigkeit (Spinath 2005; McElvany et al. 2009). In Studien zur Einschätzung von fachlichen (bspw. mathematischen) Leistungen von Kindern wird hingegen berichtet, dass Lehrkräfte eine im Durchschnitt moderate bis gute Einschätzungsgenauigkeit aufweisen (Lorenz 2011; Hosenfeld et al. 2002; Südkamp et al. 2012; Hoge und Coladarci 1989; Karing 2009; Dollinger 2013). In der Studie von Lorenz (2011) wurden die Lehrkräfte u. a. um indirekte Einschätzungen zu Arithmetikleistungen der Kinder gebeten (bspw. „*Er/sie beherrscht die Grundrechenarten.*“, „*Er/sie hat ein gutes Verständnis für Zahlen.*“). Lorenz (2011) spricht dabei von einer relativ guten Einschätzung der Lehrkräfte. Auch Hoge und Coladarci (1989) bzw. Südkamp et al. (2012) berichten in ihren Metaanalysen über 16 bzw. 75 Studien zur diagnostischen Kompetenz von Lehrkräften im Durchschnitt ein hohes Level an Übereinstimmung zwischen den Einschätzungen der Lehrkräfte und den Testleistungen der Schülerinnen und Schüler und entdecken dabei keinen Unterschied zwischen dem sprachlichen und dem mathematischen Bereich. Beide Analysen berücksichtigen dabei Studien sowohl mit direkten als auch indirekten Einschätzungen und unterschiedlicher Spezifität (*sehr unspezifisch: Lehr-*

kraft unterteilt Klassen in Leistungsfünftel und ordnet die Schülerinnen und Schüler entsprechend ein; sehr spezifisch: Lehrkraft schätzt für jedes Item eines Tests ein, ob es vom jeweiligen Kind korrekt gelöst wurde oder nicht) (Hoge und Coladarci 1989; Südkamp et al. 2012).

Verschiedene Studien zur Einschätzungsgenauigkeit weisen darüber hinaus auf unterschiedliche Ergebnisse zu Über- oder Unterschätzungstendenzen hin. Anders et al. (2010) untersuchten u. a. die Fähigkeit der Lehrkräfte, ihre Schülerinnen und Schüler hinsichtlich ihrer Leistung in mathematischen Aufgaben im PISA-Test in eine Rangreihenfolge zu bringen, während Lehrkräfte in der Studie von Hosenfeld et al. (2002) u. a. dazu aufgefordert wurden, einzuschätzen, wie viele Schülerinnen und Schüler ihrer Lerngruppe bestimmte Aufgaben eines standardisierten Mathematiktests lösen konnten. Beide Studien berichten, dass Lehrkräfte die fachlichen Leistungen von Schülerinnen und Schülern tendenziell eher überschätzen. Es wird vermutet, dass leistungsmindernde Einflüsse der Testsituation, wie begrenzte Bearbeitungszeit, Flüchtigkeitsfehler, Leistungsangst, Aufregung oder ähnliches, für eine Überschätzung ausschlaggebend sein könnten, weil Lehrkräfte die Fähigkeiten der Kinder eher unter optimalen Bedingungen einschätzen. In der Studie von Dollinger (2013) gaben Lehrkräfte (und frühpädagogische Fachkräfte) aufgabenweise die Leistung von insgesamt 633 Kindern im Übergang zwischen Kindertagesstätte und Schule bei einem Test zur Erfassung der schulrelevanten Lernausgangslage an (direkte Einschätzung). Der Test enthielt neben Items zu sprachlichen Kenntnissen u. a. auch Aufgaben zu frühen mathematischen Fähigkeiten von Kindern im Bereich Mengen und Zahlen (Anzahlbestimmung, Kenntnis von Zahlsymbolen, Rechnen mit Objekten und Punkten, erste arithmetische Fähigkeiten). In dieser Studie wird berichtet, dass Lehrkräfte die Fähigkeiten der Kinder signifikant unterschätzen. Eine Erklärung dafür sucht die Autorin in einer möglichen Überzeugung der Lehrkräfte von einer Art mathematischen „Stunde Null“ (Selter 2008), in welcher der Eintritt in die Grundschule den Startpunkt des mathematischen Lernens der Kinder markiert und bereits bestehende Kompetenzen kaum Beachtung finden (Dollinger 2013).

Die meisten Untersuchungen zur Genauigkeit von Einschätzungen durch Lehrkräfte berichten übereinstimmend, dass große individuelle Unterschiede bestehen, wie genau Lehrkräfte die Leistungen ihrer Lernenden einschätzen (Anders et al. 2010; Lorenz 2011; Spinath 2005; McElvany et al. 2009; Hosenfeld et al. 2002; Hoge und Coladarci 1989; Südkamp et al. 2012; Karing 2009; Dollinger 2013). Dies wirft die Frage auf, welche spezifischen Merkmale der Schülerinnen und Schüler sowie der Lehrkräfte die Einschätzungsgenauigkeit beeinflussen. Auf Seiten der Lehrkräfte zeigte entgegen der Erwartungen weder vertieftes fachdidaktisches Wissen noch Berufserfahrung der Lehrkräfte einen gesicherten Zusammenhang mit deren Einschätzungsgenauigkeit (Anders et al. 2010; Schrader 1989; Lorenz 2011; McElvany et al. 2009; Dollinger 2013; Muntoni et al. 2019). Anders verhält es sich bei den Merkmalen von Schülerinnen und Schülern. Muntoni et al. (2019) untersuchten indirekte Einschätzungen von Lehrkräften zu zukünftigen Leistungen der Schülerinnen und Schüler in arithmetischen Inhalten (bspw. „Wie wird dieser Schüler/diese Schülerin am Ende des Schuljahres 2 in Mathematik abschneiden?“) und fanden dabei Einflüsse des Geschlechts der Kinder auf die Einschätzungsgenauigkeit. Ehm et al. (2011) erhoben indirekte Einschätzungen der Mathematikleistungen von Erstklässle-

rinnen und Erstklässlern durch ihre Lehrkräfte (ein Item, Einschätzung der globalen Mathematikleistung auf einer fünfstufigen Skala von „*deutlich unterdurchschnittlich*“ bis „*sehr gut*“) und berichteten ebenso von Einflüssen des Kindergeschlechts. In beiden Studien wie auch in der Studie von Lorenz (2011) schätzten Lehrkräfte die Leistungen von Jungen und Mädchen im Schnitt entsprechend der oft anzutreffenden Stereotype ein, dass Jungen bessere Leistungen in mathematischen und Mädchen in sprachbezogenen Bereichen haben, wobei keine entsprechenden Unterschiede in den tatsächlichen Leistungen der Kinder berichtet werden (Lorenz 2011; Ehm et al. 2011; Muntoni et al. 2019). In der Studie von Dollinger (2013) zeigten sich in der Stichprobe der Schülerinnen und Schüler tatsächliche geschlechtsspezifische Leistungsunterschiede in den mathematischen Kompetenzen zugunsten der Jungen, die stereotypisch genauso von den Lehrkräften eingeschätzt wurden.

Einen Einfluss auf die Einschätzungsgenauigkeit von Lehrkräften nimmt auch der sozioökonomische Status der Kinder (Lorenz 2011). Die untersuchten Lehrkräfte zeigten eine größere Tendenz zur Überschätzung bei der Einschätzung von Kindern mit einem höheren sozioökonomischen Status und zur Unterschätzung bei Kindern mit niedrigem sozioökonomischen Status. Dieser Effekt zeigte sich am stärksten in der Einschätzung sprachlicher Fähigkeiten der Kinder, konnte aber in zwei von drei Messzeitpunkten auch bzgl. arithmetischer Fähigkeiten der Kinder nachgewiesen werden (Lorenz 2011). Ein weiterer Einflussfaktor auf die Einschätzungsgenauigkeit zeigte sich in eben dieser Studie in der tatsächlichen Leistung der Kinder. Leistungsstarke Kinder wurden von ihren Lehrkräften in allen Bereichen und zu allen Messzeitpunkten häufiger unter- und leistungsschwache häufiger überschätzt, wobei dieser Effekt im Bereich Arithmetik besonders stark ausgeprägt war (Lorenz 2011).

3.2 Einschätzungsgenauigkeit von frühpädagogischen Fachkräften in Kindertageseinrichtungen

Zur Einschätzungsgenauigkeit von frühpädagogischen Fachkräften gibt es nur wenige Studien im deutsch- und englischsprachigen Raum. Die Ergebnisse von Untersuchungen im vorschulischen Bereich verweisen – wie auch die Studien mit Lehrkräften – auf große individuelle Unterschiede zwischen den Fachkräften (Dollinger 2013; Kilday et al. 2012; Kowalski et al. 2018; McKeveit und Kiss 2019). Genauer betrachtet zeigen die Ergebnisse zur Genauigkeit von Einschätzungen frühpädagogischer Fachkräfte, insbesondere betreffend früher mathematischer Fähigkeiten, ein eher negatives Bild: Es wird durchschnittlich von einer eher schlechten Einschätzungsgenauigkeit früher mathematischer Fähigkeiten berichtet (Dollinger 2013; Kilday et al. 2012; McKeveit und Kiss 2019; Kowalski et al. 2018). Insbesondere werden frühe mathematische Fähigkeiten im Bereich Mengen und Zahlen (*Zählfertigkeiten, Kenntnis von Zahlsymbolen, Zahlzerlegungen, erstes Mengenverständnis, erste arithmetische Fähigkeiten*; Dollinger 2013; Kowalski et al. 2018; McKeveit und Kiss 2019), aber auch erste geometrische Fähigkeiten (*Muster legen, geometrische Formen erstellen*; Kowalski et al. 2018) weniger genau eingeschätzt als sprachliche. Die Autorinnen und Autoren erklären dies dadurch, dass den mathematischen Fä-

higkeiten im vorschulischen Bereich häufig weniger Relevanz zugeschrieben oder zumindest weniger Beachtung zuteilwird als frühen sprachlichen Fähigkeiten.

Ähnlich zu Studien im schulischen Bereich gibt es auch für die Einschätzungsgenauigkeit frühpädagogischer Fachkräfte kontroverse Ergebnisse bzgl. Über- und Unterschätzungstendenzen. In der Studie von Kowalski et al. (2018) wurden direkte Einschätzungen von 66 Fachkräften im Vorschulbereich zu 122 Kindern im Alter von drei bis sechs Jahren erhoben. Die Pädagoginnen und Pädagogen schätzten in 68 Items, darunter in 23 zu frühen mathematischen Fähigkeiten (*bspw. Mengen zählen und vergleichen, Zahlen schreiben oder einfache Formen zeichnen und benennen*), ein, ob das jeweilige Kind die Aufgaben bewältigen kann. Es zeigte sich eine Tendenz zur Überschätzung der erhobenen mathematischen Fähigkeiten der Kinder. Die Autorinnen und Autoren erklärten dies, ähnlich wie bei den Untersuchungen im schulischen Bereich, damit, dass eine Überschätzung an der ungewohnten Testsituation liegen könnte. Im Gegensatz dazu beschrieb Dollinger (2013) Tendenzen zur Unterschätzung bei der Einschätzung früher mathematischer Kompetenzen und begründete dies, wie im Schulkontext, mit einer möglichen Überzeugung der Fachkräfte von einer Art mathematischen „Stunde Null“ (Selter 2008; s. Abschn. 3.1).

Bei der Betrachtung möglicher Zusammenhänge von Merkmalen der Fachkraft mit deren Einschätzungsgenauigkeit wurde meist ein Zusammenhang mit der Berufserfahrung der Fachkräfte untersucht, der sich, ähnlich zu Untersuchungen im Schulkontext, in verschiedenen Studien nicht zeigte (Dollinger 2013; Kowalski et al. 2018; Furnari et al. 2017).

Für Merkmale auf Seiten der Kinder zeigten sich z.T. kontroverse Ergebnisse. Während die Studie von Dollinger (2013) signifikante Einflüsse des Geschlechts der Kinder dahingehend fand, dass Fachkräfte die mathematischen Fähigkeiten im Bereich Mengen, Zahlen und Zählen von Mädchen niedriger einschätzen, konnten Furnari et al. (2017) und Kowalski et al. (2018) keine Geschlechtsunterschiede feststellen. Weitere Untersuchungen und Einzelfallexperimente geben jedoch vorsichtige Hinweise darauf, dass frühpädagogische Fachkräfte Kindern stereotypische geschlechtsspezifische Stärken und Schwächen zuweisen – insb. den Jungen höhere mathematische Kompetenzen (Kuger et al. 2011; BBC Stories 2017). Die Studie von Furnari et al. (2017) untersuchte, mit welchen Personenmerkmalen auf Seiten der Fachkräfte und der Kinder die Einschätzungen frühpädagogischer Fachkräfte zusammenhängen. Es wurden frühe mathematische Fähigkeiten von 435 Kindern durch 42 frühpädagogische Fachkräfte in 12 Items in den Bereichen Mengen und Zahlen (z. B. „*identifies and understands ordinality*“ oder „*identifies and understands cardinality*“), Geometrie und Messen (z. B. „*uses instruments accurately for measuring*“) indirekt eingeschätzt (von „*not yet*“ bis „*proficient in the skill*“).

Für die Kindermerkmale Alter (Kowalski et al. 2018) und Migrationshintergrund (Dollinger 2013) konnte in den genannten Studien kein Zusammenhang mit den Einschätzungen frühpädagogischer Fachkräfte gefunden werden, während ein möglicher Zusammenhang mit dem familiären Hintergrund in amerikanischen Studien kontrovers dargestellt wird. In der Untersuchung von Furnari et al. (2017) wurde kein Zusammenhang der Einschätzungen mathematischer Fähigkeiten im Bereich Mengen und Zahlen mit dem sozioökonomischen Status der Kinder gefunden, wo-

bei sich ein Zusammenhang mit der ethnischen Zugehörigkeit gezeigt hat, der sich wiederum bei Kowalski et al. (2018) nicht herausstellte.

Als weiteres einflussnehmendes Merkmal auf die Einschätzungsgenauigkeit frühpädagogischer Fachkräfte wurde die tatsächliche mathematische Leistung der Kinder im Bereich Mengen und Zahlen untersucht. In der Studie von McKeveit und Kiss (2019) zum Einfluss verschiedener Kindermerkmale auf die Einschätzungsgenauigkeit schätzten sieben Pädagoginnen und Pädagogen die Leistungen von 126 Kindern in standardisierten Einzeltests u. a. zu frühen mathematischen Kompetenzen im Bereich Mengen und Zahlen ein (*je ein Testabschnitt zu Zählen, Zahlen, Zahlzerlegungen und Zahldarstellungen*). Die Fachkräfte wurden gebeten, einzuschätzen, inwiefern die Kinder die Aufgaben im Test lösen konnten (*keine Lösung trotz Hilfe, Lösung mit Hilfe, Lösung ohne Hilfe*). Es zeigte sich, dass die Fachkräfte zu genaueren Einschätzungen fähig waren, wenn die Kinder eine durchschnittliche oder überdurchschnittliche Leistung in Mathematik zeigten. Bei Kindern im unteren Niveau war die durchschnittliche Genauigkeit der Einschätzungen signifikant schlechter, da die Fachkräfte hier eher zur Überschätzung der kindlichen Leistung tendierten (McKeveit und Kiss 2019). Dieses Ergebnis deckte sich mit den Hypothesen der Autorinnen, da sich in vorangegangenen Studien zur Einschätzung von Lesekompetenzen ähnliche Ergebnisse zum Einfluss der tatsächlichen Leistung der Kinder gezeigt hatten. Diese Überschätzungstendenz scheint jedoch besonders relevant zu sein, wenn es um die Identifizierung von Kindern geht, die Gefahr laufen, später gravierende Schwierigkeiten bzgl. ihrer Mathematikleistung zu entwickeln. Es besteht die Gefahr, dass Kinder mit geringeren Fähigkeiten infolge fehlender Kompetenz frühpädagogischer Fachkräfte, Schwierigkeiten im Mathematiklernen zu erkennen bzw. infolge von Überschätzung, keine entsprechende frühzeitige Förderung erhalten könnten (McKeveit und Kiss 2019).

4 Zusammenfassung und Forschungsfragen

Eine gute Einschätzungsgenauigkeit ist unumgänglich, um Kindern eine passgenaue Förderung zu bieten, die an Vorkenntnisse anknüpft und zum Weiterlernen anregt. Im Hinblick auf das frühe Erkennen von Entwicklungsrisiken ist eine strukturierte Beobachtung der Kinder durch die pädagogischen Fachkräfte von Nöten (Dollinger 2013; NMW 2018), da die frühen mathematischen Fähigkeiten im Bereich Mengen und Zahlen in der Regel nicht mit standardisierten Verfahren erhoben werden, aber prädiktive Auswirkungen auf spätere mathematische Fähigkeiten haben (Duncan et al. 2007; Nguyen et al. 2016; Jordan et al. 2010; Dornheim 2013; Krajewski und Schneider 2006; Weißhaupt et al. 2006). Dazu brauchen Fachkräfte in Kindertageseinrichtungen Kompetenzen, mit denen sie individuumsbezogen im Alltag diagnostizieren und angemessen fördern können (Gasteiger und Benz 2016). Zur Einschätzungsgenauigkeit frühpädagogischer Fachkräfte gibt es jedoch nur wenige, teils widersprüchliche, Ergebnisse – vor allem auch hinsichtlich möglicher Zusammenhänge mit Personenmerkmalen. Diesbezüglich folgt eine Zusammenfassung, die das bestehende Forschungsdesiderat herausstellt.

Die Einschätzungsgenauigkeit und mögliche Zusammenhänge wurden vornehmlich in amerikanischen Studien erforscht. Die Untersuchungen von McKeveit und Kiss (2019), Furnari et al. (2017) und Kowalski et al. (2018) liefern erste konkrete Ergebnisse, wie gut frühpädagogische Fachkräfte die mathematischen Leistungen von Kindern im Bereich Mengen und Zahlen einschätzen und wie diese Einschätzungen mit Merkmalen des Kindes (Geschlecht, sozioökonomischer Status, tatsächliche Leistung) sowie Merkmalen der Fachkraft (Berufserfahrung) zusammenhängen. Die Ergebnisse dieser Studien können jedoch nicht direkt auf das deutsche System der Kindertageseinrichtungen übertragen werden: Hier steht das individuelle Weiterlernen ohne ein gemeinsam zu erreichendes Bildungsziel für alle Kinder im Fokus (JMK/KMK 2004), während amerikanische Kindergärten eher vorschulischen Angeboten mit eigenen curricularen Vorgaben entsprechen (Common Core State Standards Initiative 2021). Auch in der Ausbildung der Pädagoginnen und Pädagogen sind die Bildungssysteme nur bedingt vergleichbar, da frühpädagogische Fachkräfte in den USA in der Regel ein Studium absolvieren (Furnari et al. 2017; Kowalski et al. 2018; McKeveit und Kiss 2019), während die Ausbildung in Deutschland oft fachschulisch gestaltet ist und nur ca. ein Zehntel der fachschulisch ausgebildeten Fachkräfte ein Studium aufnimmt (Mink und Müller 2018).

Als Studie im deutschen Bildungssystem vergleicht Dollinger (2013) die Einschätzungsgenauigkeit von Lehrkräften und frühpädagogischen Fachkräften und betrachtet den Zusammenhang einiger Personenmerkmale, wie Geschlecht oder Migrationsstatus des Kindes oder die Berufserfahrung der Pädagoginnen und Pädagogen, mit deren Einschätzungsgenauigkeit. Dollinger (2013) fokussiert den Übergangsbereich und so insbesondere die Einschätzung früher mathematischer *Rechenfähigkeiten* (Rechnen mit Objekten und Punkten, erste arithmetische Aufgaben). Weitere zentrale Prädiktoren für spätere Mathematikleistung, wie Zählfertigkeiten, Verständnis von Mengenrelationen oder Teil-Ganzes, werden außer Acht gelassen.

Sowohl einige Studien im Schulbereich (Anders et al. 2010; Lorenz 2011; McElvany et al. 2009; Muntoni et al. 2019) als auch die Studien von Furnari et al. (2017), Kowalski et al. (2018) und Dollinger (2013) betrachten die Berufserfahrung der Pädagoginnen und Pädagogen und berichten keinen Zusammenhang mit deren Einschätzungen. Anders et al. (2010) folgern aus ihren Ergebnissen im Schulkontext jedoch, dass das allgemeine und fachbezogene didaktische Wissen der einschätzenden Person betrachtet werden sollte, um diagnostische Prozesse zu erklären. Es könnte vermutet werden, dass die mathematikdidaktische Kompetenz, die als eng verwoben mit der Kompetenz zur individuellen Diagnose und Förderung gesehen wird (Gasteiger und Benz 2016), mit der Fähigkeit, Kinder genau einzuschätzen, zusammenhängt. Eine Untersuchung des Zusammenhangs des fachdidaktischen Wissens von frühpädagogischen Fachkräften mit deren Einschätzungsgenauigkeit steht somit noch aus.

Die bisherigen Studien konnten zwar interessante Ergebnisse zu Zusammenhängen einzelner Merkmale mit der Einschätzungsgenauigkeit hervorbringen, jedoch erscheint die gemeinsame Berücksichtigung verschiedener Merkmale und damit einhergehend die Fragestellung, welchen Merkmalen eine hohe oder weniger hohe Gewichtung zugeschrieben werden kann, weitgehend unbeantwortet. Hauptziel dieser Arbeit ist daher, die Einschätzungsgenauigkeit frühpädagogischer Fachkräfte

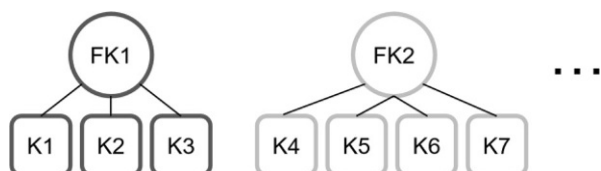
zu untersuchen und Zusammenhänge mit den potenziellen Merkmalen zu identifizieren. Dies führt zu folgenden Forschungsfragen: (1) Wie genau schätzen frühpädagogische Fachkräfte in Kindertageseinrichtungen die mathematischen Fähigkeiten von Kindern im Bereich Mengen und Zahlen im Alter von 4 bis 6 Jahren ein? (2) Inwiefern klären die Merkmale Geschlecht, sozioökonomischer Status und tatsächliche mathematische Leistung der Kinder im Bereich Mengen und Zahlen sowie das mathematikdidaktische Wissen der frühpädagogischen Fachkräfte Varianz in der Einschätzungsgenauigkeit der Fachkräfte auf? Aus den bisherigen Forschungsergebnissen ergeben sich für einzelne Merkmale erste Vermutungen: Ein Zusammenhang der kindlichen mathematischen Leistung im Bereich Mengen und Zahlen mit der Einschätzungsgenauigkeit konnte für den amerikanischen frühkindlichen und den deutschen schulischen Bereich bereits gefunden werden, sodass auch für den frühkindlichen Bereich in Deutschland entsprechende Ergebnisse erwartet werden können. In Studien zum sozioökonomischen Status zeigte sich zwar ein Zusammenhang im schulischen Bereich mit den Einschätzungen von Lehrkräften, im frühkindlichen Bereich konnte dieser jedoch nicht bestätigt werden, sodass für diese Untersuchung ebenfalls kein Zusammenhang erwartet wird. Zum Zusammenhang des Geschlechts der Kinder bzw. des mathematikdidaktischen Wissens der Fachkräfte mit der Einschätzungsgenauigkeit zeigten sich keine eindeutigen Befunde. Hier können also keine begründeten Vermutungen angestellt werden.

5 Design der Studie

5.1 Stichprobe

Zur Beantwortung der Forschungsfragen wurden die mathematischen Fähigkeiten von 160 Kindern (96 m) im Alter von durchschnittlich 5,0 Jahren (min=3 Jahre 10 Monate, max=6 Jahre 9 Monate, $SD=8$ Monate) im Bereich Mengen und Zahlen getestet. Die Leistungen der Kinder wurden von 49 frühpädagogischen Fachkräften (47 w) aus 23 Kindertagesstätten in Niedersachsen eingeschätzt, wobei pro Einrichtung eine bis vier Fachkräfte an der Studie teilnahmen ($M=2,0$). Jedes der 160 Kinder wurde einmalig von einer Bezugsfachkraft mittels eines Fragebogens eingeschätzt. Eine Fachkraft schätzte dabei im Schnitt 3,12 Kinder bzgl. ihrer mathematischen Leistung im Bereich Mengen und Zahlen ein (min=1, max=6, $SD=1,13$, s. Abb. 1). Die Kinder waren den Fachkräften bekannt und besuchten seit durchschnittlich 2 Jahren und 9 Monaten (min=4 Monate, max=6 Jahre 2 Monate, $SD=1$ Jahr 2 Monate) die jeweilige Einrichtung. Der Altersdurchschnitt der frühpädagogischen Fachkräfte lag bei 41,6 Jahren (min=24, max=64, $SD=12,15$) und

Abb. 1 Exemplarische Veranschaulichung der Zusammensetzung der Stichprobe. (FK Fachkraft, K Kind)



die Fachkräfte wiesen durchschnittlich ca. 16 Jahre Berufserfahrung auf ($\min = 1$, $\max = 38$, $SD = 11,22$).

5.2 Instrumente und Durchführung

Zur Erfassung der *mathematischen Fähigkeiten der Kinder im Bereich Mengen und Zahlen* wurde der standardisierte Test MARKO-D (Ricken et al. 2013) eingesetzt. Der Einzeltest besteht aus 55 Items zu den Inhalten Zählzahlaspekt, kardinaler und ordinaler Aspekt, Zerlegbarkeit, Enthaltensein, Klasseninklusion und Relationalität (Testdauer ca. 20–30 min). Die Items sind fünf Niveaustufen zugeordnet. Der Test wurde von geschulten Testleiterinnen in den Räumlichkeiten der Kindertagesstätten durchgeführt und hatte eine Reliabilität von 0,92 (Cronbachs Alpha).

Zur Erhebung der *Einschätzungsgenauigkeit der frühpädagogischen Fachkräfte* wurde auf direkte Einschätzungen der kindlichen Leistungen zurückgegriffen. Mit Blick auf die Operationalisierung der Qualität von Einschätzungen nach Schrader und Praetorius (2018) sollten die Fachkräfte anhand eines standardisierten Fragebogens bei 17 ausgewählten Items des MARKO-D für das jeweilige Kind entscheiden, ob es das Item lösen kann oder nicht (zusätzliche Antwortmöglichkeit: „weiß ich nicht“). Mithilfe dieses Fragebogens wurden informelle Einschätzungen der frühpädagogischen Fachkräfte zu den teilnehmenden Kindern erhoben, da sich die Fachkräfte bei ihren Antworten nicht auf ein methodisch abgesichertes, formales Diagnoseverfahren, sondern allein auf ihre subjektiven Einschätzungen zu den kindlichen Fähigkeiten stützen mussten. Da die zeitlichen Ressourcen der Fachkräfte begrenzt waren, konnten nicht alle 55 Items des MARKO-D in den Fragebogen aufgenommen werden. Bei der Auswahl der Items wurde beachtet, alle fünf Niveaustufen des MARKO-D abzudecken, um Einschätzungen von Aufgaben unterschiedlichen Schwierigkeitsgrads zu erhalten. Insbesondere sollten die ausgewählten Aufgaben die bekannten zentralen Prädiktoren für die Rechenleistung bzw. für spätere Schwierigkeiten beim Rechnen abdecken (Jordan et al. 2010; Gersten et al. 2005; Geary et al. 2007; Dornheim 2013) bzw. in Modellen zur Entwicklung des Zahl- und Mengenverständnisses als relevante Kompetenzen verankert sein (Krajewski und Schneider 2006). Die Items beinhalten daher sowohl Zählfertigkeiten und ein erstes Mengenverständnis (bspw. *Item auf Niveaustufe 1: „Vor dem Kind liegen 10 farbige Chips. Jemand sagt: ‚Gib mir bitte 6 Chips.‘ Kann das Kind die richtige Anzahl Chips herausgeben?“*) als auch tiefergehende Fähigkeiten, wie Verständnis von Mengenrelationen, Teil-Ganzes oder Differenzen zwischen Anzahlen sowie erste Rechenfertigkeiten (bspw. *Item auf Niveaustufe 3: „Ich hatte gestern 5 Bonbons und habe 3 aufgegessen. Wie viele habe ich dann noch?‘ Kann das Kind die richtige Anzahl der Bonbons bestimmen?“* oder *Item auf Niveaustufe 5: „Kann das Kind die folgende Frage richtig beantworten? ‚Wie heißt die Zahl, die um 2 größer ist als die 4?‘“*). Da die Fachkräfte darüber informiert sind, welche konkrete Aufgabe die Kinder bewältigen sollen, handelt es sich um direkte Einschätzungen, welche in der Regel genauer sind als indirekte (Südkamp et al. 2012). Insgesamt ist den Pädagoginnen und Pädagogen diese konkrete Art der Einschätzung mithilfe eines Fragebogens und die Durchführung mathematischer Kompetenztests eher unbekannt (McElvany et al. 2009; Kammermeyer 1998). Da die Items des Fragebogens

Situationen mit dem Kind beschreiben und insofern ganz gezielte Einschätzungen fordern, ist jedoch davon auszugehen, dass die Art der Erhebung für die Fachkräfte dennoch transparent und durchführbar war. Die zusätzliche Antwortmöglichkeit „weiß nicht“ wurde angeboten, um die Fachkräfte nicht zu zufälligen Antworten zu zwingen, falls sie die Kinder nicht bewusst einschätzen konnten und um fehlende Daten zu vermeiden, falls die Fachkräfte keine Einschätzung geben wollten (Moosbrugger und Kelava 2008). Es wurde mit dem Antwortformat („kann es“, „kann es nicht“, „weiß ich nicht“) bewusst eine neutral gehaltene Einschätzung und keine Beurteilung gewählt, um die Fachkräfte nicht durch verschiedene Normsetzungen zu beeinflussen (bspw. *überdurchschnittlich – unterdurchschnittlich*). Da unterschiedlich schwierige Aufgaben von den Fachkräften eingeschätzt werden sollen, handelt es sich hier um einen Fragebogen mit heterogenen Items. In diesem Fall wurde zur Prüfung des Gütekriteriums der Messgenauigkeit auf die Split-Half-Methode mit parallelisierten Testhälften zurückgegriffen (Moosbrugger und Kelava 2008). Der eingesetzte Fragebogen wies eine Reliabilität von 0,622 (Spearman-Brown) auf.¹

Die Erfassung des *sozioökonomischen Status der Kinder* erfolgte mit einem Fragebogen an die familialen Bezugspersonen der Kinder. Dabei wurden Bildungsabschluss, Ausbildung und aktuelle Berufsstellung der Sorgeberechtigten erhoben. Zur Einstufung des Ansehens der elterlichen Abschlüsse und Berufe wurden diese mithilfe der International Standard Classification of Occupations (ISCO-08) der International Labour Office (ILO) (2012) codiert und der „International Socio-Economic Index of Occupational Status“ (ISEI) gebildet. Jedem Kind wurde, wie bei den bisherigen PISA-Untersuchungen (Mang et al. 2019), der jeweils höhere Wert der Sorgeberechtigten zugeordnet (HISEI).

Zur Erfassung des *mathematikdidaktischen Wissens der frühpädagogischen Fachkräfte* liegen bislang vergleichsweise wenige validierte Instrumente vor. Eingesetzt wurde der standardisierte situationsbezogene Paper-Pencil-Test von Gasteiger et al. (2020). Er umfasst 26 Items, in denen das situationspezifische Wissen der Fachkräfte zu frühen mathematischen Fähigkeiten von Kindern in vier typischen Alltagssituationen in Kindertagesstätten gemessen wird. Die Fachkräfte sollen anhand einer ausführlichen Situationsbeschreibung entscheiden, ob ein fiktives Kind eine bestimmte Fähigkeit besitzt oder nicht (bspw. „Max kann bis fünf zählen.“, „Max kann eine eindeutige Zuordnung vornehmen.“; *zusätzliche Antwortmöglichkeiten*: „in dieser Situation nicht beobachtbar“ und „weiß ich nicht“) (Gasteiger et al. 2020). Zwei der vier Situationen beziehen sich auf kindliche Kompetenzen im Bereich Mengen und Zahlen (*Würfelspiel, Teil-Ganzes*), während zwei weitere Situationen die Themen Formen und Muster (*ein Muster aus Kreisen, Dreiecken und Quadraten fortführen*) sowie Messen (*Vergleich von Hohlmaßen*) behandeln. Der Test misst das mathematikdidaktische Wissen zu verschiedenen Inhaltsbereichen als Gesamtkonstrukt und beinhaltet daher, anders als der MARKO-D, nicht nur Fähigkeiten im

¹ Da die Fachkräfte unterschiedlich viele Kinder einschätzten, gehen unterschiedlich viele ausgefüllte Fragebögen der 49 Fachkräfte in die Reliabilitätsanalyse ein. Daher wurden zusätzlich zur Berechnung der Reliabilität der Gesamtstichprobe zehn Zufallsziehungen durchgeführt, bei denen je Fachkraft nur ein Fragebogen gezogen wurde. Auf Grundlage dieser 49 Fragebögen pro Ziehung wurden jeweils erneut die Reliabilitäten berechnet. Diese bewegten sich zwischen 0,567 und 0,732, wobei die Reliabilitäten von acht dieser zehn Ziehungen über der Gesamtreliabilität von 0,622 lagen.

Bereich Mengen und Zahlen. Der Entscheidung, dieses Testinstrument für die Untersuchung eines potenziellen Zusammenhangs der fachdidaktischen Kompetenz mit der Einschätzungsgenauigkeit zu verwenden, liegt die Erkenntnis zugrunde, dass ein hohes Maß an globalem fachdidaktischen Wissen mit einem fachlichen Blick für die Wahrnehmung von Situationen mit mathematischen Potenzial einhergeht (Duneka-cke et al. 2016), welches notwendig ist, um kindliche mathematische Fähigkeiten im Alltag überhaupt erkennen zu können. Zudem sollte ein standardisiertes Instrument zur Messung des mathematikdidaktischen Wissens eingesetzt werden. Der Test hatte eine Skalen-Reliabilität von 0,59. In der Normstichprobe zeigte sich eine Reliabilität von 0,65 (Cronbachs Alpha)² (Gasteiger et al. 2020).

Alle Fragebögen wurden in freier Bearbeitung und ohne externe Kontrolle durchgeführt.

5.3 Auswertungsmethoden

Bzgl. der *Genauigkeit von Einschätzungen* von frühpädagogischen Fachkräften interessiert insbesondere die Auswertung der deskriptiven Häufigkeitsverteilungen. Einschätzungen wurden als richtig gewertet, wenn sie mit der tatsächlichen Leistung des Kindes übereinstimmten und als falsch, wenn die Einschätzung nicht mit der tatsächlichen Leistung des Kindes übereinstimmte. Wurde von der Fachkraft die Antwortmöglichkeit „weiß nicht“ gewählt, so wurde dies gesondert aufgeführt, sodass die Prozentsätze „Übereinstimmung“, „keine Übereinstimmung“ und „weiß nicht“ gebildet wurden. Die Einschätzungen, in denen keine Übereinstimmung vorlag, wurden zusätzlich in „Überschätzung“ (Kind kann das Item nicht lösen – Fachkraft sagt „kann es“) und „Unterschätzung“ (Kind kann das Item lösen – Fachkraft sagt „kann es nicht“) kategorisiert und deren Anteile an allen Einschätzungen bestimmt. Fehlte die tatsächliche Leistung des Kindes oder die Einschätzung einer Fachkraft in einem Item, so wurde die entsprechende Aufgabe als Missing gewertet, wobei höchstens ein Missing pro Kind akzeptiert wurde. Andernfalls wurde das Kind aus der Auswertung ausgeschlossen. Aufgrund dieser Einschränkung in der Anzahl der Missings pro Kind reduzierte sich die Stichprobe auf 151 Kinder (91 m, 60 w; $n(119)=0$ Missings, $n(32)=1$ Missing). Insgesamt wurden 2535 Einschätzungen analysiert.

Zunächst wurden die Einschätzungen der frühpädagogischen Fachkräfte nach Einschätzungsfällen (d. h. nach Kindern) betrachtet. Für jedes Kind wurde also itemweise die Einschätzung der Fachkraft mit der tatsächlichen Leistung des Kindes verglichen und so ein prozentualer Übereinstimmungswert zwischen Einschätzung und tatsächlicher Leistung gebildet (im Folgenden: Übereinstimmung je Kind). Um in einem zweiten Schritt Aussagen zur Einschätzungsgenauigkeit der frühpädagogischen Fachkräfte in dieser Stichprobe treffen zu können, wurden alle Einschätzungen einer Fachkraft betrachtet. Dazu wurde für jede Fachkraft ein Mittelwert über die Übereinstimmungswerte der von dieser Fachkraft eingeschätzten Kinder gebildet. So

² Für Tests mit komplexen Operationalisierungen sind auch niedrige Reliabilitäten ($\alpha < 0,7$) akzeptabel (Schmitt 1996).

wurde ein prozentualer Übereinstimmungswert je Fachkraft erzeugt (im Folgenden: Übereinstimmung je Fachkraft).

Der *Einfluss verschiedener Merkmale auf die Einschätzungsgenauigkeit* frühpädagogischer Fachkräfte wurde mithilfe von Generalized Linear Mixed Models (GLMM) (Jiang und Nguyen 2021) analysiert. Die Modelle wurden in R mithilfe des lme4-Package (Bates et al. 2015) berechnet. GLMMs können als Verallgemeinerung von linearen Regressionen verstanden werden, die Abhängigkeiten zwischen verschiedenen Beobachtungen (Clusterungen) kontrollieren. Dazu wurde ein Zufallseffekt aufgenommen, durch den die Abhängigkeit der Beobachtungen bei Kindern einer Fachkraft berücksichtigt werden kann. Mithilfe des intraclass correlation coefficient (ICC) wurde dabei die Stärke der Datenclustering beurteilt. Der Koeffizient beschreibt, wie stark sich die Fälle innerhalb eines Clusters ähneln (0 = gar nicht, 1 = identisch), wobei man bereits ab sehr kleinen Werten (wie 0,01) mit Abhängigkeiten innerhalb der Cluster rechnen muss (Musca et al. 2011). Zur Beantwortung der Fragestellung wurden zwei Modelle gebildet. Im ersten Modell wurden die beschriebenen Merkmale Leistung des Kindes im MARKO-D (Rohwert), Geschlecht des Kindes und mathematikdidaktisches Wissen (MPCK) der Fachkräfte als mögliche Faktoren zur Varianzaufklärung berücksichtigt. Aufgrund des bekannten Zusammenhangs des sozioökonomischen Status des Kindes mit dessen kognitiver Kompetenz (Bradley et al. 2001; Baumert et al. 2001; Tucker-Drob et al. 2011) wurde der sozioökonomische Status (HISEI) erst in einem zweiten Modell als Merkmal hinzugenommen, um zu prüfen, ob darüber zusätzliche Varianz aufgeklärt werden kann.

6 Ergebnisse

Insgesamt wurden zur Beantwortung der *Fragestellung (1) zur Einschätzungsgenauigkeit* die mathematischen Fähigkeiten von 151 Kindern im Bereich Mengen und Zahlen und die zugehörigen Einschätzungen durch 49 frühpädagogische Fachkräfte erhoben. Im Hinblick auf *Fragestellung (2) zum Einfluss verschiedener Merkmale auf die Einschätzungsgenauigkeit* verringerte sich die Stichprobe aufgrund fehlender Angaben der Eltern zum sozioökonomischen Status der Kinder und aufgrund nicht auswertbarer Fragebögen zum mathematikdidaktischen Wissen der Fachkräfte auf 140 Kinder und 46 Fachkräfte.

6.1 Einschätzungsgenauigkeit frühpädagogischer Fachkräfte

Tab. 1 zeigt die gemittelten Anteile von Übereinstimmung, keine Übereinstimmung und „weiß nicht“ je Kind an allen Einschätzungen. Die prozentualen Werte geben also – gemittelt über alle 151 Kinder – den Anteil an, zu dem die Leistung der Kinder im Schnitt korrekt bzw. falsch eingeschätzt wurde. Die Kinder wurden von der jeweiligen Bezugsfachkraft im Schnitt zu ca. zwei Drittel korrekt eingeschätzt, in ca. 12% der Einschätzungen pro Kind nennt die Fachkraft „weiß nicht“ und bei knapp einem Viertel stimmt die Einschätzung nicht mit der tatsächlichen Leistung überein. Ein Wert von 64,6% entspricht – annäherungsweise nach Landis und

Tab. 1 Deskriptive Statistik zur Einschätzung der kindlichen Leistung

	<i>M</i>	<i>SD</i>	Min	Max
Übereinstimmung	64,61	14,51	5,88	100,00
Keine Übereinstimmung	23,62	11,74	0,00	52,94
„Weiß nicht“	11,77	14,97	0,00	88,24

Anteile in % an allen Einschätzungen je Kind, $n = 151$ Kinder

Tab. 2 Deskriptive Statistik zu Über- und Unterschätzungen der kindlichen Leistung

	<i>M</i>	<i>SD</i>
Überschätzung	18,25	11,79
Unterschätzung	5,37	7,54

Anteile in % an allen Einschätzungen je Kind, $n = 151$ Kinder

Koch (1977) geschätzt – einer mäßigen Übereinstimmung (Cohens $\kappa = 0,29$)³. Die Spannweiten in der prozentualen Übereinstimmung zeigen, dass die Kinder sehr unterschiedlich genau in ihrer mathematischen Kompetenz im Bereich Mengen und Zahlen eingeschätzt wurden.

Die Kategorie „Keine Übereinstimmung“ lässt sich aufteilen in die Unterkategorien „Überschätzung“ (*Kind kann es nicht – Fachkraft sagt „kann es“*) und „Unterschätzung“ (*Kind kann es – Fachkraft sagt „kann es nicht“*). Die prozentualen Angaben in Tab. 2 geben an, wie hoch die Anteile von Über- und Unterschätzung je Kind sind – gemittelt über alle Kinder.

Von den 23,6 % „Keine Übereinstimmung“ je Kind waren durchschnittlich 18,3 % Überschätzungen und 5,4 % Unterschätzungen. Es zeigt sich also (bei fehlender Übereinstimmung) eine eindeutige Tendenz zur Überschätzung der mathematischen Leistung der Kinder im Bereich Mengen und Zahlen, da bei falscher Einschätzung in durchschnittlich ca. 3 von 4 Fällen überschätzt wurde. Zusätzlich ist interessant, dass ein Teil der Kinder ausnahmslos bei allen Aufgaben, die falsch eingeschätzt wurden, immer über- (72 Kinder) bzw. immer unterschätzt (10 Kinder) wurden.

Darüber hinaus wurde der Wert der Einschätzungsgenauigkeit einer Fachkraft analysiert. Dazu wurden die Übereinstimmungswerte aller Kinder *einer* Fachkraft gemittelt über die Anzahl der Kinder, die die Fachkraft eingeschätzt hat. Abb. 2 zeigt den prozentualen Anteil an Übereinstimmung zwischen allen Einschätzungen der jeweiligen Fachkraft und den Leistungen der jeweils eingeschätzten Kinder, also die Übereinstimmung je Fachkraft. Die schwarzen Punkte geben dabei den niedrigsten Übereinstimmungswert je Kind an, den die Fachkraft erzielt hat, die weißen

³ Für eine annäherungsweise Einordnung der Einschätzungsgenauigkeit wurde Cohens Kappa verwendet. Dies gilt eigentlich als statistisches Maß für die Beurteilung von Übereinstimmungen zwischen mehreren Beobachterinnen und Beobachtern, bei der in der Regel ein nahezu perfekter Übereinstimmungswert erreicht werden soll. Da die Einschätzungen der Leistung und die Erhebung der tatsächlichen Leistung auf unterschiedlichen Situationskontexten (Test vs. Alltagsbeobachtung) beruhen, wurden in dieser Untersuchung keine perfekten Werte erwartet. Es kann daher davon ausgegangen werden, dass der Wert mit der Einordnung von Landis und Koch (1977) eher unterschätzt wird. Cohens Kappa wurde nach Landis und Koch (1977) wie folgt berechnet: $\kappa = (\text{Anteil tatsächlicher Übereinstimmung} - \text{Anteil zufälliger Übereinstimmung}) / (1 - \text{Anteil zufälliger Übereinstimmung})$.

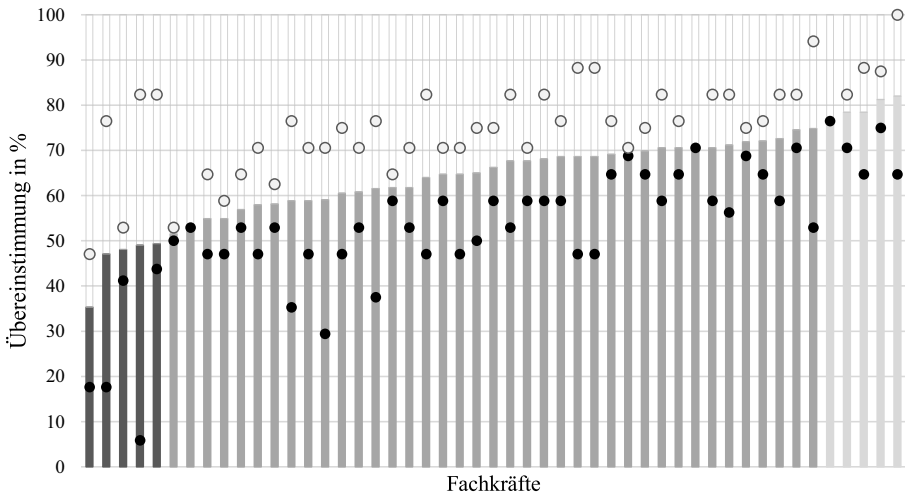


Abb. 2 Übereinstimmung je Fachkraft, aufsteigend sortiert, $N=49$, $M=64,44$, $SD=9,71$, Werte unter 50%: dunkelgrau, Werte von 50–75%: grau, Werte über 75%: hellgrau

Punkte geben den höchsten von der Fachkraft erreichten Übereinstimmungswert pro Kind an. Es besteht kein statistischer Zusammenhang zwischen der prozentualen Übereinstimmung je Fachkraft und der Anzahl eingeschätzter Kinder ($r=-0,073$, $p=0,619$).

Auch der durchschnittliche Übereinstimmungswert je Fachkraft von 64,4% ist nach Landis und Koch (1977) eine mäßige Übereinstimmung (Cohens $\kappa=0,29$). Es zeigen sich individuelle Unterschiede zwischen den Fachkräften: Den geringsten Übereinstimmungswert zeigte eine Fachkraft mit ca. 35,3% für ihre jeweilige Stichprobe, den höchsten Wert eine Fachkraft mit ca. 82,0%. Es zeigt sich jedoch auch, dass die Übereinstimmung der meisten Fachkräfte für ihre jeweiligen Stichproben zwischen 50 und 75% liegt und es nur wenige gibt, die einen höheren (s. Abb. 2: hellgrau) oder niedrigeren (s. Abb. 2: dunkelgrau) Wert erreichen.

6.2 Einfluss verschiedener Merkmale auf die Einschätzungsgenauigkeit

Zur Untersuchung der Frage, welche Merkmale Varianz an der Einschätzungsgenauigkeit aufklären, wurden zwei Generalized Linear Mixed Models mit dem Anteil der Übereinstimmung je Kind als abhängige Variable gerechnet. In das erste Modell wurden das Geschlecht des Kindes, der vom Kind erreichte Rohwert im MARKO-D und das mathematikdidaktische Wissen der Fachkraft als unabhängige Variablen aufgenommen, in das zweite Modell zusätzlich der sozioökonomische Status des Kindes. Um die Clusterung der Daten zu berücksichtigen (mehrere Kinder pro Fachkraft), wurde in beiden Modellen zusätzlich ein Zufallseffekt aufgenommen.

Für die GLMMs mussten zwei Ausreißer ausgeschlossen werden (standardisiertes Residuum $>|3|$; Baltes-Götz 2019), sodass sich die Fallzahl um zwei von 140 auf 138 verringert und der Wert der Übereinstimmung in % leicht von den berichteten Ergebnissen zu Forschungsfrage (1) abweicht (das Geschlecht der Kinder teilt sich

Tab. 3 Mittelwerte und Standardabweichungen zur abhängigen und zu den unabhängigen Variablen

	<i>N</i>	<i>M</i>	<i>SD</i>
Übereinstimmung je Kind ^a	138	65,42	12,78
Rohwert MARKO-D ^b	138	25,37	9,13
HISEI ^c	138	62,55	19,68
MPCK ^d	46	69,04	12,36

^aIn % an allen Einschätzungen

^bTatsächliche mathematische Leistung, mögl. Min/Max: 0/55

^cSozioökonomischer Status, mögl. Min/Max: 16/90

^dMathematikdidaktisches Wissen der Fachkraft, korrekte Antworten in %

Tab. 4 Korrelationen nach Pearson der in die Analysen einbezogenen Variablen

	Geschlecht ^b	Üb. je Kind ^a	Rohwert	HISEI
Übereinstimmung je Kind ^a	0,090	–	–	–
Rohwert MARKO-D	-0,030	0,254**	–	–
HISEI	0,077	0,003	0,308**	–
MPCK	-0,086	0,049	0,088	-0,006

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

^aAbhängige Variable

^b0 = m, 1 = w

nach Ausschluss der Ausreißer wie folgt auf: 86 m, 52 w). Um einen ersten Überblick über die erhobenen Merkmale für die Einschätzungsgenauigkeit zu bekommen, werden in Tab. 3 zunächst deren Ausprägungen dargestellt.

Tab. 4 stellt die paarweisen Korrelationen der abhängigen und der unabhängigen Variablen nach Pearson dar. Die mathematischen Fähigkeiten im Bereich Mengen und Zahlen (Rohwert) korrelierten sowohl mit dem sozioökonomischen Status der Kinder (HISEI) als auch mit der prozentualen Übereinstimmung je Kind (abhängige Variable).⁴

Die Ergebnisse der GLMMs sind in Tab. 5 dargestellt. Es wurden zwei Modelle gerechnet: Modell 1 zeigt die Abhängigkeit der Einschätzungsgenauigkeit von den Variablen Geschlecht, erreichter Rohwert im MARKO-D und dem mathematikdidaktischen Wissen der Fachkraft. Modell 2 berücksichtigt zusätzlich den sozioökonomischen Status des Kindes, der mit dem Rohwert korreliert, ggf. aber zusätzliche Varianz an der Einschätzungsgenauigkeit aufklären kann.

Der hohe ICC (0,2) beider Modelle weist darauf hin, dass Abhängigkeiten zwischen den Einschätzungen einer Fachkraft bestehen, die einer erwartbaren Clustering in einer Schulklasse entsprechen (Musca et al. 2011). Die Wahl eines GLMMs zur Berücksichtigung der Datenclustering erwies sich daher als sinnvoll. Der Chi-Quadrat-Test zeigt, dass sich die beiden Modelle nicht signifikant unterscheiden

⁴ Um sicherzustellen, dass die 17 MARKO-D-Items, die dem Einschätzungsfragebogen zugrunde lagen, die Leistung der Kinder in ähnlicher Breite widerspiegeln wie der Gesamttest MARKO-D, und um Decken- und Bodeneffekte auszuschließen, wurden zusätzlich Lage- und Streumaße des Rohwerts in den entsprechenden 17 MARKO-D-Items berechnet. Diese sind vergleichbar mit denen des gesamten MARKO-D-Test ($M = 8,38$, $SD = 2,92$, $\min = 1$, $\max = 15$). Der von den Kindern in den 17 Items erreichte Rohwert korrelierte zudem stark mit dem erreichten Rohwert im Gesamttest ($r = 0,878$, $p = 0,000$).

Tab. 5 Einflussmerkmale auf die Einschätzungsgenauigkeit frühpädagogischer Fachkräfte (GLMMs)

	Modell 1				Modell 2			
	<i>B</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>B</i>	<i>SE</i>	<i>df</i>	<i>t</i>
<i>Intercept</i>	51,67	7,78	47,53	6,64***	54,08	8,19	54,54	6,61***
<i>Geschlecht</i>	2,90	2,10	123,32	1,38	3,05	2,10	121,80	1,45
<i>Rohwert</i>	0,42	0,11	131,04	3,65***	0,456	0,12	130,47	3,76***
<i>MARKO-D</i>								
<i>MPCK</i>	3,03	10,51	38,39	0,29	3,13	10,51	37,72	0,30
<i>HISEI</i>	–	–	–	–	–0,06	0,06	132,68	–0,95
<i>Pseudo-R²</i>								
Marginal	0,096				0,101			
Conditional	0,279				0,283			
<i>AIC</i>	1081,32				1086,28			
<i>BIC</i>	1098,88				1106,77			
<i>ICC</i>	0,20				0,20			

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

($\chi^2(1) = 0,9371$, $p = 0,333$), der sozioökonomische Status der Kinder klärt also keine weitere Varianz an der Einschätzungsgenauigkeit der frühpädagogischen Fachkräfte auf. Daher wird im Folgenden das Modell 1 analysiert (Wolf und Best 2010).

Bei Betrachtung der Variablen Geschlecht, Rohwert und MPCK klärt das Modell eine Varianz von 9,6% auf (marginale R^2), mit Einbezug der durch die Clustering bedingten Varianz wird eine Varianz von 27,9% aufgeklärt (konditionales R^2). Das Geschlecht ($p = 0,169$) sowie das mathematikdidaktische Wissen der Fachkräfte ($p = 0,774$) klären keine Varianz bezüglich der Einschätzungsgenauigkeit auf. Lediglich die tatsächliche mathematische Leistung der Kinder im Bereich Mengen und Zahlen zeigt ein signifikantes Gewicht zur prozentualen Übereinstimmung je Kind ($p = 0,000$). Um die Stärke des Effekts grob einschätzen zu können, lässt sich ein standardisierter Koeffizient über die Standardabweichungen der abhängigen bzw. unabhängigen Variable von $\beta = 0,30$ für den Rohwert berechnen (Wolf und Best 2010)⁵, der vorsichtig wie bei einer Regressionsanalyse zu deuten ist.

7 Diskussion

Ziel der vorliegenden Studie war es, die Einschätzungsgenauigkeit mathematischer Fähigkeiten von Kindern im Bereich Mengen und Zahlen durch frühpädagogische Fachkräfte zu untersuchen und mögliche Einflussmerkmale auf Seiten der Kinder oder der Pädagoginnen und Pädagogen aufzudecken.

⁵ Diese Berechnung eines standardisierten Koeffizienten β berücksichtigt die Clusterung in den Daten nicht. Andere Effektstärkemaße sind in der statistischen Forschung zu linearen Mischmodellen derzeit jedoch noch nicht abschließend diskutiert.

7.1 Genauigkeit der Einschätzung frühpädagogischer Fachkräfte

Die Einschätzungen frühpädagogischer Fachkräfte bei den hier geprüften Items zu Mengen und Zahlen wiesen eine durchschnittliche Übereinstimmung von ca. 64,6% auf Ebene jedes Kindes bzw. von ca. 64,4% auf Ebene der Fachkräfte auf. Hoge und Coladarci (1989) interpretieren in ihrer Metaanalyse mehrerer Studien zur Einschätzungsgenauigkeit einen Wert von 66% als moderat, Südkamp et al. (2012) einen Wert von 63% sogar als „ziemlich hoch“. Mithilfe von Cohens κ lassen sich die Werte dieser Untersuchung nach Landis und Koch (1977) annäherungsweise als mäßig bezeichnen ($\kappa=0,29$). Eine inhaltliche Betrachtung lässt jedoch annehmen, dass diese Einordnung die Übereinstimmung eher unterschätzt: Eine perfekte Übereinstimmung zwischen Einschätzung und tatsächlicher Leistung war, anders als für die Übereinstimmung zwischen mehreren Beobachterinnen und Beobachtern und bei der Interpretation nach Landis und Koch (1977) in der Regel erwünscht, nicht zu erwarten (s. Abschn. 6.1). Die Untersuchung mit dem MARKO-D spiegelt lediglich die situationsabhängige Performanz der Kinder im Test wider. Diese kann bspw. durch Tagesverfassung, Bearbeitungszeit, Flüchtigkeitsfehler, Leistungsangst, mangelnde Anstrengung oder Aufregung leistungsmindernd beeinflusst werden (Anders et al. 2010; Karing 2009; Hosenfeld et al. 2002; Kowalski et al. 2018). Im Gegensatz dazu beruhen die Einschätzungen der Fachkräfte vermutlich eher auf Beobachtungen über einen längeren Zeitraum in verschiedenen Kontexten und Situationen, in denen die Kinder ihr volles Potenzial zeigen können (Kowalski et al. 2018). Es ist zudem davon auszugehen, dass dem frühen mathematischen Lernen im Gegensatz zu sprachlichen Fähigkeiten im Alltag der Kindertagesstätte eher eine untergeordnete Rolle zugeschrieben wird (Kuratli Geeler 2019; Blevins-Knabe et al. 2000), sodass Beobachtungen in diesem Bereich ggf. in den Hintergrund rücken und eine Einschätzung der kindlichen Leistung für die Fachkräfte deshalb besonders anspruchsvoll ist.

Anders als in den zitierten Studien zeigen die Mittelwerte der Einschätzungsgenauigkeit je Fachkraft für diese Stichprobe eine eher geringe Streuung (Dollinger 2013; Kilday et al. 2012; Kowalski et al. 2018; McKeveit und Kiss 2019). Es zeigen sich zum Teil jedoch große Unterschiede, wie genau verschiedene Kinder von einer einzelnen frühpädagogischen Fachkraft eingeschätzt werden (s. Abb. 2). Besonders die Fachkräfte mit einem niedrigen durchschnittlichen Übereinstimmungswert je Fachkraft zeigen deutliche Unterschiede in den Einschätzungen ihrer Kinder (bspw. eine einzelne Fachkraft mit einer Übereinstimmung von 5,88% für ein Kind mit einem Rohwert von 15 Punkten und einer Übereinstimmung von 82,35% für ein anderes Kind mit einem Rohwert von 13 Punkten).

Die frühpädagogischen Fachkräfte neigten, wie bei Kowalski et al. (2018), *tendenziell* eher zur Überschätzung der Kinder in ihren mathematischen Fähigkeiten im Bereich Mengen und Zahlen. Auch hier könnten als mögliche Begründung leistungsmindernde Einflüsse in der Testsituation angeführt werden. Die Fragestellung „Kann das Kind ...“ könnte die Fachkräfte dazu tendieren lassen, das prinzipiell mögliche Leistungspotential des Kindes anzugeben, welches jedoch nicht von jedem Kind in einer ungewohnten Testsituation gezeigt werden kann. Darüber hinaus herrscht in der Kindertagesstätte oftmals eine sehr wertschätzende Einstellung mit

Blick auf das, was Kinder *schon* können (Gasteiger 2010; Selter 2008). Interessant ist diese Überlegung auch im Hinblick auf bekannte *Unterschätzungstendenzen* von Grundschullehrkräften im Sinne einer mathematischen „Stunde Null“ (Selter 2008). Eine tendenzielle Überschätzung der Leistungen könnte für die Kinder auch einen positiven Effekt auf deren Lernentwicklung haben, da Kinder in der Regel von hohen Erwartungen und demzufolge stärkeren Anregungen und Herausforderungen profitieren können (McElvany et al. 2009; Muntoni et al. 2019).

7.2 Einfluss verschiedener Merkmale auf die Einschätzungsgenauigkeit

Die *tatsächliche mathematische Leistung der Kinder im Bereich Mengen und Zahlen* scheint, wie vermutet, einen Einfluss auf die Einschätzungsgenauigkeit mathematischer Fähigkeiten im Bereich Mengen und Zahlen durch frühpädagogische Fachkräfte dahingehend zu haben, dass Fachkräfte leistungsstarke Kinder genauer einschätzen als leistungsschwache Kinder. Diese Resultate decken sich mit der Untersuchung von McKeveit und Kiss (2019) und den Ergebnissen von PISA 2000 (Baumert et al. 2001). Der Einfluss könnte damit zusammenhängen, dass leistungsstarke Kinder ihre möglichen Fähigkeiten in alltäglichen Situationen der Kindertagesstätte (bspw. beim Zählen im Morgenkreis, in Würfelspielen mit der Fachkraft oder beim Decken des Tisches) eher zeigen als leistungsschwächere, die in solchen Situationen vielleicht weniger in Erscheinung treten. Dieses Ergebnis deutet jedoch vorsichtig auf eine mögliche Gefahr hin: Werden leistungsschwächere Kinder nicht genau eingeschätzt, so kann dies bedeuten, dass mögliche Entwicklungsrückstände von frühpädagogischen Fachkräften nicht erkannt werden und die Kinder auch keine entsprechende Förderung erhalten. Eine weniger genaue Einschätzung leistungsschwächerer Kinder könnte somit dazu beitragen, dass sich längerfristig die Unterschiede zwischen leistungsstarken und -schwachen Kindern verstärken (Muntoni et al. 2019).

Das *Geschlecht der Kinder* zeigte, wie bei Furnari et al. (2017) und Kowalski et al. (2018), keinen Einfluss auf die Einschätzungsgenauigkeit der frühpädagogischen Fachkräfte. Auch der *sozioökonomische Status der Kinder* klärte, analog zu Furnari et al. (2017), keine zusätzliche Varianz an der Einschätzungsgenauigkeit der frühpädagogischen Fachkräfte auf. Da Lorenz (2011) in seiner Studie zur Einschätzungsgenauigkeit von *Lehrkräften* hingegen von einem zum Teil deutlichen Einfluss des sozioökonomischen Status der Kinder auf die Einschätzungsgenauigkeit berichtet, stellt sich die Frage, ob sich frühpädagogische Fachkräfte im Gegensatz zu Lehrkräften in der Schule weniger stark von Herkunftsmerkmalen der Kinder beeinflussen lassen.

Dass die tatsächliche Leistung des Kindes im MARKO-D jedoch nicht allein die Einschätzungsgenauigkeit der frühpädagogischen Fachkräfte beeinflusst, lässt ein detaillierter Blick in die Daten vermuten: Die Fachkraft mit der größten Range in den Einschätzungen ihrer Kinder schätzte ein Kind mit einem Rohwert von 15 Punkten nur zu 5,88 % richtig ein, während sie ein anderes Kind mit einem noch niedrigeren Rohwert von 13 Punkten zu 82,35 % korrekt einschätzte. Hier könnten (qualitative) Interviewstudien mit frühpädagogischen Fachkräften Einblicke in die Einschätzungen von Kindern und in mögliche Gründe dafür gewähren.

Insgesamt veranschaulicht der ICC eine nicht zu vernachlässigende Clusterung der Daten (eine Fachkraft schätzte ein bis sechs Kinder ein), was nahelegt, dass durch die Fachkraft ein beträchtlicher Anteil an Varianz der Einschätzungsgenauigkeit aufgeklärt werden könnte. Das *mathematikdidaktische Wissen der frühpädagogischen Fachkräfte* scheint jedoch – so die Ergebnisse der hier vorliegenden Studie – keinen Einfluss auf die Einschätzungsgenauigkeit zu haben. Vor dem Hintergrund, dass z. B. Anders et al. (2010) explizit nahelegten, das allgemeine und fachbezogene didaktische Wissen als möglichen Erklärungsansatz bei der Erforschung diagnostischer Prozesse zu berücksichtigen, muss dieses Ergebnis kritisch diskutiert werden. Die Situationswahrnehmung frühpädagogischer Fachkräfte wird häufig als zentraler Mittler zwischen Wissen und kompetentem Handeln bzw. für individuumsbezogene Diagnose und Förderung im Elementarbereich genannt (Fröhlich-Gildhoff et al. 2014; Gasteiger und Benz 2016). Es könnte also sein, dass frühpädagogische Fachkräfte, die über gute Beobachtungs- und Wahrnehmungsfähigkeiten verfügen, durch die konkreten Formulierungen der Items im Einschätzungsfragebogen zumindest zum Teil genaue Einschätzungen geben können, ohne umfassendes fachdidaktisches Wissen zu benötigen (z. B. bei den Einschätzungen zu den Zählfähigkeiten). Interessant wäre daher eine Untersuchung der Beobachtungs- und Wahrnehmungsfähigkeiten der Fachkräfte dahingehend, welche Situationen frühpädagogische Fachkräfte als Situationen mit mathematischem Potenzial und mit Beobachtungsmöglichkeiten zu mathematischen Fähigkeiten der Kinder im Bereich Mengen und Zahlen wahrnehmen und ob die Wahrnehmung mathematischen Potentials im Zusammenhang mit der Einschätzungsgenauigkeit steht.

Da sich die Datenclusterung als offensichtlich bedeutsam gezeigt hat, bleibt die Frage offen, ob es weitere Merkmale auf Ebene der Fachkraft gibt, die die Einschätzungsgenauigkeit beeinflussen. Vorstellbar wäre evtl. eine Betrachtung von Einstellungen der Fachkräfte zum Fach Mathematik. Erfahrungsgemäß zeigt sich hier eine sehr hohe Heterogenität. Durch eine fehlende Ausbildung zur frühen mathematischen Bildung in Verbindung mit ggf. eigenen negativen Erfahrungen zur Mathematik aus der Schulzeit kann bei manchen Fachkräften von einer Art Mathematikängstlichkeit gesprochen werden (Gresham 2007; Gasteiger 2010; Benz et al. 2017; Jenßen et al. 2020), welche ggf. Auswirkungen auf die Wahrnehmung mathematischer Lerngelegenheiten durch frühpädagogische Fachkräfte haben könnte. Jenßen et al. (2020) verweisen in ihren Untersuchungen zur Mathematikangst von angehenden frühpädagogischen Fachkräften auf einen Zusammenhang der Mathematikangst sowohl mit mathematischem und mathematikdidaktischem Wissen als auch mit mathematikbezogener Situationswahrnehmung, welche wiederum von Gasteiger und Benz (2016) als eine Voraussetzung für kompetentes Handeln (und somit auch für die Diagnose) im frühpädagogischen Kontext herausgestellt wird.

Unabhängig von Merkmalen auf Seiten der Fachkräfte oder der Kinder wurde in dieser Studie nicht im Detail analysiert, ob es Inhalte gibt, die von den Fachkräften genauer eingeschätzt werden als andere oder ob es sogar Typen von Fachkräften gibt, die verschiedene Items unterschiedlich genau einschätzen. Ein Blick in die Daten offenbart erste Anzeichen, dass Aufgaben zum Zählen genauer eingeschätzt wurden als komplexe Aufgaben zu weiterführenden Kompetenzen. Inwiefern Itemmerkmale

jedoch wirklich weitere Varianz in der Einschätzungsgenauigkeit frühpädagogischer Fachkräfte aufklären, muss noch genauer untersucht werden.

7.3 Limitationen

Bevor ein Fazit gezogen werden kann, werden einige Limitationen der Studie aufgeführt. Diese Studie kann keine Aussagen über die allgemeine Einschätzungskompetenz der Fachkräfte dieser Stichprobe angeben. Es wurde ein Einschätzungswert je Fachkraft (gemittelt über alle von ihr eingeschätzten Kinder) gebildet. Dadurch, dass die Einschätzung eines Kindes mit niedriger mathematischer Leistung im MARKO-D scheinbar schwieriger zu sein scheint als die eines Kindes mit guter Leistung im Test, können die Einschätzungsleistungen der Fachkräfte untereinander nicht verglichen werden. Um die vergleichbare Einschätzungskompetenz einer Fachkraft zu ermitteln, müssten alle Fachkräfte dieselben Kinder einschätzen. Da der methodische Ansatz dieser Studie jedoch eine hohe Praxisnähe aufweist (jede Fachkraft schätzt die Kinder ein, die sie täglich betreut), wurden lediglich Aussagen zur Einschätzungsgenauigkeit der Fachkräfte bzgl. dieser Stichprobe getätigt und auf vergleichende Aussagen zur Einschätzungskompetenz einzelner frühpädagogischer Fachkräfte wurde verzichtet. Ein detaillierterer Blick in die Daten zeigt jedoch, dass das Treffen genauer Einschätzungen unabhängig von der tatsächlichen Leistung der Kinder durchaus eine Kompetenz der Fachkraft sein kann: Das am besten eingeschätzte Kind (100 % Übereinstimmung) hatte einen Rohwert von lediglich 16 Punkten. Bei den Einschätzungen handelt es sich nicht ausschließlich um „kann es nicht“-Angaben (11 von 16 „kann es nicht“, fünf „kann es“, ein Missing durch das Kind).

Eine weitere Limitation ergibt sich dadurch, dass die aus dem MARKO-D ausgewählten Items des Fragebogens zur Einschätzungsgenauigkeit nicht zwangsläufig den im Alltag der Kindertagesstätte relevanten Situationen entsprechen, da bei der Auswahl der Items insbesondere beachtet wurde, die bekannten Prädiktoren abzubilden. Vor allem bei den Aufgaben auf Niveaustufe 1 handelt es sich um Basisfähigkeiten, die sich oft in typischen Alltagssituationen der Kindertagesstätte zeigen (bspw. Zahlwortreihe aufsagen – Kinder im Morgenkreis zählen, Eins-zu-Eins-Zuordnung – Frühstückstisch decken, kleinere Mengen abzählen – Würfelspiele spielen). Bei den anderen Niveaustufen handelt es sich hauptsächlich um Aufgaben zu weitergehenden mathematischen Kompetenzen und damit um Fähigkeiten der höheren Ebenen nach Krajewski und Schneider (2006) (Ordinalität, Additionsaufgaben im Zahlenraum bis 10, sicherer Umgang mit Mengen, Umgang mit Teilmengen, Differenzen bestimmen). Je nach Gestaltung des Gruppenalltags könnte es also sein, dass sich für frühpädagogische Fachkräfte keine Möglichkeit bietet, Kinder in ihren über die Basisfähigkeiten hinausgehenden frühen mathematischen Kompetenzen zu beobachten – es sei denn, Kinder fallen durch ihre hohen Kompetenzen, die sie explizit in Gesprächen oder Spielsituationen von sich aus zeigen, auf. Werden über die Basisfähigkeiten hinausgehende Anforderungen nicht von allen Kindern gefordert, fällt es ggf. besonders schwer, (noch) fehlende Fähigkeiten zu identifizieren. Die vorliegende Tendenz zur Überschätzung bekräftigt diese Vermutung. Interessant wäre daher eine umfassende Beobachtung und Analyse des Gruppenalltags dahingehend, ob

dieser auch weitergehende Fähigkeiten und damit bekannte Prädiktoren für späteres Mathematiklernen aufgreift.

Einschätzungen sind von komplexer Natur und die untersuchten Einflussmerkmale auf die Genauigkeit der Einschätzung konnten in dieser Studie nur zu einem eher geringen Anteil Varianz erklären. Es bleibt daher nach wie vor ungeklärt, womit die Einschätzungsgenauigkeit – außer mit den mathematischen Fähigkeiten der Kinder im Bereich Mengen und Zahlen – noch zusammenhängt. Diesbezüglich können weiterhin nur Vermutungen geäußert werden. Hilfreich wäre hier ggf. ein tieferer Einblick mittels qualitativer Beobachtungen von Interaktionen im Alltag der Kindertagesstätte und Interviews mit den frühpädagogischen Fachkräften, wodurch der Entstehungsprozess von Einschätzungen ggf. zunächst noch genauer charakterisiert werden könnte. Zudem könnte auch die Erhebung von Mathematikangst oder Einstellungen zum Mathematiklernen zu detaillierteren Ergebnissen führen.

Weiterhin ist limitierend zu beachten, dass alle Fragebögen selbstadministrativ ausgefüllt wurden und dadurch keine Kontrolle über die Bearbeitung gewährleistet werden konnte. Da der Paper-Pencil-Test zum mathematikdidaktischen Wissen der Fachkräfte jedoch situationsbasiertes implizites Wissen prüft, erscheint es kaum möglich, zu den richtigen Antworten durch die Verwendung von Hilfsmitteln zu gelangen. Dass der Test in dieser Stichprobe keine ungewöhnlich hohe Lösungsquote ($M=69\%$) aufwies und es auch nicht zu Deckeneffekten kam, unterstützt diese Vermutung.

Zusätzlich scheinen einige Items des MARKO-D sprachlich anspruchsvoll zu sein (z. B. Niveaustufe 5: „Wie heißt die Zahl, die um 2 größer ist als die 4?“). Es kann nicht sichergestellt werden, dass die frühpädagogischen Fachkräfte in ihren Einschätzungen hier nicht die sprachlichen Fähigkeiten der Kinder im Blick hatten. Da Mengenrelationen, Zahlzerlegungen und explizit auch Differenzen zwischen Anzahlen jedoch die dritte Ebene des Entwicklungsmodells früher mathematischer Kompetenzen nach Krajewski und Schneider (2006) darstellen, sollten entsprechende Items – die zwangsläufig auch sprachliche Kompetenzen erfordern – nicht ausgeschlossen werden. Den Fachkräften wurde außerdem mitgeteilt, dass in dieser Untersuchung die frühen *mathematischen* Fähigkeiten von Kindern interessant sind.

Das gewählte Testinstrument zum fachdidaktischen Wissen frühpädagogischer Fachkräfte von Gasteiger et al. (2020) erhebt das mathematikdidaktische Wissen der frühpädagogischen Fachkräfte zur Entwicklung von Zähl- und Mengenerfassungskompetenzen sowie zu geometrischen Fähigkeiten und zu Fähigkeiten zum Messen von Kindern. Im eingesetzten Fragebogen zur Einschätzung, der in Anlehnung an das Testinstrument MARKO-D (Ricken et al. 2013) entwickelt wurde, ist von den Fachkräften jedoch nur Wissen zu Zähl- und Mengenkompetenzen der Kinder gefragt. Auch wenn wohlüberlegt ein Globalmaß für das fachdidaktische Wissen herangezogen wurde, wäre es denkbar, dass ein Test, der spezifischer auf die Inhaltsbereiche ausgerichtet ist, zu anderen Ergebnissen geführt hätte. Jedoch zeigte sich auch im Schulkontext kein Einfluss des fachdidaktischen Wissens (McElvany et al. 2009).

7.4 Fazit

Auch wenn sich überwiegend erfreuliche Ergebnisse dahingehend zeigen, dass sich frühpädagogische Fachkräfte in ihren Einschätzungen in erster Linie nicht vom Geschlecht und vom sozioökonomischen Status der Kinder beeinflussen lassen, bleibt die Frage offen, welche Merkmale außer der tatsächlichen mathematischen Leistung des Kindes die Einschätzung mathematischer Kompetenzen im Bereich Mengen und Zahlen beeinflussen. Unter anderem könnte es also zielführend sein, mit weiteren (u. a. explorativen) Erhebungen erste Tendenzen dahingehend zu überprüfen, ob Einstellungen der frühpädagogischen Fachkräfte zur Mathematik und ihre Beobachtungs- und Wahrnehmungsfähigkeiten zur Aufklärung weiterer Varianz in der Einschätzungsgenauigkeit mathematischer Fähigkeiten im Bereich Mengen und Zahlen beitragen könnten. Auch eine genaue Betrachtung der einzuschätzenden Aufgaben könnte Hinweise auf weitere Zusammenhänge mit der Einschätzungsgenauigkeit aufdecken.

Je genauer das Verständnis von Einschätzungen und Einflussfaktoren geklärt wird, umso eher können Ideen entwickelt werden, wie die Einschätzungsgenauigkeit gefördert werden kann. Unabhängig von allen Einflussmerkmalen auf die Einschätzungsgenauigkeit frühpädagogischer Fachkräfte bleibt ein Forschungsdesiderat, inwiefern sich die Einschätzungsgenauigkeit tatsächlich im pädagogischen Handeln der Fachkräfte niederschlägt (Gasteiger und Benz 2016).

Funding Niedersächsisches Ministerium für Wissenschaft und Kultur (11-76251-14-1/18 (ZN3393))

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Anders, Y., Kunter, M., Brunner, M., Krauss, S., & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 57(3), 175–193. <https://doi.org/10.2378/peu2010.art13d>.
- Artelt, C., & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(34), 157–160. <https://doi.org/10.1024/1010-0652.23.34.157>.
- Avenarius, H., Ditton, H., Döbert, H., Klemm, K., Klieme, E., Rürup, M., Tenorth, H.-E., Weishaupt, H., & Weiß, M. (2003). *Bildungsbericht für Deutschland: Erste Befunde*. Wiesbaden: VS.
- Baltes-Götz, B. (2019). Lineare Regressionsanalyse mit SPSS. <https://www.uni-trier.de/fileadmin/urt/doku/linreg/linreg.pdf>. Zugegriffen: 21. Mai 2021.

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). lme4: linear mixed-effects models using Eigen and S4. <http://CRAN.R-project.org/package=lme4>. Zugegriffen: 21. Mai 2021. R package version 1.1-10.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J., & Weiß, M. (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Wiesbaden: VS.
- BBC Stories (Hrsg.). (2017). Girl toys vs boy toys: the experiment. <https://www.youtube.com/watch?v=nWu44AqF0iI>. Zugegriffen: 3. Nov. 2020.
- Benz, C., Peter-Koop, A., & Grübing, M. (2015). *Frühe mathematische Bildung: Mathematiklernen der Drei- bis Achtjährigen*. Berlin, Heidelberg: Springer Spektrum.
- Benz, C., Grübing, M., Lorenz, J.H., Reiss, K., Selter, C., & Wollring, B. (2017). *Frühe mathematische Bildung: Ziele und Gelingensbedingungen für den Elementar- und Primarbereich*. Berlin, Toronto: Verlag Barbara Budrich.
- Blevins-Knabe, B., Austin, A.B., Musun, L., Eddy, A., & Jones, R.M. (2000). Family home care providers' and parents' beliefs and practices concerning mathematics with young children. *Early Child Development and Care*, 165(1), 41–58. <https://doi.org/10.1080/0300443001650104>.
- Bradley, R.H., Corwyn, R.F., Burchinal, M., McAdoo, H.P., & Coll, C.G. (2001). The home environments of children in the United States part II: relations with behavioral development through age thirteen. *Child development*, 72(6), 1868–1886. <https://doi.org/10.1111/1467-8624.t01-1-00383>.
- Common Core State Standards Initiative (Hrsg.). (2021). Standards for mathematical practice: kindergarten. www.corestandards.org/Math/Content/K/. Zugegriffen: 20. Aug. 2021.
- Dollinger, S. (2013). *Diagnosegenauigkeit von ErzieherInnen und LehrerInnen*. Wiesbaden: Springer VS.
- Dornheim, D. (2013). *Prädiktion von Rechenleistung und Rechenschwäche: der Beitrag von Zahlen-Vorwissen und allgemein-kognitiven Fähigkeiten*. Berlin: Logos.
- Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L.S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental psychology*, 43(6), 1428–1446. <https://doi.org/10.1037/0012-1649.43.6.1428>.
- Dunekacke, S., Jenßen, L., Eilerts, K., & Blömeke, S. (2016). Epistemological beliefs of prospective preschool teachers and their relation to knowledge, perception, and planning abilities in the field of mathematics: a process model. *ZDM – Mathematics Education*, 48(1–2), 125–137. <https://doi.org/10.1007/s11858-015-0711-6>.
- Ehm, J.-H., Duzy, D., & Hasselhorn, M. (2011). Das akademische Selbstkonzept bei Schulanfängern. *Frühe Bildung*, 2011(0), 37–45. <https://doi.org/10.1026/2191-9186/a000008>.
- Fisher, C.W., Berliner, D.C., Filby, N.N., Marliave, R., Cahen, L.S., & Dishaw, M.M. (1981). Teaching behaviors, academic learning time, and student achievement: an overview. *The Journal of Classroom Interaction*, 17(1), 2–15.
- Fröhlich-Gildhoff, K., Weltzien, D., Kirstein, N., Pietsch, S., & Rauh, K. (2014). Kompetenzen früh-/kindheitspädagogischer Fachkräfte im Spannungsfeld von normativen Vorgaben und Praxis: Expertise. <https://www.bmfsfj.de/resource/blob/94174/6a47780ac348d80e4718f38d52230ae8/kompetenzen-frueh-kindheitspaedagogischer-fachkraefte-im-spannungsfeld-von-normativen-vorgaben-und-praxis-data.pdf>. Zugegriffen: 3. Mai 2021.
- Furnari, E.C., Whittaker, J., Kinzie, M., & DeCoster, J. (2017). Factors associated with accuracy in prekindergarten teacher ratings of students' mathematics skills. *Journal of Psychoeducational Assessment*, 35(4), 410–423. <https://doi.org/10.1177/0734282916639195>.
- Gasteiger, H. (2010). *Elementare mathematische Bildung im Alltag der Kindertagesstätte: Grundlegung und Evaluation eines kompetenzorientierten Förderansatzes*. Münster: Waxmann.
- Gasteiger, H., & Benz, C. (2016). Mathematikdidaktische Kompetenz von Fachkräften im Elementarbereich – ein theoriebasiertes Kompetenzmodell. *Journal für Mathematik-Didaktik*, 37(2), 263–287. <https://doi.org/10.1007/s13138-015-0083-z>.
- Gasteiger, H., Bruns, J., Benz, C., Brunner, E., & Sprenger, P. (2020). Mathematical pedagogical content knowledge of early childhood teachers: a standardized situation-related measurement approach. *ZDM – Mathematics Education*, 52(2), 193–205. <https://doi.org/10.1007/s11858-019-01103-2>.
- Geary, D.C., Hoard, M.K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child development*, 78(4), 1343–1359. <https://doi.org/10.1111/j.1467-8624.2007.01069.x>.
- Gersten, R., Jordan, N.C., & Flojo, J.R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of learning disabilities*, 38(4), 293–304. <https://doi.org/10.1177/00222194050380040301>.

- Gresham, G. (2007). A study of mathematics anxiety in pre-service teachers. *Early Childhood Education Journal*, 35(2), 181–188. <https://doi.org/10.1007/s10643-007-0174-7>.
- Helmke, A. (2021). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze-Velber: Klett.
- Hoge, R. D., & Coladarsi, T. (1989). Teacher-based judgments of academic achievement: a review of literature. *Review of Educational Research*, 59(3), 297–313. <https://doi.org/10.2307/1170184>.
- Hosenfeld, I., Helmke, A., & Schrader, F.-W. (2002). Diagnostische Kompetenz: Unterrichts- und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Unterrichtsstudie SALVE. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 65–82). Weinheim: Beltz.
- International Labour Office (Hrsg.). (2012). *International standard classification of occupations: structure, group definitions and correspondence tables*. Geneva: International Labour Office.
- Jenßen, L., Thiel, O., Dunekacke, S., & Blömeke, S. (2020). Mathematikangst bei angehenden frühpädagogischen Fachkräften: Bedeutsam für professionelles Wissen und Wahrnehmung von mathematischen Inhalten im Kita-Alltag? *Journal für Mathematik-Didaktik*, 41(2), 301–327. <https://doi.org/10.1007/s13138-019-00151-1>.
- Jiang, J., & Nguyen, T. (2021). *Linear and generalized linear mixed models and their applications*. New York: Springer New York.
- JMK, & KMK (2004). *Gemeinsamer Rahmen der Länder für die frühe Bildung in Kindertageseinrichtungen*.
- Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and individual differences*, 20(2), 82–88. <https://doi.org/10.1016/j.lindif.2009.07.004>.
- Kammermeyer, G. (1998). *Schulfähigkeit: Kriterien und diagnostische, prognostische Kompetenz von Lehrerinnen, Lehrern und Erzieherinnen*. Erlangen-Nürnberg: Julius Klinkhardt.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbe- reich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 197–209. <https://doi.org/10.1024/1010-0652.23.34.197>.
- Kilday, C. R., Kinzie, M. B., Mashburn, A. J., & Whittaker, J. V. (2012). Accuracy of teacher judgments of preschoolers' math skills. *Journal of Psychoeducational Assessment*, 30(2), 148–159. <https://doi.org/10.1177/0734282911412722>.
- Kowalski, K., Brown, R. D., Pretti-Frontczak, K., Uchida, C., & Sacks, D. F. (2018). The accuracy of teachers' judgments for assessing young children's emerging literacy and math skills. *Psychology in the Schools*, 55(9), 997–1012. <https://doi.org/10.1002/pits.22152>.
- Krajewski, K., & Schneider, W. (2006). Mathematische Vorläuferfertigkeiten im Vorschulalter und ihre Vorhersagekraft für die Mathematikleistungen bis zum Ende der Grundschulzeit. *Psychologie in Erziehung und Unterricht*, 53(4), 246–262.
- Kuger, S., Kluczniok, K., Sechtig, J., & Smidt, W. (2011). Gender im Kindergarten. Empirische Datenlage zu Unterschieden zwischen Mädchen und Jungen. *Zeitschrift für Pädagogik*, 57(2), 269–288.
- Kuratli Geeler, S. (2019). *Mathematische Kompetenzen von Kindergartenkindern: Überprüfung eines Testinstruments und Analyse von Unterschieden in der numerischen Leistungsentwicklung*. Zurich: University of Zurich.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>.
- Lorenz, C. (2011). *Diagnostische Kompetenz von Grundschullehrkräften: Strukturelle Aspekte und Bedingungen*. Bamberg: Univ. of Bamberg Press.
- Mang, J., Ustjanzew, N., Leßke, I., Schiepe-Tiska, A., & Reiss, K. (Hrsg.). (2019). *PISA 2015 Skalenhandbuch: Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., Horz, H., & Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(3), 223–235. <https://doi.org/10.1024/1010-0652.23.34.223>.
- McKevett, N. M., & Kiss, A. J. (2019). The influence of data on teachers' judgments of students' early reading and math skills. *Psychology in the Schools*, 56(7), 1157–1172. <https://doi.org/10.1002/pits.22256>.
- Mink, N., & Müller, S. (2018). Weiterqualifizierung nach Ausbildung und Studium. In K. Fuchs-Rechlin & I. Züchner (Hrsg.), *Was kommt nach dem Berufsstart? Mittelfristige berufliche Platzierung von Erzieherinnen und Erziehern sowie Kindheitspädagoginnen und Kindheitspädagogen* (S. 50–60). München: Deutsches Jugendinstitut e. V.

- MKFFI NRW (2018). *Bildungsgrundsätze: Mehr Chancen durch Bildung von Anfang an*. Freiburg im Breisgau: Herder.
- Moosbrugger, H., & Kelava, A. (2008). *Testtheorie und Fragebogenkonstruktion*. Berlin, Heidelberg: Springer.
- Muntoni, F., Dunekacke, S., Heinze, A., & Retelsdorf, J. (2019). Geschlechtsspezifische Erwartungseffekte in Mathematik. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 51(2), 84–96. <https://doi.org/10.1026/0049-8637/a000212>.
- Musca, S. C., Kamiejski, R., Nugier, A., Méot, A., Er-Rafiy, A., & Brauer, M. (2011). Data with hierarchical structure: impact of intraclass correlation and sample size on type-I error. *Frontiers in psychology*. <https://doi.org/10.3389/fpsyg.2011.00074>.
- Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., & Spitler, M. E. (2016). Which preschool mathematics competencies are most predictive of fifth grade achievement? *Early childhood research quarterly*, 36, 550–560. <https://doi.org/10.1016/j.ecresq.2016.02.003>.
- NMW (2018). Orientierungsplan für Bildung und Erziehung. https://www.mk.niedersachsen.de/startseite/fruhkindliche_bildung/orientierungsplan/orientierungsplan-fuer-bildung-und-erziehung-86998.html. Zugegriffen: 3. Nov. 2020.
- Ostermann, A., Leuders, T., & Philipp, K. (2019). Fachbezogene diagnostische Kompetenzen von Lehrkräften – Von Verfahren der Erfassung zu kognitiven Modellen zur Erklärung. In T. Leuders, M. Nückles, S. Mikelskis-Seifert & K. Philipp (Hrsg.), *Pädagogische Professionalität in Mathematik und Naturwissenschaften* (S. 93–116). Wiesbaden: Springer Vieweg.
- Ricken, G., Fritz-Stratmann, A., & Balzer, L. (2013). MARKO-D: Mathematik- und Rechenkonzepte im Vorschulalter – Diagnose. <https://www.testzentrale.de/shop/mathematik-und-rechenkonzepte-im-vorschulalter-diagnose.html>. Zugegriffen: 3. Nov. 2020.
- Rogalla, M., & Vogt, F. (2008). Förderung adaptiver Lehrkompetenz: eine Interventionsstudie. *Unterrichtswissenschaft*, 36(1), 17–36.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353. <https://doi.org/10.1037/1040-3590.8.4.350>.
- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Frankfurt am Main: Peter Lang.
- Schrader, F.-W., & Praetorius, A.-K. (2018). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost, J. R. Sparfeldt & S. Buch (Hrsg.), *Handwörterbuch pädagogische Psychologie* (S. 92–98). Weinheim, Basel: Beltz.
- Selter, C. (2008). Wie junge Kinder rechnen lernen. In L. Fried, G. Büttner, I. Melle, N. Mette & C. Selter (Hrsg.), *Das wissbegierige Kind: Neue Perspektiven in der Früh- und Elementarpädagogik* (S. 37–55). Weinheim: Juventa.
- 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. <https://doi.org/10.1024/1010-0652.19.12.85>.
- StMAS (Hrsg.). (2016). *Der Bayerische Bildungs- und Erziehungsplan für Kinder in Tageseinrichtungen bis zur Einschulung*. Berlin: Cornelsen Verlag.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: a meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>.
- Tucker-Drob, E. M., Rhemtulla, M., Harden, K. P., Turkheimer, E., & Fask, D. (2011). Emergence of a Gene x socioeconomic status interaction on infant mental ability between 10 months and 2 years. *Psychological science*, 22(1), 125–133. <https://doi.org/10.1177/0956797610392926>.
- Weißhaupt, S., Peucker, S., & Wirtz, M. (2006). Diagnose mathematischen Vorwissens im Vorschulalter und Vorhersage von Rechenleistungen und Rechenschwierigkeiten in der Grundschule. *Psychologie in Erziehung und Unterricht*, 53(4), 236–245.
- Wolf, C., & Best, H. (2010). *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS.