

Mathes 3

Formative Erfassung der mathematischen
Kompetenzen von Drittklässlern

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 Was wird von Schülerinnen und Schüler am Ende der Grundschulzeit erwartet?.....	6
2.2 Konstruktion inhaltsvalider Testanforderungen.....	7
2.3 Itempilottierung und -auswahl.....	9
2.4 Überlegungen zur Steigerung der Testökonomie.....	11
3 Testanwendung	11
3.1 Anwendungszeitraum und Zielgruppe	11
3.2 Testmaterial	12
3.3 Hinweise zur Testdurchführung	12
3.4 Auswertung und Interpretation.....	13
3.4.1 Manuelle Auswertung.....	13
3.4.2 Automatisierte Auswertung.....	13
3.4.3 Interpretation der Ergebnisse	13
4 Testgütekriterien	14
4.1 Itemkennwerte	15
4.2 Skalenstruktur.....	16
4.3 Reliabilität.....	17
4.4 Validität	18
4.4.1 Konstruktvalidität.....	18
4.4.2 Prognostische Validität	20
4.5 Gültigkeit des Messmodells	21
4.6 Änderungssensitivität	22
5 Literaturverzeichnis	24
6 Anhang	28

Abbildungsverzeichnis

Abbildung 1: Allgemeine und inhaltsbezogene mathematische Kompetenzen im Mathematikunterricht der Grundschule (nach KMK, 2005, S. 7)	7
Abbildung 2: Referenzniveaus als Interpretationshilfen für die erzielte Testleistung	14
Abbildung 3: Verteilungen von Aufgabenschwierigkeiten und Kompetenzwerten zu den vier Messzeitpunkten ...	16
Abbildung 4: Verteilung der Itemfit-Statistiken (<i>wMNSQ</i>) der vier Testskalen des „Mathes 3“	22

Tabellenverzeichnis

Tabelle 1: inhaltliche Anforderungen des "Mathes 3"	10
Tabelle 2: Durchführungszeiträume des "Mathes 3"	11
Tabelle 3: Modellfit-Indizes der konfirmatorischen Faktorenanalysen	17
Tabelle 4: Ergebnisse der Analysen hinsichtlich der EAP/PV-Reliabilität der Skalen des "Mathes 3"	17
Tabelle 5: Ergebnisse der Analysen hinsichtlich der Internen Konsistenz (Cronbachs α) der Subtests des "Mathes 3"	18
Tabelle 6: Zusammenhänge der jeweiligen Testskala des „Mathes 3“ zu konstruktähnlichen Testverfahren	19
Tabelle 7: Zusammenhänge der jeweiligen Testskala des „Mathes 3“ zu konstruktfernen Testverfahren.....	19
Tabelle 8: Zusammenhänge der jeweiligen Testskala des „Mathes 3“ mit zukünftigen Mathematikleistungen.....	20
Tabelle 9: Ergebnisse zur klassifikatorischen Güte des "Mathes 3" bezüglich Risiken im Fach Mathematik	21
Tabelle 10: Deskriptive Statistiken der Skalen des „Mathes 3“ über die vier Messzeitpunkte	23
Tabelle 11: Ergebnisse der inferenzstatistischen Prüfung der Mittelwertunterschiede über die vier Messzeitpunkte	23

Kurzinformation

Titel, Autoren, Jahr	Mathes 3, Simon Sikora & Bodo Hartke, 2012
Diagnostische Zielsetzung	Erfassung der Entwicklung der mathematischen Kompetenzen im Verlauf der dritten Klassenstufe bei allen Kindern
Anwendungsbereiche	Grundschulmathematikunterricht in der dritten Klasse
Aufbau	vier inhaltlich parallele Testformen A, B, C und D mit jeweils etwa 60 Aufgaben, verteilt auf vier Testskalen (Zahlen und Operationen, Größen und Messen, Daten, Häufigkeit und Wahrscheinlichkeit, Raum und Form)
Anwendungszeitraum	dritte Klassenstufe: 10./11. Schulwoche (vor den Herbstferien), 20./21. Schulwoche (vor den Winterferien), 30./31. Schulwoche (nach den Osterferien) vierte Klassenstufe: 3./4. Schulwoche (Anfang Klasse 4)
Durchführung	Gruppentestung im Klassenkontext durch die Lehrkraft bei freier Testbearbeitung innerhalb einer Unterrichtsstunde
Auswertung	skalenweise Auswertung mithilfe von Auswertungsvorlagen
Normen	Quotenstichprobe Rostock, N pro Testzeitpunkt zwischen 451 und 492 Repräsentativität nachgewiesen durch Vergleiche der VERA-Leistungen der Region Rostock mit Durchschnitt M-V sowie Bundesdurchschnitt
Reliabilität	Interne Konsistenz (EAP/PV-Reliabilität) der testheftübergreifenden Skalen, $N = 432 - 479$: Zahlen und Operationen: $EAP/PV = .83$ Größen und Messen: zwischen $EAP/PV = .72$ und $EAP/PV = .74$ Daten, Häufigkeit und Wahrscheinlichkeit: zwischen $EAP/PV = .69$ und $EAP/PV = .74$ Raum und Form: zwischen $EAP/PV = .71$ und $EAP/PV = .80$
Validität	Konstruktvalidität: konstruktnahe Korrelationen (mit KEKS Mathematik, DEMAT 3+): zwischen $r = .41^{**}$ und $r = .70^{**}$ (MW: $r = .56$) ($N = 366-427$) konstruktferne Korrelationen (mit DERET 3-4+, HSP 3, CFT 20-R): zwischen $r = .25^*$ und $r = .58^{**}$ (MW: $r = .41$) ($N = 372-426$) prognostische Validität: RATZ-Indizes Mathes 3 (2. Halbjahr Klasse 3) zur Zeugnisnote (Ende Klasse 4): 0.81 ($N = 306$); zur Erreichung der Bildungsstandards (Ende Klasse 4): 0.77 ($N = 353$)
Änderungssensitivität	durchgängiger Anstieg der Mittelwerte der Personenparameter (WLE) in allen Testskalen ANOVA mit Messwiederholung: pro Schuljahr zwischen $d = .69$ und $d = 1.14$ ($N = 298$)

1. Theoretische Grundlagen

1.1 Zielstellung der Mathes-Testreihe

Taglich stellen sich Lehrkrafte Fragen wie die folgenden:

„Lernen alle Kinder in meinem Unterricht erfolgreich oder kommt jemand nicht mit?“
„Kann ich das aktuelle Themengebiet abschlieen und im Stoff weitermachen oder brauchen die Schulerinnen und Schuler noch mehr Lernzeit?“
„Was genau hat die Schulerin bzw. der Schuler noch nicht verstanden?“

Solche Fragen haben das Ziel, den Unterricht bestmoglich an die individuellen Lernausgangslagen und Forderbedurfnisse der Kinder anzupassen. Die Mathes-Testreihe soll Lehrkrafte bei vertretbarem Aufwand dabei unterstutzen, zu einer prazisen Einschatzung der aktuellen Lernstande sowie der Leistungsentwicklung der Schulerinnen und Schuler zu gelangen. Die Testverfahren konnen in dreierlei Hinsicht helfen, namlich

- beim genauen Einschatzen des Spektrums der Leistungen der Schulerinnen und Schuler, um den Unterricht daran bestmoglich anpassen zu konnen,
- beim rechtzeitigen Erkennen derjenigen Schulerinnen und Schuler mit Risiken bzw. bereits ausgepragten Schwierigkeiten im Kompetenzerwerb sowie
- bei der Planung effektiver Fordermanahmen, insbesondere fur diejenigen Schulerinnen und Schuler mit besonderen Unterstutzungsbedarfen.

Zu diesem Zweck wurden Mathes-Tests fur jede Klassenstufe entwickelt, welche jeweils zum Beginn und in der Mitte des Schuljahres eingesetzt werden konnen.

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

Schwierigkeiten beim Rechnen- bzw. mathematischen Lernen gehoren fur viele Schulerinnen und Schuler (und deren Lehrerinnen und Lehrer) zum schulischen Alltag. Allerdings erhalt nicht jedes Kind mit schwachen Leistungen in Mathematik die Diagnose „Dyskalkulie“ bzw. „Rechenstorung“. Nach den Kriterien der Weltgesundheitsorganisation leidet ein Kind nur dann unter einer Dyskalkulie, wenn seine Beeintrachtigung 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

2.1 Was wird von Schülerinnen und Schülern am Ende der Grundschulzeit erwartet?

Im Zusammenhang mit der Fragestellung, was Schülerinnen und Schüler bis zum Ende der Grundschulzeit lernen sollen, trat der Begriff der Kompetenz in den Fokus pädagogischer Überlegungen (Klieme et al., 2007; Blum, 2010). Die Kompetenz eines Individuums umfasst nach Weinert (2001) zusammenwirkende Aspekte wie Wissen, Fähigkeiten, Verständnis, Können, Handeln, Erfahrungen und Motivation, welche eine Person befähigen, konkrete

Anforderungen zu bewältigen. Dieses Verständnis spiegelt sich in den Rahmenplänen aller deutschen Bundesländer wider, welche sich an den länderübergreifenden Bildungsstandards (KMK, 2005) orientieren. Die Standards für den Primarbereich im Fach Mathematik beschreiben die mathematischen Kompetenzen, die Schülerinnen und Schüler am Ende der vierten Klassenstufe erreicht haben sollen, wobei zwischen inhaltsbezogenen und allgemeinen, eher prozessbezogenen, Kompetenzen unterschieden wird. Beide Kompetenzfelder sind in enger Verbindung zu sehen, wie die nachfolgende Abbildung 1 veranschaulicht.

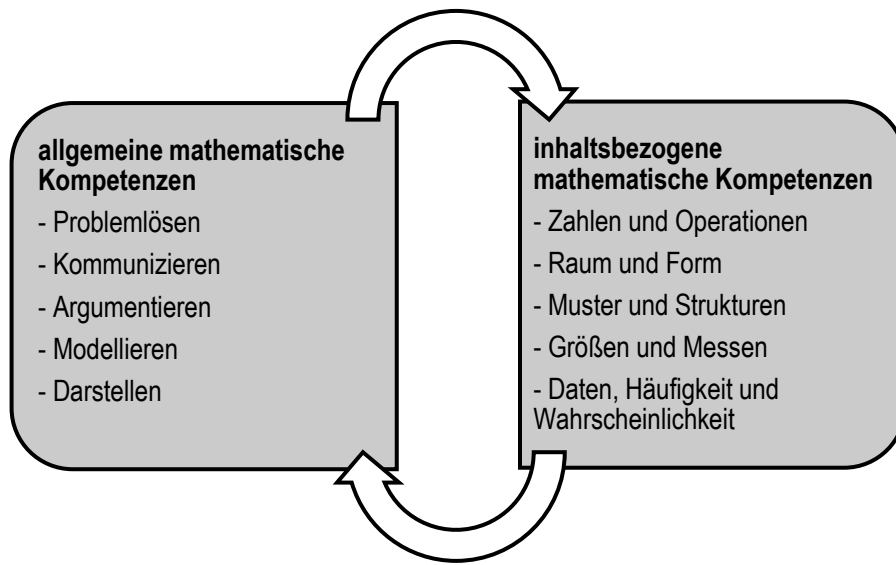


Abbildung 1: Allgemeine und inhaltsbezogene mathematische Kompetenzen im Mathematikunterricht der Grundschule (nach KMK, 2005, S. 7)

Aus den dem Kompetenzbegriff inhärenten Leistungen erwachsen neue Erwartungen und Anforderungen an den modernen Mathematikunterricht. Die Entwicklung von Kompetenzen „hängt nicht nur davon ab, welche Inhalte unterrichtet wurden, sondern in mindestens gleichem Maße davon, wie sie unterrichtet wurden“ (KMK, 2005, S. 6; Hervorhebungen im Original). Die Standards sollen einen Unterricht unterstützen, der „nicht auf die Aneignung von Kenntnissen und Fertigkeiten reduziert“ (KMK, 2005, S. 6) ist, sondern vielmehr auf „die Entwicklung eines gesicherten Verständnisses“ (ebd.; Hervorhebung im Original) abzielt.

Um die in Abschnitt 1.1 genannten Ziele für Lehrkräfte erfüllen zu können, müssen Mathematiktests am Ende der Grundschulzeit abbilden, welche inhalts- und prozessbezogenen Kompetenzen die Schülerinnen und Schüler bereits erworben haben und welche noch nicht.

2.2 Konstruktion inhaltsvalider Testanforderungen

Die Bildungsstandards der KMK (2005) beinhalten eine Zielvorgabe des Mathematikunterrichts in der Grundschule, welche ebenfalls den aktuellen Stand der mathematikdidaktischen Fachdiskussion widerspiegelt (u.a. Walther, van den Heuvel-Panhuizen et al., 2011; Grassmann et al., 2014). Kritik wurde allerdings an der Leitidee Muster und Strukturen geübt. Wittmann und Müller stellen heraus, dass es sich bei diesem Bereich „nicht einfach um einen Aspekt des Mathematikunterrichts unter anderen handelt, sondern dass dieser Aspekt grundlegend ist“ (2011,

S. 42). In Anlehnung an Devlin (1998) können Muster und Strukturen gar als das Wesen der Mathematik gesehen werden. Folglich ist die Berücksichtigung mathematischer Gesetzmäßigkeiten in den Bildungsstandards keineswegs falsch, vielmehr wird ihr Stellenwert höher eingeschätzt.

Die Bildungsstandards der KMK (2005) sind damit auch aus wissenschaftlicher Perspektive ein geeigneter Referenzpunkt für die Erarbeitung eines formativen Schulleistungstests für das Fach Mathematik in dritten Klassen. Die in den Standards formulierten Leitideen liefern einen sinnvollen inhaltlichen Rahmen für ein solches Verfahren: Indem regelmäßig Einblick in die Entwicklung in den inhaltlichen Schwerpunkten der dritten Klassenstufe genommen wird, kann der „Weg“ des Kompetenzerwerbs bezogen auf Zahlen und Operationen, Raum und Form, Größen und Messen sowie Daten, Häufigkeit und Wahrscheinlichkeit verfolgt werden. Zu jedem Messzeitpunkt kann so ein Leistungsprofil der Schülerinnen und Schüler erstellt werden, aus dem leicht ersichtlich ist, in welchem Inhaltsbereich eine gezielte Förderung am dringlichsten wäre, aber auch, sofern das Verfahren mehrfach eingesetzt wird, wie die Entwicklung im jeweiligen Bereich verläuft.

Diese Überlegungen werden durch empirische Erkenntnisse zur Struktur mathematischer Kompetenzen im späten Grundschulalter gestützt. Winkelmann, Robitzsch, Stanat und Köller (2012) untersuchten die fachspezifischen Kompetenzen anhand von zwei Datensätzen mit $N = 10\,328$ und $N = 6\,638$ Dritt- und Viertklässlern, die bei der Entwicklung von Messinstrumenten zur Überprüfung der Bildungsstandards im Fach Mathematik in der Grundschule generiert wurden. Ihre Dimensionsanalysen belegen, dass die Daten am besten durch ein Modell mit fünf inhaltsbezogenen Faktoren entsprechend den Leitideen der Bildungsstandards (KMK, 2005) abgebildet werden.

Die Bildungsstandards der KMK (2005) bilden nicht nur einen sinnvollen inhaltlichen Rahmen für einen Test zur Lernverlaufsdiagnostik, sondern legen darüber hinaus implizit Anforderungen an die Aufgaben eines solchen Verfahrens fest. Die Aufgaben dürfen somit nicht nur die Wiedergabe von Kenntnissen oder ein rezeptartiges Nachmachen und Ausführen von Algorithmen erfordern, sondern sollten komplexere Probleme an die Schülerinnen und Schüler stellen, deren Lösungen der Aktivierung allgemeiner mathematischer Kompetenzen bedürfen. Dennoch darf nicht unerwähnt bleiben, dass auch technische Grundfertigkeiten wie Rechenkompetenzen oder Zahlvorstellungen in dem Verfahren enthalten sein müssen. Schließlich kann eine komplexere Aufgabenlösung auch an der Reproduktion von Wissen, Fertigkeiten und Grundvorstellungen scheitern. Ein förderrelevantes Verfahren muss demnach auch solche Basiskompetenzen umfassen, aber eben nicht ausschließlich.

Da die Bildungsstandards der KMK (2005) diejenigen Kompetenzen ausweisen, welche von Schülerinnen und Schüler *am Ende der Grundschulzeit* erwartet werden, mussten zunächst diejenigen Inhalte herausgearbeitet werden, die *in der dritten Klassenstufe* zu behandeln sind. Dazu wurde die mathematikdidaktische Fachliteratur umfassend analysiert. Die Ergebnisse dieser Analyse werden bei Sikora (2017) ausführlich dargestellt. Zudem wurden, wenn möglich, die Testinhalte an fundamentalen Ideen der Mathematik (zusammenfassend Vohns, 2007) ausgerichtet. Diese mathematischen Grundideen sind die zentralen, fachtypischen „Fragestellungen, Bezeichnungsweisen, Strukturierungs- und Begründungsformen“ (Führer, 1997, S. 83), welche sich wie ein roter Faden durch das mathematische Denken ziehen. Für die Grundschulmathematik existieren zum aktuellen

Zeitpunkt zwei etablierte Aufstellungen fundamentaler Ideen, und zwar von Wittmann (1995, 1999) bzw. Wittmann und Müller (2012) sowie Winter (2001). Beide wurden bei der Entwicklung der Testaufgaben berücksichtigt.

Zusammenfassend ergeben sich aus den dargestellten mathematikdidaktischen Überlegungen folgende Konsequenzen für einen formativ ausgerichteten Schulleistungstest bei Schülerinnen und Schülern dritter Klassen:

- Zuordnung der Aufgaben zu Testskalen nach Leitideen der Bildungsstandards der KMK (2005)
- Integration des Kompetenzbereichs Muster und Strukturen in die anderen Testskalen
- Konstruktion von Aufgaben, deren Lösung für vorhandene inhaltsbezogene und allgemeine mathematische Kompetenzen gemäß den Bildungsstandards der KMK (2005) sprechen
- Konstruktion von Aufgaben, welche mathematische Basiskompetenzen erfassen
- Konstruktion von Aufgaben, welche die fundamentalen Ideen der Mathematik berücksichtigen und so das Verständnis der zugrundeliegenden mathematischen Konzepte und Strukturen abprüfen

2.3 Itempilotierung und -auswahl

Insgesamt wurden 308 Items konstruiert und auf vier Itempools verteilt. Alle Items wurden so formuliert, dass nur zwei Ausprägungen (richtig oder falsch) möglich sind. Die entwickelten Aufgaben wurden mit allen Drittklässlerinnen und Drittklässlern Rügener Grundschulen (zwischen $N = 364$ und $N = 400$) zu vier Messzeitpunkten im Abstand von etwa 10 Schulwochen im Schuljahr 2011/12 pilotiert. Dabei bearbeitete jedes Kind zu jedem Messzeitpunkt zwei der Itempools innerhalb von zwei Unterrichtsstunden. Der Datenausfall aufgrund von Nichtteilnahme an einem Messzeitpunkt, z. B. durch Krankheit, war relativ gering und unsystematisch. 327 Drittklässlerinnen und Drittklässler nahmen an allen vier Erhebungen der Untersuchung teil.

Mit den so gewonnenen Daten wurden die Items einer kritischen Prüfung unterzogen. Unpassende Aufgaben wurden in Anlehnung an Prenzel und Blum (2007) in einem stufenweisen Prozess nach verschiedenen Kriterien (fehlerhafte Items, zu hoher [$< 5\%$ Lösungsrate] oder niedriger Schwierigkeitsgrad [$> 95\%$ Lösungsrate], ungenügender Modellfit [weighted Mean Square $wMNSQ < 0.7$ bzw. $wMNSQ > 1.3$], geringe Trennschärfe [$r_{pbis} < .20$], durch Befragungsergebnisse der Lehrkräfte als ungeeignet eingeschätzt) ausgeschlossen.

Insgesamt wurden 178 der pilotierten Items in das endgültige Verfahren mit dem Titel „Mathes 3“ aufgenommen und auf vier Testformen so verteilt, dass die wesentlichen Inhalte aus jedem Kompetenzbereich abgedeckt werden. Dabei wurde ein Multi-Matrix-Design (Mislevy, Beaton, Kaplan & Sheehan, 1992) verwendet, sodass die Ergebnisse in den vier Testformen vergleichbar sind. In Übereinstimmung mit den Empfehlungen von Kolen und Brennan (2004) wurden in jeden Subtest anteilmäßig etwa 20 % Anker-Items mit verschiedenen Schwierigkeitsgraden aufgenommen. Jede Form des Verfahrens erhielt etwa 60 Items, sodass der Test innerhalb einer Schulstunde durchführbar ist.

Durch das gewählte Testdesign mit Anker-Items resultieren folgende Vorteile für die Durchführung des „Mathes 3“ im Klassenverband:

- Bei Klassentestungen können nebeneinandersitzenden Schülerinnen und Schülern verschiedene Testformen des „Mathes 3“ dargeboten werden, sodass Abschreibeffekte ausgeschlossen werden können.
- Das auf Klassenebene gemessene Konstrukt ist somit umfangreicher (höhere inhaltliche Validität), da mehr Facetten durch die Mehrzahl von Aufgaben erfasst werden können.
- Bei wiederholten Messungen können einem Kind verschiedene Testformen des „Mathes 3“ dargeboten werden, sodass Erinnerungseffekte ausgeschlossen werden können.
- Bei wiederholten Messungen ist dementsprechend auch das auf Schülerebene gemessene Konstrukt umfangreicher (höhere inhaltliche Validität der Lernverlaufsmessungen).

Die Tabelle 1 ordnet die inhaltlichen Anforderungen des „Mathes 3“ den vier Testformen/-versionen A bis D zu.

Tabelle 1: inhaltliche Anforderungen des "Mathes 3"

	Aufgabengruppe bzw. Anforderung	Testform des „Mathes 3“			
Zahlen und Operationen	Stellentafel	A			
	Zahlenstrahlen		B		D
	Hundertertafel			C	
	formales Rechnen (Kopfrechnen)	A	B	C	D
	Halbieren	A		C	
	Verdoppeln			C	
	Reihen ergänzen		B		
	Vergleiche von Termen			C	
	arithmetische Muster				D
	Zahlenrätsel (Zahlenwissen)	A	B		
	Rechenstrategien				D
	abzählende Kombinatorik	A		C	
	halbschriftliches Rechnen		B	C	
	Größen und Messen	Schätzen	A	B	C
Größenvergleiche		A	B	C	D
Umrechnungen		A	B	C	D
Einheiten zuordnen					D
Längen schätzen		A	B		
Zeitspannen ablesen				C	D
Argumentieren mit Geld		A	B	C	D
Argumentieren mit Massen		A	B		D
Modellieren		A	B	C	D
Daten, Häufigkeit und Wahrscheinlichkeit		Kreisdiagramm	A	B	C
	Säulendiagramm und Strichlisten	A		C	
	Fahrplan	A	B	C	D
	Wahrscheinlichkeiten beim Lösen			C	
	Raum und Form	Begriffswissen geometrischer Körper	A	B	C
visuelle Diskrimination im Raum		A	B	C	D
Figur-Grund-Diskrimination			B	C	
Spiegelachsen ebener Figuren		A	B	C	D
Begriffswissen ebener Figuren (Figuren ergänzen)		A	B	C	D
Begriffswissen ebener Figuren (Quadrat zerlegen)		A	B	C	D
Spiegelsymmetrisch ergänzen					D

2.4 Überlegungen zur Steigerung der Testökonomie

Zur Erreichung einer hohen Testökonomie und Zumutbarkeit der Durchführung des „Mathes 3“ 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
- Testheffformat A5 mit Ermöglichung eines schwarz/weiß-Ausdrucks
- computergestützte Aufbereitung der Testergebnisse (Leistungsprofil entsprechend den Inhaltsbereichen, Lernverlaufgraphen) auf Kind- und Klassenebene

3 Testanwendung

Hinweis: Alle für die Durchführung und Auswertung von „Mathes 3“ 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 (inkluisiven) Grundschulklasse konzipiert.

„Mathes 3“ kann zu drei Messzeitpunkten in der dritten Klassenstufe sowie am Anfang der vierten Klassenstufe durchgeführt werden. Die nachfolgende Tabelle 2 stellt die Durchführungszeiträume dar.

Tabelle 2: Durchführungszeiträume des "Mathes 3"

	Schulwoche	Kommentar
MZP 1	10./11.	vor den Herbstferien
MZP 2	20./21.	vor dem Halbjahreszeugnis
MZP 3	30./31.	nach den Osterferien
MZP 4	3./4	Anfang Klasse 4

3.2 Testmaterial

Für die Durchführung des „Mathes 3“ 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 3“ 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 3“ 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.
- Nebeneinandersitzende Schülerinnen und Schüler sollten unterschiedliche Testformen erhalten, um Abschreiben zu verhindern.
- 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 3“ sollte möglichst nicht in der letzten Unterrichtsstunde und nicht im Nachmittagsunterricht durchgeführt werden.
- Bei mehrfachem Einsatz des „Mathes 3“ sollte jedes Kind eine andere Testform erhalten, um Erinnerungseffekte auszuschließen.

3.4 Auswertung und Interpretation

3.4.1 Manuelle Auswertung

Die [Auswertungsvorlagen](#) unterstützen eine objektive und ökonomische Auswertung des „Mathes 3“. Alle richtig gelösten Aufgaben werden mit einem Punkt, falsch gelöste mit null Punkten bewertet. Die erreichten Punkte werden für jede Testskala aufsummiert. Mithilfe der Normtabellen 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 alle Messzeitpunkte zur Verfügung (s. Anhang D, S. 40ff.). Für Einschätzungen des Lernverlaufes können diese Vergleichswerte jedoch nur bedingt herangezogen werden. Zu diesem Zweck werden die im Rasch-Modell geschätzten Personenparameter angegeben (s. Anhang E, S. 56ff.), welche zur besseren Anschaulichkeit auf die in der Probabilistischen Testtheorie gebräuchliche Kompetenzskala (MW = 500, SD = 100) transformiert wurden. Indem die abgelesenen Kompetenzwerte in die vorbereiteten Lernverlaufgraphen (s. Anhang F, S. 60f.) eingetragen werden, kann eine normorientierte Einschätzung des aktuellen Leistungsstandes sowie der Leistungsentwicklung des Kindes vorgenommen werden. Dabei helfen die in Abschnitt 3.4.3 aufgeführten Referenzniveaus.

Differenzierte Informationen zur Auswertung des „Mathes 3“ 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 3“ angeboten. Bei dieser Variante müssen lediglich die erreichten Rohwerte in den vier Testskalen (*Zahlen und Operationen; Größen und Messen; Daten, Häufigkeit und Wahrscheinlichkeit; Raum und Form*) mithilfe der [Auswertungsvorlagen](#) 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 3“ stellt die Internetplattform ebenfalls den Lernverlauf 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.

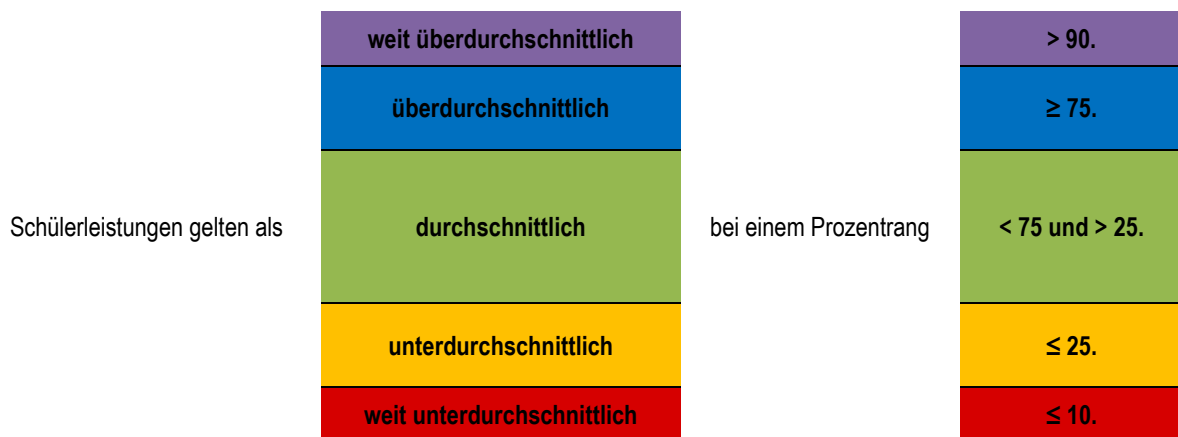


Abbildung 2: Referenzniveaus als Interpretationshilfen für die erzielte Testleistung

4 Testgütekriterien

Zur Einschätzung der psychometrischen Güte des „Mathes 3“ wurden zwei Teilstudien im Zeitraum von 2012 bis 2014 durchgeführt.

In einer ersten Teilstudie wurden die vier entwickelten Formen des „Mathes 3“ in einer Rostocker Stichprobe ($N = 545$; ♂ = 49 %, ♀ = 51 %) zu vier Messzeitpunkten im Abstand von etwa 10 Schulwochen im Schuljahr 2012/2013 erprobt. Die Kinder waren zum ersten Messzeitpunkt im Mittel 9;1 Jahre alt und kamen aus 27 Klassen von zehn Schulen. Die Untersuchungsgruppe verteilt sich prozentual gemäß ihrem wahren Anteil auf das Rostocker Stadtgebiet und entspricht somit einer Quotenstichprobe für die Hansestadt. Eine Analyse der VERA-Leistungen (ausführlich in Sikora, 2017) erbrachte keine bedeutsamen Unterschiede zwischen Rostocker Schülerinnen und Schülern und allen anderen Drittklässlerinnen und Drittklässlern in Mecklenburg-Vorpommern, sodass die ermittelten Vergleichswerte auch über die Stadtgrenzen Rostocks hinaus bedeutsam sind. Jedes Kind bearbeitete zu jedem Messzeitpunkt eine andere Version des Verfahrens innerhalb von einer Unterrichtsstunde, wobei die Reihenfolge der Testformen zufällig festgelegt wurde. Der Datenausfall war relativ gering und unsystematisch und ist vermutlich auf Krankheit einzelner Kinder zu einem der Messzeitpunkte zurückzuführen. Fehlende Werte wurden daher nicht ersetzt und die betreffenden Kinder verblieben zu den übrigen Messzeitpunkten im Datensatz.

In einer zweiten Teilstudie konnten die Daten der Evaluationsstudie zum „Rügener Inklusionsmodell“ (zusammenfassend Voß et al., 2016) genutzt werden, um Hinweise zur Validität des „Mathes 3“ zu erhalten. Dabei handelte es sich um eine Totalerhebung ($N = 448$; ♂ = 50 %, ♀ = 50 %) über anderthalb Schuljahre, in die alle Schülerinnen und Schüler staatlicher Grundschulen Rügens ($N = 12$), welche 2012/2013 eine dritte Klasse ($N = 23$)

besuchten, einbezogen wurden. Die Kinder waren zum ersten Messzeitpunkt im Mittel 9;4 Jahre alt. Es war ein geringer, aber unsystematischer Datenausfall zu verzeichnen, mit fehlenden Werten wurde analog der ersten Teilstudie umgegangen. Eine differenzierte Dropoutanalyse ist bei Voß et al. (2016) aufgeführt. Die Schülerinnen und Schüler bearbeiteten „Mathes 3“ im zweiten Halbjahr der dritten Klassenstufe (30. Schulwoche) innerhalb von einer Unterrichtsstunde.

Im Rahmen des „Rügener Inklusionsmodells“ kamen mit dem Deutschen Mathematiktest für dritte Klassen (DEMAT 3+; Roick, Göllitz & Hasselhorn, 2004) am Ende der dritten Klassenstufe, der Lernverlaufsdagnostik – Mathematik für zweite bis vierte Klassen (LVD-M 2-4; Strathmann & Klauer, 2012) im Verlauf des zweiten Halbjahres sowie der Kompetenzerfassung in Kindergarten und Schule – Mathematik Klasse 4 (KEKS 4; Ricken, Hildenbrand & May, 2013) zum Anfang der vierten Klasse konstruktähnliche Messverfahren zum Einsatz. Zudem wurden Tests mit konstruktfernen Instrumenten durchgeführt: ein Rechtschreibtest, der Deutsche Rechtschreibtest für das dritte und vierte Schuljahr (DERET 3-4+; Stock & Schneider, 2008), ein Lesetest, die Würzburger Leise-Leseprobe – Revision (WLLP-R Schneider, Blanke, Faust & Küspert, 2011) sowie ein Intelligenztest, der Grundintelligenztest Skala 2 – Revision (CFT 20-R; Weiß, 2008).

Um zu überprüfen, inwiefern sich die Ergebnisse im „Mathes 3“ zur Vorhersage der mathematischen Leistungsentwicklung eignen, wurden das Verfahren Bildungsstandards: Kompetenzen überprüfen. Mathematik Klasse 3/4 (Granzer et al., 2008) am Ende der vierten Klassenstufe sowie die Note auf dem Endjahreszeugnis von Klasse 4 herangezogen.

Für die Testgüte des „Mathes 3“ liegen weitere Befunde vor, beispielsweise zum Nutzen und zur Anwendbarkeit des Verfahrens. Diese sind bei Sikora (2017) zu finden.

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 3“ für Schülerinnen und Schüler dritter Klassen angemessen ist, wurde mittels der Übereinstimmung der im Rasch-Modell (s. Abschnitt 4.5) geschätzten Item- und Personenparameter geprüft (Carstensen & Taskinen, 2007). Die nachfolgende Abbildung 3 stellt die Verteilungen der Item- und Personenparameter der vier testheftübergreifenden Skalen des „Mathes 3“ zu den vier Messzeitpunkten dar.

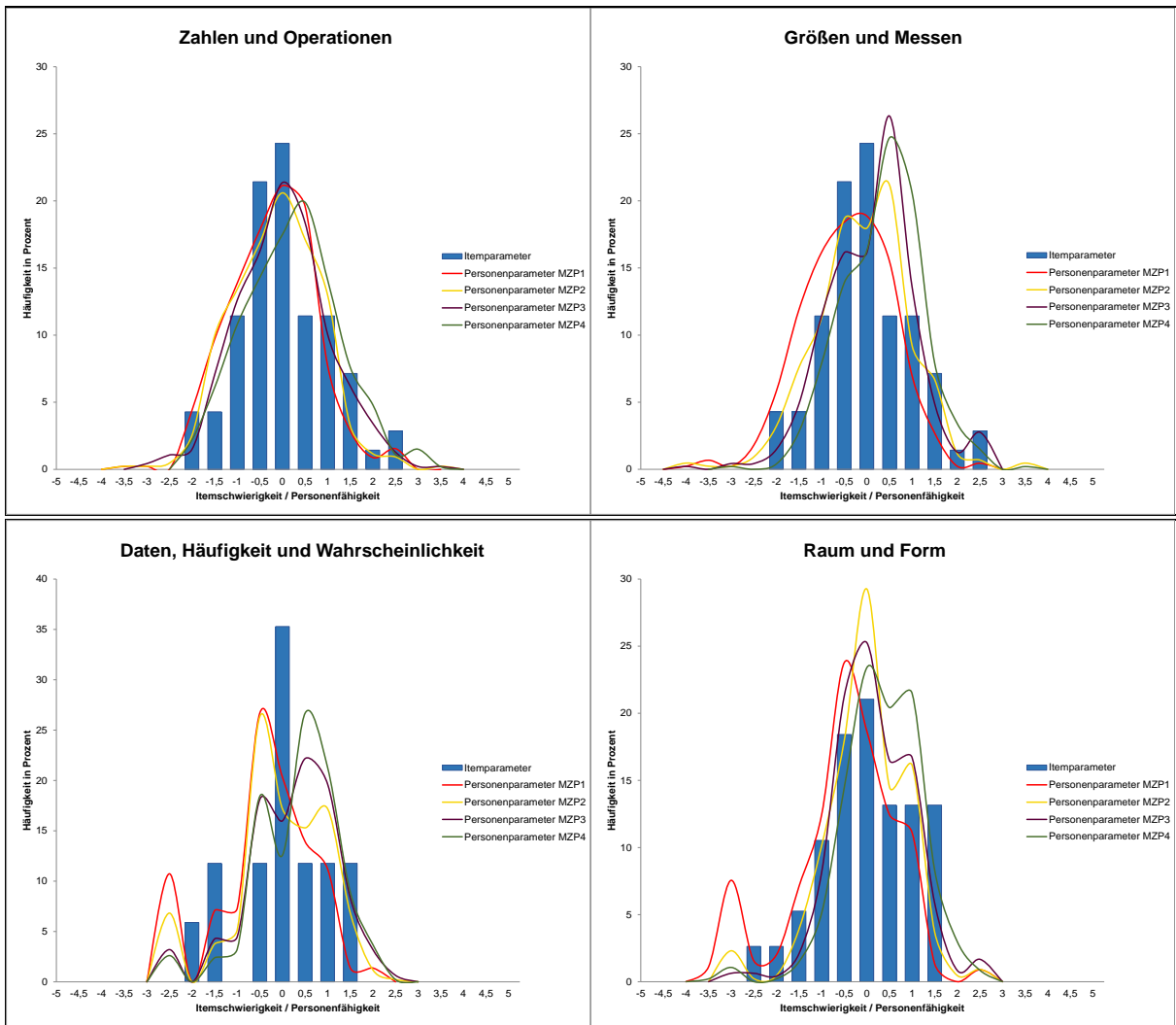


Abbildung 3: Verteilungen von Aufgabenschwierigkeiten und Kompetenzwerten zu den vier Messzeitpunkten

Aus den Darstellungen geht hervor, dass „Mathes 3“ in allen vier Inhaltsbereichen den mittleren und oberen Leistungsbereich gut bis sehr gut abdeckt, den unteren Leistungsbereich in den Skalen Zahlen und Operationen sowie Größen und Messen akzeptabel und in den Skalen Daten, Häufigkeit und Wahrscheinlichkeit sowie Raum und Form zumindest ausreichend Items vertreten sind.

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 3“ zu den vier Messzeitpunkten befindet sich im Anhang A auf Seite 28ff.

4.2 Skalenstruktur

Mittels konfirmatorischer Faktorenanalysen (CFA) wurde geprüft, ob die Skalen des Verfahrens die theoretisch angenommene Faktorenstruktur abbilden (s. Abschnitt 2.2). Für jede der vier Testformen des „Mathes 3“ wurde eine CFA mit vier Faktoren mithilfe der Software Mplus 7 (Muthén & Muthén, 1998-2012) gerechnet. Die latenten Variablen wurden frei geschätzt. Die Parameter wurden mithilfe der WLSMV-Methode (robust weighted least squares) ermittelt. Für die Modellevaluation stehen verschiedene Fit-Indizes zur Verfügung. In Übereinstimmung

mit den Empfehlungen von Kline (2005) wurden der χ^2 -Wert und der dazugehörige p-Wert, der RMSEA (Root Mean-Square-Error of Approximation) sowie der CFI (Comparative-Fit-Index) berechnet. Bei Stichproben von $N > 200$ sollte „der χ^2 -Test nur als erster Indikator der Modellpassung gesehen werden“ (Wilbert & Linnemann, 2011, S. 233). Für einen guten Modellfit sollte der aus dem Modelltest resultierende χ^2 -Wert kleiner als zweimal die Anzahl der Freiheitsgrade sein, bis $\chi^2 / df \leq 3$ ist von einem akzeptablen Modellfit auszugehen (Moosbrugger & Schermelleh-Engel, 2012). Werte von $RMSEA < .08$ gelten als akzeptabel, Werte von $RMSEA < .05$ als gut (Moosbrugger & Schermelleh-Engel, 2012). Beim CFI sprechen Werte von $CFI > .95$ für einen guten Modellfit, Werte von $CFI > .90$ gelten als akzeptabel (Hu & Bentler, 1999). Die Tabelle 3 stellt die Ergebnisse der CFA für die vier Testformen des „Mathes 3“ dar.

Tabelle 3: Modellfit-Indizes der konfirmatorischen Faktorenanalysen

„Mathes 3“ Testform	N	χ^2	df	RMSEA	CFI
A	474	3431.57**	1704	0.046	0.88
B	480	3401.06**	1946	0.043	0.89
C	473	3161.07**	1704	0.043	0.91
D	469	2460.77**	1763	0.029	0.94

Anmerkungen: N – Stichprobenumfang; χ^2 – Anpassungstest; ** – Modellanpassung ist auf dem Niveau von 0.01 (2-seitig) signifikant; df – Anzahl der Freiheitsgrade des χ^2 -Anpassungstests; RMSEA – Root Mean-Square-Error of Approximation; CFI – Comparative Fit-Index

Der χ^2 -Anpassungstest als Maß für den globalen Modellfit fällt in allen vier Testformen des „Mathes 3“ signifikant aus. Das Verhältnis aus χ^2 -Wert und den Freiheitsgraden liegt bei den Formen B, C und D unter 2, bei Form A beträgt $\chi^2 / df = 2.01$. Der RMSEA zeigt in allen vier Formen des „Mathes 3“ einen guten Modellfit an. Der CFI liegt bei den Formen A und B knapp unterhalb des Cut-Off Wertes von $CFI = .90$, was gegen den Modellfit spricht. Die Werte der Formen C und D weisen auf einen akzeptablen Modellfit hin.

4.3 Reliabilität

Zur Schätzung der Zuverlässigkeit des Verfahrens im Sinne der Internen Konsistenz wurde die sogenannte EAP/PV-Reliabilität (Rost, 2004) berechnet. Die resultierenden Koeffizienten sind von der Höhe vergleichbar mit Cronbachs α (Rost, 2004), welcher zusätzlich für jeden Subtest der vier Testformen des „Mathes 3“ ermittelt wurde. Die nach der EAP/PV-Methode berechneten Reliabilitätskoeffizienten liegen zwischen $EAP/PV = .69$ und $EAP/PV = .83$. Die Tabelle 4 stellt die Internen Konsistenzen der vier Testskalen zu den vier Messzeitpunkten dar.

Tabelle 4: Ergebnisse der Analysen hinsichtlich der EAP/PV-Reliabilität der Skalen des "Mathes 3"

EAP/PV-Reliabilität von				
	Zahlen und Operationen	Größen und Messen	Daten, Häufigkeit und Wahrscheinlichkeit	Raum und Form
MZP 1	.83 N = 454	.74 N = 445	.70 N = 438	.80 N = 450
MZP 2	.83 N = 432	.76 N = 435	.74 N = 425	.73 N = 434
MZP 3	.83 N = 468	.73 N = 471	.71 N = 469	.71 N = 479
MZP 4	.83 N = 462	.72 N = 466	.69 N = 461	.71 N = 470

Anmerkungen: MZP – Messzeitpunkt; N – Stichprobenumfang

Die interne Konsistenz der Subtests des „Mathes 3“ liegt überwiegend im hohen Bereich mit Koeffizienten zwischen $\alpha = .54$ und $\alpha = .91$ ($MW: \alpha = .78$, $SD = .08$). Die nachfolgende Tabelle 5 stellt die Alphakoeffizienten der Subtests des „Mathes 3“ getrennt nach Testform zu den vier Messzeitpunkten dar.

Tabelle 5: Ergebnisse der Analysen hinsichtlich der Internen Konsistenz (Cronbachs α) der Subtests des "Mathes 3"

Reliabilitätskoeffizient Cronbachs α von						
Form	MZP	N	Zahlen und Operationen	Größen und Messen	Daten, Häufigkeit und Wahrscheinlichkeit	Raum und Form
A	1	120	.87	.73	.74	.74
	2	112	.85	.72	.72	.67
	3	123	.86	.76	.76	.54
	4	119	.90	.76	.79	.75
B	1	118	.86	.77	.60	.85
	2	111	.87	.80	.77	.66
	3	126	.91	.80	.81	.79
	4	125	.91	.76	.73	.64
C	1	116	.87	.80	.74	.78
	2	119	.89	.84	.80	.79
	3	124	.86	.72	.76	.66
	4	114	.84	.73	.72	.71
D	1	119	.88	.81	.76	.85
	2	109	.85	.74	.71	.76
	3	122	.82	.71	.69	.79
	4	119	.87	.74	.77	.74

Anmerkungen: MZP – Messzeitpunkt; N – Stichprobenumfang, α – Koeffizient nach Cronbach

4.4 Validität

4.4.1 Konstruktvalidität

Zur Einschätzung der Konstruktvalidität wurden die Ergebnisse der Kinder zum Ende der dritten Klassenstufe in der jeweiligen Testskala des „Mathes 3“ mit konstruktähnlichen und konstruktfernen Verfahren korreliert (Produkt-Moment-Korrelation nach Pearson; Rasch, Friese, Hofmann & Naumann, 2014). Alle Korrelationen wurden einem Signifikanztest unterzogen. Gemäß der gängigen Konvention sollten die Zusammenhänge mindestens auf dem Niveau $\alpha = .05$ signifikant sein (Rasch et al., 2014). Wenn diese Bedingung erfüllt war, wurde die Höhe der Korrelation anhand der Klassifikation von Cohen (1988) eingeschätzt.

Die Zusammenhänge mit konstruktähnlichen Verfahren fallen durchweg signifikant aus und liegen im mittleren bis hohen Bereich. Die Korrelationskoeffizienten streuen in einem Intervall von $r = .41$ bis $r = .77$. Die höchsten Korrelationen bestehen zur Skala Zahlen und Operationen. Der geringste Zusammenhang zeigt sich zwischen der Skala Raum und Form und dem Untertest Geometrie im DEMAT 3+ (Roick et al., 2004). Da die zur Validierung genutzten Verfahren nicht alle Inhaltsbereiche des Mathematikunterrichts erfassen, wurden nur die inhaltlich übereinstimmenden Testskalen des „Mathes 3“ korreliert. Beispielsweise prüft die LVD-M 2-4 (Strathmann & Klauer,

2012) nur Rechenfähigkeiten ab, weshalb das Verfahren nur zur Validierung der Skala Zahlen und Operationen herangezogen wurde.

Tabelle 6: Zusammenhänge der jeweiligen Testskala des „Mathes 3“ zu konstruktähnlichen Testverfahren

	Korrelation r mit			
	Zahlen und Operationen	Größen und Messen	Daten, Häufigkeit und Wahrscheinlichkeit	Raum und Form
KEKS 4	.70** $N = 376$.58** $N = 369$.48** $N = 366$	
DEMAT 3+ UT Arithmetik	.67** $N = 427$			
DEMAT 3+ UT Sachrechnungen		.49** $N = 420$		
DEMAT 3+ UT Geometrie				.41** $N = 421$
LVD-M 2-4	.77** $N = 417$			

Anmerkungen: KEKS 4 – Kompetenzerfassung in Kindergarten und Schule – Mathematik Klasse 4 (Ricken, et al., 2013); DEMAT 3+ – Deutscher Mathematiktest für dritte Klassen (Roick et al., 2004); LVD-M 2-4 – Lernverlaufdiagnostik – Mathematik für zweite bis vierte Klassen (Strathmann & Klauer, 2012); UT – Untertest; N – Stichprobenumfang; r – Korrelationskoeffizient nach Pearson; ** – Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant

Die Zusammenhänge der Testergebnisse in den vier Skalen des „Mathes 3“ mit den Lese-, rechtschriftlichen sowie kognitiven Leistungen der Schülerinnen und Schüler werden in der Tabelle 7 dargestellt. Da beim DERET 3-4+ (Stock & Schneider, 2008) die im Diktat gemachten Fehler ausgezählt werden, bedeuten höhere Werte eine schlechtere Testleistung.

Tabelle 7: Zusammenhänge der jeweiligen Testskala des „Mathes 3“ zu konstruktfernen Testverfahren

	Korrelation r mit			
	Zahlen und Operationen	Größen und Messen	Daten, Häufigkeit und Wahrscheinlichkeit	Raum und Form
DERET 3-4+	-.58** $N = 424$	-.44** $N = 412$	-.35** $N = 412$	-.26** $N = 418$
WLLP-R	.48** $N = 426$.37** $N = 418$.36** $N = 413$.30** $N = 420$
CFT 20-R	.51** $N = 426$.45** $N = 419$.42** $N = 414$.41** $N = 420$

Anmerkungen: DERET 3-4+ – Deutscher Rechtschreibtest für das dritte und vierte Schuljahr (Stock & Schneider, 2008); WLLP-R – Würzburger Leise Leseprobe – Revision (Schneider et al., 2011); CFT 20-R – Grundintelligenztest Skala 2 – Revision (Weiß, 2008); N – Stichprobenumfang; r – Korrelationskoeffizient nach Pearson; ** – Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant

Erwartungsgemäß fallen die Zusammenhänge zum DERET 3-4+ (Stock & Schneider, 2008) negativ aus. Die meisten Korrelationskoeffizienten liegen im mittleren Bereich, einer ist als gering zu klassifizieren und zwei als hoch. Für die einzelnen Testskalen gilt, dass die Zusammenhänge zu den jeweils konstruktfernen Verfahren durchweg geringer sind als zu den jeweils konstruktähnlichen Verfahren, was für die Konstruktvalidität des „Mathes 3“ 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 3“ von besonderer Bedeutung. Zur Prüfung der prognostischen Validität wurden in einem ersten Schritt die Zusammenhänge der Testskalen des „Mathes 3“ mit denen im Bildungsstandards-Test (Granzer et al., 2008) am Ende der vierten Klassenstufe sowie mit der Note auf dem Endjahreszeugnis von Klasse 4 ermittelt. Bei den Korrelationen mit Zeugnisnoten wurde der Rangkorrelationskoeffizient Spearmans Rho berechnet (Rasch et al., 2014). In der Tabelle 8 werden Ergebnisse der Korrelationsanalysen dargestellt.

Tabelle 8: Zusammenhänge der jeweiligen Testskala des „Mathes 3“ mit zukünftigen Mathematikleistungen

	Korrelation r mit			
	Zahlen und Operationen	Größen und Messen	Daten, Häufigkeit und Wahrscheinlichkeit	Raum und Form
Zeugnisnote Mathematik	-.64** $N = 360$	-.52** $N = 354$	-.38** $N = 345$	-.34** $N = 355$
Bildungsstandards	.66** $N = 415$.61** $N = 406$.51** $N = 401$.38** $N = 407$

Anmerkungen: Bildungsstandards – Bildungsstandards: Kompetenzen überprüfen. Mathematik Klasse 3/4 (Granzer et al., 2008); N – Stichprobenumfang; r – Korrelationskoeffizient nach Spearman bzw. Pearson; ** – Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant

Die Korrelationen mit der Zeugnisnote fallen erwartungsgemäß negativ aus, d. h. ein hoher Testwert geht tendenziell mit einer besseren Zeugnisnote einher. Die hohen Zusammenhänge sprechen für eine hohe Vorhersagegüte der Skala Zahlen und Operationen, die anderen Skalen weisen eine geringere prognostische Aussagekraft auf.

Als Ergänzung zu den Ergebnissen der korrelativen Analysen wurde die klassifikatorische Güte des „Mathes 3“ ermittelt (Marx & Lenhard, 2010). Zur Berechnung der Kennwerte wurde Kindern, welche am Ende der dritten Klasse im „Mathes 3“ in mindestens einer Skala einen Testwert entsprechend einem Prozentrang kleiner bzw. gleich 10 oder in mindestens zwei Skalen einem Prozentrang kleiner bzw. gleich 25 erzielten (s. Abschnitt 3.4.3), ein Risikostatus in ihrer mathematischen Entwicklung zugeordnet. Diese relativ milden Cut-Off-Werte wurden 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 Kriterien dienen die Zeugnisnoten, operationalisiert als eine ausreichende Zeugnisnote oder schlechter, sowie die Verfehlung der Regelstandards, gemessen mit dem Bildungsstandards-Test (Granzer et al., 2008) am Ende der vierten Klassenstufe. Die Tabelle 9 stellt die Befunde der klassifikatorischen Analysen dar.

Tabelle 9: Ergebnisse zur klassifikatorischen Güte des "Mathes 3" bezüglich Risiken im Fach Mathematik

	Stichproben- umfang	n richtig positiv	n falsch positiv	n falsch negativ	n richtig negativ	Sensitivität	Spezifität	α - Fehlerquote	β - Fehlerquote	positiver prä- diktiver Wert	negativer prä- diktiver Wert	RATZ-Index
Zeugnisnote Mathematik	306	26	113	3	164	0.90	0.59	0.41	0.10	0.19	0.98	0.81
Bildungs- standards	353	47	110	7	189	0.87	0.63	0.37	0.13	0.30	0.96	0.77

Anmerkungen: n – Anzahl; Bildungsstandards – Bildungsstandards: Kompetenzen überprüfen. Mathematik Klasse 3/4 (Granzer et al., 2008).

Die Ergebnisse zur Sensitivität zeigen, dass 90 % bzw. 87 % der Kinder mit schwachen Mathematikleistungen am Ende der Grundschulzeit durch „Mathes 3“ bereits in der dritten Klassenstufe korrekt identifiziert werden. Für diese Trefferquote muss allerdings ein relativ hoher α -Fehler akzeptiert werden, was die eher geringen Werte der Spezifität anzeigen. Besonders erwähnenswert sind zudem die RATZ-Indizes, welche mit 81 % bzw. 77 % im sehr hohen Bereich liegen und auf eine sehr gute Klassifikationsleistung des „Mathes 3“ hinweisen (Marx & Lenhard, 2000).

4.5 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 für jede Skala des „Mathes 3“ eingeschätzt. Dazu wurden Fit-Statistiken (weighted Mean Square [wMNSQ]) für jedes Item mit der Software ConQuest® (Wu, Adams & Wilson, 2007) berechnet, sodass beurteilt werden kann, wie gut die geschätzten Itemparameter auf das Rasch-Modell passen. In Anlehnung an Bond und Fox (2015) zeigen bei gewöhnlichen Stichprobengrößen Werte von $0.7 \leq wMNSQ \leq 1.3$ an, dass das Item den Annahmen des Rasch-Modells entspricht.

Nachfolgend werden die Ergebnisse der Raschskalierung präsentiert. Die wMNSQ-Werte wurden aus Gründen der Übersichtlichkeit in der Abbildung 4 mittels Boxplots graphisch aufbereitet, sodass schnell ersichtlich ist, in welchem Wertebereich die wMNSQs der Items des „Mathes 3“ streuen.

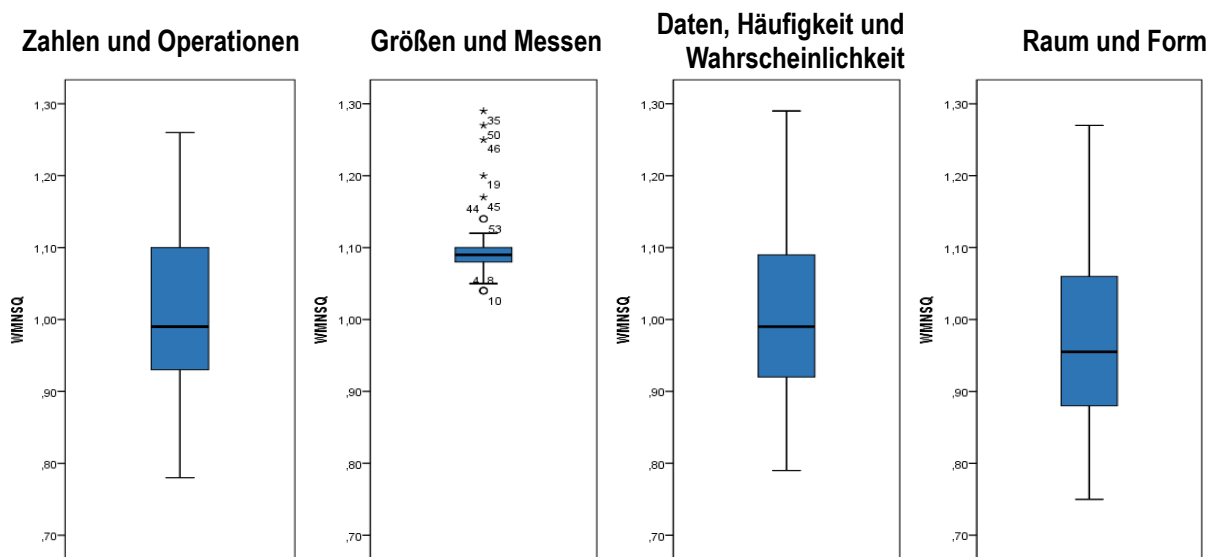


Abbildung 4: Verteilung der Itemfit-Statistiken ($wMNSQ$) der vier Testskalen des „Mathes 3“

Über alle Skalen des „Mathes 3“ hinweg liegen die $wMNSQ$ -Werte der Aufgaben innerhalb des zulässigen Wertebereiches. Deutlich wird auch, dass der Großteil der Items in den Skalen Zahlen und Operationen, Daten, Häufigkeit und Wahrscheinlichkeit sowie Raum und Form in der Nähe des Erwartungswertes von 1 liegt. In der Skala Größen und Messen gibt es tendenziell einen Underfit. Eine Auflistung der im Rasch-Modell geschätzten Fitstatistiken zu jedem Item kann im Anhang B auf S. 33ff. eingesehen werden.

Eine zentrale Voraussetzung für Veränderungsmessungen ist die Stichprobeninvarianz der Itemparameter. Diese wurde analog zu Vorarbeiten mit ähnlicher Zielstellung (z. B. Gebhardt, Heine, Zeuch & Förster, 2015) auf verschiedene Messzeitpunkte bezogen und mittels graphischer Modelltests geprüft. Die Itemparameter erweisen sich – insgesamt betrachtet – zwischen den Messzeitpunkten als annähernd konstant, d. h. messinvariant über die Zeit (s. Anhang C, S. 38f. für weitere Informationen). Nur bei wenigen Aufgaben sind diesbezüglich kleinere Abweichungen festzustellen.

4.6 Änderungssensitivität

Zur Prüfung der Änderungssensitivität wurden deskriptivstatistische Angaben (Mittelwert, Standardabweichung, Minimum, Maximum) zu den im Rasch-Modell geschätzten Personenparametern für jede Skala des „Mathes 3“ ermittelt. Zeigte sich dabei ein Anstieg über die Messzeitpunkte, wurde geprüft, ob die Veränderung der Messwerte systematisch entstanden ist. Dazu wurde eine Varianzanalyse mit Messwiederholung für jede Skala des Verfahrens durchgeführt. Alle Analysen beziehen sich auf die geschätzten Personenparameter in den Skalen des „Mathes 3“, welche zur besseren Anschaulichkeit auf die, beispielsweise in Schulleistungsstudien gebräuchliche, Kompetenzskala ($M = 500$, $SD = 100$) transformiert wurden (s. Abschnitt 3.4.1).

Tabelle 10: Deskriptive Statistiken der Skalen des „Mathes 3“ über die vier Messzeitpunkte

Skala	MZP	N	M	SD	Min	Max
Zahlen und Operationen	1	454	482.3	95.3	154.2	773.3
	2	432	489.3	96.3	154.2	841.4
	3	468	504.0	100.3	177.3	841.4
	4	462	523.3	103.0	276.4	841.4
Größen und Messen	1	445	457.1	99.5	110.9	742.5
	2	434	491.2	102.1	110.9	833.5
	3	471	512.3	93.9	123.4	757.3
	4	466	536.8	87.4	221.7	833.5
Daten, Häufigkeit und Wahrscheinlichkeit	1	438	459.3	100.9	264.9	694.0
	2	425	491.2	99.8	264.9	736.3
	3	469	518.1	95.7	264.9	736.3
	4	461	528.4	89.3	264.9	736.3
Raum und Form	1	450	451.5	116.3	164.1	753.4
	2	434	499.0	89.5	117.8	753.4
	3	479	513.1	86.4	177.8	745.2
	4	470	534.0	86.7	164.1	745.2

Anmerkungen: MZP – Messzeitpunkt; N – Stichprobenumfang; M – Mittelwert; SD – Standardabweichung; Min – Minimum; Max – Maximum

In allen vier Skalen steigen die geschätzten Personenparameter im Mittel kontinuierlich an. Die Ergebnisse der Varianzanalyse mit Messwiederholung werden in der Tabelle 11 im Überblick dargestellt.

Tabelle 11: Ergebnisse der inferenzstatistischen Prüfung der Mittelwertunterschiede über die vier Messzeitpunkte

Skala	F	df	Mittel der Quadrate	η^2	Paarweise Vergleiche
Zahlen und Operationen	34.69**	2.97	93060.89	.11	** (außer 1-2)
Größen und Messen	92.05**	2.91	341737.29	.24	**
Daten, Häufigkeit und Wahrscheinlichkeit	71.34**	2.82	309636.25	.19	** (außer 3-4)
Raum und Form	75.33**	2.73	372481.70	.20	**

Anmerkungen: kursiv – Verletzung der Sphäritätsannahme: Berechnung mittels Huynh-Feldt Korrekturformel; ** – F-Test bzw. Post-Hoc-Test ist auf dem Niveau von 0.01 (2-seitig) signifikant; df – Freiheitsgrade; η^2 – Partielles Eta-Quadrat

Die F-Tests zeigen in allen Skalen einen statistisch signifikanten Einfluss des Faktors Zeit auf die Leistung an. Die Varianzaufklärung durch die Messwiederholung beträgt in der Skala Zahlen und Operationen 11 %, was einem mittleren Effekt entspricht (Cohen, 1988), in den übrigen Testskalen ist der Effekt groß mit Varianzaufklärungen von 24 %, 19 % bzw. 20 %.

Die nach Bonferroni korrigierten paarweisen Vergleiche zeigen, dass die Mittelwertdifferenzen zwischen den Messzeitpunkten überwiegend statistisch signifikant ausfallen. In der Skala Zahlen und Operationen erreicht die Mittelwertdifferenz zwischen erster und zweiter Messung keine statistische Signifikanz, in der Skala Daten, Häufigkeit und Wahrscheinlichkeit betrifft dies den Unterschied zwischen dritter und vierter Messung.

5 Literaturverzeichnis

- Blum, W. (2010). Die Bildungsstandards Mathematik. In W. Blum, C. Drücke-Noe, R. Hartung & O. Köller (Hrsg.), *Bildungsstandards Mathematik: konkret. Sekundarstufe I. Aufgabenbeispiele, Unterrichts Anregungen, Fortbildungsideen*. 4. Auflage (S. 14-32). Berlin: Cornelsen.
- 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.
- 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.
- Devlin, K. (1998). *Muster der Mathematik*. Heidelberg: Spektrum.
- Dilling H., Mombour, W. & Schmidt, M.H. (2011). Internationale Klassifikation psychischer Störungen. ICD-10 Kapitel 5 (F). Klinisch-diagnostische Leitlinien. Bern: Huber.
- Führer, L. (1997). *Pädagogik des Mathematikunterrichts. Eine Einführung in die Fachdidaktik für Sekundarstufen*. Braunschweig & Wiesbaden: Vieweg.
- Gaidoschik, M. (2011). *Rechenschwäche – Dyskalkulie. Eine unterrichtspraktische Einführung für LehrerInnen und Eltern*. 6. Auflage. Hamburg: Persen.
- Gebhardt, M. Heine, J.-H., Zeuch, N. & Förster, N. (2015). Lernverlaufsdiagnostik im Mathematikunterricht der zweiten Klasse: Raschanalysen und Empfehlungen zur Adaptation eines Testverfahrens für den Einsatz in inklusiven Klassen. *Empirische Sonderpädagogik*, 3, 206–222.
- Granzer, D., Reiss, K., Winkelmann, H., Robitzsch, A., Köller, O. & Walther, G. (2008). *Bildungsstandards: Kompetenzen überprüfen. Mathematik Grundschule Klasse 3/4*. Berlin: Cornelsen.
- Grassmann, M., Eichler, K.-P., Mirwald, E. & Nitsch, B. (2014). *Mathematikunterricht. Kompetent im Unterricht der Grundschule*. 3. korrigierte und veränderte Auflage. Baltmannsweiler: Schneider Verlag Hohengehren.
- Hartke, B. & Diehl, K. (2013). *Schulische Prävention im Bereich Lernen. Problemlösungen mit dem RTI-Ansatz*. Stuttgart: Kohlhammer.
- 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.
- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6 (1), 1-55.
- Jacobs, C. & Petermann, F. (2012). *Diagnostik von Rechenstörungen. 2., überarbeitete und erweiterte Auflage*. Göttingen: Hogrefe.
- Klieme, E., Avenarius, H., Blum, W., Gruber, H., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E., Vollmer, H.J. (2007). *Zur Entwicklung nationaler Bildungsstandards. Expertise*. Berlin: BMBF.
- Kline, R.B. (2005). *Principles and Practice of Structural Equation Modeling (2nd Edition ed.)*. New York: The Guilford.
- 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.
- Kolen, M.J. & Brennan, R.J. (2004). *Test equating, scaling, and linking: methods and practice* S. 2nd edition. New York: Springer.
- Krajewski, K. (2003): Vorhersage von Rechenschwäche in der Grundschule. Hamburg: Dr. Kovač.
- 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.
- 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.
- Mislevy, R.J., Beaton, A.E., Kaplan, B. & Sheehan, K.M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29, 133-161.
- Moosbrugger, H. & Schemmelleh-Engel, K. (2012). Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion. 2., aktualisierte und überarbeitete Auflage* (S. 325-343). Berlin: Springer.

- 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.
- Muthén, L.K. & Muthén, B.O. (1998-2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.
- Prenzel, M. & Blum, W. (Hrsg.) (2007). *Entwicklung eines Testverfahrens zur Überprüfung der Bildungsstandards in Mathematik für den Mittleren Schulabschluss. Technischer Bericht*. Kiel: IPN.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen & Lydiche.
- Rasch, B., Frieze, M., Hofmann, W. & Naumann, E. (2014a). *Quantitative Methoden 1. Einführung in die Statistik für Psychologen und Sozialwissenschaftler. 4., überarbeitete Auflage*. Berlin: Springer.
- Ricken, G., Hildenbrand, C. & May, P. (2013). *Kompetenzerfassung in Kindergarten und Schule – Mathematik Klasse 4*. Berlin: Cornelsen.
- Roick, T., Göllitz, D. & Hasselhorn, M. (2004). *Deutscher Mathematiktest für dritte Klassen (DEMAT 3+)*. Göttingen: Hogrefe.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion. 2. vollständig überarbeitete und erweiterte Auflage*. Bern: Huber.
- Schneider, W. Blanke, I., Faust, V. & Küspert, P. (2011). *Würzburger Leise Leseprobe – Revision (WLLP-R)*. Göttingen: Hogrefe.
- 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.
- Strathmann, A. M. & Klauer, K. J. (2012). *Lernverlaufdiagnostik – Mathematik für zweite bis vierte Klassen (LVD-M 2-4)*. Göttingen: Hogrefe.
- Vohns, A. (2007). *Grundlegende Ideen und Mathematikunterricht. Entwicklung und Perspektiven eines fachdidaktischen Prinzips*. Norderstedt: Books on Demand.
- Von Aster, M., Schweiter, M. & Weinhold Zulauf, M. (2007). Rechenstörungen bei Kindern. Vorläufer, Prävalenz und psychische Symptome. *Zeitschrift für Pädagogische Psychologie*, 39, 85-96.
- Voß, S., Blumenthal, Y., Mahlau, K., Marten, K., Diehl, K., Sikora, S. et al. (2016). *Der Response-to-Intervention-Ansatz in der Praxis. Evaluationsergebnisse zum Rügener Inklusionsmodell*. Münster: Waxmann.
- Walther, G., van den Heuvel-Panhuizen, M., Granzer, D. & Köller, O. (Hrsg.) (2011). *Bildungsstandards für die Grundschule. Mathematik konkret. 5. Auflage*. Berlin: Cornelsen.

- Weinert, F.E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17-31). Weinheim: Beltz.
- Weiß, R. H. (2008). *Grundintelligenztest Skala 2 - Revision (CFT 20-R)*. Göttingen: Hogrefe.
- Wilbert, J. & Linnemann, M. (2011). Kriterien zur Analyse eines Tests zur Lernverlaufdiagnostik. *Empirische Sonderpädagogik*, 3, 225-242.
- Winkelmann, H., Robitzsch, A., Stanat, P. & Köller, O. (2012). Mathematische Kompetenzen in der Grundschule. Struktur, Validierung und Zusammenspiel mit allgemeinen kognitiven Fähigkeiten. *Diagnostica*, 58, 15-30.
- Winter, H. (2001). *Inhalte mathematischen Lernens. Expertise für das Ministerium für Bildung, Wissenschaft, Jugend und Kultur des Landes Rheinland-Pfalz*. Abruf am 15.4.20. Online verfügbar unter: http://grundschule.bildung-rp.de/fileadmin/user_upload/grundschule.bildung-rp.de/Downloads/Mathematik/Winter_Inhalte_math_Lernens.pdf
- Wittmann, E.Ch. (1995). Aktiv-entdeckendes und soziales Lernen im Arithmetikunterricht. In G.N. Müller & E.Ch. Wittmann (Hrsg.), *Mit Kindern rechnen* (S. 10-41). Frankfurt am Main: Arbeitskreis Grundschule – Der Grundschulverband – e.V.
- Wittmann, E.Ch. (1999). Konstruktion eines Geometrieunterrichts ausgehend von Grundideen der Elementargeometrie. In H. Henning (Hrsg.), *Mathematik lernen durch Handeln und Erfahrung. Festschrift zum 75. Geburtstag von Heinrich Besuden* (S. 205-223). Oldenburg: Bültmann und Gerriets.
- Wittmann, E.Ch. & Müller, G.N. (2011). Muster und Strukturen als fachliches Grundkonzept. In G. Walther, M. van den Heuvel-Panhuizen, D. Granzer & O. Köller (Hrsg.), *Bildungsstandards für die Grundschule: Mathematik konkret. 5. Auflage* (S. 42-65). Berlin: Cornelsen.
- Wittmann, E.Ch. & Müller, G.N. (2012a). *Das Zahlenbuch 1 (Neubearbeitung). Begleitband*. Stuttgart: Klett.
- Wu, M.L., Adams, R.J., & Wilson, M.R. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software*. Melbourne: ACER Press.

6 Anhang

Anhang A: Schwierigkeiten und Trennschärfen der Items

Zahlen und Operationen

Item				Schwierigkeit P_i zum Messzeitpunkt				Trennschärfe r_{pbis} zum Messzeitpunkt			
Form A	Form B	Form C	Form D	1	2	3	4	1	2	3	4
1.1				0.44	0.46	0.51	0.52	0.49	0.50	0.36	0.45
1.2				0.37	0.39	0.43	0.47	0.41	0.41	0.51	0.57
2.1				0.91	0.99	0.95	0.9	0.38	0.20	0.36	0.37
2.2				0.89	0.90	0.86	0.84	0.31	0.44	0.39	0.33
2.3	2.2	2.1	2.3	0.82	0.79	0.84	0.83	0.50	0.54	0.50	0.55
2.4	2.3	2.2	2.4	0.64	0.62	0.60	0.62	0.58	0.61	0.56	0.62
2.5	4.4	2.5	2.7	0.59	0.62	0.68	0.78	0.52	0.59	0.56	0.55
2.6				0.86	0.79	0.80	0.85	0.45	0.50	0.62	0.54
2.7				0.54	0.61	0.71	0.71	0.56	0.57	0.50	0.59
2.8	4.5	2.8	2.8	0.57	0.58	0.63	0.71	0.62	0.66	0.67	0.64
3.1				0.57	0.69	0.79	0.74	0.58	0.48	0.36	0.52
3.2				0.18	0.24	0.32	0.38	0.50	0.54	0.59	0.65
4.1	2.4	2.3	4.1	0.70	0.73	0.78	0.78	0.64	0.60	0.61	0.61
4.2				0.42	0.72	0.69	0.63	0.66	0.64	0.65	0.55
4.3	2.6	2.4	4.2	0.48	0.60	0.61	0.67	0.59	0.60	0.64	0.66
4.4				0.29	0.31	0.32	0.39	0.61	0.58	0.55	0.64
4.5				0.32	0.29	0.33	0.32	0.60	0.58	0.58	0.64
4.6				0.35	0.34	0.36	0.43	0.55	0.58	0.52	0.78
5.1				0.18	0.29	0.29	0.30	0.48	0.49	0.62	0.58
5.2				0.09	0.15	0.26	0.30	0.48	0.45	0.59	0.59
6.1				0.22	0.24	0.28	0.27	0.53	0.42	0.41	0.30
	1.1			0.70	0.71	0.72	0.72	0.34	0.48	0.49	0.44
	1.2			0.57	0.50	0.67	0.79	0.56	0.55	0.64	0.45
	1.3			0.67	0.67	0.72	0.79	0.68	0.57	0.61	0.51
	1.4			0.61	0.66	0.63	0.74	0.63	0.39	0.69	0.48
	2.1			0.86	0.81	0.74	0.77	0.45	0.43	0.58	0.48
	2.5			0.92	0.88	0.79	0.87	0.33	0.46	0.57	0.50
	3.1			0.86	0.84	0.84	0.88	0.34	0.45	0.46	0.40
	3.2			0.74	0.56	0.66	0.75	0.47	0.66	0.62	0.45
	3.3			0.71	0.66	0.69	0.77	0.52	0.55	0.50	0.54
	3.4			0.38	0.35	0.44	0.58	0.49	0.45	0.62	0.57
	4.1			0.28	0.32	0.36	0.54	0.53	0.41	0.62	0.72
	4.2			0.23	0.25	0.32	0.48	0.47	0.44	0.63	0.73
	4.3			0.11	0.09	0.15	0.32	0.50	0.33	0.53	0.68
	4.6			0.43	0.50	0.41	0.60	0.53	0.68	0.59	0.66
	5.1			0.37	0.29	0.38	0.43	0.53	0.37	0.55	0.53
	5.2			0.03	0.05	0.12	0.16	0.28	0.30	0.38	0.52
	6.1			0.39	0.50	0.59	0.50	0.34	0.34	0.40	0.58

6.2		0.33	0.45	0.42	0.55	0.42	0.48	0.48	0.57
	1.1	0.72	0.77	0.75	0.80	0.46	0.52	0.40	0.30
	1.2	0.63	0.65	0.64	0.68	0.38	0.52	0.48	0.26
	1.3	0.51	0.54	0.47	0.65	0.48	0.43	0.43	0.29
	1.4	0.53	0.58	0.54	0.59	0.47	0.45	0.46	0.24
	2.6	0.64	0.60	0.70	0.69	0.74	0.63	0.60	0.66
	2.7	0.55	0.51	0.60	0.58	0.53	0.65	0.48	0.56
	3.1	0.61	0.55	0.62	0.61	0.73	0.60	0.55	0.63
	3.2	0.55	0.53	0.58	0.54	0.52	0.53	0.47	0.64
	3.3	0.30	0.35	0.36	0.41	0.40	0.37	0.36	0.37
	4.1	0.48	0.49	0.50	0.63	0.49	0.61	0.51	0.53
	4.2	0.43	0.44	0.47	0.62	0.58	0.64	0.59	0.60
	4.3	0.35	0.41	0.43	0.48	0.59	0.69	0.59	0.59
	5.1	0.15	0.19	0.20	0.22	0.42	0.55	0.49	0.31
	6.1	0.69	0.62	0.67	0.78	0.50	0.45	0.34	0.43
	6.2	0.3	0.24	0.37	0.34	0.39	0.48	0.45	0.44
	1.1	0.62	0.61	0.62	0.68	0.51	0.52	0.55	0.28
	1.2	0.61	0.57	0.74	0.74	0.57	0.47	0.48	0.40
	2.1	0.91	0.94	0.96	0.97	0.42	0.37	0.26	0.11
	2.2	0.81	0.84	0.88	0.83	0.50	0.56	0.37	0.48
	2.5	0.90	0.94	0.94	0.91	0.47	0.44	0.30	0.46
	2.6	0.75	0.84	0.87	0.86	0.55	0.52	0.44	0.64
	3.1	0.82	0.71	0.81	0.83	0.38	0.53	0.43	0.46
	3.2	0.59	0.54	0.50	0.44	0.59	0.53	0.47	0.46
	3.3	0.05	0.04	0.04	0.05	0.18	0.29	0.24	0.15
	3.4	0.07	0.06	0.07	0.08	0.27	0.30	0.39	0.26
	4.3	0.62	0.69	0.73	0.77	0.59	0.51	0.52	0.56
	4.4	0.50	0.50	0.55	0.64	0.74	0.73	0.66	0.74
	4.5	0.50	0.48	0.56	0.61	0.67	0.55	0.50	0.62
	4.6	0.21	0.16	0.22	0.33	0.51	0.48	0.62	0.64
	5.1	0.48	0.57	0.61	0.68	0.56	0.47	0.47	0.44
	5.2	0.15	0.27	0.19	0.25	0.17	0.20	0.33	0.32

Größen und Messen

Item				Schwierigkeit P_i zum Messzeitpunkt				Trennschärfe r_{pbis} zum Messzeitpunkt			
Form A	Form B	Form C	Form D	1	2	3	4	1	2	3	4
7.1	7.2	7.1	9.1	0.86	0.86	0.88	0.93	0.39	0.41	0.29	0.23
7.2				0.40	0.43	0.54	0.52	0.40	0.39	0.39	0.30
7.3				0.43	0.39	0.44	0.43	0.25	0.39	0.38	0.27
7.4	7.4	7.4	9.4	0.44	0.44	0.55	0.50	0.37	0.38	0.43	0.30
8.1	8.1			0.76	0.79	0.80	0.87	0.36	0.37	0.48	0.41
8.2				0.49	0.55	0.66	0.66	0.39	0.26	0.44	0.55
8.3				0.62	0.75	0.75	0.82	0.47	0.30	0.46	0.37
9.1	9.1	9.2	6.1	0.38	0.53	0.60	0.73	0.57	0.60	0.52	0.47
9.2				0.26	0.43	0.44	0.57	0.42	0.55	0.55	0.47
9.3	9.2	9.3	6.2	0.27	0.36	0.38	0.46	0.54	0.49	0.51	0.55
10.1	10.1	10.1	7.1	0.54	0.63	0.72	0.77	0.51	0.56	0.52	0.54
10.2	10.2	10.2	7.2	0.64	0.68	0.74	0.79	0.57	0.55	0.51	0.48
10.3				0.11	0.42	0.52	0.52	0.33	0.49	0.56	0.50
11.1				0.48	0.60	0.67	0.58	0.48	0.41	0.47	0.54
12.1				0.39	0.44	0.53	0.46	0.53	0.46	0.39	0.48
13.1				0.58	0.74	0.67	0.74	0.44	0.50	0.42	0.51
13.2				0.28	0.34	0.41	0.42	0.28	0.24	0.27	0.24
13.3				0.54	0.67	0.63	0.68	0.44	0.36	0.32	0.45
13.4				0.04	0.14	0.17	0.15	0.12	0.41	0.34	0.35
13.5				0.25	0.33	0.33	0.31	0.29	0.31	0.26	0.34
	7.1			0.52	0.66	0.70	0.70	0.34	0.54	0.17	0.18
	7.3			0.25	0.27	0.26	0.31	0.30	0.52	0.20	0.33
	8.2			0.41	0.40	0.39	0.48	0.39	0.15	0.33	0.41
	8.3			0.34	0.24	0.33	0.39	0.23	0.33	0.43	0.19
	9.3			0.35	0.34	0.35	0.62	0.53	0.61	0.59	0.54
	10.3		7.3	0.30	0.49	0.49	0.67	0.46	0.55	0.55	0.57
	11.1			0.40	0.44	0.41	0.57	0.55	0.57	0.57	0.40
	12.1			0.28	0.27	0.33	0.44	0.42	0.54	0.53	0.41
	13.1			0.38	0.44	0.39	0.49	0.52	0.38	0.47	0.48
	13.2			0.18	0.24	0.27	0.28	0.43	0.33	0.38	0.49
	13.3			0.36	0.48	0.40	0.53	0.55	0.44	0.49	0.55
	13.4			0.19	0.30	0.32	0.34	0.36	0.24	0.43	0.38
	13.5			0.49	0.69	0.52	0.67	0.47	0.42	0.5	0.56
	14.1			0.19	0.24	0.33	0.43	0.36	0.49	0.48	0.38
		7.2		0.91	0.89	0.96	0.98	0.13	0.41	0.37	0.23
		7.3		0.52	0.43	0.55	0.59	0.49	0.42	0.41	0.48
		8.1		0.35	0.49	0.54	0.56	0.44	0.57	0.37	0.43
		8.2		0.39	0.61	0.55	0.71	0.64	0.54	0.50	0.46
		9.1		0.55	0.65	0.80	0.79	0.67	0.67	0.62	0.57
		10.3		0.53	0.61	0.77	0.83	0.69	0.66	0.56	0.57
		11.1		0.41	0.42	0.64	0.73	0.55	0.54	0.54	0.60
		12.1		0.39	0.44	0.47	0.53	0.36	0.44	0.35	0.46

6.3	0.16	0.23	0.31	0.38	0.55	0.43	0.60	0.61
8.1	0.62	0.86	0.82	0.91	0.60	0.39	0.45	0.28
8.2	0.77	0.80	0.90	0.88	0.46	0.36	0.43	0.45
8.3	0.82	0.94	0.90	0.97	0.42	0.25	0.43	0.20
9.2	0.36	0.60	0.52	0.73	0.60	0.41	0.28	0.26
9.3	0.55	0.49	0.53	0.55	0.51	0.42	0.53	0.37
10.1	0.36	0.43	0.50	0.53	0.42	0.44	0.42	0.46
10.2	0.06	0.09	0.05	0.11	0.27	0.29	0.22	0.31
11.1	0.33	0.47	0.42	0.47	0.34	0.45	0.39	0.34
12.1	0.21	0.31	0.33	0.40	0.40	0.48	0.35	0.49
13.1	0.12	0.18	0.26	0.24	0.44	0.39	0.35	0.40

Daten, Häufigkeit und Wahrscheinlichkeit

Item				Schwierigkeit P_i zum Messzeitpunkt				Trennschärfe r_{pbis} zum Messzeitpunkt			
Form A	Form B	Form C	Form D	1	2	3	4	1	2	3	4
14.1	15.1	13.1	14.1	0.76	0.84	0.91	0.92	0.65	0.62	0.60	0.60
14.2	15.2	13.2	14.2	0.76	0.87	0.89	0.93	0.67	0.62	0.58	0.50
14.3	15.3	13.3	14.3	0.79	0.86	0.89	0.93	0.66	0.61	0.62	0.57
15.1				0.34	0.44	0.71	0.66	0.69	0.71	0.64	0.69
15.2				0.32	0.41	0.63	0.56	0.68	0.75	0.68	0.76
15.3		15.3		0.16	0.21	0.26	0.27	0.48	0.50	0.60	0.62
15.4				0.15	0.22	0.30	0.42	0.67	0.62	0.63	0.65
16.1	16.1	16.1	15.1	0.41	0.56	0.62	0.71	0.54	0.65	0.66	0.65
16.2				0.27	0.46	0.47	0.46	0.53	0.48	0.56	0.61
	16.2			0.44	0.48	0.63	0.68	0.63	0.69	0.69	0.78
		14.1		0.12	0.11	0.19	0.18	0.36	0.39	0.45	0.26
		14.2		0.07	0.16	0.19	0.22	0.16	0.37	0.43	0.42
		15.1		0.3	0.37	0.50	0.49	0.58	0.67	0.74	0.60
		15.2		0.41	0.43	0.48	0.50	0.64	0.76	0.75	0.72
		15.4		0.45	0.44	0.60	0.58	0.62	0.71	0.67	0.60
		16.2		0.30	0.49	0.57	0.69	0.45	0.61	0.54	0.55
			15.2	0.41	0.50	0.62	0.62	0.78	0.78	0.77	0.81

Raum und Form

Item				Schwierigkeit P_i zum Messzeitpunkt				Trennschärfe r_{pbis} zum Messzeitpunkt			
Form A	Form B	Form C	Form D	1	2	3	4	1	2	3	4
17.1				0.52	0.70	0.67	0.75	0.59	0.51	0.55	0.57
17.2				0.51	0.64	0.71	0.77	0.52	0.53	0.54	0.62
18.1	18.2			0.60	0.82	0.86	0.90	0.70	0.54	0.47	0.49
18.2	18.1	18.2	17.2	0.73	0.89	0.91	0.94	0.61	0.54	0.43	0.46
18.3	18.3	18.1	17.1	0.36	0.52	0.62	0.68	0.48	0.55	0.46	0.52
19.1				0.79	0.91	0.92	0.94	0.65	0.56	0.47	0.53
19.2	20.2	20.1	20.1	0.50	0.65	0.72	0.76	0.65	0.53	0.57	0.56
19.3	20.3	20.3	20.3	0.59	0.68	0.73	0.79	0.60	0.48	0.41	0.43
19.4	20.4	20.4	20.4	0.23	0.28	0.29	0.38	0.24	0.16	0.28	0.25
20.1				0.32	0.38	0.47	0.49	0.50	0.41	0.42	0.54
21.1				0.16	0.14	0.25	0.29	0.45	0.45	0.54	0.58
	17.1			0.81	0.96	0.94	0.99	0.72	0.40	0.47	0.23
	17.2			0.38	0.43	0.48	0.48	0.42	0.35	0.48	0.41
	19.1	19.1		0.71	0.89	0.91	0.89	0.72	0.55	0.45	0.51
	19.2			0.53	0.58	0.67	0.82	0.65	0.51	0.62	0.54
	19.3			0.33	0.50	0.45	0.45	0.54	0.51	0.53	0.51
	20.1			0.27	0.14	0.26	0.30	0.50	0.19	0.43	0.50
	21.1			0.06	0.13	0.26	0.34	0.29	0.35	0.59	0.49
	22.1			0.56	0.65	0.74	0.90	0.62	0.61	0.63	0.35
		17.1		0.72	0.79	0.68	0.81	0.50	0.51	0.36	0.43
		17.2		0.49	0.53	0.58	0.65	0.48	0.57	0.43	0.51
		18.3		0.16	0.26	0.23	0.25	0.44	0.43	0.31	0.42
		19.2		0.74	0.83	0.89	0.90	0.70	0.52	0.46	0.53
		19.3		0.26	0.34	0.05	0.03	0.57	0.55	0.08	0.08
		20.2		0.54	0.66	0.76	0.81	0.63	0.70	0.64	0.58
		21.1		0.62	0.69	0.58	0.68	0.60	0.51	0.42	0.59
		22.1		0.15	0.29	0.38	0.50	0.39	0.51	0.53	0.58
			16.1	0.89	0.94	0.99	0.99	0.45	0.38	0.15	0.30
			16.2	0.61	0.68	0.76	0.79	0.43	0.39	0.34	0.26
			17.3	0.27	0.29	0.37	0.34	0.37	0.32	0.40	0.21
			18.1	0.74	0.90	0.82	0.86	0.57	0.39	0.53	0.48
			19.1	0.54	0.68	0.74	0.74	0.49	0.38	0.50	0.37
			20.2	0.22	0.33	0.37	0.43	0.56	0.46	0.60	0.51
			21.1	0.48	0.59	0.59	0.61	0.71	0.64	0.61	0.61
			21.2	0.41	0.51	0.41	0.56	0.75	0.70	0.68	0.67
			21.3	0.57	0.63	0.57	0.57	0.75	0.72	0.66	0.63
			21.4	0.41	0.52	0.36	0.47	0.73	0.73	0.69	0.67
			21.5	0.27	0.21	0.28	0.35	0.66	0.50	0.71	0.55

Anhang B: Itemkennwerte der Rasch-Analyse
Zahlen und Operationen

Form A	Form B	Form C	Form D	N	Schwierig- keit	Standard- fehler	wMNSQ	t-wMNSQ
1.1				456	0.52	0.085	1.21	3.9
1.2				456	0.91	0.086	1.14	2.6
2.1				456	-3.20	0.111	0.99	0.0
2.2				456	-2.23	0.101	1.12	1.2
2.3	2.2	2.1	2.3	1817	-1.77	0.062	0.94	-1.6
2.4	2.3	2.2	2.4	1817	-0.37	0.053	0.98	-0.9
2.5	4.4	2.5	2.7	1817	-0.67	0.054	0.99	-0.4
2.6				456	-1.75	0.096	0.95	-0.6
2.7				456	-0.41	0.087	1.01	0.2
2.8	4.5	2.8	2.8	1817	-0.41	0.053	0.87	-5.1
3.1				456	-0.76	0.089	1.10	1.7
3.2				456	1.76	0.089	0.91	-1.5
4.1	2.4	2.3	4.1	1817	-1.24	0.057	0.86	-4.5
4.2				456	-0.24	0.086	0.91	-1.7
4.3	2.6	2.4	4.2	1817	-0.20	0.052	0.91	-3.4
4.4				456	1.45	0.088	0.93	-1.3
4.5				456	1.52	0.088	0.94	-1.1
4.6				456	1.20	0.087	0.94	-1.2
5.1				456	1.87	0.09	0.97	-0.4
5.2				456	2.37	0.094	0.89	-1.5
6.1				456	1.95	0.091	1.16	2.4
	1.1			450	-1.13	0.09	1.17	2.6
	1.2			450	-0.65	0.087	1.02	0.4
	1.3			450	-1.16	0.09	0.94	-1.0
	1.4			450	-0.81	0.088	1.01	0.2
	2.1			450	-1.72	0.094	1.06	0.8
	2.5			450	-2.40	0.100	0.95	-0.6
	3.1			450	-2.33	0.099	1.08	0.9
	3.2			450	-0.92	0.088	1.00	0.0
	3.3			450	-1.11	0.089	1.03	0.5
	3.4			450	0.50	0.086	1.03	0.6
	4.1			450	0.87	0.087	0.94	-1.2
	4.2			450	1.21	0.089	0.88	-2.2
	4.3			450	2.42	0.098	0.80	-2.4
	4.6			450	0.24	0.086	0.94	-1.2
	5.1			450	0.94	0.087	1.09	1.6
	5.2			450	3.35	0.107	0.89	-0.9
	6.1			450	0.19	0.086	1.26	4.8
	6.2			450	0.51	0.086	1.12	2.2
		1.1		461	-1.33	0.340	1.15	2.1
		1.2		461	-0.62	0.086	1.20	3.4
		1.3		461	0.01	0.084	1.22	4.2

1.4	461	-0.10	0.084	1.22	4.2
2.6	461	-0.66	0.086	0.85	-2.9
2.7	461	-0.08	0.084	1.02	0.4
3.1	461	-0.29	0.085	0.91	-1.8
3.2	461	-0.03	0.084	1.03	0.7
3.3	461	1.06	0.086	1.19	3.5
4.1	461	0.12	0.084	1.03	0.7
4.2	461	0.32	0.084	0.93	-1.6
4.3	461	0.71	0.085	0.90	-2.2
5.1	461	2.20	0.094	0.98	-0.2
6.1	461	-0.84	0.087	1.15	2.5
6.2	461	1.33	0.087	1.11	1.9
1.1	450	-0.33	0.086	1.13	2.4
1.2	450	-0.56	0.088	1.08	1.5
2.1	450	-3.36	0.114	0.99	0.0
2.2	450	-1.84	0.098	0.98	-0.2
2.5	450	-2.91	0.109	0.93	-0.5
2.6	450	-1.76	0.097	0.92	-1.0
3.1	450	-1.45	0.094	1.12	1.5
3.2	450	0.33	0.085	1.10	2.0
3.3	450	4.25	0.115	1.04	0.3
3.4	450	3.78	0.110	0.97	-0.2
4.3	450	-0.78	0.089	0.99	-0.2
4.4	450	0.14	0.085	0.78	-4.8
4.5	450	0.21	0.085	0.97	-0.7
4.6	450	2.04	0.091	0.85	-2.4
5.1	450	-0.06	0.086	1.09	1.8
5.2	450	2.19	0.074	1.24	3.2

Anmerkungen: wMNSQ – Weighted Mean Square Residual; t-wMNSQ – T-Wert des wMNSQ

Größen und Messen

Form A	Form B	Form C	Form D	N	Schwierigkeit	Standardfehler	wMNSQ	t-wMNSQ
7.1	7.2	7.1	9.1	1818	-2.45	0.067	1.10	0.1
7.2				456	0.19	0.080	1.07	2.4
7.3				456	0.42	0.081	1.08	3.2
7.4	7.4	7.4	9.4	1818	0.12	0.049	1.04	5.6
8.1	8.1			917	-1.68	0.074	1.10	-0.1
8.2				456	-0.39	0.081	1.08	1.0
8.3				456	-1.18	0.086	1.11	0.0
9.1	9.1	9.2	6.1	1818	-0.28	0.049	1.04	-4.9
9.2				456	0.42	0.081	1.08	-1.4
9.3	9.2	9.3	6.2	1818	0.71	0.050	1.04	-3.0
10.1	10.1	10.1	7.1	1818	-0.83	0.051	1.05	-3.7
10.2	10.2	10.2	7.2	1818	-1.09	0.052	1.05	-3.2
10.3				456	0.57	0.081	1.08	-1.4
11.1				456	-0.34	0.081	1.08	-0.2
12.1				456	0.28	0.080	1.07	-0.2
13.1				456	-0.86	0.083	1.10	-0.1
13.2				456	0.75	0.082	1.08	3.7
13.3				456	-0.59	0.082	1.09	1.5
13.4				456	2.38	0.097	1.20	-0.1
13.5				456	1.05	0.084	1.10	2.4
	7.1			461	-0.76	0.082	1.09	3.3
	7.3			461	1.20	0.085	1.11	1.8
	8.2			461	0.38	0.081	1.08	3.4
	8.3			461	0.88	0.083	1.09	3.0
	9.3			461	0.40	0.081	1.08	-2.9
	10.3		7.3	911	0.12	0.064	1.06	-3.0
	11.1			461	0.22	0.081	1.08	-1.5
	12.1			461	0.85	0.083	1.09	-0.5
	13.1			461	0.35	0.081	1.08	0.6
	13.2			461	1.39	0.087	1.12	0.4
	13.3			461	0.26	0.081	1.08	-0.7
	13.4			461	1.10	0.085	1.10	1.4
	13.5			461	-0.46	0.081	1.08	-0.2
	14.1			461	1.00	0.084	1.10	-0.2
		7.2		451	-3.28	0.107	1.29	-0.1
		7.3		451	-0.13	0.082	1.09	2.2
		8.1		451	0.04	0.082	1.08	1.4
		8.2		451	-0.40	0.083	1.09	-0.2
		9.1		451	-1.13	0.086	1.11	-3.0
		10.3		451	-1.06	0.085	1.11	-3.1
		11.1		451	-0.30	0.082	1.09	-1.6
		12.1		451	0.20	0.082	1.08	2.8
			6.3	450	1.35	0.086	1.11	-2.0

8.1	450	-1.58	0.091	1.14	-0.6
8.2	450	-1.90	0.094	1.17	-0.6
8.3	450	-2.68	0.103	1.25	-0.4
9.2	450	-0.13	0.082	1.08	1.5
9.3	450	-0.03	0.081	1.08	2.2
10.1	450	0.35	0.081	1.08	1.1
10.2	450	3.07	0.105	1.27	0.1
11.1	450	0.53	0.082	1.08	2.4
12.1	450	1.11	0.084	1.10	0.8
13.1	450	1.82	0.059	1.14	0.0

Anmerkungen: wMNSQ – Weighted Mean Square Residual; t-wMNSQ – T-Wert des wMNSQ

Daten, Häufigkeit und Wahrscheinlichkeit

Form A	Form B	Form C	Form D	N	Schwierigkeit	Standardfehler	wMNSQ	t-wMNSQ
14.1	15.1	13.1	14.1	1794	-3.13	0.073	0.80	-4.0
14.2	15.2	13.2	14.2	1794	-3.19	0.073	0.87	-2.4
14.3	15.3	13.3	14.3	1794	-3.28	0.074	0.79	-4.0
15.1				449	-0.22	0.089	0.99	-0.1
15.2				449	0.18	0.089	0.92	-1.3
15.3		15.3		882	2.08	0.079	1.03	0.6
15.4				449	1.71	0.094	0.95	-0.6
16.1	16.1	16.1	15.1	1794	-0.52	0.057	1.07	2.1
16.2				449	0.67	0.090	1.29	4.2
	16.2			466	-0.67	0.091	1.09	1.3
		14.1		433	2.77	0.101	1.25	2.5
		14.2		433	2.65	0.100	1.22	2.3
		15.1		433	0.67	0.090	0.99	-0.2
		15.2		433	0.41	0.089	0.93	-1.2
		15.4		433	-0.02	0.090	1.07	1.1
		16.2		433	0.03	0.090	1.21	3.3
			15.2	446	-0.12	0.035	0.84	-2.6

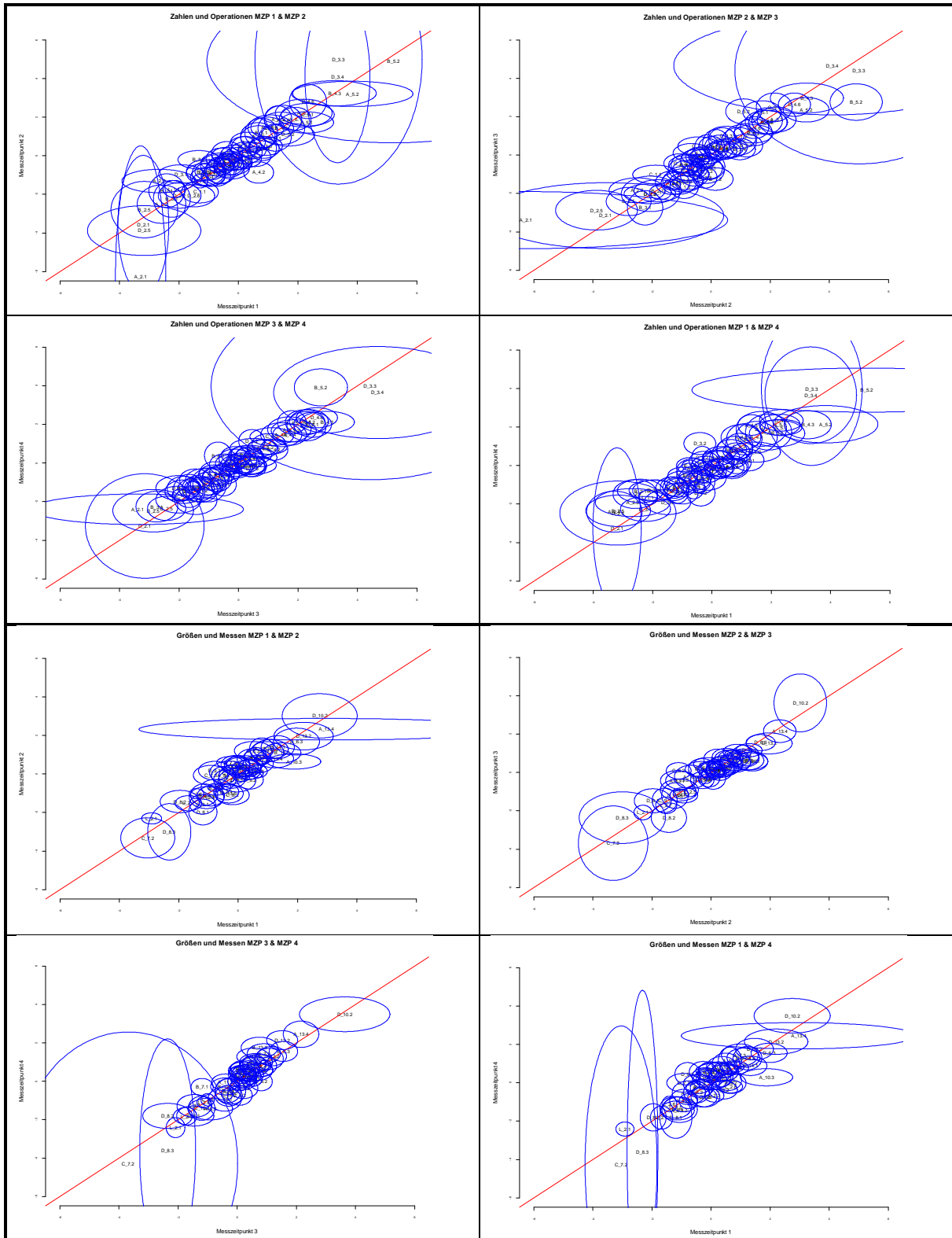
Anmerkungen: wMNSQ – Weighted Mean Square Residual; t-wMNSQ – T-Wert des wMNSQ

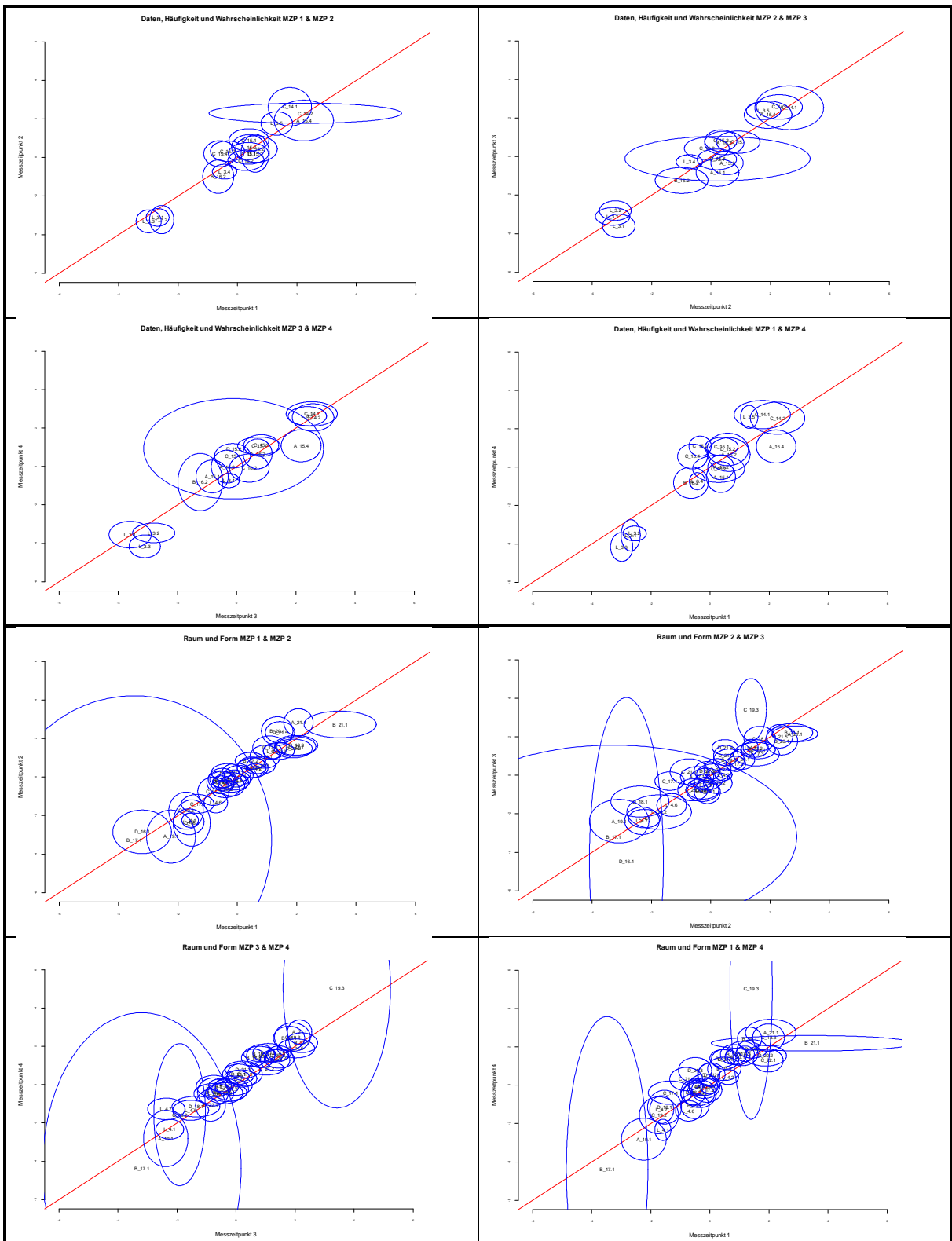
Raum und Form

Form A	Form B	Form C	Form D	N	Schwierigkeit	Standardfehler	wMNSQ	t-wMNSQ
17.1				468	-0.36	0.082	1.00	0.0
17.2				468	-0.34	0.082	1.01	0.1
18.1	18.2			937	-1.33	0.074	0.90	-1.9
18.2	18.1	18.2	17.2	1834	-1.98	0.065	0.88	-2.4
18.3	18.3	18.1	17.1	1834	0.23	0.049	1.02	1.1
19.1				468	-2.21	0.097	0.84	-1.5
19.2	20.2	20.1	20.1	1834	-0.39	0.051	0.94	-2.5
19.3	20.3	20.3	20.3	1834	-0.62	0.052	1.04	1.4
19.4	20.4	20.4	20.4	1834	1.62	0.052	1.27	9.5
20.1				468	0.96	0.080	1.08	1.8
21.1				468	2.12	0.087	0.92	-1.2
	17.1			469	-2.87	0.102	0.75	-1.9
	17.2			469	0.80	0.080	1.15	3.4
	19.1	19.1		921	-1.86	0.079	0.85	-2.2
	19.2			469	-0.36	0.082	0.96	-0.8
	19.3			469	0.87	0.080	1.04	1.0
	20.1			469	1.95	0.085	1.05	0.8
	21.1			469	2.26	0.087	0.92	-1.2
	22.1			469	-0.76	0.084	0.92	-1.3
		17.1		452	-0.96	0.086	1.12	1.8
		17.2		452	0.14	0.080	1.06	1.4
		18.3		452	2.04	0.086	1.03	0.5
		19.2		452	-1.73	0.092	0.87	-1.4
		19.3		452	2.48	0.090	1.11	1.3
		20.2		452	-0.62	0.083	0.86	-2.6
		21.1		452	-0.32	0.082	1.06	1.1
		22.1		452	1.37	0.082	0.95	-1.0
			16.1	445	-3.38	0.107	0.90	-0.5
			16.2	445	-0.73	0.084	1.18	2.9
			17.3	445	1.46	0.083	1.19	3.5
			18.1	445	-1.63	0.091	0.93	-0.9
			19.1	445	-0.50	0.083	1.09	1.7
			20.2	445	1.34	0.083	0.97	-0.6
			21.1	445	0.11	0.081	0.87	-2.8
			21.2	445	0.60	0.081	0.81	-4.6
			21.3	445	0.02	0.081	0.85	-3.3
			21.4	445	0.79	0.081	0.82	-4.4
			21.5	445	1.70	0.050	0.88	-2.1

Anmerkungen: wMNSQ – Weighted Mean Square Residual; t-wMNSQ – T-Wert des wMNSQ

Anhang C: Messinvarianz der Testskalen des „Mathes 3“ über die Messzeitpunkte





Anmerkungen: Zur Prüfung der Messinvarianz der Aufgaben über die Zeit wurden die Itemparameter im Rahmen von graphischen Modelltests jeweils an der X- und Y-Achse abgetragen. Die Konfidenzintervalle (95 %) werden durch Ellipsen dargestellt.

Anhang D: Statusdiagnostische Normtabellen des „Mathes 3“

Mathes 3 Form A (10./11. Schulwoche)

	Rohwert	f _{cum}	Prozentrang	T-Wert		Rohwert	f _{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	0	0	<33	Größen und Messen	0	2	2	26
	1	0	0	<33		1	2	2	26
	2	8	7	33		2	7	6	32
	3	13	11	35		3	10	9	35
	4	19	16	37		4	18	16	37
	5	19	16	37		5	23	20	40
	6	25	21	41		6	30	26	43
	7	31	27	43		7	43	38	45
	8	43	37	45		8	51	45	48
	9	51	44	47		9	63	55	51
	10	62	53	49		10	78	68	53
	11	66	56	51		11	86	75	56
	12	75	64	53		12	93	82	59
	13	84	72	55		13	103	90	62
	14	93	80	57		14	108	95	64
	15	100	86	59		15	112	98	67
	16	103	88	61		16	114	100	70
	17	107	92	63		17	114	100	>70
	18	111	95	65		18	114	100	>70
	19	112	96	67		19	114	100	>70
	20	113	97	69		20	114	100	>70
21	117	100	71						

	Rohwert	f _{cum}	Prozentrang	T-Wert		Rohwert	f _{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	8	7	32	Raum und Form	0	10	9	31
	1	13	12	36		1	11	10	34
	2	22	20	41		2	2	19	38
	3	47	43	45		3	31	27	42
	4	65	59	50		4	39	34	45
	5	85	77	54		5	56	49	49
	6	94	86	58		6	72	63	53
	7	100	91	63		7	85	75	56
	8	104	95	67		8	100	88	60
	9	110	100	72		9	110	97	63
						10	114	100	67
				11	114	100	>67		

Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form A (20./21. Schulwoche)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	0	0	<27	Größen und Messen	0	0	0	<23
	1	1	1	27		1	1	1	23
	2	2	2	30		2	1	1	23
	3	4	4	31		3	1	1	23
	4	9	8	34		4	6	6	32
	5	13	12	36		5	8	7	34
	6	17	16	38		6	13	12	37
	7	26	24	40		7	21	19	40
	8	30	28	42		8	34	31	42
	9	34	32	44		9	40	37	45
	10	40	37	46		10	50	46	48
	11	48	45	49		11	58	53	51
	12	57	53	51		12	73	67	53
	13	67	63	53		13	86	79	56
	14	70	65	55		14	91	84	59
	15	81	76	57		15	97	89	61
	16	91	85	59		16	103	95	64
	17	98	92	61		17	106	97	67
	18	101	95	64		18	108	99	70
	19	104	97	66		19	108	99	70
	20	105	98	68		20	109	100	75
21	107	100	70						

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	3	3	27	Raum und Form	0	3	3	22
	1	7	7	32		1	3	3	22
	2	10	10	36		2	6	5	30
	3	27	26	41		3	11	10	35
	4	40	38	45		4	23	21	39
	5	68	65	50		5	31	28	43
	6	75	71	55		6	43	39	47
	7	88	84	59		7	69	62	52
	8	100	95	64		8	88	79	56
	9	105	100	68		9	101	91	60
						10	109	98	64
				11	111	100	69		

Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form A (30./31. Schulwoche)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	0	0	<26	Größen und Messen	0	0	0	<25
	1	1	1	26		1	0	0	<25
	2	1	1	26		2	1	1	25
	3	3	3	31		3	2	2	27
	4	10	9	33		4	5	4	30
	5	13	11	35		5	11	9	33
	6	17	15	37		6	15	13	35
	7	22	19	39		7	19	16	38
	8	26	22	41		8	26	22	40
	9	30	26	43		9	35	30	43
	10	41	35	45		10	38	33	46
	11	50	43	47		11	43	37	48
	12	58	50	49		12	59	50	51
	13	65	56	51		13	74	63	53
	14	76	65	54		14	88	75	56
	15	83	71	56		15	99	85	58
	16	96	82	58		16	107	92	61
	17	97	83	60		17	113	97	64
	18	105	90	62		18	115	98	66
	19	110	94	64		19	117	100	69
	20	114	97	66		20	117	100	>69
21	117	100	68						

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	2	2	24	Raum und Form	0	0	0	<22
	1	4	3	29		1	0	0	<22
	2	11	9	33		2	1	1	22
	3	21	18	38		3	4	3	28
	4	29	24	42		4	10	8	33
	5	53	45	47		5	23	19	38
	6	71	60	51		6	39	32	43
	7	89	75	56		7	60	49	48
	8	104	87	60		8	86	70	53
	9	119	100	64		9	108	88	59
				10	109	98	10		
				11	111	100	11		

Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form A (3./4. Schulwoche Klasse 4)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	0	0	<30	Größen und Messen	0	0	0	<28
	1	0	0	<30		1	0	0	<28
	2	3	3	30		2	0	0	<28
	3	7	6	32		3	1	1	28
	4	10	9	34		4	4	4	30
	5	13	11	36		5	8	7	33
	6	20	18	38		6	12	10	35
	7	25	22	40		7	17	15	38
	8	28	25	42		8	26	23	40
	9	37	33	43		9	38	33	43
	10	42	37	45		10	45	39	46
	11	50	44	47		11	56	49	48
	12	54	47	49		12	63	55	51
	13	58	51	51		13	71	62	53
	14	65	57	53		14	78	68	56
	15	73	64	55		15	90	78	58
	16	81	71	57		16	105	91	61
	17	90	79	59		17	110	96	64
	18	101	89	60		18	114	99	66
	19	105	92	62		19	115	100	69
	20	111	97	64		20	115	100	>69
21	114	100	66						

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	3	3	26	Raum und Form	0	2	2	18
	1	5	4	30		1	3	3	22
	2	9	8	34		2	6	5	27
	3	23	20	38		3	7	6	31
	4	36	32	42		4	11	9	35
	5	49	43	47		5	17	14	39
	6	63	55	51		6	30	25	43
	7	78	68	55		7	49	41	47
	8	97	85	59		8	71	60	51
	9	114	100	63		9	87	73	55
						10	109	98	59
				11	111	100	63		

Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form B (10./11. Schulwoche)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	0	0	<29	Größen und Messen	0	1	1	29
	1	0	0	<29		1	4	4	31
	2	1	1	29		2	6	5	33
	3	2	2	30		3	14	12	36
	4	8	7	32		4	17	15	38
	5	12	11	34		5	26	23	41
	6	17	15	36		6	35	31	43
	7	19	17	38		7	44	39	46
	8	23	21	40		8	52	46	48
	9	30	27	42		9	65	58	51
	10	37	33	44		10	76	67	53
	11	40	36	46		11	87	77	56
	12	45	41	48		12	94	83	58
	13	50	45	50		13	98	87	60
	14	65	59	52		14	102	80	63
	15	76	69	54		15	105	93	65
	16	81	73	56		16	108	96	68
	17	89	80	58		17	112	99	70
	18	95	86	60		18	113	100	73
	19	101	91	62		19	113	100	>73
	20	103	93	64		20	113	100	>73
	21	108	97	66		21	113	100	>73
	22	109	98	68					
	23	111	100	70					
24	111	100	>70						

	Rohwert	f_{cum}	Rohwert	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	9	8	30	Raum und Form	0	16	14	33
	1	23	21	36		1	21	18	36
	2	33	30	43		2	25	22	38
	3	58	52	50		3	27	24	41
	4	94	84	56		4	32	28	43
	5	112	100	63		5	40	35	46
						6	47	41	49
						7	54	47	51
						8	71	62	54
						9	84	74	56
						10	98	86	59
						11	106	93	62
						12	114	100	64
						13	114	100	>64
						14	114	100	>64

Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form B (20./21. Schulwoche)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	1	1	26	Größen und Messen	0	1	1	28
	1	2	2	28		1	4	4	31
	2	4	4	30		2	5	5	33
	3	4	4	30		3	7	7	35
	4	6	6	34		4	13	12	37
	5	14	14	35		5	21	19	40
	6	16	16	37		6	27	25	42
	7	19	18	39		7	36	33	44
	8	25	24	41		8	48	44	46
	9	29	28	43		9	53	49	49
	10	35	34	45		10	67	62	51
	11	38	37	46		11	73	68	53
	12	41	40	48		12	79	73	55
	13	53	52	50		13	83	77	58
	14	57	55	52		14	93	86	60
	15	64	62	54		15	97	90	62
	16	69	67	56		16	100	93	65
	17	80	78	57		17	105	97	67
	18	87	85	59		18	106	98	69
	19	89	86	61		19	107	99	71
	20	98	95	63		20	107	99	71
	21	100	97	65		21	108	100	76
	22	102	99	67					
	23	103	100	68					
24	103	100	>68						

	Rohwert	f_{cum}	Rohwert	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	11	10	30	Raum und Form	0	1	1	17
	1	17	16	36		1	1	1	17
	2	27	25	42		2	3	3	25
	3	53	49	48		3	7	7	29
	4	70	65	54		4	8	8	33
	5	108	100	60		5	14	13	37
						6	23	22	41
						7	31	30	45
						8	52	50	49
						9	70	67	53
						10	88	84	57
						11	101	96	61
						12	104	99	65
						13	105	100	70
						14	105	100	>70

 Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form B (30./31. Schulwoche)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	2	2	30	Größen und Messen	0	0	0	<32
	1	5	4	31		1	1	1	32
	2	7	6	33		2	6	5	34
	3	8	7	34		3	9	8	36
	4	15	13	36		4	17	14	38
	5	17	15	37		5	27	23	40
	6	20	17	39		6	39	33	42
	7	24	21	40		7	46	38	45
	8	29	25	42		8	55	46	47
	9	36	31	43		9	60	50	49
	10	41	35	45		10	68	57	51
	11	48	41	47		11	79	66	53
	12	51	44	48		12	88	73	56
	13	57	49	50		13	97	81	58
	14	62	53	51		14	104	87	60
	15	68	59	53		15	107	89	62
	16	74	64	54		16	112	93	64
	17	80	69	56		17	114	95	66
	18	86	74	57		18	115	96	69
	19	93	80	59		19	117	98	71
	20	99	85	60		20	120	100	73
	21	102	88	62		21	120	100	>73
	22	107	92	64					
	23	115	99	65					
24	116	100	67						

	Rohwert	f_{cum}	Rohwert	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	9	7	27	Raum und Form	0	3	2	22
	1	18	15	33		1	4	3	25
	2	24	20	39		2	5	4	29
	3	45	37	46		3	7	6	32
	4	62	50	52		4	12	10	35
	5	123	100	58		5	16	13	38
						6	32	26	41
						7	39	31	45
						8	52	42	48
						9	72	58	51
						10	93	74	54
						11	103	82	58
						12	111	89	61
						13	119	95	64
						14	125	100	67

 Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form B (3./4. Schulwoche Klasse 4)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	0	0	<30	Größen und Messen	0	0	0	<28
	1	0	0	<30		1	0	0	<28
	2	0	0	<30		2	0	0	<28
	3	2	2	30		3	1	1	28
	4	6	5	31		4	4	3	30
	5	9	8	33		5	10	8	33
	6	14	12	35		6	13	11	35
	7	18	15	36		7	18	15	37
	8	22	18	38		8	24	20	40
	9	25	21	40		9	30	25	42
	10	30	25	41		10	42	35	45
	11	33	28	43		11	49	41	47
	12	39	33	44		12	63	53	50
	13	46	38	46		13	74	62	52
	14	47	39	48		14	87	73	55
	15	53	44	49		15	93	78	57
	16	62	52	51		16	102	85	60
	17	70	58	53		17	108	90	62
	18	76	63	54		18	114	95	65
	19	80	67	56		19	119	99	67
	20	88	73	58		20	119	99	67
	21	97	81	59		21	120	100	72
	22	103	86	61					
	23	113	94	62					
24	120	100	64						

	Rohwert	f_{cum}	Rohwert	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	3	2	17	Raum und Form	0	0	0	<21
	1	6	5	25		1	0	0	<21
	2	13	11	33		2	0	0	<21
	3	32	26	41		3	2	2	21
	4	51	42	49		4	4	3	25
	5	123	100	57		5	8	6	29
						6	10	8	34
						7	17	14	38
						8	32	26	43
						9	53	42	47
						10	78	62	51
						11	98	78	56
						12	112	90	60
						13	122	98	64
				14	125	100	69		

 Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form C (10./11. Schulwoche)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	1	1	29	Größen und Messen	0	1	1	30
	1	4	4	31		1	2	2	33
	2	10	9	33		2	11	10	36
	3	11	10	35		3	21	20	39
	4	14	12	37		4	35	33	42
	5	20	17	38		5	46	43	45
	6	28	24	40		6	54	51	47
	7	34	30	42		7	56	53	50
	8	38	33	44		8	66	62	53
	9	41	36	46		9	73	69	56
	10	48	42	48		10	81	76	59
	11	56	49	50		11	98	93	61
	12	60	52	52		12	100	94	64
	13	72	63	53		13	104	98	67
	14	79	69	55		14	106	100	70
	15	88	77	57					
	16	96	84	59					
	17	102	89	61					
	18	106	92	63					
	19	111	97	65					
	20	113	98	67					
21	115	100	68						

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	17	17	33	Raum und Form	0	8	7	29
	1	20	19	37		1	8	7	29
	2	22	21	41		2	12	11	35
	3	37	36	45		3	20	18	38
	4	45	44	48		4	29	27	41
	5	59	57	52		5	31	28	45
	6	77	75	56		6	43	39	48
	7	93	90	60		7	59	54	51
	8	101	98	64		8	73	67	54
	9	103	100	67		9	88	81	57
	10	103	100	>67		10	96	88	60
	11	103	100	>67		11	101	93	63
				12	107	98	66		
				13	108	99	69		
				14	109	100	72		

Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form C (20./21. Schulwoche)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert	
Zahlen und Operationen	0	0	0	<33	Größen und Messen	0	2	2	30	
	1	5	4	33		1	3	3	32	
	2	10	9	34		2	14	12	35	
	3	17	15	36		3	19	17	38	
	4	21	18	38		4	28	25	40	
	5	27	23	40		5	36	32	43	
	6	31	27	41		6	45	40	46	
	7	35	30	43		7	52	46	48	
	8	39	33	45		8	62	55	51	
	9	45	39	47		9	74	66	54	
	10	53	45	48		10	81	72	56	
	11	60	51	50		11	90	80	59	
	12	69	59	52		12	98	87	61	
	13	74	63	53		13	111	98	64	
	14	81	69	55		14	113	100	67	
	15	88	75	57						
	16	91	78	59						
	17	94	80	60						
	18	104	89	62						
	19	110	94	64						
	20	115	98	66						
21	117	100	67							

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	9	8	31	Raum und Form	0	6	5	25
	1	12	11	35		1	6	5	25
	2	16	15	38		2	7	6	31
	3	32	29	42		3	12	11	35
	4	43	39	45		4	14	12	37
	5	56	51	49		5	21	19	41
	6	68	62	52		6	32	28	44
	7	76	70	56		7	40	35	47
	8	91	84	59		8	55	49	50
	9	105	96	63		9	73	65	53
	10	108	99	66		10	84	74	56
	11	109	100	70		11	96	85	59
				12	106	94	62		
				13	111	98	65		
				14	113	100	68		

Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form C (30./31. Schulwoche)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	0	0	<28	Größen und Messen	0	0	0	<22
	1	3	3	28		1	2	2	22
	2	5	4	30		2	2	2	22
	3	7	6	32		3	5	4	28
	4	12	10	34		4	8	7	32
	5	14	12	36		5	13	11	35
	6	22	19	38		6	21	18	39
	7	25	21	40		7	32	27	42
	8	32	27	42		8	42	35	45
	9	34	29	44		9	54	45	49
	10	40	34	46		10	71	60	52
	11	51	43	48		11	87	73	56
	12	58	49	50		12	107	90	59
	13	65	55	52		13	113	95	63
	14	75	64	54		14	119	100	66
	15	80	68	56					
	16	89	75	58					
	17	98	83	59					
	18	108	92	61					
	19	111	94	63					
	20	117	99	65					
21	118	100	67						

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	2	2	26	Raum und Form	0	0	0	<19
	1	5	4	30		1	1	1	19
	2	8	7	34		2	2	2	23
	3	17	15	38		3	5	4	28
	4	31	27	42		4	7	6	32
	5	48	43	45		5	15	13	36
	6	61	54	49		6	25	21	40
	7	73	65	53		7	39	33	44
	8	86	76	57		8	58	48	48
	9	101	89	61		9	77	64	53
	10	110	97	65		10	97	81	57
	11	113	100	68		11	110	92	61
						12	119	99	65
						13	120	100	69
				14	120	100	>69		

Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form C (3./4. Schulwoche Klasse 4)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	0	0	<25	Größen und Messen	0	0	0	<19
	1	2	2	25		1	1	1	19
	2	3	3	27		2	2	2	23
	3	4	4	29		3	4	4	26
	4	6	5	31		4	6	5	30
	5	9	8	33		5	12	11	33
	6	13	12	35		6	16	14	37
	7	18	16	37		7	22	20	40
	8	19	17	39		8	33	30	44
	9	25	23	42		9	44	39	47
	10	32	29	44		10	55	49	51
	11	38	34	46		11	73	65	54
	12	47	42	48		12	96	86	57
	13	54	49	50		13	108	96	61
	14	60	54	52		14	112	100	64
	15	70	63	54					
	16	80	72	56					
	17	89	80	58					
	18	99	89	60					
	19	105	95	62					
	20	110	99	65					
21	111	100	67						

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	1	1	22	Raum und Form	0	3	3	16
	1	3	3	26		1	3	3	16
	2	5	5	30		2	4	4	23
	3	12	11	35		3	4	4	23
	4	22	20	39		4	4	4	23
	5	32	30	43		5	8	7	35
	6	49	45	48		6	15	14	39
	7	69	64	52		7	26	24	42
	8	85	79	56		8	39	36	46
	9	98	91	61		9	60	55	50
	10	107	99	65		10	75	68	54
	11	108	100	69		11	92	84	58
				12	105	96	61		
				13	110	100	65		
				14	110	100	>65		

Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form D (10./11. Schulwoche)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	1	1	25	Größen und Messen	0	0	0	<32
	1	1	1	25		1	5	5	32
	2	4	4	29		2	10	9	35
	3	8	7	31		3	15	13	37
	4	9	8	33		4	22	20	40
	5	12	11	35		5	34	30	43
	6	16	14	37		6	45	40	45
	7	20	18	39		7	55	49	48
	8	29	26	41		8	65	58	50
	9	34	31	43		9	79	71	53
	10	40	36	45		10	84	75	56
	11	45	41	47		11	89	80	58
	12	50	45	49		12	91	81	61
	13	53	48	51		13	101	90	63
	14	63	57	53		14	108	96	66
	15	71	64	55		15	111	99	69
	16	78	70	57		16	112	100	71
	17	92	83	59		17	112	100	>71
	18	104	94	61		18	112	100	>71
	19	109	98	63					
	20	110	99	65					
	21	111	100	67					
22	111	100	>67						

	Rohwert	f_{cum}	Rohwert	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	13	12	31	Raum und Form	0	5	4	31
	1	22	20	37		1	7	6	33
	2	33	29	43		2	11	10	36
	3	59	52	49		3	18	16	38
	4	78	69	55		4	29	26	41
	5	113	100	61		5	33	29	43
						6	42	37	45
						7	58	51	48
						8	65	58	50
						9	68	60	53
						10	77	68	55
						11	87	77	58
						12	96	85	60
						13	102	90	62
						14	109	97	65
						15	110	97	67
						16	113	100	70

 Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form D (20./21. Schulwoche)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	1	1	22	Größen und Messen	0	0	0	<31
	1	1	1	22		1	0	0	<31
	2	1	1	22		2	0	0	<31
	3	4	4	29		3	2	2	31
	4	6	6	31		4	9	9	34
	5	11	11	33		5	17	16	37
	6	12	11	35		6	20	19	39
	7	15	14	37		7	32	31	42
	8	20	19	40		8	36	35	45
	9	26	25	42		9	50	48	48
	10	32	31	44		10	59	57	51
	11	37	35	46		11	73	70	54
	12	43	41	48		12	81	78	57
	13	49	47	50		13	88	85	60
	14	55	52	53		14	94	90	63
	15	71	68	55		15	101	97	66
	16	81	77	57		16	103	99	68
	17	88	84	59		17	104	100	71
	18	100	95	61		18	104	100	>71
	19	101	96	63					
	20	104	99	65					
	21	104	99	65					
22	105	100	70						

	Rohwert	f_{cum}	Rohwert	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	6	6	24	Raum und Form	0	0	0	<25
	1	9	9	31		1	1	1	25
	2	14	14	38		2	1	1	25
	3	43	42	45		3	5	5	31
	4	58	56	52		4	10	10	34
	5	103	100	59		5	16	15	37
						6	22	21	40
						7	29	28	43
						8	38	36	46
						9	50	48	49
						10	67	64	52
						11	75	71	55
						12	82	78	58
						13	96	91	61
						14	101	96	64
						15	103	98	67
						16	105	100	70

 Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form D (30./31. Schulwoche)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	0	0	<19	Größen und Messen	0	1	1	19
	1	1	1	19		1	2	2	22
	2	1	1	19		2	2	2	22
	3	2	2	24		3	3	3	28
	4	2	2	24		4	5	4	31
	5	3	3	29		5	8	7	34
	6	5	4	31		6	14	12	37
	7	7	6	34		7	22	19	40
	8	12	10	36		8	33	29	43
	9	23	20	39		9	45	39	46
	10	27	23	41		10	65	57	49
	11	35	30	44		11	75	65	53
	12	44	38	46		12	86	75	56
	13	53	45	48		13	97	84	59
	14	62	53	51		14	103	90	62
	15	75	64	53		15	111	97	65
	16	85	73	56		16	113	98	68
	17	94	80	58		17	115	100	71
	18	105	90	61		18	115	100	>71
	19	110	94	63					
	20	114	97	65					
	21	116	99	68					
22	117	100	70						

	Rohwert	f_{cum}	Rohwert	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	2	2	19	Raum und Form	0	0	0	<26
	1	8	7	27		1	2	2	26
	2	13	11	34		2	2	2	26
	3	36	32	42		3	5	5	32
	4	53	47	50		4	11	10	35
	5	114	100	58		5	16	14	38
						6	26	23	40
						7	30	27	43
						8	42	38	46
						9	60	54	49
						10	71	64	52
						11	77	69	54
						12	83	75	57
						13	92	83	60
						14	101	91	63
						15	109	98	66
						16	111	100	68

 Anmerkungen: f_{cum} – kumulierte Häufigkeit

Mathes 3 Form D (3./4. Schulwoche Klasse 4)

	Rohwert	f_{cum}	Prozentrang	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Zahlen und Operationen	0	0	0	<23	Größen und Messen	0	0	0	<21
	1	0	0	<23		1	0	0	<21
	2	3	3	23		2	1	1	21
	3	3	3	23		3	1	1	21
	4	3	3	23		4	2	2	27
	5	4	3	39		5	6	5	31
	6	8	7	32		6	12	10	34
	7	12	10	34		7	19	16	37
	8	14	12	36		8	23	19	40
	9	20	17	38		9	34	29	43
	10	27	23	41		10	45	38	46
	11	33	28	43		11	58	49	49
	12	34	29	45		12	71	60	52
	13	43	37	47		13	87	73	56
	14	50	43	49		14	102	86	59
	15	59	50	52		15	110	92	62
	16	69	59	54		16	117	98	65
	17	85	73	56		17	119	100	68
	18	100	86	58		18	119	100	>68
	19	109	93	61					
	20	115	98	63					
	21	116	99	65					
22	117	100	67						

	Rohwert	f_{cum}	Rohwert	T-Wert		Rohwert	f_{cum}	Prozentrang	T-Wert
Daten, Häufigkeit und Wahrscheinlichkeit	0	5	4	20	Raum und Form	0	1	1	18
	1	9	8	27		1	1	1	18
	2	11	10	35		2	1	1	18
	3	33	28	42		3	2	2	27
	4	49	42	49		4	4	3	30
	5	116	100	57		5	11	10	34
						6	19	16	37
						7	23	20	40
						8	33	28	43
						9	47	41	46
						10	54	47	50
						11	68	59	53
						12	88	76	56
						13	102	88	59
						14	109	94	62
						15	115	99	65
						16	116	100	69

 Anmerkungen: f_{cum} – kumulierte Häufigkeit

Anhang E: Verlaufsdagnostische Normtabellen des „Mathes 3“

Mathes 3 Form A

	Rohwert	Y-Wert		Rohwert	Y-Wert		Rohwert	Y-Wert		Rohwert	Y-Wert
	0	<257		0	145		0	265		0	208
	1	257		1	243		1	333		1	293
	2	301		2	294		2	380		2	341
	3	333		3	331		3	428		3	379
	4	360		4	361		4	475		4	412
	5	382		5	386		5	512		5	443
	6	403		6	409		6	546		6	473
	7	422		7	430		7	582		7	505
Zahlen und Operationen	8	440	Größen und Messen	8	450	Daten, Häufigkeit und Wahrscheinlichkeit	8	625	Raum und Form	8	540
	9	458		9	469		9	694		9	581
	10	475		10	488		10			10	633
	11	492		11	506		11			11	722
	12	509		12	525						
	13	527		13	545						
	14	545		14	566						
	15	564		15	588						
	16	584		16	613						
	17	605		17	643						
	18	629		18	680						
	19	658		19	732						
20	698	20	832								
21	773										

Mathes 3 Form B

	Rohwert	Y-Wert	Rohwert	Y-Wert	Rohwert	Y-Wert	Rohwert	Y-Wert
	0	177	0	153	0	265	0	178
	1	252	1	254	1	334	1	264
	2	289	2	307	2	380	2	312
	3	317	3	344	3	433	3	349
	4	338	4	374	4	491	4	380
	5	357	5	400	5	566	5	408
	6	374	6	423			6	435
	7	389	7	444			7	462
	8	404	8	463			8	488
Zahlen und Operationen	9	418	9	482			9	516
	10	431	10	499			10	545
	11	445	11	517			11	577
	12	458	12	534			12	614
	13	472	13	552			13	661
	14	485	14	570			14	745
	15	499	15	589				
	16	514	16	610				
	17	530	17	633				
	18	548	18	660				
	19	567	19	693				
	20	589	20	740				
	21	616	21	833				
	22	650						
	23	698						
	24	785						

Größen und Messen

Daten, Häufigkeit und Wahrscheinlichkeit

Raum und Form

Mathes 3 Form C

	Rohwert	Y-Wert	Rohwert	Y-Wert	Rohwert	Y-Wert	Rohwert	Y-Wert
Zahlen und Operationen	0	216	0	111	0	265	0	202
	1	289	1	222	1	333	1	284
	2	326	2	283	2	378	2	328
	3	353	3	328	3	426	3	361
	4	374	4	364	4	469	4	390
	5	392	5	395	5	502	5	416
	6	408	6	422	6	531	6	441
	7	423	7	449	7	561	7	466
	8	437	8	475	8	593	8	492
	9	450	9	501	9	628	9	519
	10	464	10	528	10	669	10	548
	11	477	11	559	11	736	11	581
	12	490	12	596			12	619
	13	504	13	646			13	668
Größen und Messen	14	519	14	742			14	753
	15	534						
	16	552						
	17	571						
	18	593						
	19	622						
	20	662						
	21	739						

Daten, Häufigkeit und Wahrscheinlichkeit

Raum und Form

Mathes 3 Form D

	Rohwert	Y-Wert	Rohwert	Y-Wert	Rohwert	Y-Wert	Rohwert	Y-Wert
	0	154	0	123	0	266	0	164
	1	234	1	225	1	335	1	259
	2	276	2	281	2	382	2	312
	3	308	3	323	3	438	3	350
	4	333	4	358	4	503	4	381
	5	354	5	388	5	580	5	407
	6	374	6	416			6	430
	7	391	7	443			7	452
	8	408	8	467			8	474
	9	424	9	491			9	494
	10	440	10	515			10	516
	11	456	11	539			11	538
	12	472	12	564			12	561
	13	488	13	591			13	587
	14	506	14	621			14	618
	15	525	15	655			15	660
	16	546	16	698			16	740
	17	571	17	575				
	18	601	18	>575				
	19	639						
	20	687						
	21	748						
	22	841						

Zahlen und Operationen

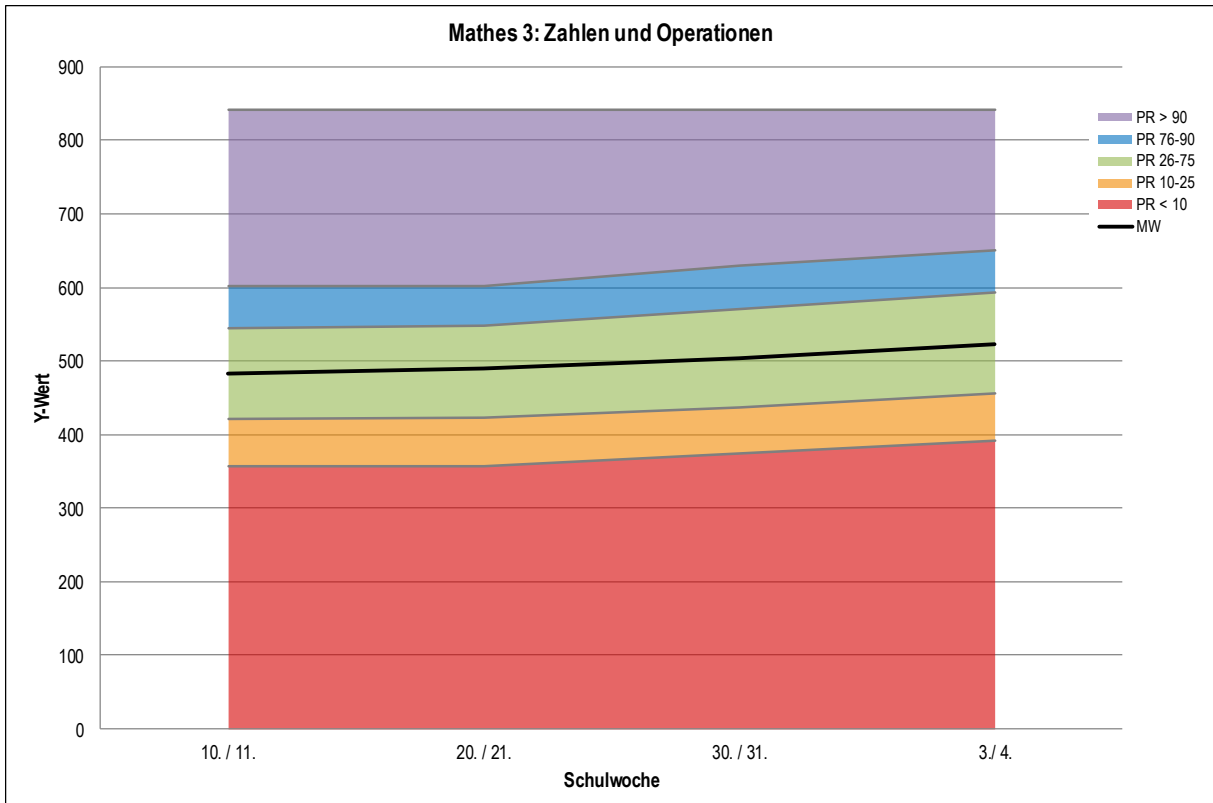
Größen und Messen

Daten, Häufigkeit und Wahrscheinlichkeit

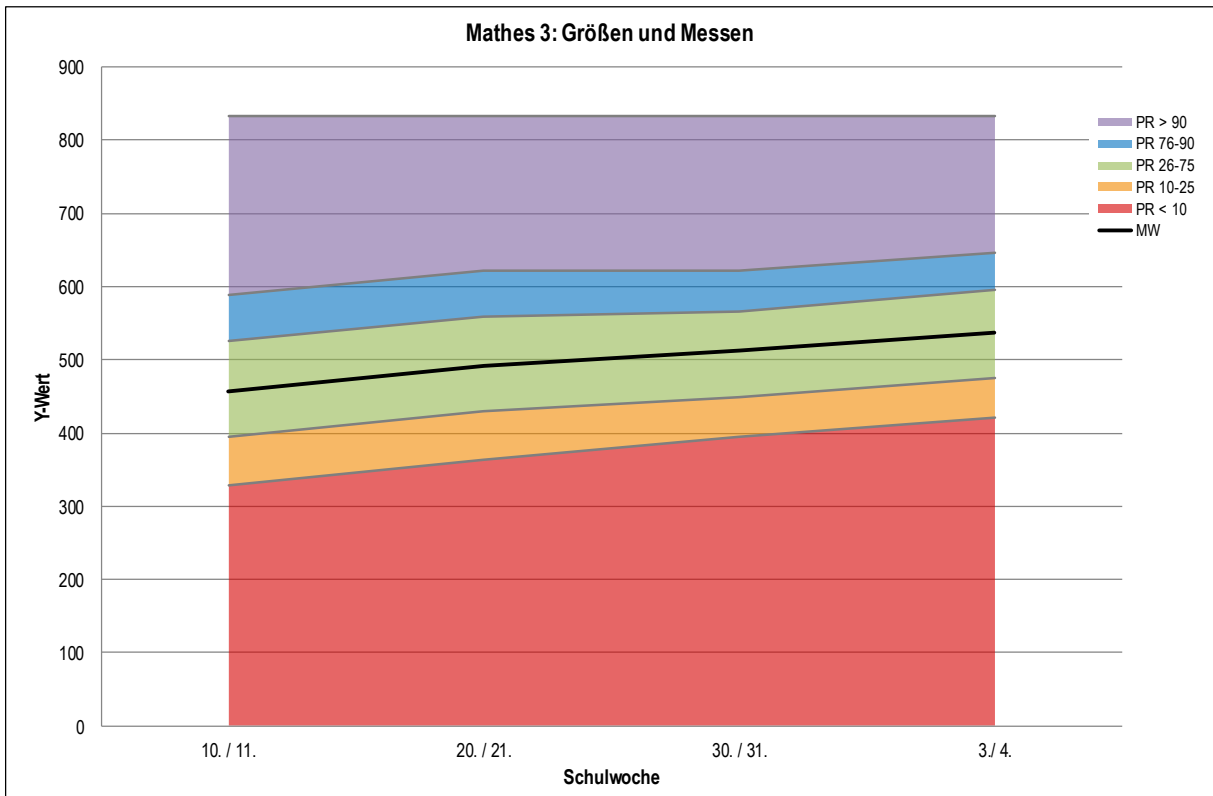
Raum und Form

Anhang F: Vorlagen für Lernverlaufsgraphen des „Mathes 3“

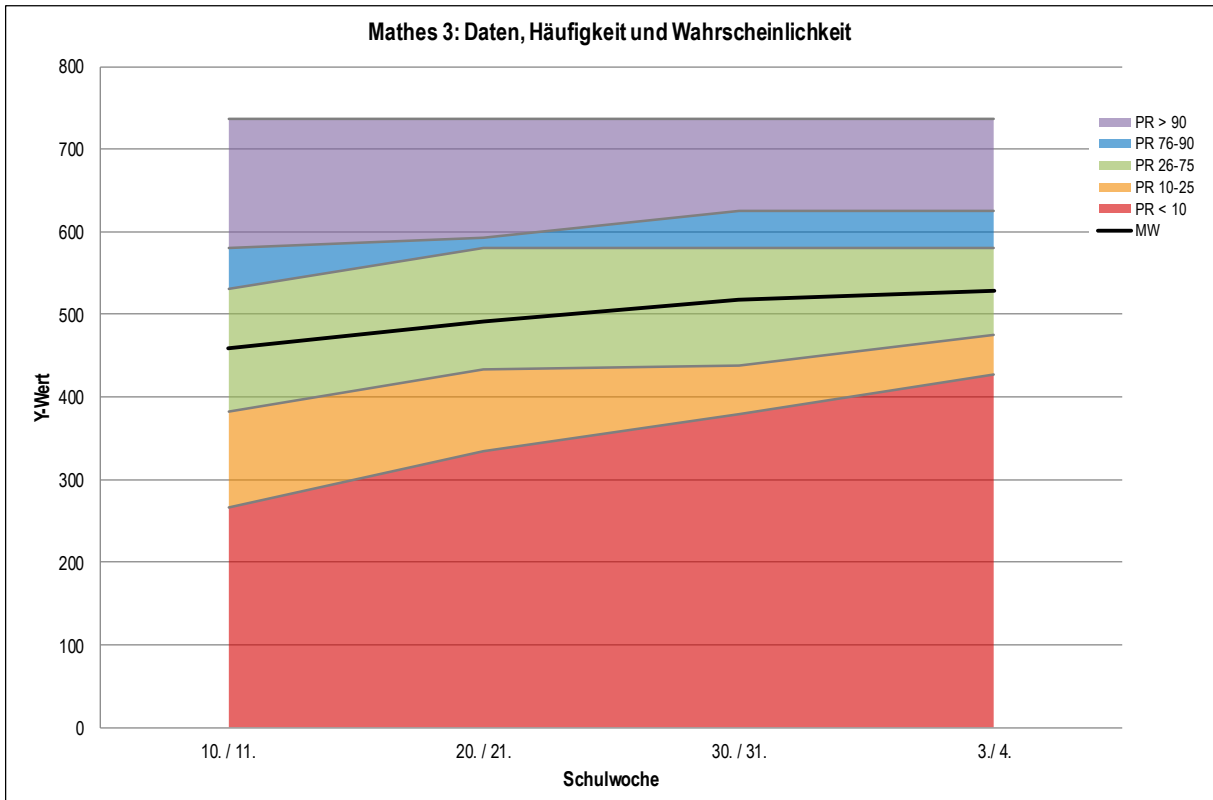
Mathes 3: Zahlen und Operationen



Mathes 3: Größen und Messen



Mathes 3: Daten, Häufigkeit und Wahrscheinlichkeit



Mathes 3: Raum und Form

