

Mathes 2

Screening zur Erfassung der Mathematikleistungen in
Klasse 2

Manual



Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen 4.0 International Lizenz](https://creativecommons.org/licenses/by-nc-sa/4.0/).



Inhalt

Abbildungsverzeichnis	3
Tabellenverzeichnis	3
Kurzinformation	4
1. Theoretische Grundlagen	5
1.1 Zielstellung der Mathes-Testreihe.....	5
1.2 Exkurs: Klassifikation und Epidemiologie von mathematischen Lernschwierigkeiten - Was soll verhindert werden?.....	5
2 Theoriebasierte Testentwicklung	6
2.1 Welche Kompetenzen erwerben Schülerinnen und Schüler in der zweiten Klasse?.....	7
2.2 Konstruktion inhaltsvalider Testanforderungen.....	8
2.3 Überlegungen zur Steigerung der Testökonomie.....	10
3 Testanwendung	10
3.1 Anwendungszeitraum und Zielgruppe	10
3.2 Testmaterial	11
3.3 Hinweise zur Testdurchführung	11
3.4 Auswertung und Interpretation.....	12
3.4.1 Manuelle Auswertung.....	12
3.4.2 Automatisierte Auswertung.....	12
3.4.3 Interpretation der Ergebnisse	12
4 Testgütekriterien	13
4.1 Itemkennwerte	13
4.2 Gültigkeit des Messmodells.....	14
4.3 Reliabilität.....	15
4.4 Validität.....	16
4.4.1 Konstruktvalidität.....	16
4.4.2 Prognostische Validität	16
5 Literaturverzeichnis	18
6 Anhang	21

Abbildungsverzeichnis

Abbildung 1: Referenzniveaus als Interpretationshilfen für die erzielte Testleistung 13
Abbildung 2: Verteilungen von Aufgabenschwierigkeiten und Schülerfähigkeiten zu den beiden Messzeitpunkten
..... 14
Abbildung 3: Verteilung der Itemfit-Statistiken des „Mathes 2“ 15

Tabellenverzeichnis

Tabelle 1: inhaltliche Anforderungen des "Mathes 2" 10
Tabelle 2: Durchführungszeiträume des "Mathes 2" 11
Tabelle 3: Zusammenhänge des „Mathes 2“ zu konstruktähnlichen und konstruktfernen Testverfahren..... 16
Tabelle 4: Ergebnisse zur klassifikatorischen Güte des "Mathes 2" bezüglich Risiken im Fach Mathematik 17

Kurzinformation

Titel, Autoren, Jahr	Mathes 2, Simon Sikora & Stefan Voß, 2016
Diagnostische Zielsetzung	Erfassung der arithmetischen Kompetenzen der zweiten Klassenstufe bei allen Kindern
Anwendungsbereiche	Grundschulmathematikunterricht in der zweiten Klasse
Aufbau	56 Aufgaben aus den Bereichen Zahlbegriffsverständnis sowie den vier Grundrechenarten
Anwendungszeitraum	Mitte Klasse 2 (20. / 21. Schulwoche) Anfang Klasse 3 (3. / 4. Schulwoche)
Durchführung	Gruppentestung im Klassenkontext durch die Lehrkraft bei freier Testbearbeitung innerhalb einer Unterrichtsstunde
Auswertung	Auswertung mithilfe von Auswertungsvorlagen
Normen	Klumpenstichprobe aus Mecklenburg-Vorpommern ($N = \text{ca. } 580$)
Reliabilität	Interne Konsistenz: $\alpha = .94$ zu beiden Messzeitpunkten Retest-Reliabilität: $r_{tt} = .66^{**}$ ($N = 531$)
Validität	Konstruktvalidität: konstruktnahe Korrelation mit DEMAT 2+: $r = .74^{**}$ ($N = 548$) konstruktferne Korrelationen: mit DERET 1-2+: $r = -.47^{**}$ ($N = 563$); mit FE-L 2: $r = .34^{**}$ ($N = 566$) prognostische Validität: Korrelation Mathes 2 (Mitte Klasse 2) zum DEMAT 2+ (Ende Klasse 2): $r = .71^{**}$ ($N = 520$) RATZ-Index Mathes 2 (Mitte Klasse 2) zum DEMAT 2+ (Ende Klasse 2): 0.87 ($N = 520$)

1. Theoretische Grundlagen

1.1 Zielstellung der Mathes-Testreihe

Täglich stellen sich Lehrkräfte Fragen wie die folgenden:

„Lernen alle Kinder in meinem Unterricht erfolgreich oder kommt jemand nicht mit?“
„Kann ich das aktuelle Themengebiet abschließen und im Stoff weitermachen oder brauchen die Schülerinnen und Schüler noch mehr Lernzeit?“
„Was genau hat die Schülerin bzw. der Schüler noch nicht verstanden?“

Solche Fragen haben das Ziel, den Unterricht bestmöglich an die individuellen Lernausgangslagen und Förderbedürfnisse der Kinder anzupassen. Die Mathes-Testreihe soll Lehrkräfte bei vertretbarem Aufwand dabei unterstützen, zu einer präzisen Einschätzung der aktuellen Lernstände sowie der Leistungsentwicklung der Schülerinnen und Schüler zu gelangen. Die Testverfahren können in dreierlei Hinsicht helfen, nämlich

- beim genauen Einschätzen des Spektrums der Leistungen der Schülerinnen und Schüler, um den Unterricht daran bestmöglich anpassen zu können,
- beim rechtzeitigen Erkennen derjenigen Schülerinnen und Schüler mit Risiken bzw. bereits ausgeprägten Schwierigkeiten im Kompetenzerwerb sowie
- bei der Planung effektiver Fördermaßnahmen, insbesondere für diejenigen Schülerinnen und Schüler mit besonderen Unterstützungsbedarfen.

Zu diesem Zweck wurden Mathes-Tests für jede Klassenstufe entwickelt, welche jeweils zum Beginn und in der Mitte des Schuljahres eingesetzt werden können.

1.2 Exkurs: Klassifikation und Epidemiologie von mathematischen Lernschwierigkeiten - Was soll verhindert werden?

Schwierigkeiten beim Rechnen- bzw. mathematischen Lernen gehören für viele Schülerinnen und Schüler (und deren Lehrerinnen und Lehrer) zum schulischen Alltag. Allerdings erhält nicht jedes Kind mit schwachen Leistungen in Mathematik die Diagnose „Dyskalkulie“ bzw. „Rechenstörung“. Nach den Kriterien der Weltgesundheitsorganisation leidet ein Kind nur dann unter einer Dyskalkulie, wenn seine Beeinträchtigung der Rechenfertigkeiten im Gegensatz sowohl zur allgemeinen Intelligenz als auch zu anderen schulischen Leistungen, z. B. dem Lesen und der Rechtschreibung, steht (sog. doppeltes Diskrepanzkriterium):

„Diese Störung beinhaltet eine umschriebene Beeinträchtigung von Rechenfertigkeiten, die nicht allein durch eine allgemeine Intelligenzminderung oder eine unangemessene Beschulung erklärbar ist. Das Defizit betrifft die Beherrschung grundlegender Rechenfertigkeiten wie Addition, Subtraktion, Multiplikation und Division, weniger die höheren mathematischen Fertigkeiten, die für Algebra, Trigonometrie, Geometrie oder Differential- und Integralrechnungen benötigt werden.“ (ICD-10, Dilling, Mombour & Schmidt, 2011, S. 338)

Das Verwenden des Diskrepanzkriteriums zur Bestimmung von Kindern mit Rechenschwierigkeiten sowie eine daran gebundene Zuweisung von Fördermaßnahmen wird in der Fachliteratur aus verschiedenen Gründen kritisiert (u. a. Gaidoschik, 2011; Hartke & Diehl, 2013; Krajewski, 2003; Lorenz, 2005; Moser Opitz, 2004; Koch, 2005; Koch & Knopp, 2010). Gaidoschik stellt berechtigterweise die Frage: „Verdient denn ein Kind, das nicht nur im Rechnen, sondern auch beim Lesen Probleme hat, weniger Förderung in Mathematik als jenes, welches dem ‚Diskrepanz-Kriterium‘ genügt?“ (2011, S. 12). „Es erscheint hingegen sinnvoller, all jene Kinder in die Förderung aufzunehmen, deren Lernfortschritte, durch welche Gründe auch immer, als unzureichend angesehen werden“ (Lorenz, 2005, S. 15).

Während nur etwa 4 % bis 8 % aller Schülerinnen und Schüler den Kriterien einer Dyskalkulie bzw. Rechenstörung entsprechen (von Aster, Schweiter & Weinhold Zulauf, 2007; Lorenz, 2014), gehen Hasselhorn, Marx und Schneider (2005) vor dem Hintergrund der Befunde einschlägiger Prävalenz- und Schulleistungsstudien wie IGLU (Bos et al., 2003) oder TIMSS (Bos et al., 2008; Bos, Wendt, Köller & Selter, 2012; Selter, Walter, Walther & Wendt, 2016) davon aus, dass etwa 20 % aller Viertklässlerinnen und Viertklässler im Fach Mathematik Leistungsrückstände im Umfang von zwei Schuljahren aufweisen.

Schwierigkeiten im Fach Mathematik kommen bei Jungen und Mädchen etwa gleich häufig vor (Jacobs & Petermann, 2012; Landerl & Kaufmann, 2008). Bei vielen Schülerinnen und Schülern treten mathematische Lernschwierigkeiten nicht isoliert auf, sondern in Kombination mit Lese-Rechtschreibschwächen und psychischen Auffälligkeiten, insbesondere ADHS, Ängste und Depressionen (zusammenfassend Lambert, 2015; Sikora & Voß, 2018).

Die berichteten Befunde machen deutlich, dass statistisch gesehen etwa jedes fünfte Kind besondere Unterstützung im Fach Mathematik benötigt. „Frühzeitig zu erkennen, wenn Kinder Schwierigkeiten beim Erlernen mathematischer Begriffe haben, ist vermutlich der wichtigste Schritt auf dem Weg zur Förderung“ (Hasemann & Gasteiger, 2014, S. 151). Die Verfahren der Mathes-Testreihe sollen Lehrkräfte dabei unterstützen.

2 Theoriebasierte Testentwicklung

Aufgrund der Relevanz der Arithmetik für den weiteren mathematischen Kompetenzerwerb wurde beschlossen, die Mathes-Testreihe in den ersten beiden Grundschuljahren auf den Inhaltsbereich *Zahlen und Operationen* zu begrenzen, wenngleich auch im Anfangsunterricht bereits geometrische und stochastische Fragestellungen sowie der Umgang mit Größen thematisiert werden sollten (KMK, 2005). Diese Entscheidung war notwendig, um dem Anspruch der Verfahren einer zeitökonomischen Leistungsmessung ganzer Klassen gerecht werden zu können.

2.1 Welche Kompetenzen erwerben Schülerinnen und Schüler in der zweiten Klasse?

Nachdem in der ersten Klasse typischerweise im Zahlenraum bis 20 gearbeitet wird, wird im zweiten Schuljahr traditionell der Hunderterraum erschlossen (Krauthausen, 2018). Um auch in diesem erweiterten Zahlenraum sicher rechnerisch operieren zu können, gilt es, aspektreiche Vorstellungen davon aufzubauen. Grundlegend sind die *kardinale* und die *ordinale* Sichtweise auf Zahlen, welche nach Wittmann und Müller (2012) die erste fundamentale Idee der Arithmetik darstellen.

Dass Zahlen eine Reihe bilden, wissen bereits die meisten dreijährigen Kinder, zumindest implizit. Dieser sogenannte *Ordinalzahlaspekt* wird typischerweise noch einmal unterteilt, und zwar in den Zählaspekt, also die Folge der Zahlen, welche beim Zählen durchlaufen wird, und in den Ordnungszahlaspekt, welcher den Platz einer Zahl in der Zahlenreihe angibt (z. B. 7 als Vorgänger von 8 und als Nachfolger von 6, aber auch der drittgrößte Junge der Klasse, etc.; Padberg & Benz, 2011).

Beim *Kardinalzahlaspekt* (auch Anzahlaspekt bezeichnet) wird eine Menge von Elementen in den Blick genommen (ebd.). Für die Zähl- und Rechenentwicklung ist die Einsicht von fundamentaler Bedeutung, dass strukturierte, d. h. regelhafte, Mengendarstellungen gegenüber unstrukturiert bzw. unübersichtlich präsentierten effektiver erfasst werden können. Die Ausnutzung von Strukturen wird umso wichtiger, je größer die Mengen werden, mit denen im Unterricht gearbeitet wird. In den einschlägigen didaktischen Lehrbüchern zum mathematischen Anfangsunterricht wird deshalb empfohlen, „gezielt zu einer *strukturierten, nicht-zählenden* Anzahlerfassung hinzuführen (Gaidoschik, 2010, S. 216; Hervorhebungen im Original). Diese erfordert „ein *bewusstes Wahrnehmen* der Teilanzahlen und ein *wissendes Zusammensetzen* dieser Teilanzahlen zum Zahlganzen“ (ebd., S. 217; Hervorhebungen im Original). Damit die Kinder Strukturen erkennen und nutzen können, müssen diese natürlich in den Mengendarstellungen auch vorhanden sein. Als günstig wurde die sogenannte Kraft der Fünf (Flexer, 1986) herausgearbeitet, d. h. die Bündelung der Einzelelemente zu Fünfern. Unser Stellenwertsystem gibt zudem eine weitere grundlegende Struktur vor, die sogenannte Kraft der Zehn, also die Zusammenfassung von zehn Einern zu einem Zehner.

Die Rechenoperationen *Addition* und *Subtraktion* wurden bereits in der ersten Klassenstufe ausgiebig behandelt (Krauthausen, 2018). Bereits Schulanfängerinnen und Schulanfänger verfügen über beachtliche Kompetenzen beim Plus- und Minusrechnen (zusammenfassend Sikora, 2017), in der Regel nutzen Kinder in diesem Alter jedoch noch Zählstrategien zur Lösung der gestellten Aufgaben. Das zentrale Ziel des mathematischen Anfangsunterrichts ist, neben dem Aufgreifen und Vertiefen des Verständnisses für beide Rechenoperationen, der Erwerb höherwertiger, sogenannter heuristischer, Strategien zur Lösung der Aufgaben (Häsel-Weide, Nührenböcker, Moser Opitz & Wittich, 2019). Diese Ableitungsstrategien machen sich die Strukturen des Zahlenraumes und die Beziehungen zwischen Aufgaben zunutze. Unter Rückgriff auf einfache oder bereits bekannte Aufgaben wird so das Rechnen erleichtert, beispielsweise indem Verdopplungen bzw. Halbierungen, Analogien oder Nachbaraufgaben für die Lösung einer unbekannteren Aufgabe genutzt werden. Solche Strategien sind effizienter und eleganter als zählendes Rechnen, das zwar bei kleineren Zahlen noch gut funktionieren mag, in den immer größer werdenden Zahlenräumen der nachfolgenden Schuljahre jedoch versagt (Padberg & Benz, 2011).

Sobald der Hunderterraum erarbeitet ist und die Operationen Addition und Subtraktion geläufig ausgeführt werden können, lernen die Kinder im zweiten Schuljahr die anderen beiden Grundrechenarten kennen, die *Multiplikation* und die *Division* (Krauthausen, 2018). Bei der Erarbeitung der neuen Operationen sollte an die Vorerfahrungen der Kinder angeknüpft werden (Hasemann & Gasteiger, 2014). So werden die Grundvorstellungen der *Multiplikation* typischerweise über zeitlich-sukzessive Handlungen sowie räumlich-simultane Anordnungen vermittelt (Padberg & Benz, 2011). Erstere betonen den dynamischen Aspekt der Operation, indem die Multiplikation als mehrmalige Wiederholung der gleichen Handlung aufgefasst wird (z. B. „Ich bin dreimal zum Schrank gegangen und habe je zwei Gläser herausgenommen.“). Bei letzterem, dem statischen Aspekt, wird die Operation als mehrmalige Wiederholung der gleichen Anordnung dargestellt (z. B. beim Würfeln derselben Augenzahl mit mehreren Würfeln). Nachdem Grundvorstellungen aufgebaut wurden, gilt es, die neue Operation auch formal zu erarbeiten. Dazu wird in aktuellen Veröffentlichungen ein ganzheitlicher Zugang empfohlen (Gaidoschik, 2014). Das bedeutet, dass zunächst nur einfache, sogenannte Kernaufgaben (*Verdopplungen* bzw. *Halbierungen* sowie Einer-, Fünfer- und Zehnerreihe), des kleinen Einmaleins bzw. Einsdurcheins behandelt werden, die Lösung schwierigerer Aufgaben erfolgt anschließend durch Ableitungen (z. B. Tauschaufgaben) bzw. unter Anwendung von Zerlegungsstrategien (z. B. sieben mal acht als fünf mal acht plus zwei mal acht).

Die *Division* stützt sich auf die Grundvorstellung des Aufteilens sowie des Verteilens (Hasemann & Gasteiger, 2014). Beim Aufteilen wird die Anzahl der Teilmengen gesucht, wobei der Divisor die Anzahl der Elemente in einer Teilmenge repräsentiert (z. B. „Zwölf Kinder werden auf vierer Gruppen aufgeteilt.“). Diese Vorstellung kann auch als wiederholte Subtraktion gleicher Subtrahenden dargestellt werden, indem, um beim Beispiel zu bleiben, die Frage gestellt wird, wie viele Gruppen entstehen. Die Vorstellung des Verteilens mag den Kindern hingegen intuitiv näherliegen (z. B. „Zwölf Bonbons werden gerecht an vier Kinder verteilt.“), allerdings wird dabei der Zusammenhang zur Multiplikation schwerer ersichtlich. Obwohl beide Rechenoperationen eng zusammenhängen, „sollte ihre grundlegende Erschließung *nicht* gleichzeitig erfolgen. So ist *zunächst* ein gutes Verständnis der Multiplikation sowie einige rechnerische Sicherheit im Hunderterraum erforderlich, bevor die Division erfolgreich eingeführt werden kann“ (Padberg & Benz, 2011, S. 152; Hervorhebungen im Original).

Sobald tragfähige Vorstellungen beider Rechenoperationen und flexible Lösungsstrategien erworben wurden, gilt es anschließend, die Aufgaben des kleinen Einmaleins und Einsdurcheins zu automatisieren (Krauthausen, 2018; Padberg & Benz, 2011; Gaidoschik, 2014), denn der gedächtnismäßige Abruf der Grundaufgaben ist wiederum die Voraussetzung für die Lösung schwierigerer Multiplikations- und Divisionsaufgaben im Kopfrechenverfahren in den nachfolgenden Schuljahren.

2.2 Konstruktion inhaltsvalider Testanforderungen

Die im vorherigen Abschnitt 2.1 durchgeführte Analyse der mathematikdidaktischen Fachliteratur hat im Bereich der Arithmetik die folgenden inhaltlichen Schwerpunkte in der zweiten Klassenstufe aufgezeigt:

1. kardinales und ordinales Zahlbegriffsverständnis
2. Addition und Subtraktion

3. kleines Einmaleins und kleines Einsdurcheins

Zu 1) Der *kardinale Zahlaspekt* wird durch Aufgaben zur Anzahlerfassung abgeprüft. Die Mengen werden in Hunderterfeldern angeordnet, wodurch eine strukturierte Erfassung unterstützt wird. Dieses Darstellungsmittel ist im Anfangsunterricht gebräuchlich (Käpnick, 2014), zudem berücksichtigt es die Fünfer- und die Zehnerbündelung. Sofern ein Kind jedes Element der vorgegebenen Menge einzeln zählt, ist dieser Vorgang bereits im Hunderterraum sehr fehleranfällig und zeitaufwendig (Häsel-Weide et al., 2019), insbesondere bei großen Mengen. Daher sind sowohl kleinere als auch größere Mengen zu erfassen, wodurch der Schwierigkeitsgrad der Aufgaben systematisch variiert wird.

Um das Wissen über den Aufbau des Zahlenraumes abzubilden, werden jeweils sechs Zahlen vorgegeben, welche der Größe nach geordnet werden müssen. Diese Anforderung entspricht dem *Ordnungsaspekt*. Der *Zählaspekt* wird im „Mathes 2“ hingegen nicht mehr erfasst.

Zu 2) Ob die Kinder sicher im Kopf *addieren* und *subtrahieren* können, wird durch eine Vielzahl von Aufgaben verschiedener Schwierigkeitsgrade geprüft. Dabei wird die gesamte Bandbreite der möglichen Gleichungen in diesem Zahlenraum abgebildet. So werden sowohl sehr einfache Aufgaben vom Typ $Z \pm Z$ oder $ZE \pm E$ dargeboten, welche sicherstellen, dass „Mathes 2“ auch bei Kindern mit größerem Lernrückstand im Fach Mathematik hinreichend differenziert. Außerdem wurden auch schwierige Aufgaben in das Verfahren aufgenommen. Am schwierigsten sind Rechnungen vom Typ $ZE \pm ZE$ mit Zehnerüberschreitung. Die Aufgaben werden zudem nicht nur in der gewohnten Form präsentiert, bei denen die Summe bzw. Differenz zu ermitteln ist, sondern auch als Gleichungen mit Platzhaltern. Bei diesen müssen verschiedene Grundvorstellungen aktiviert und vernetzt werden, insbesondere das Gleichungsverständnis und das Teil-Ganzes-Konzept (Fritz, Ricken & Gerlach, 2007), was viele Kinder vor besondere Herausforderungen stellt, insbesondere zählend rechnende (Gaidoschik, 2011).

Zu 3) Bezüglich der Multiplikation ist das *Verdoppeln* die elementarste Handlung und daher typischerweise der Ausgangspunkt bei der Erarbeitung der Rechenoperation (Gaidoschik, 2014). Analog dazu ist das *Halbieren* grundlegend für die Division. Daher wurden Verdopplungs- und Halbierungsaufgaben in das vorliegende Verfahren aufgenommen. Der Schwierigkeitsgrad dieser wurde systematisch variiert: Am leichtesten sind Verdopplungen bzw. Halbierungen von Einer- oder Zehnerzahlen, am schwierigsten sind Zahlen mit Zehnern und Einern, bei denen die Verdopplung bzw. Halbierung einen Zehnerübergang erfordert.

Ob das Kind über *grundlegende Operationsvorstellungen bezüglich der Multiplikation* verfügt, wird durch flächig angeordnete Mengen erfasst, welche in Einmaleinsaufgaben übersetzt werden müssen. Dies entspricht dem räumlich-simultanen Aspekt der Multiplikation.

Ebenso wie in der Addition und Subtraktion auch, werden zahlreiche *Multiplikations- und Divisionsaufgaben* vorgegeben, welche im Kopf zu lösen sind und verschiedene Schwierigkeitsgrade aufweisen. So wurden sowohl Kernaufgaben in beiden Grundrechenarten in das Verfahren aufgenommen, aber auch solche Aufgaben, welche Ableitungs- oder Zerlegungsstrategien erfordern.

Insgesamt wurden 56 Aufgaben konstruiert. Die Tabelle 1 fasst die inhaltlichen Anforderungen des „Mathes 2“ zusammen.

Tabelle 1: inhaltliche Anforderungen des "Mathes 2"

Aufgabengruppe bzw. Anforderung	Anzahl der Items
Mengenerfassung	2
Zahlen ordnen	2
Kopfrechnen Addition	12
Kopfrechnen Subtraktion	12
Verdoppeln	6
Halbieren	6
Grundvorstellung Multiplikation	4
Kopfrechnen Multiplikation	6
Kopfrechnen Division	6

2.3 Überlegungen zur Steigerung der Testökonomie

Zur Erreichung einer hohen Testökonomie und Zumutbarkeit der Durchführung des „Mathes 2“ wurden folgende Maßnahmen getroffen:

- freie Testbearbeitung ohne Zeitbegrenzung
- kurze und prägnante Aufgabenstellungen bzw. Operatoren
- weitgehend sprachfreie Gestaltung der Aufgaben, soweit zweckmäßig mit einem Beispiel und/oder grafischer Visualisierung
- kindlich angemessene, ansprechende Gestaltung des Testhefts mit Identifikationsfigur („Mathes – der Matheaffe“)
- freie Verfügbarkeit des Verfahrens durch Veröffentlichung unter Creative-Commons-Lizenz
- Testheftformat A5 mit Ermöglichung eines schwarz/weiß-Ausdrucks
- computergestützte Aufbereitung der Testergebnisse auf Kind- und Klassenebene

3 Testanwendung

Hinweis: Alle für die Durchführung und Auswertung des „Mathes 2“ benötigten Informationen sowie die Testinstruktionen finden sich in übersichtlicher Form in den [Durchführungshinweisen](#).

3.1 Anwendungszeitraum und Zielgruppe

Vor dem Hintergrund der in Abschnitt 1.1 dargestellten Zielstellungen der Mathes-Testreihe, sind alle Mathes-Verfahren für die zeitökonomische Erfassung der Mathematikleistungen aller Schülerinnen und Schüler einer (inklusive) Grundschulklasse konzipiert.

„Mathes 2“ kann zum Halbjahr der zweiten und zu Beginn der dritten Klassenstufe eingesetzt werden. Die nachfolgende Tabelle 2 stellt die Durchführungszeiträume dar.

Tabelle 2: Durchführungszeiträume des "Mathes 2"

	Schulwoche	Kommentar
MZP 1	20./21.	vor dem Halbjahreszeugnis
MZP 2	3./4.	Anfang Klasse 3

3.2 Testmaterial

Für die Durchführung des „Mathes 2“ werden folgende Materialien benötigt:

- 1 Testheft pro Kind,
- 1 Füller, 1 Bleistift,
- 1 Testheft für die Lehrkraft zur Demonstration,
- 1 [Durchführungsanleitung](#) für die Lehrkraft.

3.3 Hinweise zur Testdurchführung

Die Durchführung des „Mathes 2“ erfolgt in Gruppen (Klassenverband). Um eine objektive Testanwendung zu gewährleisten, müssen folgende Punkte beachtet werden:

- Es muss gewährleistet sein, dass die Schülerinnen und Schüler in einer ruhigen, störungsfreien Atmosphäre die Aufgaben bearbeiten.
- Die Schülerinnen und Schülern haben insgesamt eine Unterrichtsstunde Zeit für die Aufgabenbearbeitung. Falls es Kinder gibt, die es in dieser Zeit nicht schaffen, alle Aufgaben zu bearbeiten, sollten diese in einer anderen Stunde die Möglichkeit erhalten, den Test zu beenden.
- „Mathes 2“ wird ohne Pause durchgeführt.
- Die Durchführungshinweise sind zu berücksichtigen und die [Testinstruktionen](#) wörtlich vorzulesen.
- Die zu lösenden Aufgaben dürfen vorab nicht mit den Schülerinnen und Schülern geübt werden.
- Die Testhefte werden erst ausgeteilt, nachdem der erste Abschnitt der wörtlichen Instruktionen vorgelesen wurde.
- Die Schülerinnen und Schüler dürfen während der Durchführung keine Hinweise und Hilfestellungen erhalten. Ermutigungen sind erlaubt.
- Es ist darauf zu achten, dass die Kinder nicht voneinander abschreiben.
- „Mathes 2“ sollte möglichst nicht in der letzten Unterrichtsstunde und nicht im Nachmittagsunterricht durchgeführt werden.

3.4 Auswertung und Interpretation

3.4.1 Manuelle Auswertung

Die [Auswertungsvorlage](#) unterstützt eine objektive und ökonomische Auswertung des „Mathes 2“. Alle richtig gelösten Aufgaben werden mit einem Punkt, falsch gelöste mit null Punkten bewertet. Die erreichten Punkte werden aufsummiert. Mithilfe der Normtabelle im Anhang kann die Testleistung des Kindes mit denen gleichaltriger Schülerinnen und Schüler verglichen werden. Dazu stehen die für statusdiagnostische Einschätzungen gängigen Standardwerte (Prozentrang und T-Wert) für beide Messzeitpunkte zur Verfügung (s. Anhang C, S. 24f.).

Differenzierte Informationen zur Auswertung des „Mathes 2“ liefern die [Durchführungshinweise](#) des Verfahrens.

3.4.2 Automatisierte Auswertung

Für Lehrkräfte aus Mecklenburg-Vorpommern wird über die Homepage www.lernlinie.de eine internetgestützte Auswertung des „Mathes 2“ angeboten. Bei dieser Variante müssen lediglich die erreichten Rohwerte der Kinder mithilfe der [Auswertungsvorlage](#) wie in Abschnitt 3.4.1 beschrieben ermittelt und auf der Internetplattform eingetragen werden. Anschließend werden automatisiert Ergebnisübersichten für jedes Kind erstellt, sodass auf einen Blick ersichtlich ist, wie seine Leistungen im Vergleich zu gleichaltrigen Schülerinnen und Schülern einzuschätzen sind. Bei mehrmaligem Einsatz des „Mathes 2“ stellt die Internetplattform ebenfalls die Lernentwicklung des Kindes graphisch dar. Zudem besteht die Möglichkeit, die Ergebnisse aller Schülerinnen und Schüler einer Klasse im Überblick anzuzeigen.

Lehrkräfte außerhalb Mecklenburg-Vorpommerns können die Testergebnisse ihrer Schülerinnen und Schüler in die vorbereitete [Klassenübersicht](#) eintragen, welche automatisch den erreichten Rohwerten die Prozentränge und T-Werte zuordnet und die in Abschnitt 3.4.3 aufgeführten Referenzniveaus graphisch veranschaulicht.

3.4.3 Interpretation der Ergebnisse

Bei der Einschätzung der Testleistung eines Kindes helfen sogenannte Referenzniveaus, welche auf den Prozentrang Bezug nehmen und diesen vereinfachend interpretieren, indem die Testleistung des Kindes im Vergleich zur Referenzgruppe in fünf Stufen eingeordnet wird. Referenzniveaus stellen Marker an empirisch bedeutsamen Grenzen dar (Prozentrang 10, 25, 75 und 90). Ein Prozentrang von 10 bedeutet, dass 10 Prozent der gleichaltrigen Schülerinnen und Schüler gleiche oder schlechtere Leistungen aufweisen, ein Prozentrang von 25, dass ein Viertel der Kinder ein gleiches oder schlechteres Testergebnis erzielt, usw. Auf diese Weise entstehen fünf Leistungsbereiche, sodass einfach ersichtlich ist, wie das Kind im Vergleich zu Gleichaltrigen abgeschnitten hat.

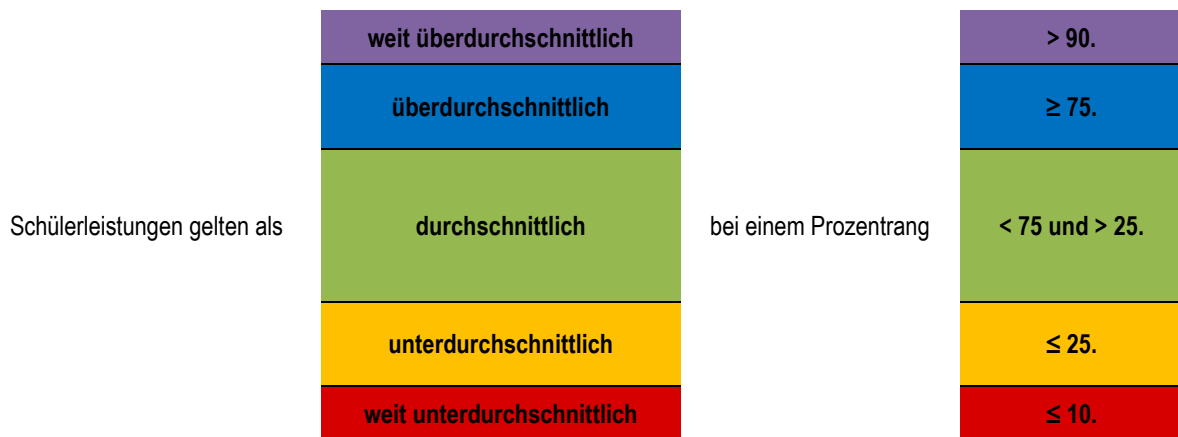


Abbildung 1: Referenzniveaus als Interpretationshilfen für die erzielte Testleistung

4 Testgütekriterien

„Mathes 2“ ist eine Kurzversion des im Rahmen der Dissertation von Kuhlmann (2013) entwickelten „Formative Erfassung der arithmetischen Fähigkeiten im 2. Schuljahr“ (FE-AF 2; Kuhlmann & Hartke, 2009a). Die Daten aus der Dissertationsstudie wurden reanalysiert, um Informationen zur Testgüte des „Mathes 2“ zu gewinnen.

Die Untersuchung wurde im Schuljahr 2009 / 2010 im Schulamtsbereich Neubrandenburg (Mecklenburg-Vorpommern) durchgeführt. Alle 18 Grundschulen aus der Stadt Neubrandenburg und dem Landkreis Mecklenburg-Strelitz wurden um Teilnahme gebeten. 13 Grundschulen mit 29 Klassen konnten dafür gewonnen werden.

Für die Evaluation der psychometrischen Güte des „Mathes 2“ wurden die Daten von zwei Messzeitpunkten im Januar 2009 (20. Schulwoche) sowie im Juni 2010 (40. Schulwoche) herangezogen. Zum zweiten Erhebungszeitpunkt wurden mit dem „Deutscher Mathematiktest für zweite Klassen“ (DEMAT 2+; Krajewski, Liehm & Schneider, 2004), dem „Deutscher Rechtschreibtest für das erste und zweite Schuljahr“ (Stock & Schneider, 2008) sowie der „Formative Erfassung der Lesefertigkeit im zweiten Schuljahr“ (FE-L 2; Kuhlmann & Hartke, 2009b) drei weitere Messverfahren durchgeführt, sodass Einschätzungen zur Validität (s. Abschnitt 4.4) möglich sind.

Die Stichprobe umfasste zum ersten Messzeitpunkt 574 Schülerinnen und Schüler und zum zweiten 578 Kinder. 531 Schülerinnen und Schüler nahmen an beiden Erhebungen der Untersuchung teil. Zu beiden Messzeitpunkten erwies sich das Geschlechterverhältnis als annähernd ausgeglichen (♂: 48 %; ♀: 52 %). Das Durchschnittsalter betrug zum ersten Messzeitpunkt 8;1 Jahre und zum zweiten 8;5 Jahre.

4.1 Itemkennwerte

Für eine optimale Differenzierungsfähigkeit bei hoher Messeffizienz ist eine gute Passung zwischen den Aufgabenschwierigkeiten und den Personenfähigkeiten wünschenswert (Prenzel & Blum, 2007). Dazu sollte der Test über das gesamte Schuljahr hinweg alle Leistungsbereiche abdecken und möglichst zu keinem Messzeitpunkt Boden- bzw. Deckeneffekte aufweisen. Ob die Schwierigkeit der Aufgaben des „Mathes 2“ für Schülerinnen und

Schüler zweiter Klassen angemessen ist, wurde mittels der Übereinstimmung der im Rasch-Modell (s. Abschnitt 4.2) geschätzten Item- und Personenparameter geprüft (Carstensen & Taskinen, 2007). Die nachfolgende Abbildung 2 stellt die Verteilungen der Item- und Personenparameter des „Mathes 2“ zu den beiden Messzeitpunkten gegenüber.

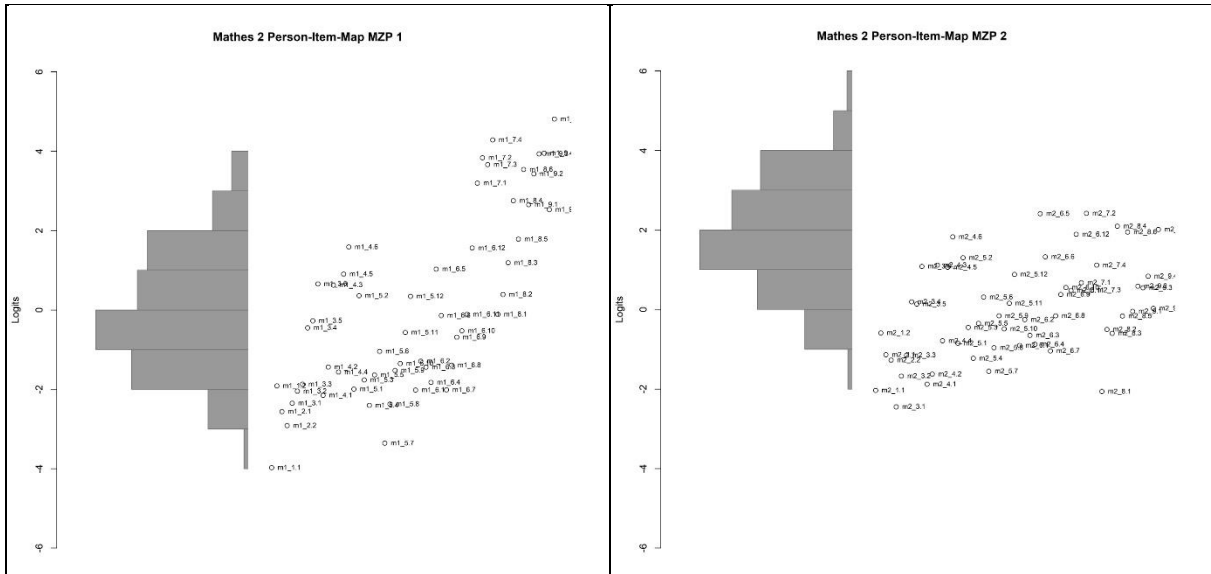


Abbildung 2: Verteilungen von Aufgabenschwierigkeiten und Schülerfähigkeiten zu den beiden Messzeitpunkten

Aus den Darstellungen geht hervor, dass „Mathes 2“ den unteren und mittleren Leistungsbereich gut bis sehr gut abdeckt. Zum ersten Messzeitpunkt ist auch noch eine ausreichende Anzahl an schwierigen Items vertreten, zum zweiten ist die Differenzierungsfähigkeit für leistungsstarke Schülerinnen und Schüler eingeschränkt.

Aus Gründen der Lesbarkeit werden die Trennschärfekoeffizienten an dieser Stelle nicht berichtet. Eine tabellarische Übersicht mit Angaben zur Schwierigkeit und Trennschärfe aller Aufgaben des „Mathes 2“ zu beiden Messzeitpunkten befindet sich im Anhang A auf Seite 21f.

4.2 Gültigkeit des Messmodells

Dem Verfahren wurde das dichotome Rasch-Modell (Rasch, 1960) als Messmodell zugrunde gelegt. Dabei wurde die Technik virtueller Personen genutzt (Rost, 2004). Die Gültigkeit des Rasch-Modells wurde durch eine Analyse lokaler Modellverletzungen eingeschätzt. Die Rasch-Analysen wurden mit dem Statistikprogramm R (R Core Team, 2013) mithilfe des Pakets pairwise (Heine, 2020) durchgeführt. Die Modellpassung der Items wurde anhand ihrer geschätzten Infit-Werte beurteilt. Da die Outfit-Statistiken deutlich durch Ausreißerwerte beeinflusst werden, Infit-Werte hingegen sensitiver im Bereich mittlerer Fähigkeitsausprägungen ausfallen (Linacre, 2002), wurden für den „Mathes 2“ in erster Linie die Infit-Statistiken auf Abweichungen vom Erwartungswert 1 untersucht. In Anlehnung an Bond und Fox (2015) zeigen bei gewöhnlichen Stichprobengrößen Werte von $0.7 \leq \text{Infit} \leq 1.3$ an, dass das Item den Annahmen des Rasch-Modells entspricht.

In der Abbildung 3 werden die Ergebnisse der Raschskalierung präsentiert. Die Infit-Werte wurden aus Gründen der Übersichtlichkeit mittels Boxplots graphisch aufbereitet, sodass die Streuung der Werte des „Mathes 2“ schnell ersichtlich ist. Es zeigt sich, dass alle Werte innerhalb des zulässigen Wertebereiches liegen, der Großteil der Items weist einen Infit in der Nähe des Erwartungswertes von 1 auf. Lediglich das zweite Item in der ersten Aufgabengruppe (Erfassung der Menge 78) sowie das fünfte Item in der sechsten Aufgabengruppe ($83 - 33 = \underline{\quad}$) weichen etwas stärker vom Erwartungswert ab.

Die Passung der Items auf das (eindimensionale) Rasch-Modell ist ein Beleg dafür, dass die Aufgaben des „Mathes 2“ dasselbe Merkmal messen, sie also eine gemeinsame Skala bilden (Bühner, 2011).

Eine Auflistung der im Rasch-Modell geschätzten Fitstatistiken zu jedem Item kann im Anhang B auf S. 23f. eingesehen werden.

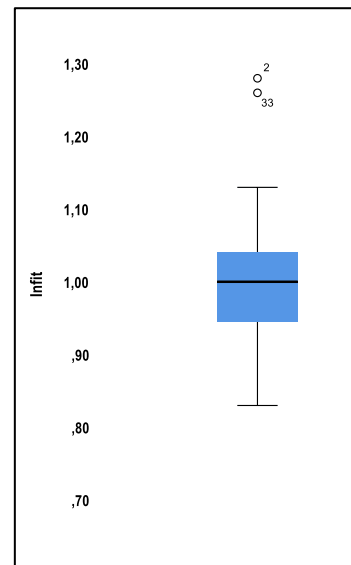


Abbildung 3: Verteilung der Itemfit-Statistiken des „Mathes 2“

4.3 Reliabilität

Zur Schätzung der Zuverlässigkeit des Verfahrens im Sinne der Internen Konsistenz wurde Cronbachs α (Cronbach, 1951) zu beiden Messzeitpunkten ermittelt (Rost, 2004). Die Alphakoeffizienten liegen mit jeweils $\alpha = .94$ (MZP 1: $N = 574$; MZP 2: $N = 578$) im hohen Bereich.

Zudem wurde die Retest-Reliabilität als Maß für die Stabilität der Messungen berechnet (Schermelleh-Engel & Werner, 2012). Dazu wurden die Testergebnisse der Schülerinnen und Schüler, welche „Mathes 2“ zu beiden Messzeitpunkten bearbeiteten (s. Kapitel 4), miteinander korreliert (Rasch, Friese, Hofmann & Naumann, 2014). Die Korrelation wurde einem Signifikanztest unterzogen. Gemäß der gängigen Konvention sollten die Zusammenhänge mindestens auf dem Niveau $\alpha = .05$ signifikant sein (ebd.). Wenn diese Bedingung erfüllt war, wurde die Höhe der Korrelation anhand der Klassifikation von Cohen (1988) eingeschätzt.

Der Zusammenhang zwischen den Testleistungen zur Mitte sowie zum Ende der zweiten Klassenstufe fällt signifikant aus und liegt mit $r_{tt} = .66^{**}$ ($N = 531$) im hohen Bereich. Die Retest-Reliabilität ist als geringer zu bewerten (Rost, 2013) als die Interne Konsistenz, was einen, wenn auch unspezifischen, Hinweis auf die Änderungssensitivität des „Mathes 2“ liefert. Ein änderungssensibler Test sollte nach Klauer (2011, 2014) nur mäßige Retest-Reliabilitäten aufweisen und die Zusammenhänge zwischen den Messungen sollten abnehmen, je mehr Zeit dazwischen liegt.

4.4 Validität

4.4.1 Konstruktvalidität

Zur Einschätzung der Konstruktvalidität wurden die Ergebnisse der Kinder im „Mathes 2“ zum zweiten Messzeitpunkt mit konstruktähnlichen und konstruktfernen Verfahren korreliert. Das Analyseverfahren ist analog dem im vorherigen Abschnitt 4.3 beschriebenen.

Die Zusammenhänge der Testergebnisse im „Mathes 2“ mit den mathematischen, rechtschriftlichen sowie Leseleistungen der Schülerinnen und Schüler werden in der Tabelle 3 dargestellt. Da beim DERET 1-2+ (Stock & Schneider, 2008) die im Diktat gemachten Fehler ausgezählt werden, bedeuten höhere Werte eine schlechtere Testleistung.

Tabelle 3: Zusammenhänge des „Mathes 2“ zu konstruktähnlichen und konstruktfernen Testverfahren

	Korrelation r mit		
	DEMAT 2+	DERET 1-2+	FE-L 2
Mathes 2	.74** $N = 548$	-.47** $N = 563$.34** $N = 566$

Anmerkungen: DEMAT 2+ – Deutscher Mathematiktest für zweite Klassen (Krajewski et al., 2004; DERET 1-2+ – Deutscher Rechtschreibtest für das erste und zweite Schuljahr (Stock & Schneider, 2008); FE-L 2 – Formative Erfassung der Lesefertigkeit im zweiten Schuljahr (Kuhlmann & Hartke, 2009b); N – Stichprobenumfang; r – Korrelationskoeffizient nach Pearson; ** – Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant

Die Zusammenhänge fallen durchweg signifikant aus. Die Korrelation zwischen „Mathes 2“ und DEMAT 2+ (Krajewski et al., 2004) ist als hoch zu klassifizieren, die anderen beiden Koeffizienten liegen nach Cohen (1988) im mittleren Bereich, wobei der Zusammenhang zum DERET 1-2+ (Stock & Schneider, 2008) erwartungsgemäß negativ ausfällt. Somit bleibt festzuhalten, dass die Zusammenhänge zu den jeweils konstruktfernen Verfahren geringer sind als zum konstruktähnlichen Verfahren, was für die Konstruktvalidität des „Mathes 2“ spricht.

4.4.2 Prognostische Validität

Vor dem Hintergrund der im Abschnitt 1.1 ausgewiesenen Zielstellungen des Verfahrens (Identifikation von Risikoschülerinnen und -schülern) ist die Vorhersagegüte des „Mathes 2“ von besonderer Bedeutung. Zur Prüfung der prognostischen Validität wurde in einem ersten Schritt der Zusammenhang der Ergebnisse im „Mathes 2“ zum Halbjahr der zweiten Klassenstufe mit denen im DEMAT 2+ (Krajewski et al., 2004) am Ende des Schuljahres ermittelt. Das Analyseverfahren ist analog dem im Abschnitt 4.3 beschriebenen. Der Zusammenhang zwischen den Testleistungen fällt signifikant aus und liegt mit $r = .71^{**}$ ($N = 520$) im hohen Bereich, was für eine große prognostische Aussagekraft des „Mathes 2“ spricht.

Als Ergänzung zu den Ergebnissen der korrelativen Analysen wurde die klassifikatorische Güte des „Mathes 2“ ermittelt (Marx & Lenhard, 2010). Zur Berechnung der Kennwerte wurde Kindern, welche im „Mathes 2“ zur Mitte der zweiten Klasse eine unterdurchschnittliche Leistung erzielten (Prozentrang < 25), ein Risikostatus in ihrer mathematischen Entwicklung zugeordnet. Dieser relativ milde Cut-Off-Wert wurde vor dem Hintergrund der

Ergebnisse jüngerer Schulleistungsstudien gewählt, auf deren Grundlage davon auszugehen ist, dass etwa ein Fünftel bis ein Viertel aller Grundschülerinnen und -schüler Lernrückstände im Umfang von mehreren Schuljahren aufweist (z. B. TIMSS 2015: 23.3 %; Selter et al., 2016; s. Abschnitt 1.2). Da ein wesentliches Ziel des Verfahrens ist, Kinder mit Schwierigkeiten beim Mathematiklernen für eine weiterführende Diagnostik und präventiv ausgerichtete Förderung zu erfassen, sollen möglichst keine falsch-negativen Klassifizierungen erfolgen. Als Kriterium diente ein Risikostatus im DEMAT 2+ (Krajewski et al., 2004), operationalisiert als ein Prozentrang < 16 am Ende der zweiten Klassenstufe. Die Tabelle 4 stellt die Befunde der klassifikatorischen Analysen dar.

Tabelle 4: Ergebnisse zur klassifikatorischen Güte des "Mathes 2" bezüglich Risiken im Fach Mathematik

	Stichproben- umfang	<i>n</i> richtig positiv	<i>n</i> falsch positiv	<i>n</i> falsch negativ	<i>n</i> richtig negativ	Sensitivität	Spezifität	α - Fehlerquote	β - Fehlerquote	positiver prä- dikativer Wert	negativer prä- dikativer Wert	RATZ-Index
DEMAT 2+	520	25	91	21	383	0.54	0.81	0.19	0.46	0.22	0.95	0.87

Anmerkungen: *n* – Anzahl; Deutscher Mathematiktest für zweite Klassen (Krajewski et al., 2004)

Die Ergebnisse zur Sensitivität zeigen, dass 54 % der Kinder mit schwachen Mathematikleistungen am Ende der zweiten Klassenstufe durch „Mathes 2“ bereits zum Halbjahr korrekt identifiziert werden. Die Spezifität gibt an, dass 81 % der Schülerinnen und Schüler mit zufriedenstellenden Mathematikleistungen im DEMAT 2+ (Krajewski et al., 2004) kein Risikostatus zugewiesen wird. Besonders erwähnenswert ist zudem der RATZ-Index, welcher mit 87 % im sehr hohen Bereich liegt und auf eine sehr gute Klassifikationsleistung des „Mathes 2“ hinweist (Marx & Lenhard, 2000).

5 Literaturverzeichnis

- Bond, T. G. & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York & London: Routledge.
- Bos, W., Bensen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (Hrsg.) (2008). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Lankes, E.M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (Hrsg.) (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Wendt, H., Köller, O. & Selter, C. (Hrsg.) (2012). *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. 3., aktualisierte und erweiterte Auflage. München: Pearson Studium.
- Carstensen, C.H. & Taskinen, P. (2007). Feldtest. In M. Prenzel & W. Blum (Hrsg.), *Entwicklung eines Testverfahrens zur Überprüfung der Bildungsstandards in Mathematik für den Mittleren Schulabschluss. Technischer Bericht* (S. 16-23). Kiel: IPN.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Erlbaum.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Dilling H., Mombour, W. & Schmidt, M.H. (2011). *Internationale Klassifikation psychischer Störungen. ICD-10 Kapitel 5 (F). Klinisch-diagnostische Leitlinien*. Bern: Huber.
- Flexer, R.J. (1986). The Power of Five: The Step Before the Power of Ten. *The Arithmetic Teacher*, 34 (3), 5-9.
- Fritz, A., Ricken, G. & Gerlach, M. (2007). *Kalkulie. Handreichung zur Durchführung der Diagnose*. Berlin: Cornelsen.
- Gaidoschik, M. (2010). *Die Entwicklung von Lösungsstrategien zu den additiven Grundaufgaben im Laufe des ersten Schuljahres*. Dissertation. Universität Wien. Abruf am 28.4.20. Online verfügbar unter: http://othes.univie.ac.at/9155/1/2010-01-18_8302038.pdf
- Gaidoschik, M. (2011). *Rechenschwäche – Dyskalkulie. Eine unterrichtspraktische Einführung für LehrerInnen und Eltern*. 6. Auflage. Hamburg: Persen.
- Hartke, B. & Diehl, K. (2013). *Schulische Prävention im Bereich Lernen. Problemlösungen mit dem RTI-Ansatz*. Stuttgart: Kohlhammer.
- Häsel-Weide, U., Nührenböcker, M., Moser Opitz, E. & Wittich, C. (2019). *Ablösung vom zählenden Rechnen. Förderseinheiten für heterogene Lerngruppen*. 5. Auflage. Seelze: Klett-Kallmeyer.
- Hasemann, K. & Gasteiger, H. (2014). *Anfangsunterricht Mathematik*. 3., überarbeitete und erweiterte Auflage. Berlin: Springer.
- Hasselhorn, M., Marx, H. & Schneider, W. (Hrsg.) (2005). *Diagnostik von Mathematikleistungen. Tests und Trends N.F. Bd. 4*. Göttingen: Hogrefe.

- Heine, J.H. (2020). *pairwise: Rasch Model Parameters by Pairwise Algorithm*. R package version 0.4.4-7. Abruf am 28.4.20. Online verfügbar unter: <http://CRAN.R-project.org/package=pairwise>
- Jacobs, C. & Petermann, F. (2012). *Diagnostik von Rechenstörungen. 2., überarbeitete und erweiterte Auflage*. Göttingen: Hogrefe.
- Käpnick, F. (2014). *Mathematiklernen in der Grundschule*. Berlin: Springer.
- Klauer, K.J. (2011). Lernverlaufsdiagnostik – Konzept, Schwierigkeiten und Möglichkeiten. *Empirische Sonderpädagogik*, 3, 207-224.
- Klauer, K.J. (2014). Formative Leistungsdiagnostik: Historischer Hintergrund und Weiterentwicklung zur Lernverlaufsdiagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik. Tests und Trends, N.F. Bd. 12* (S. 1-18). Göttingen: Hogrefe.
- Koch, K. (2005). Probleme im Bereich des mathematischen Lernens. In S. Ellinger & M.C. Wittrock (Hrsg.), *Sonderpädagogik in der Regelschule. Konzepte, Forschung, Praxis* (S. 279-298). Stuttgart: Kohlhammer.
- Koch, K. & Knopp, E. (2010). Mathematisches Lernen. In B. Hartke, K. Koch & K. Diehl (Hrsg.), *Förderung in der schulischen Eingangsstufe* (S. 91-118). Stuttgart: Kohlhammer.
- Krajewski, K. (2003). *Vorhersage von Rechenschwäche in der Grundschule*. Hamburg: Dr. Kovač.
- Krajewski, K., Liehm, S., Schneider, W. (2004). *Deutscher Mathematiktest für zweite Klassen (DEMAT 2+)*. Göttingen: Beltz Test GmbH.
- Krauthausen, G. (2018): *Einführung in die Mathematikdidaktik – Grundschule. 4. Auflage*. Berlin: Springer.
- Kuhlmann, K. & Hartke, B. (2009a). *Formative Erfassung der arithmetischen Fähigkeiten im 2. Schuljahr (FE-AF 2)*. Rostock: Universität Rostock.
- Kuhlmann, K. & Hartke, B. (2009b). *Formative Erfassung der Lesefertigkeit im zweiten Schuljahr (FE-L 2)*. Rostock: Universität Rostock.
- Kultusministerkonferenz (2005). *Bildungsstandards im Fach Mathematik für den Primarbereich*. München: Luchterhand.
- Lambert, K. (2015). *Rechenschwäche. Grundlagen, Diagnostik und Förderung*. Göttingen: Hogrefe.
- Landerl, K. & Kaufmann, L. (2008). *Dyskalkulie. Modelle, Diagnostik, Intervention*. München: Reinhardt.
- Linacre, J.M. (2002). What do Infit and Outfit, Mean-square and Standardized mean? *Rasch Measurement Transactions*, 16 (2), 878. Abruf am 28.4.20. Verfügbar unter: <http://www.rasch.org/rmt/rmt162f.htm>
- Lorenz, J.H. (2005). *Lernschwache Rechner fördern. 2. Auflage*. Berlin: Cornelsen.
- Lorenz, J.H. (2014). Rechenschwäche. In G.W. Lauth, M. Grünke & J.C. Brunstein (Hrsg.), *Interventionen bei Lernstörungen. 2., überarbeitete und erweiterte Auflage* (S. 43-55). Göttingen: Hogrefe.
- Marx, H., Jansen, H. & Skowronek, H. (2000). Prognostische, differentielle und konkurrente Validität des Bielefelder Screenings zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC). In M. Hasselhorn, W. Schneider & H. Marx (Hrsg.), *Diagnostik von Lese-Rechtschreibschwierigkeiten. Tests und Trends, N.F. Bd. 1* (S. 9-34). Göttingen: Hogrefe.
- Marx, H. & Lenhard, W. (2010). Diagnostische Merkmale von Screeningverfahren. In M. Hasselhorn & W. Schneider (Hrsg.), *Frühprognose schulischer Kompetenzen. Tests und Trends, N.F. Bd. 9* (S. 68-84). Göttingen: Hogrefe.

- Moser Opitz, E. (2004). Dyskalkulie: Krankheit, Erfindung, Mythos, Etikett ... ? Auseinandersetzung mit einem geläufigen, aber ungeklärten Begriff. *Vierteljahrszeitschrift für Heilpädagogik und ihre Nachbargebiete*, 72, 179-190.
- Padberg, F. & Benz, C. (2011). *Didaktik der Arithmetik für Lehrerbildung und Lehrerfortbildung*. 4. erweiterte, stark überarbeitete Auflage. Berlin: Springer.
- Prenzel, M. & Blum, W. (Hrsg.) (2007). *Entwicklung eines Testverfahrens zur Überprüfung der Bildungsstandards in Mathematik für den Mittleren Schulabschluss. Technischer Bericht*. Kiel: IPN.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Abruf am 28.4.20 Online verfügbar unter: <http://www.R-project.org/>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen & Lydiche.
- Rasch, B., Frieze, M., Hofmann, W. & Naumann, E. (2014). *Quantitative Methoden 1. Einführung in die Statistik für Psychologen und Sozialwissenschaftler*. 4., überarbeitete Auflage. Berlin: Springer.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. 2. vollständig überarbeitete und erweiterte Auflage. Bern: Huber.
- Rost, D.H. (2013). *Interpretation und Bewertung pädagogisch-psychologischer Studien*. 3., vollständig überarbeitete und erweiterte Auflage. Bad Heilbrunn: Klinkhardt.
- Schermelleh-Engel, K. & Werner, C.S. (2012). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion*. 2., aktualisierte und überarbeitete Auflage (S. 119-142). Berlin: Springer.
- Selter, C., Walter, D., Walther, G. & Wendt, H. (2016). Mathematische Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 79-136). Münster: Waxmann.
- Sikora, S. (2017). *Lernverlaufdiagnostik im Mathematikunterricht. Theoretische Grundlagen, Konzeption und Güte eines formativen Schulleistungstests für dritte Klassen*. Hamburg: Dr. Kovač.
- Sikora, S. & Voß, S. (2018). *Mathematikunterricht in der inklusiven Grundschule*. Stuttgart: Kohlhammer.
- Stock, C. & Schneider, W. (2008). *Deutscher Rechtschreibtest für das dritte und vierte Schuljahr (DERET 3-4+)*. Göttingen: Hogrefe.
- Von Aster, M., Schweiter, M. & Weinhold Zulauf, M. (2007). Rechenstörungen bei Kindern. Vorläufer, Prävalenz und psychische Symptome. *Zeitschrift für Pädagogische Psychologie*, 39, 85-96.
- Wittmann, E.Ch. & Müller, G.N. (2012). *Das Zahlenbuch 1 (Neubearbeitung). Begleitband*. Stuttgart: Klett.

6 Anhang

Anhang A: Schwierigkeiten und Trennschärfen der Items

Item	Schwierigkeit P_i zum Messzeitpunkt		Trennschärfe r_{pbis} zum Messzeitpunkt	
	1	2	1	2
1.1	0.96	0.96	0.18	0.15
1.2	0.84	0.88	0.19	0.23
2.1	0.90	0.92	0.21	0.17
2.2	0.91	0.93	0.29	0.23
3.1	0.86	0.97	0.38	0.19
3.2	0.82	0.94	0.43	0.25
3.3	0.80	0.92	0.44	0.27
3.4	0.57	0.77	0.61	0.52
3.5	0.53	0.78	0.63	0.54
3.6	0.39	0.63	0.59	0.53
4.1	0.82	0.95	0.44	0.28
4.2	0.73	0.93	0.51	0.36
4.3	0.40	0.63	0.66	0.56
4.4	0.75	0.88	0.50	0.39
4.5	0.35	0.64	0.63	0.58
4.6	0.27	0.51	0.64	0.59
5.1	0.81	0.89	0.43	0.34
5.2	0.43	0.60	0.55	0.60
5.3	0.79	0.85	0.42	0.42
5.4	0.84	0.90	0.48	0.46
5.5	0.76	0.84	0.53	0.47
5.6	0.67	0.75	0.56	0.55
5.7	0.93	0.94	0.36	0.22
5.8	0.86	0.90	0.41	0.33
5.9	0.75	0.82	0.51	0.44
5.10	0.72	0.85	0.54	0.49
5.11	0.59	0.77	0.62	0.54
5.12	0.45	0.67	0.66	0.53
6.1	0.80	0.89	0.51	0.40
6.2	0.71	0.82	0.57	0.50
6.3	0.74	0.86	0.48	0.50
6.4	0.78	0.88	0.58	0.47
6.5	0.31	0.40	0.49	0.53
6.6	0.52	0.60	0.63	0.59
6.7	0.81	0.90	0.48	0.41
6.8	0.73	0.81	0.51	0.51
6.9	0.62	0.74	0.58	0.56
6.10	0.59	0.72	0.57	0.52
6.11	0.51	0.73	0.56	0.57
6.12	0.27	0.50	0.59	0.58

7.1	0.08	0.69	0.40	0.40
7.2	0.05	0.33	0.27	0.30
7.3	0.05	0.72	0.35	0.45
7.4	0.04	0.61	0.33	0.40
8.1	0.48	0.95	0.51	0.29
8.2	0.40	0.85	0.51	0.39
8.3	0.29	0.87	0.53	0.34
8.4	0.12	0.44	0.48	0.52
8.5	0.18	0.82	0.38	0.32
8.6	0.08	0.47	0.46	0.54
9.1	0.10	0.80	0.41	0.47
9.2	0.06	0.71	0.37	0.58
9.3	0.05	0.71	0.41	0.55
9.4	0.05	0.67	0.41	0.54
9.5	0.09	0.79	0.37	0.45
9.6	0.01	0.45	0.18	0.52

Anhang B: Itemkennwerte der Rasch-Analyse

Item	Schwierigkeit	Infit	Item	Schwierigkeit	Infit
1.1	-2.80	0.90	6.1	-1.37	0.94
1.2	-1.16	1.28	6.2	-0.69	0.96
2.1	-1.73	1.08	6.3	-0.96	0.97
2.2	-1.94	1.03	6.4	-1.26	0.90
3.1	-2.22	1.04	6.5	1.80	1.26
3.2	-1.74	1.02	6.6	0.69	1.09
3.3	-1.41	1.04	6.7	-1.44	0.97
3.4	-0.03	0.97	6.8	-0.69	1.02
3.5	0.04	0.93	6.9	-0.07	1.00
3.6	0.97	1.04	6.10	0.10	1.07
4.1	-1.91	1.03	6.11	0.27	1.00
4.2	-1.37	1.02	6.12	1.78	1.03
4.3	0.96	0.99	7.1	1.55	1.00
4.4	-1.08	1.05	7.2	2.79	1.13
4.5	1.06	0.93	7.3	1.56	0.94
4.6	1.77	0.98	7.4	1.93	1.01
5.1	-1.32	1.04	8.1	-0.58	1.06
5.2	0.92	1.08	8.2	0.11	1.02
5.3	-1.00	1.09	8.3	0.42	0.97
5.4	-1.73	0.98	8.4	2.36	0.99
5.5	-0.88	1.00	8.5	0.85	1.06
5.6	-0.26	1.05	8.6	2.43	0.92
5.7	-2.28	0.90	9.1	1.16	0.91
5.8	-1.54	1.00	9.2	1.64	0.83
5.9	-0.73	1.06	9.3	1.63	0.83
5.10	-0.84	0.98	9.4	1.80	0.85
5.11	-0.11	0.95	9.5	1.18	0.94
5.12	0.70	0.99	9.6	2.66	0.90

Anhang C: Normtabelle des „Mathes 2“

Rohwert	Mitte Klasse 2 (20. / 21. Schulwoche) N = 574		Anfang Klasse 3 (3. / 4. Schulwoche) N = 578	
	T-Wert	Prozentrang	T-Wert	Prozentrang
0	<24	0	<15	0
1	24	0	15	0
2	25	0	15	0
3	26	1	15	0
4	27	1	15	0
5	28	1	15	0
6	29	2	15	0
7	30	2	15	0
8	31	3	16	0
9	32	4	17	1
10	32	5	17	1
11	33	6	19	1
12	34	7	20	1
13	35	8	21	1
14	36	10	22	2
15	37	12	23	2
16	38	14	24	2
17	39	16	25	2
18	40	19	26	3
19	41	21	27	3
20	42	23	28	4
21	43	26	29	4
22	43	29	30	5
23	44	32	31	6
24	45	35	32	7
25	46	38	33	9
26	47	42	34	9
27	48	45	35	10
28	49	49	36	12
29	50	51	37	12
30	51	54	38	14
31	52	57	39	16
32	53	59	40	18
33	54	62	41	19
34	55	65	42	20
35	55	68	43	22
36	56	72	43	24
37	57	76	44	27
38	58	78	45	29
39	59	80	46	31
40	60	84	47	34
41	61	87	48	37

42	62	89	49	40
43	63	92	50	43
44	64	93	51	48
45	65	95	52	51
46	66	95	53	56
47	67	97	54	59
48	67	97	55	64
49	68	99	56	68
50	69	99	57	72
51	69	99	58	77
52	71	99	59	82
53	72	100	60	88
54	73	100	61	94
55	73	100	62	98
56	73	100	63	100