Empirische Sonderpädagogik, 2010, Nr. 3, S. 5-25

Das Inventar Rechenfische – Anwendung, Reliabilität und Validität eines Verfahrens zur Erfassung des Leistungsstandes von Erstklässlern in Mathematik

Eva Knopp[1], Bodo Hartke[2]

[1]Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN), Kiel;
[2]Universität Rostock

Schulalltagstaugliche Verfahren zur formativen Evaluation von Lernprozessen als wesentliche Komponenten erfolgreicher Prävention werden aktuell an verschiedenen Stellen gefordert wie z.B. bei Klauer (2006), Diehl und Hartke (2007), Strathmann und Klauer (2008), Walter (2008), Diehl, Hartke und Knopp (2009), Koch und Knopp (2010). Das Inventar „Rechenfische“ wurde als Diagnoseverfahren zur formativen Evaluation von Lernprozessen im Anfangsunterricht Mathematik konzipiert. Inwieweit hiermit die Erstellung eines reliablen und validen Verfahrens zur Dokumentation von Lernfortschritten im Anfangsunterricht Mathematik gelungen ist, wird in dem vorliegenden Beitrag diskutiert. Dafür wird die Konzeption des Verfahrens vorgestellt, das Design der Erprobungsstudie mit N=1688 Erstklässlern skizziert und es werden einige Ergebnisse präsentiert. In einem abschließenden Fazit werden Grenzen des Verfahrens aufgezeigt und es wird auf mögliche zukünftige Forschungsperspektiven hingewiesen.

Schlüsselwörter: Dokumentation von Lernfortschritten, Anfangsunterricht Mathematik, arithmetische Kenntnisse, Diagnose von Lernschwierigkeiten, formative Evaluation

The „Rechenfische“ Test – Appliance, Reliability and Validity of an Instrument to Measure Math Proficiency of First Graders

There is a strong need for instruments which make it possible to evaluate students' learning progress in a formative way and also in a way that can be accomplished in primary school settings (Klauer, 2006; Diehl & Hartke, 2007; Strathmann & Klauer, 2008; Walter, 2008; Diehl, Hartke & Knopp, 2009; Koch & Knopp, 2010). Instruments meeting these criteria can be considered as one crucial element of effectively preventing learning difficulties. In this article, the test “Rechenfische“ is discussed as one possibility for evaluating the learning progress in a formative way. It tests students' knowledge in first-year arithmetic. First the test itself and the design of a study (N=1688) used to implement this test for the first time are described and, afterwards, findings are reported and discussed.

Key words: document progress in learning, first-year mathematics instruction, learning arithmetics, diagnosis of learning difficulties, formative evaluation

Das Inventar „Rechenfische“ – eine Möglichkeit der Dokumentation arithmetischer Kenntnisse von Erstklässlern?! Vorstellung ausgewählter Ergebnisse einer Erprobungsstudie

Forschungsstand

Für eine präzise Erfassung der Lernausgangslage als elementare Komponente einer wirksamen Förderung im Unterricht erscheinen insbesondere Verfahren interessant, die regelmäßige, standardisierte Messwiederholungen beinhalten. Mit ihnen lassen sich Fortschritte, Stagnation und Rückschritte im Sinne einer formativen Evaluation des Unterrichts abbilden. Es wird nicht das Endergebnis des Lernens abgeprüft, sondern der Lernprozess im Sinne des „response-to-intervention“ bzw. „ response-to-instruction“-Paradigmas (RTI), wie es bspw. Vellutino, Scanlon, Sipay, Small, Pratt, Chen und Denckla (1996), Fuchs und Fuchs (1998) oder auch Speece und Case (2001) beschreiben, kontinuierlich beobachtet und ausgewertet. In diesem Sinne sprechen einige Autoren auch von Lernverlaufsdiagnostik (Strathmann & Klauer, 2008).

Warum ist es nun wichtig, Verfahren zur formativen Evaluation von Lernprozessen insbesondere im Anfangsunterricht Mathematik einzusetzen? Unter der Annahme, dass Unterrichtsprozesse den aktuellen Lern- und Wissensstand von Lernenden berücksichtigen sollen, damit Lernen optimal erfolgen kann, stellt sich die Frage, inwieweit der individuelle Lern- und Wissensstand jedes Kindes von seinen Lehrkräften eingeschätzt werden kann. Das heißt, können Lehrende für alle ihre Schülerinnen und Schüler folgende Fragen beantworten: Wo steht die Schülerin/der Schüler jetzt und wie sind seine Lernfortschritte bis dato? Bei Klassengrößen von mehr als 25 Kindern sind diese Fragen nicht leicht zu beantworten. Auch stellt sich die Frage, ob Einschätzungen, die „nebenbei“ im Schulalltag vorgenommen werden, nicht sehr subjektiv sind und der Täuschung unterliegen können. Insbesondere bei Lehrkräften, die das betreffende Fach fachfremd unterrichten und/oder noch nicht auf langjährige Erfahrungen zurückgreifen können, ist zu hinterfragen, ob sie tatsächlich die Kenntnisse ihrer Schülerinnen und Schüler präzise einschätzen können.

Viele Studien haben die diagnostische Kompetenz von Lehrkräften untersucht. Zusammenfassend stellt Bromme (1997, S. 200) fest, dass erfahrene Lehrkräfte die Schülerleistungen in einem Test, der sich auf aktuelle Unterrichtsinhalte bezog, gut vorhersagen konnten, wobei Schrader (2006) anmerkt, dass „die Werte in den einzelnen Studien [...] außerordentlich stark streuen“ (S. 97). Außerdem seien erhebliche Unterschiede von Lehrkraft zu Lehrkraft zu verzeichnen. Hinzu kommt, dass die Güte der Einschätzungen, welche im regulären Unterricht stattfinden, anscheinend deutlich schlechter ausfällt als die Güte von Einschätzungen, die im Rahmen von Versuchssettings abgegeben werden. So berichtet Bromme (1997, S. 201), dass Mathematiklehrkräfte, die direkt nach dem Unterricht befragt wurden, kaum etwas zu den Lernfortschritten einzelner Kinder sagen konnten. Zudem sind Effekte, die die Einschätzung von Lehrkräften verzerren können, bereits seit längerem bekannt. Exemplarisch seien hier nur der Strenge-, Milde-, Kontrast- und der Halo-Effekt genannt. Im Schulalltag kommt noch verfälschend hinzu, dass die Beobachtungen immer verbunden sind „mit den gespeicherten Bildern und Erfahrungen (und den damit verbundenen Emotionen), die man sich von einem Schüler oder von einer häufig wiederkehrenden Situation gemacht hat“ (Werning, 2006, S. 11). Darüber hinaus werden Lehrkräfte beim Diagnostizieren ihrer Schülerinnen und Schüler mit den Ergebnissen ihrer geleisteten Arbeit konfrontiert. Horstkemper (2006) beschreibt dieses

Dilemma folgendermaßen: „Eben diese in der Charakteristik pädagogischer Arbeit liegende Verstrickung, das Scheitern meiner Schüler ist teilweise vielleicht auch mein Scheitern – macht pädagogische Diagnostik doppelt schwierig, sie kann zur Bedrohung des eigenen Selbstwertgefühls werden" (S. 5). Inwieweit Lehrende die Kenntnisse ihrer Schülerinnen und Schüler speziell im Anfangsunterricht Mathematik einschätzen können, untersuchten Grassmann, Klunter, Köhler, Mirwald, Raudies und Thiel (2002). Sie stellten fest, dass die Einschätzungen der Lehrkräfte durchaus wesentlich von den in der Studie gezeigten Leistungen ihrer Schülerinnen und Schüler (N=830) abwichen. Nur die eigene, subjektive Einschätzung reicht als Basis für die Konzeption von Unterrichtsangeboten somit offensichtlich nicht aus (S. 15).

Dass man zudem nicht von dem Durchschnittskind bzw. der Durchschnittsklasse ausgehen kann, sondern immer wieder den Lernstand seiner Schülerinnen und Schüler neu und objektiv erheben sollte, belegen auch die Ergebnisse verschiedener Studien, welche die Lernstände von Erstklässlern untersuchten. Z.B. testeten Grassmann et al. (2002) im Rahmen der bereits erwähnten Studie Kinder kurz nach ihrem Schulanfang und am Ende der ersten Klasse. Schon während der Durchführung der ersten Testung konnten die Autoren beobachten, dass die Erstklässler die vorgegebenen Aufgaben unterschiedlich schnell bearbeiteten. Auch traten bereits zu Beginn der ersten Klasse erhebliche Unterschiede in den Lösungsraten der einzelnen Kinder innerhalb einer Klasse und zwischen verschiedenen Klassen auf (S. 49). Mit einer weiteren Studie zum Ende des ersten Schuljahres dokumentierten Grassmann, Klunter, Köhler, Mirwald, Raudies und Thiel (2003) die auch zu diesem Zeitpunkt weiterhin enorm große Spannbreite zwischen den Kenntnisständen der einzelnen Kinder (S. 7). Ähnliches zeigt sich in den Ergebnissen der Studie von Rinkens (1997) zu den arithmetischen Fähigkeiten von Schulanfängern. Rinkens überprüfte insgesamt 76 Klassen (N = 2013) und kam analog zu Grassmann et al. (2002) zu dem Ergebnis, dass sowohl die Unterschiede zwischen einzelnen Kindern als auch zwischen den Klassen erheblich waren (Rinkens, 1997, S. 2). Auch Stamm (2005) stellte in ihrer Untersuchung (N = 2711) zur Lernausgangslage von Schulanfängern fest, „dass zwar der Durchschnitt der untersuchten Kinder sechs Wochen nach Schuleintritt bereits über einige Kompetenzen verfügte, ein relativ großer Anteil jedoch entweder mit gar keinen oder dann mit weit überdurchschnittlichen Kenntnissen zur Schule kam" (S. 77). Es lässt sich somit feststellen, dass die Spannbreite der Kenntnisse von Erstklässlern sehr weit ist. Hinzu kommt, dass sie sich vermutlich in Zukunft noch einmal maximieren wird, und zwar wenn die veränderte Schuleingangsphase eingeführt wird, die derzeit in vielen Bundesländern im Gespräch ist (Koch & Ellinger, 2007). Verschiedene Varianten sind diesbezüglich angedacht bzw. bereits umgesetzt. Ein in einigen Bundesländern gesetzlich verankertes Modell ist, alle Kinder ohne die Möglichkeit einer Zurückstellung einzuschulen und ihnen dann für die ersten beiden Schuljahre ein, zwei oder drei Jahre Zeit zu geben. In diesem veränderten Rahmen müssen Unterrichtsangebote so konzipiert sein, dass sie die vielen verschiedenen Kenntnisstände innerhalb einer Klasse berücksichtigen, um allen gerecht zu werden. Damit dies gelingen kann, sollten Lehrkräfte möglichst frühzeitig in der ersten Klasse auf Messverfahren zurückgreifen können, die dies durch eine große Spannweite der Schwierigkeitsgrade der verwendeten Items ermöglichen.

Neben den genannten Argumenten für eine Durchführung regelmäßiger Diagnostik im Anfangsunterricht Mathematik sprechen auch zahlreiche Forschungsergebnisse dafür, Verfahren zur formativen Evaluation von Lernprozessen bereits in dieser Klassenstufe einzusetzen. So konnte gezeigt werden, dass sich Lücken im Lernprozess gravierend auf den weiteren Lernerfolg in Mathematik auswirken. Dieses ist insbesondere in dem im ersten Schuljahr behandelten Zahlenraum bis 20 der Fall, „da hier zentrale Aspekte thematisiert werden, die in allen weiteren Zahlenräumen wieder auftauchen und die späteren Lernprozesse erleichtern können“ (Scherer, 2007, S. 593). Wie entscheidend sich das Vorwissen auf den weiteren Lernerfolg im Mathematikunterricht auswirkt, demonstrieren bspw. die Studien von Krajewski (2003), Aunola, Leskinen, Lerkkannen und Nurmi (2004), Mazzocco und Thompson (2005), Grube und Hasselhorn (2006) sowie Weberschock und Grube (2006). Umso entscheidender ist es somit, dass Lehrerinnen und Lehrer Verfahren an die Hand bekommen, die es ihnen ermöglichen, rechtzeitig Lernrückstände bzw. Stagnationen im Lernprozess zu erkennen, Fördermaßnahmen einzuleiten, den Erfolg der Fördermaßnahmen zu evaluieren und ggf. die Maßnahmen zu verändern. Geschieht dieses nicht, werden Lernschwierigkeiten von Schülerinnen und Schülern in Mathematik nicht rechtzeitig erkannt, verfestigen sich die Lücken, und der weitere Lernerfolg in Mathematik wird erschwert. Folge ist, dass es auch in höheren Klassen Kinder bzw. Jugendliche gibt, die die Lerninhalte der ersten Schuljahre nicht beherrschen. Dies belegen u.a. die Studien von Gaupp, Zoelch und Schumann-Hengsteler (2004), Mittelberg (2004), Balzer, Fritz, Ricken und Jäger (2007), Moser Opitz (2007) sowie die Untersuchung von Wartha (2009).

Zusammenfassend lässt sich festhalten, dass sich Lehrkräfte beim Planen und Durchführen des mathematischen Anfangsunterrichts nicht nur auf ihre langjährige Berufserfahrung bzw. die Vorgaben des Schulbuchs verlassen sollten. Es erscheint notwendig, auf Verfahren zurückzugreifen, die es ermöglichen, die Lernausgangslage der Kinder immer wieder neu in kurzen Abständen zu erfassen und in Beziehung zum Leistungsverhalten von Vergleichsgruppen zu setzen. Dadurch könnte die Basis für eine Unterrichtsgestaltung gewonnen werden, die dann eine möglichst optimale Passung zwischen Lernangebot und Lernstand der Kinder bietet. Im Folgenden wird ein Verfahren für den Anfangsunterricht Mathematik vorgestellt, welches dieses ermöglicht.

Das Inventar „Rechenfische“

Das Inventar „Rechenfische“ (Knopp, 2010) wurde an der Universität Rostock 2006/07 konzipiert und erprobt. Mit seiner Hilfe soll es ermöglicht werden, die Lernfortschritte beim Erwerb arithmetischer Kenntnisse in der ersten Klasse entsprechend den testtheoretischen Gütekriterien objektiv, reliabel und valide über drei Messzeitpunkte (MZP) im zweiten Halbjahr des ersten Schuljahres zu dokumentieren. Das Inventar „Rechenfische“ soll somit, bevor massive Schwierigkeiten beim weiteren Lernen arithmetischer Lerninhalte entstehen, Lehrende auf Lücken im Lernprozess ihrer Schülerinnen und Schüler hinweisen. Entsprechende Ergebnisse können dann frühzeitig signalisieren, ob die Lernverläufe wie erwünscht sind oder Fördermaßnahmen eingeleitet werden müssen.

Das Inventar „Rechenfische“ enthält insgesamt 169 Items, die systematisch auf Grundlage des fachdidaktischen und entwicklungspsychologischen Forschungsstandes zum Erwerb erster arithmetischer Kenntnisse (u.a. Fuson, 1982, 1988; Resnick, 1989; Dehaene, 1992, 1999; Radatz, Schipper, Ebeling & Dröge, 1996; Ostad, 1997; Stern, 1998, 2003; Torbeyns, Verschaffel & Ghesquière, 2002, 2004; Carpenter, Franke & Levi, 2003; Jordan, Hanich & Kaplan, 2003; Geary, Hoard, Byrd-Craven & DeSoto, 2004; Gerster & Schultz, 2004; Padberg, 2004; von Aster, 2005; Baroody, 2006; Grube, 2006; Krajewski & Schneider, 2006; Siegler, Deloache & Eisenberg, 2006; Fritz, Ricken & Gerlach, 2007; Scherer, 2007) und unter Berücksichtigung der curricularen Vorgaben zusammengestellt wurden. Detaillierte Ausführungen zu den einzelnen Items finden sich in Knopp

(2010). Sie sind den folgenden Aufgabenbereichen zugeordnet.

- „Zahl-Mengen-Zuordnung“
- „Addition: $a+b=x$, $x\leq10$“
- „Addition: $a+b=x$, $10<x\leq20$“
- „Subtraktion: $a-b=x$, $a\leq10$“
- „Subtraktion: $a-b=x$, $10<a\leq20$“
- „Addition: $a+x=c$, $c\leq20$“
- „Addition: $x+b=c$, $c\leq20$“
- „Subtraktion: $x-b=c$, $x\leq20$“
- „Subtraktion: $a-x=c$, $a\leq20$“
- „Addition: $a+b=x$, $x\leq10$ in Päckchen“
- „Addition: $a+b=x$, $10<x\leq20$ in Päckchen“
- „Subtraktion: $a-b=x$, $a\leq10$ in Päckchen“
- „Subtraktion: $a-b=x$, $10<a\leq20$ in Päckchen“
- „Zahlen zerlegen“
- „Bild-Rechenoperation-Zuordnung“
- „Zahlreihen ergänzen“
- „Textaufgaben“
- „Größer-Kleiner-Vergleiche“
- „Zahlenstrahl“
- „Kettenaufgaben“

Eine Beispielaufgabe des Inventars zeigt die Abbildung 1.

Da das Inventar zu drei MZP im Laufe des ersten Schuljahres differenziert den Leistungsstand von Kindern dokumentieren soll und somit die entsprechende Bandbreite der arithmetischen Lerninhalte abdecken muss, wurden in das Inventar auch Items mit einbezogen, von denen anzunehmen ist, dass die Erstklässler ihnen zum 1. MZP – zu Beginn des zweiten Schulhalbjahres – und zum 2. MZP – kurz nach den Osterferien – noch nicht im Unterricht begegnet sind. Zusätzlich wurde darauf geachtet, dass die einzelnen Items ohne schriftliche oder mündliche Anweisungen von Erstklässlern verstanden und bearbeitet werden können. Dieses ermöglicht es, das Inventar „Rechenfische“ durchzuführen, ohne mit einer Stoppuhr einzelne Zeitbegrenzungen vorzugeben. Die Erstklässler können somit die Items in ihrer eigenen Geschwindigkeit bearbeiten. Das Inventar „Rechenfische“ ist darüber hinaus für die Durchführung im Klassenverband als Gruppentest konzipiert. Dieses kommt der Forderung nach einem diagnostischen Verfahren nach, das der „Praktikabilitäts-Ethik“ (Hofer, 1986, S. 385) der Lehrerinnen und Lehrer entspricht. Es erleichtert die Durchführung des Inventars im Schulalltag, da keine wertvollen Doppelbesetzungen bzw. Förderstunden verwendet werden müssen und deutlich weniger Zeit benötigt wird, als dieses bei einer Einzeltestung nötig wäre. Da das Inventar zweigeteilt ist (Teil A und Teil B), besteht die Möglichkeit, es an zwei aufeinander folgen-

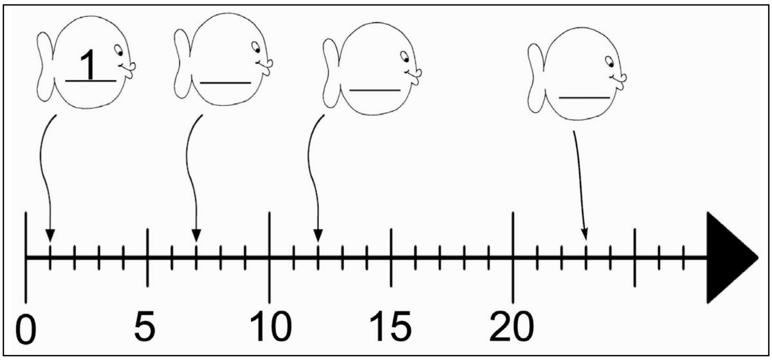

Abb. 1: Beispielaufgabe aus dem Inventar "Rechenfische" (Wagner & Hartke, 2006)

den Tagen durchzuführen. Für jeden Teil haben die Schülerinnen und Schüler eine Schulstunde Zeit.

Fragestellung

Im Anschluss an die Konzeption des Inventars wurde im Rahmen einer Erprobungsstudie der Fragestellung nachgegangen:

– Lässt sich ein vorwiegend unter didaktisch-curricularen Aspekten modelliertes sowie auf drei MZP normiertes Verfahren zur Dokumentation der Lernfortschritte im Anfangsunterricht Mathematik reliabel und valide gestalten?

Methode

Das Inventar konnte im Rahmen der wissenschaftlichen Begleitstudie des Projektes „Primarstufe“, gefördert vom Ministerium für Bildung, Wissenschaft und Kultur des Landes Mecklenburg-Vorpommern (Koch, Hartke & Blumenthal, 2008), im Schuljahr 2006/07 an einer Stichprobe von N=1688 Erstklässlern erprobt werden. In die Stichprobe wurden alle Erstklässler der Hansestadt Rostock und der Insel Rügen einbezogen (Ausnahmen: Krankheit, Umzug oder Eltern verweigern die Teilnahme ihrer Kinder an der Studie). Zusätzlich zu den Grundschulklassen wurden Schülerinnen und Schüler aus Diagnoseförderklassen beider Regionen mit in die Stichprobe aufgenommen. In Diagnoseförderklassen werden Kinder unterrichtet, bei denen vermutet wird, dass sie im Anfangsunterricht der regulären Grundschule nicht effektiv lernen können. Sie bekommen für das Erlernen der Unterrichtsinhalte der ersten beiden Schuljahre drei Jahre Zeit. Eine spezifische Klassifikationsdiagnostik, mittels derer bestimmt wird, wer in eine Diagnoseförderklasse und wer in eine Grundschulklasse eingeschult wird, liegt nicht vor. Eine ausführliche Beschreibung des Konzepts „Diagnoseförderklasse“ findet sich in Koch, Hartke und Blumenthal (2008). Zusätzlich zum Inventar wurden verschiedene andere Diagnoseverfahren eingesetzt. So wurden u.a. der Kognitive Fähigkeitstest für 1. und 2. Klassen – KFT 1-2 R (Kawthar & Perleth, 2005) im November 2006 und die Würzburger Leise Leseprobe – WLLP (Küspert, Schneider, 1998) sowie der DEMAT 1+ (Krajewski, Küspert, Schneider & Visé, 2002) im Juni/Juli 2007 durchgeführt. Die Durchführung der einzelnen Testungen in der Stadt Rostock übernahmen studentische Hilfskräfte. Für die Region Rügen konnten Lehrende des Förderzentrums Bergen als Testleiterinnen und Testleiter gewonnen werden. Alle Testleiterinnen und Testleiter wurden vorab im Rahmen einer mehrstündigen Fortbildung geschult. Die Testungen fanden im Rahmen des regulären Schulalltags statt. Zu jedem MZP bearbeiteten die Kinder sowohl Teil A als auch Teil B des Inventars „Rechenfische“. Eine detaillierte Anweisung gewährleistete dabei die Durchführungsobjektivität.

Stichprobe

Von den insgesamt 1688 Erstklässlern bearbeiteten im 1. Erhebungszeitraum 1453, im 2. 1455 und zum 3. MZP, Ende des Schuljahres, 1352 Kinder das Inventar „Rechenfische“. Es gab Erstklässler, die zu keinem der drei MZP an der Studie partizipierten, andere die nur zu einem MZP die Items lösten, und andere Kinder, die an zwei bzw. an allen drei MZP teilnahmen. Zu den drei Erhebungszeitpunkten fiel dementsprechend jeweils eine bestimmte Anzahl von Erstklässlern aus bzw. versuchten sich zum Teil andere Schülerinnen und Schüler an den Items des Inventars. Von 1117 Kindern liegen zu allen drei MZP Daten vor. Da die Verschiebungen in der Zusammensetzung der Stichproben zu den einzelnen Testungen v.a. auf Krankheitsausfälle, Umzüge von Schülerinnen und Schülern bzw. organisatorische Gründe zurückzuführen sind, erscheint ein gerichteter Einfluss auf

Tab. 1: Anzahl der Schülerinnen/Schüler, die an der Erhebung mit dem Inventar „Rechenfische“ zu den drei MZP teilnahmen, aufgeteilt in Kinder in Grundschulklassen und Kinder in Diagnoseförderklassen

	Anzahl der Erstklässler, die das Inventar "Rechenfische" bearbeiteten	
	in Grundschulklassen	in Diagnoseförderklassen
1. MZP	1321	132
2. MZP	1325	117
3. MZP	1218	114

Anmerkung: MZP = Messzeitpunkt

einzelne Ergebnisse unwahrscheinlich. Um diese Annahme abzusichern, wurden die Stichproben der drei Erhebungen u.a. hinsichtlich ihrer Zusammensetzung bestehend aus Kindern, die in Grundschulklassen beschult werden, und Lernenden in Diagnoseförderklassen betrachtet. Hier gab es über die drei Testungen hinweg nur leichte Verschiebungen.

Ergebnisse

Kennwerte der Verteilungen

Die Abbildungen 2-4 veranschaulichen die Häufigkeitsverteilungen der Auswertung des gesamten Inventars zu den drei Erhebungszeitpunkten. Die Angabe zur Schiefe der Verteilung (-0,3) weist darauf hin, dass bereits zur 1. Testung eine leicht rechtssteile Verteilung vorliegt. Der Kurtosiswert (-0,7) spricht ebenfalls schon zum 1. MZP gegen eine Normalverteilung. Schließlich muss auch auf Basis der Ergebnisse des Kolmogorov-Smirnov-Anpassungstests die Annahme, dass zum 1. MZP eine Normalverteilung der Daten vorliegt, negiert werden ($p < .01$, 2-seitig). Gleiches zeigt sich bei den Ergebnissen zum 2. und 3. MZP. Hier müssen die Annahmen einer Normalverteilung mit $p < .01$ (2-seitig) ebenfalls zurückgewiesen werden. In der graphischen Darstellung der Daten wird deutlich, dass die Verteilung zum Zeitpunkt des zweiten Einsatzes des Inventars bereits stark rechtssteil ist. Im unteren Leistungsbereich kann ein leichter, aber stetiger Anstieg der Punktezahlen vermerkt werden. Zum 3. MZP verstärkt sich die Verschiebung der Werte nach rechts. Darüber hinaus steigen die Daten zwischen dem Rohwertpunkt von 120 und dem Rohwertpunkt von 130 sprunghaft an. Bis zu einem Rohwertpunkt von 120 ist immer noch ein relativ leichter, aber stetiger Anstieg zu verzeichnen.

Im Durchschnitt lösten die Erstklässler von insgesamt 169 Items zum 1. MZP 90,8 Items richtig. Die Standardabweichung beträgt $SD = 39.6$ Rohwertpunkte und der Median liegt bei 92,0. Zum 2. MZP lösten die Erstklässler im Mittel bereits 120,2 Items richtig. Die Standardabweichung ist mit $SD = 39.3$ vergleichbar zu der Standardabweichung der ersten Testung. Der Median steigt auf 132,0 richtig gelöste Items. Zum 3. MZP lösten die Kinder bei weiterhin gleichbleibender Standardabweichung ($SD = 38.4$) schließlich im Mittel 128.1 Items richtig. Dementsprechend ist auch beim Median im Vergleich zu den Daten vom 2. Erhebungszeitpunkt ein weiterer Anstieg zu verzeichnen. Er liegt zum 3. MZP bei 143.0. Angaben zu den Häufigkeitsverteilungen in den einzelnen Aufgabenbereichen finden sich in Knopp (2010).

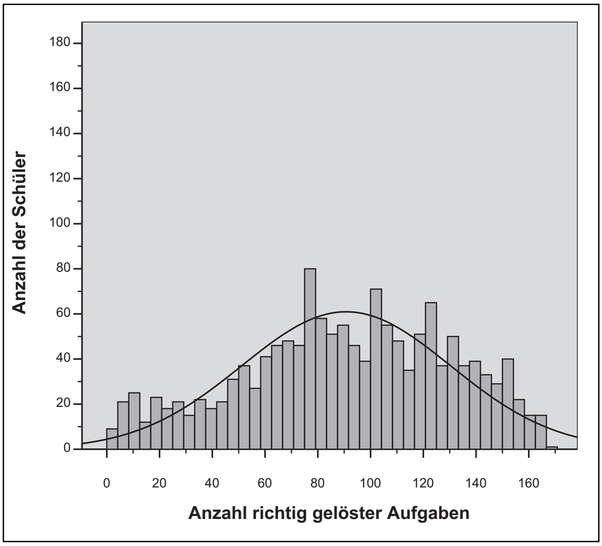

Abb. 2: Häufigkeitsverteilung der Ergebnisse im Inventar "Rechenfische" zum 1. MZP (Rohwertpunkte)

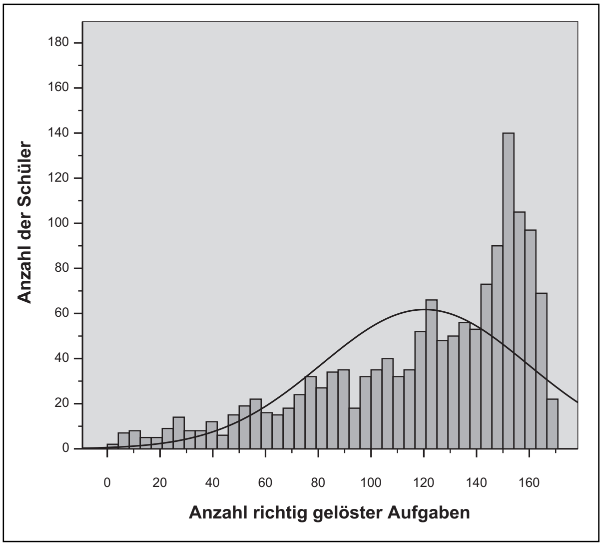

Abb. 3: Häufigkeitsverteilung der Ergebnisse im Inventar "Rechenfische" zum 2. MZP (Rohwertpunkte)

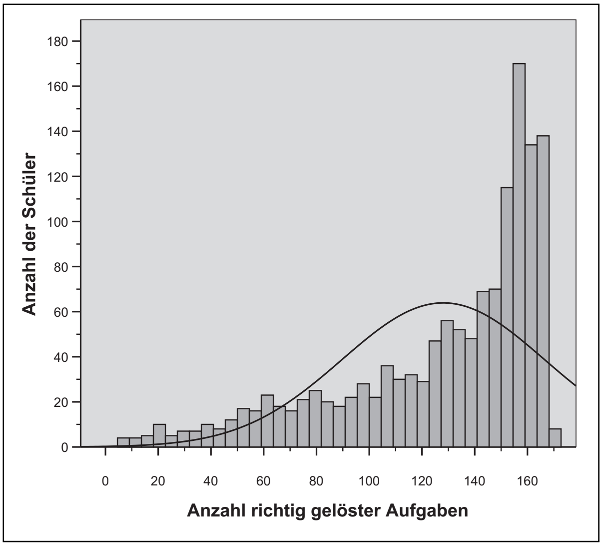

Abb. 4: Häufigkeitsverteilung der Ergebnisse im Inventar "Rechenfische" zum 3. MZP (Rohwertpunkte)

Angaben zu Itemschwierigkeiten und Trennschärfen

Zum 1. MZP liegen 139 der 169 Items innerhalb der von Krohne und Hock (2006, S. 47) empfohlenen Spanne der Schwierigkeitsgrade von *P* = .80 - .20. Zur zweiten Erhebung Ende April/Anfang Mai liegen 122 Items innerhalb des empfohlenen Schwierigkeitsbereichs und 47 Items weisen einen Schwierigkeitsindex größer als *P* = .80 auf. Zum 3. MZP zum Ende des Schuljahres zeigen 70 Items der insgesamt 169 Items einen Schwierigkeitsindex *P* > .80. Alle anderen Items fallen in den Bereich von *P* = .30 - .80. Bezogen auf die Trennschärfe von Items in einem Diagnoseverfahren gehen Krohne und Hock (2006, S. 48) davon aus, dass diese bei mindestens r_{it}=.30 liegen sollte. Als gut bezeichnen sie Items mit einer Trennschärfe zwischen r_{it}=.50 und r_{it}=.60. Der Tabelle 2 ist zu entnehmen, bei wie vielen der insgesamt 169 Items des Inventars „Rechenfische" die Trennschärfen zu den drei MZP innerhalb der als gut befundenen Spanne liegen.

Ergebnisse zur Einschätzung der Reliabilität des Inventars

Zwischen dem 1. und dem 2. MZP ist eine Test-Retest-Reliabilität von r_{tt}=.73 zu verzeichnen. Die Daten vom 1. und 3. MZP korrelieren noch zu r_{tt}=.68 miteinander und zwischen den Ergebnissen des 2. und 3. MZPs zeigt sich eine Korrelation von r_{tt}=.76. Cronbachs alpha ist zu allen drei MZP ≥.98 (1. MZP: r_{tt}=.98, 2. MZP: r_{tt}=.99, 3. MZP: r_{tt}=.99). Die Testhalbierungsreliabilität des Gesamttests liegt zu den drei Messzeitpunkten bei r_{tt}=.91.

Ergebnisse zur Einschätzung der Validität des Inventars

Zur Einschätzung der Validität des Inventars „Rechenfische" wurden u.a. die Höhen der Korrelationen (Pearson) zu den Ergebnissen der Erstklässler im KFT 1-2 R und in der WLLP berechnet. Die Korrelation zwischen der erzielten Rohwertpunkteanzahl im KFT 1-2 R und dem Ergebnis (Rohwertpunkte) im Inventar „Rechenfische" zum 1. MZP beträgt r_{ct}=.57. Zum 2. MZP korrelieren die Er-

Tab. 2: Anzahl der Items des Inventars "Rechenfische" zu den drei MZP mit den jeweiligen Trennschärfen

	Trennschärfen liegen innerhalb der Spanne von r_{it}=.50 - .60	**Trennschärfen liegen über r_{it}=.60**	**Trennschärfen liegen in der Spanne von r_{it}=.30 - .50**	**Trennschärfen liegen unterhalb von r_{it}=.30**
Anzahl der Items des Inventars "Rechenfische" zum 1. MZP	95	15	55	4
Anzahl der Items des Inventars "Rechenfische" zum 2. MZP	70	40	48	11
Anzahl der Items des Inventars "Rechenfische" zum 3. MZP	71	37	49	12

Anmerkung: MZP = Messzeitpunkt

gebnisse im Inventar „Rechenfische" mit den erzielten Rohwertpunkten im KFT 1-2 R zu r_{ct}=.54 und zum 3. MZP zu r_{ct}=.51. Geringer fallen die Korrelationen zwischen den Ergebnissen im Inventar „Rechenfische" und denen in der WLLP aus. Zum 1. MZP betragen sie r_{ct}=.40, zum 2. r_{ct}=.41 und zum 3. MZP r_{ct}=.41. Anzumerken ist hierbei, dass die Verteilung der Häufigkeiten der Ergebnisse in der WLLP in der vorliegenden Stichprobe im Unterschied zu der Normierungsstichprobe von Küspert und Schneider (1998, S. 26) nicht einer Normalverteilung entspricht. Die Verteilung ist nach links verschoben.

Zur Einschätzung der Übereinstimmungsvalidität als eine Möglichkeit, die Kriteriumsvalidität eines diagnostischen Verfahrens zu erkunden, wurden die Korrelationen der Ergebnisse im Inventar „Rechenfische" mit den Ergebnissen der Schülerinnen und Schüler im DEMAT 1+ berechnet. Bevor die Korrelationen berichtet werden, soll jedoch vorher noch dargelegt werden, wie die Schülerinnen und Schüler in der vorliegenden Stichprobe im DEMAT 1+ abschnitten und ob ihre Ergebnisse vergleichbar sind mit denen der Normierungsstudie von Krajewski et al. (2002). Es zeigte sich, dass die Werte der Gesamtstichprobe der vorliegenden Studie, d.h. inklusive der Kinder aus Diagnoseförderklassen, am ehesten denen der Normierungsstichprobe des DEMAT 1+ entsprechen, wobei sich die Mittelwerte auf dem 0,01 Niveau signifikant voneinander unterscheiden (T-Test für unabhängige Stichproben). Zum 1. MZP konnte eine Korrelation zwischen den Ergebnissen der Erstklässler im Inventar „Rechenfische" und ihren Ergebnissen im DEMAT 1+ von r_{ct}=.73 festgestellt werden. Zum 2. und 3. MZP fallen die Korrelationen noch enger aus (r_{ct}=.78 bzw. r_{ct}=.79).

Eine weitere Möglichkeit, die Validität eines Verfahrens einzuschätzen, stellt die Durchführung von Regressionsanalysen dar. Mit ihrer Hilfe können die Zusammenhänge zwischen den Ergebnissen von zwei Variablen, in diesem Fall die Ergebnisse zu jeweils zwei MZP, analysiert werden und somit kann die Güte von Prognosen ermittelt werden, die mit dem Verfahren formuliert werden können. Die Prognose der Ergebnisse im DEMAT 1+ auf Grundlage der Ergebnisse im Inventar „Rechenfische" zum 1. MZP kann mit einem R^2 von .61 aufgestellt werden. Der Zusammenhang zwischen den Ergebnissen im Inventar „Rechenfische" zum 2. MZP und den Ergebnissen im DEMAT 1+ lässt sich mit einem R^2 von .64 beschreiben. Die beste Prognose auf die DEMAT 1+ Ergebnisse lässt sich schließlich wagen, wenn die Ergebnisse im Inventar „Rechenfische" zum 1. MZP mit den Ergebnissen zum 2. MZP summiert werden. Mit einem R^2 von .70 kann hier eine Prognose formuliert werden. Weitere Ergebnisse hierzu sind auch in Knopp (2010) zu finden.

Neben der Durchführung von Regressionsanalysen wurde auch der Extremgruppenvergleich als eine Möglichkeit genutzt, der Validität des Inventars „Rechenfische" auf die Spur zu kommen. Die Tabelle 3 zeigt die Ergebnisse der einzelnen Teilstichproben im Inventar „Rechenfische" im Überblick. Hier deutet sich bereits an, zwischen welchen Teilstichproben ein signifikanter Unterschied bestehen könnte. So lösten die Erstklässler aus Grundschulklassen im Mittel zu jedem MZP wesentlich mehr Aufgaben (>61.5 Items) richtig als die Kinder aus den Diagnoseförderklassen bei vergleichbaren Standardabweichungen. Die Schülerinnen und Schüler aus den Diagnoseförderklassen erreichten durchschnittlich noch nicht einmal zum 3. MZP den Mittelwert, den die Kinder aus Grundschulklassen im Mittel bereits zum 1. MZP erreichten.

Der T-Test für unabhängige Stichproben wurde eingesetzt, um zu prüfen, ob zwischen den Ergebnissen der Grundschülerinnen und Grundschüler und der Kinder aus Diagnoseförderklassen ein signifikanter Unterschied vorliegt. Es ist festzustellen, dass sich zu allen drei MZP die Ergebnisse der Kinder in den Diagnoseförderklassen höchst signifikant von

Tab. 3: Kennwerte der Ergebnisse ausgewählter Teilstichproben im Inventar "Rechenfische" zu den drei MZP

	MZP	N	Mittelwert	Median	SD	Signifikanz
Kennwerte der Ergebnisse der Grundschülerinnen/-schüler	1	1321	96.8	96.0	34.9	0.00
	2	1325	126.2	136.0	33.5	
	3	1218	133.6	146.5	33.2	
Kennwerte der Ergebnisse der Schülerinnen/Schüler in Diagnoseförderklassen	1	132	30.1	19.0	32.9	
	2	117	56.6	52.0	40.0	
	3	114	72.1	65.0	42.3	
Kennwerte der Ergebnisse der Schülerinnen/Schüler mit einem RW ≤55 im KFT 1-2 R	1	424	63.6	65.0	37.5	0.00
	2	377	91.5	94.0	41.9	
	3	348	101.2	111.0	43.5	
Kennwerte der Ergebnisse der Schülerinnen/Schüler mit einem RW >55 im KFT 1-2 R	1	965	103.4	105.0	33.8	
	2	1003	131.7	142.0	31.1	
	3	927	138.5	150.0	30.3	

Anmerkung: MZP = Messzeitpunkt, KFT 1-2 R = Kognitiver Fähigkeitstest für 1. und 2. Klassen (Kawthar & Perleth, 2005)

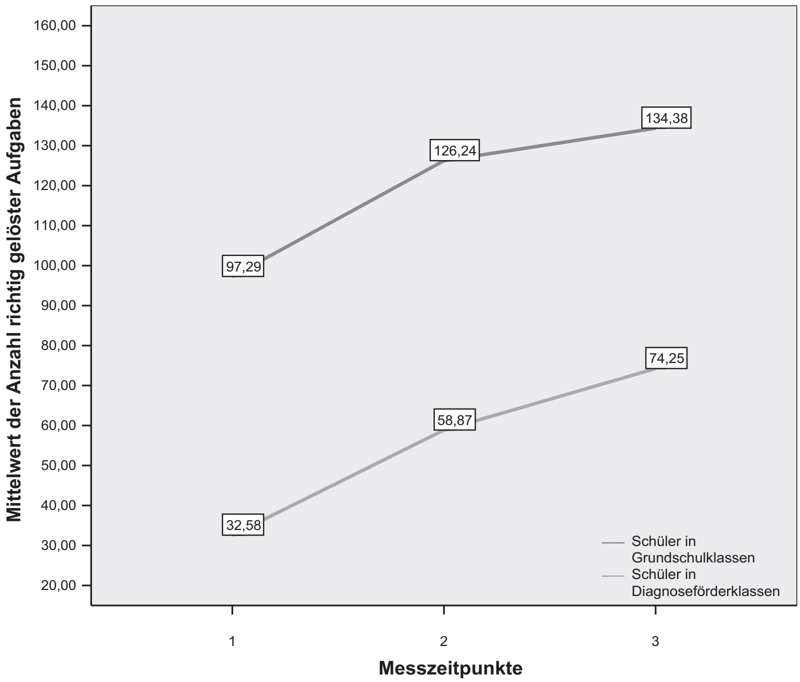

Abb. 5: Mittelwerte der Anzahl richtig gelöster Items im Inventar "Rechenfische" der Gruppe von Schülerinnen und Schülern aus Grundschulklassen im Vergleich zu den Mittelwerten der Gruppe von Kindern aus Diagnoseförderklassen zu den drei MZP

den Ergebnissen der Erstklässler in den Grundschulklassen unterscheiden ($p < .01$, 2-seitig), wobei die Mittelwertdifferenzen und auch die Differenzen der Angaben zu den Medianen zwischen den MZP in beiden Gruppen ähnlich hoch ausfallen.

Die Analyse der Ergebnisse im Inventar „Rechenfische" von Erstklässlern mit einer Gesamtanzahl an erreichten Rohwertpunkten im KFT 1-2 R >55 im Vergleich zu den Ergebnissen von Kindern mit einem Ergebnis ≤55 Rohwertpunkte ergibt zu allen drei MZP ebenfalls höchst signifikante Unterschiede ($p < .01$, 2-seitig) zugunsten der Kinder mit einem Ergebnis im KFT 1-2 R >55 Rohwertpunkte (T-Test für unabhängige Stichproben). Auch hier zeigen die Graphen weitgehend parallel verlaufende Kurven.

Ergebnisse zur Annahme, dass das Inventar Lernfortschritte dokumentiert

Um einen Hinweis darüber zu erhalten, ob mit dem Inventar „Rechenfische" Lernfortschritte dokumentiert werden können, wurden die Mittelwerte der Häufigkeiten richtig gelöster Items zu den drei MZP miteinander verglichen. Dabei wurden, wenn es um den Vergleich der Werte zu allen drei MZP ging, jeweils nur die Kinder in die Berechnungen einbezogen, die an allen drei Erhebungen teilnahmen. Folglich sind die Schüleranzahlen in den einzelnen Berechnungen leicht unterschiedlich.

Die Unterschiede zwischen den Mittelwerten sind alle signifikant (p=0,00, 2-seitig). Die Kinder konnten vom 1. zum 2. sowie vom 1. zum 3. und vom 2. zum 3. MZP im

Tab. 4: Veränderung der Kennwerte der Ergebnisse im Inventar "Rechenfische" über die drei MZP

	Vergleich der MZP	N	Mittelwerte	SD	Signifikanz
Anzahl richtig gelöster Items	1 → 2	1287	91,6 → 119,2	39.3 → 39.4	.00
	1 → 3	1179	91,1 → 127,3	39.3 → 38.5	.00
	2 → 3	1262	121,3 → 129,8	38.6 → 37.3	.00

Anmerkung: MZP = Messzeitpunkt, N = Anzahl, SD = Standardabweichung

Mittel deutlich mehr Items richtig lösen. Es zeigt sich, dass darüber hinaus die Differenz zwischen dem 1. und dem 3. MZP mit durchschnittlich 36,2 mehr richtig gelösten Items (*SD* = 31.2) am größten ist. Die Verteilung der Differenzhäufigkeiten ist eingipflig, wobei es deutlich mehr Kinder gibt, die sich vom 1. zum 3. MZP verbesserten, als Schülerinnen und Schüler, die zum 3. MZP weniger Items lösen konnten als zum 1. MZP. Die Differenz zwischen dem 1. und 2. MZP fällt leicht geringer aus als die zwischen dem 1. und dem 3. MZP. Der Mittelwert liegt bei 27,6 Items (*SD* = 28.9). Werden nur die Mittelwerte der Häufigkeiten richtig gelöster Items der Kinder aus Diagnoseförderklassen zu den drei MZP verglichen, lassen sich ebenfalls signifikante ($p < .05$) Unterschiede konstatieren. Auch ergeben sich ebenfalls signifikante Unterschiede zwischen den Häufigkeiten richtig gelöster Items, wenn nur die Gruppe der Erstklässler in die Berechnung einbezogen wird, die im KFT 1-2 R einen Rohwertpunkt ≤55 erreichten.

Diskussion der Ergebnisse

Ist die Stichprobe repräsentativ?

Diagnostische Verfahren sollten an einer möglichst repräsentativen Stichprobe erprobt werden. Die Stichprobe (N=1688) erfüllt das Kriterium einer ausreichenden Größe. Ob die Stichprobe für die Bundesrepublik Deutschland repräsentativ ist und damit als Eichstichprobe zur Ableitung von Normen verwendet werden kann, lässt sich nur schwer abschätzen. Bortz und Döring (2006, S. 397) unterscheiden zwischen einer spezifischen Repräsentativität und einer globalen Repräsentativität. Letztere lässt sich nur über eine Zufallsstichprobe erreichen, welche in der vorliegenden Studie nicht gegeben ist. Die spezifische Repräsentativität ist erfüllt, wenn die Stichprobe der Population, in diesem Fall den Erstklässlern in der Bundesrepublik Deutschland, bezüglich der interessierenden Merkmale entspricht. Nun stellt sich die Frage: Welche Merkmale wirken sich auf den mathematischen Lernerfolg von Schülerinnen und Schülern aus, d.h. welche Faktoren müssen auf ihre Repräsentativität hin geprüft werden? Es gibt ein Konglomerat an verschiedenen Determinanten, die sich auf den Lernerfolg in Mathematik auswirken können, wobei die jeweiligen Effekte einzelner Determinanten nicht abschließend geklärt sind (vgl. Roick, 2006). Folglich ist auch nicht abschließend zu beantworten, welche Eigenschaften bzw. Fähigkeiten erfasst werden müssen und auf eine repräsentative Verteilung hin zu prüfen sind. Als Anhaltspunkt für eine mögliche Repräsentativität der Stichprobe wurde die Schulleistung in Mathematik mit dem DEMAT 1+ abgeprüft und mit seiner Normierungsstichprobe verglichen. Es konnten vergleichbare Ausprägungen der Kenntnisse der Erstklässler in der vorliegenden Studie und in der Normierungsstudie des DEMAT 1+ vermerkt werden, wobei sich hier, wie bereits erwähnt, ein signifikanter Unterschied (p < .01, 2-seitig) zwischen den Mittelwerten zeigt. Als Fazit ist festzuhalten, dass eine Repräsentativität der Stichprobe zumindest in Hinblick auf die mathematische Schulleistung auf Basis der vergleichbaren Verteilungsformen vermutet werden kann. Neben der Frage der Repräsentativität der Stichprobe interessiert auch in Hinblick auf die weiteren Auswertungen bzw. die Interpretation der Auswertungsergebnisse, ob die Stichproben zu den drei MZP unterschiedlich zusammengesetzt sind. Es konnte, wie bereits angeführt, gezeigt werden, dass die Verschiebungen in der Zusammensetzung der drei Stichproben v.a. auf unsystematische Ausfälle bei der Durchführung der Studie zurückzuführen sind. Insofern ist zu vermuten, dass die Ausfälle zu den drei MZP die Daten nicht in eine Richtung verzerren.

Sind die Häufigkeiten zu allen drei Erhebungszeitpunkten normalverteilt?

Die Häufigkeitsverteilungen der Ergebnisse (Rohwertpunkte) der Schülerinnen und Schüler im Inventar „Rechenfische" sind zu allen drei MZP nicht normal verteilt. Die Angaben zu Mittelwert, Median, Schiefe und Kurtosis zeigen, dass sich die Verteilung bereits zum 2. MZP stark nach rechts verschiebt. Zum 3. MZP verstärkt sich dieser Effekt. Das Inventar „Rechenfische" ermöglicht es demnach bereits zum 2. MZP nicht mehr, die arithmetischen Kenntnisse der leistungsstärkeren Erstklässler differenziert zu erfassen. Gleichwohl lässt sich der leichte, aber stetige Anstieg der richtig gelösten Itemanzahlen auch noch zum 3. MZP als ein Hinweis für eine gute Differenzierung zu allen drei MZP bei Kindern mit Schwierigkeiten beim Erlernen arithmetischer Kenntnisse interpretieren.

Decken die Items des Inventars „Rechenfische" ein breites Spektrum unterschiedlicher Schwierigkeitsgrade ab und sind sie ausreichend trennscharf?

Die Items decken, wie gefordert, ein breites Spektrum an Schwierigkeitsgraden ab. Sogar zum dritten MZP enthält das Inventar noch Aufgaben unterschiedlicher Schwierigkeitsgrade, so dass die Voraussetzung für eine gute Differenzierungsfähigkeit geboten ist. Dass sich das Inventar insbesondere dafür eignet, die Kenntnisse von Erstklässlern mit Schwierigkeiten beim Erlernen arithmetischer Unterrichtsinhalte zu dokumentieren, zeigt sich an der großen Anzahl an relativ leichten Items. Da diese leichten Items jedoch nur im unteren Leistungsbereich zwischen den Leistungen der Kinder differenzieren und nicht im mittleren und oberen Leistungsbereich, fällt ihre Trennschärfe ebenfalls geringer aus. Insofern sind die angegebenen Anzahlen von Items, die im Bereich einer guten Trennschärfe liegen (zum 1. MZP 95 der insgesamt 169 Items, zum 2. MZP 70 und zum 3. MZP 71 Items), noch mal höher zu bewerten.

Erfasst das Inventar die arithmetischen Kenntnisse von Erstklässlern reliabel?

Grundvoraussetzung für eine reliable Erfassung der Ausprägung eines Merkmals ist, dass das eingesetzte Verfahren das Gütekriterium der Objektivität erfüllt. Dies ist in dem Inventar „Rechenfische" gewährleistet. Sowohl eine Durchführungsanweisung als auch eine Anleitung zur Codierung der Ergebnisse sind detailliert vorgegeben, so dass davon auszugehen ist, dass sowohl die Durchführungs- als auch die Auswertungsobjektivität sicher gestellt sind und dementsprechend diese Voraussetzung für die Reliabilität eines Verfahrens erfüllt ist. In Bezug auf die Test-Retest-Reliabilität des Inventars „Rechenfische" konnte vom 1. zum 2. und vom 2. zum 3. MZP jeweils eine hohe Korrelation (r_{tt}=.73 bzw. r_{tt}=.76) vermerkt werden. Auch zwischen dem 1. und 3. MZP konnte, obwohl hier mindestens 16 Schulwochen lagen, noch eine Korrelation von r_{tt}=.68 festgestellt werden. Anzumerken ist an dieser Stelle allerdings, dass die Test-Retestmethode problematisch zur Bestimmung der Reliabilität eines diagnostischen Verfahrens ist, wenn instabile und/oder zeitabhängige Merkmale erfasst werden, da geringe Werte entweder auf eine geringe Reliabilität des Tests oder auf eine geringe Stabilität des Merkmals zurückgeführt werden können (Bortz & Döring, 2006, S.197). Diese Problematik ergibt sich bei der Berechnung der Testhalbierungsreliabilität (split-half-Reliabilität) und der Einschätzung der internen Konsistenz (Cronbachs alpha) nicht. Daher bot es sich an diese Verfahren zur Reliabilitätsbestimmung ebenfalls zu berechnen und ihre Ergebnisse neben den Angaben zur Test-Retest-Reliabilität für eine Einschätzung der Reliabilität des Inventars „Rechenfische“ zu nutzen. Cronbachs alpha fiel für die Gesamttestwerte mit über r_{tt}=.90 hoch aus, ebenso der split-half-Koeffizient (jeweils bei =.91). Diese Befunde sprechen dafür, dass das Inventar „Rechenfische“ reliabel misst. Gleichwohl ist einschränkend darauf hinzuweisen, dass eine hohe Itemanzahl, wie sie das Inventar beinhaltet, Cronbachs alpha verzerrt und es höher ausfallen lässt. Des Weiteren ist nicht auszuschließen, dass der von Bortz und Döring (2006, S. 199) beschriebene „Transient Error“ sich auf die Höhe von Cronbachs alpha positiv ausgewirkt hat. Insgesamt ist jedoch auf Grundlage der angeführten Reliabilitätswerte davon auszugehen, dass das Inventar „Rechenfische“ die arithmetischen Kenntnisse der Erstklässler reliabel erfasst.

Erfasst das Inventar valide die arithmetischen Kenntnisse von Erstklässlern?

Um die Validität des Inventars „Rechenfische“ einzuschätzen, wurden verschiedene Berechnungen durchführt, von denen in diesem Beitrag nur einige angeführt werden konnten. Eine Möglichkeit, die Konstruktvalidität zu bestimmen, stellt die Interpretation der Korrelationen der Ergebnisse in dem zu evaluierenden Verfahren zu denen in diagnostischen Verfahren, die vorgeben andere Konstrukte zu erfassen, im Sinne der diskriminanten Validität dar. Um Aussagen darüber treffen zu können, inwieweit das Inventar „Rechenfische“ ein anderes Konstrukt erfasst als bspw. ein Intelligenztest oder ein Verfahren zur Erfassung der Lesekompetenz, wurden u.a. die Korrelationen der Ergebnisse im Inventar „Rechenfische“ mit den Ergebnissen im KFT 1-2 R sowie in der WLLP berechnet. Die Korrelation mit den Ergebnissen im KFT 1-2 R liegen zwischen r_{ct}=.51 und r_{ct}=.57, so dass aufgrund der immer noch geringen gegenseitigen Varianzaufklärung angenommen werden kann, dass der KFT 1-2 R und das Inventar „Rechenfische“ unterschiedliche Fähigkeiten bzw. Kompetenzen abprüfen. Zumal hier zu berücksichtigen ist, dass der KFT 1-2 R als schulnahes bzw. für die Schulleistungsprognostik konzipiertes Verfahren gilt, so dass die Korrelationen daher möglicherweise höher ausfallen, als dies mit einem eher schulfernen Intelligenztest der Fall gewesen wäre. Mit der WLLP korrelieren die Ergebnisse im Inventar nur gering (r_{ct}=.40 - .41), so dass davon auszugehen ist, dass das Inventar „Rechenfische“ andere Kenntnisse abprüft als die WLLP. Für die Validität des Inventars spricht des Weiteren, dass die Ergebnisse im Inventar „Rechenfische“ mindestens zu r_{ct}=.73 zu allen drei MZP mit dem DEMAT 1+ korrelieren. Die Korrelation fällt somit relativ hoch aus. Der DEMAT 1+ und das Inventar „Rechenfische“ scheinen ähnliche Konstrukte abzuprüfen. Bei einer Korrelation von mindestens r_{ct}=.73. zu allen drei MZP ist jedoch auch zu hinterfragen, ob die Existenz beider Verfahren noch berechtigt erscheint oder ob sie nicht im Wesentlichen das gleiche abprüfen und daher eines das andere ersetzt. Für eine Existenz beider Verfahren spricht die Annahme, dass das Inventar „Rechenfische“ durch die höhere Anzahl an Items und die Möglichkeit, die Auswertung anhand der Gesamtanzahl richtig gelöster Items und der erreichten Rohwertpunkte in den einzelnen Aufgabenbereichen vorzunehmen, einen detaillierten Einblick in die Kenntnisstände der Schülerinnen und Schüler bietet und darüber hinaus es ermöglicht, die Veränderungen in den Lernständen über einen längeren Zeitraum im Laufe des ersten Schuljahres hinweg zu dokumentieren. Mit dem DEMAT 1+ werden dagegen schnell und effizient am Ende des ersten Schuljahres die wesentlichen Lerninhalte in Mathematik über insgesamt 36 Items abgeprüft. Bei der Diskussion der Ergebnisse in Bezug auf die Annahmen zu den Extremgruppenvergleichen:

- Ergebnisse der Schülerinnen und Schüler in Grundschulklassen vs. Ergebnisse der Kinder in Diagnoseförderklassen,
- Ergebnisse der Kinder mit einer Anzahl an Rohwertpunkten im KFT 1-2 R >55 vs. Ergebnisse der Kinder mit einer Anzahl an Rohwertpunkten im KFT 1-2 R ≤55,

muss der Einfluss des Regressionseffektes beachtet werden. So ist bei den Ergebnissen der Gruppen „Schülerinnen und Schüler aus Diagnoseförderklassen“ sowie „Kinder mit einem KFT 1-2 R Rohwertpunkt ≤55“ aufgrund der Ergebnisse zum 1. MZP zu erwarten, dass sich bei ihnen eher ein Bodeneffekt zeigt als bei den Kindern der Gruppen „Grundschülerinnen/Grundschüler“ bzw. „Kinder mit einem KFT 1-2 R Rohwertpunkt >55“. Bei diesen Gruppen ist hingegen die Wahrscheinlichkeit eines Deckeneffektes höher. Es ist anzunehmen, dass sich die Ergebnisse beider Gruppen aufgrund des Regressionseffektes über die drei MZP hinweg annähern. Die Frage ist, ob sie sich trotzdem noch signifikant voneinander unterscheiden. Dieses konnte bestätigt werden. Insofern ist zu konstatieren, dass das Inventar die aufgrund der Zusammensetzung der Extremgruppen anzunehmenden Unterschiede tatsächlich abbildet und damit ein weiterer Beleg dafür gegeben ist, dass das Inventar die arithmetischen Kenntnisse valide erfasst. Die höchste Prognosesicherheit lässt sich erzielen, wenn die Ergebnisse vom 1. und 2. MZP summiert werden. Die Prognose auf die Ergebnisse im DEMAT 1+ zum 3. MZP kann dann mit einem R^2 von .70 als hoch eingeschätzt werden. Trotz dieser hohen prognostischen Aussagekraft lassen sich aber mit dem Inventar keine sicheren Aussagen über die Lernentwicklung einer einzelnen Schülerin/eines einzelnen Schülers treffen. Die Güte der Prognosen lässt sich lediglich als Zeichen dafür werten, dass das Inventar anscheinend die für den weiteren Lernerfolg entscheidenden Inhalte abprüft.

Zusammenfassend lässt sich konstatieren: Werden die Angaben zur Validität insgesamt betrachtet, kann das Inventar „Rechenfische“ als valides Verfahren zur Erfassung der Lernstände von Erstklässlern zu drei MZP im zweiten Halbjahr der ersten Klasse angesehen werden.

Diskussion der Ergebnisse in Hinblick auf die Möglichkeit, mit dem Inventar Lernfortschritte zu dokumentieren

Das Inventar „Rechenfische“ erfasst die Lernstände von Erstklässlern in Mathematik zu drei MZP im zweiten Halbjahr der ersten Klasse objektiv, reliabel und valide. Die Lernfortschritte einer Schülerin/eines Schülers lassen sich mit dem Verfahren jedoch trotzdem nur vorsichtig über den Vergleich zur Entwicklung der Gesamtstichprobe schätzen. Diese Einschränkung ist zum einen in dem von Rost (2004, S. 280) beschriebenen Validitätsdilemma der Veränderungsmessung begründet als auch in der Problematik der Interpretation von Differenzwerten aufgrund des Einflusses verschiedener Größen auf ihre Ausprägungen. Um tatsächlich Lernfortschritte der Erstklässler auch inhaltlich zu bestimmen, müsste das Inventar so gestaltet sein, dass es die Lernentwicklung auf Basis der probabilistischen Testtheorie als einheitliche Kurve abbildet, wie dieses bspw. der vorgestellte Osnabrücker Test zur Zahlbegriffsentwicklung – OTZ (van Luit, van de Rijt & Hasemann, 2001) für den vorschulischen Bereich leistet. Dass das Inventar „Rechenfische" es trotzdem grundsätzlich ermöglicht, Lernstandsveränderungen in den Ausprägungen der arithmetischen Kenntnisse der Schülerinnen und Schüler über das zweite Halbjahr der ersten Klasse hinweg zu erfassen, wird an der beschriebenen Verschiebung der Mittelwerte der Häufigkeiten richtig gelöster Items über die drei MZP hinweg deutlich. Hieran zeigt sich die Eignung des Verfahrens für die formative Evaluation des Unterrichts besonders deutlich: Es erweist sich als ausreichend sensibel, um Lernfortschritte einer Lerngruppe im 2. Halbjahr des ersten Schuljahres zu erfassen.

Diskussion

Das Inventar „Rechenfische" wurde für den Anfangsunterricht Arithmetik konzipiert und zu drei MZP anhand einer Stichprobe von 1453 (1. MZP), 1455 (2. MZP) bzw. 1352 (3. MZP) Erstklässlern evaluiert. Es zeigte sich, dass das Inventar die klassischen testtheoretischen Gütekriterien Objektivität, Reliabilität und Validität erfüllt. Folglich können mit ihm die arithmetischen Kenntnisse von Erstklässlern zu drei MZP im zweiten Halbjahr des ersten Schuljahres dokumentiert und über den Vergleich mit der Entwicklung der Gesamtstichprobe erste Hinweise auf die Lernfortschritte gewonnen werden.

Zu Beginn dieses Beitrags wurden mehrere Gründe für die Konzeption eines Inventars und für seinen Einsatz in der Schule angeführt. So wurde dargelegt, wie wichtig eine rechtzeitige und präzise Erfassung der Lernstände von Kindern im Anfangsunterricht Mathematik ist. Ebenfalls wurde die generelle Notwendigkeit einer Verbesserung der Prävention von Lernschwierigkeiten in Mathematik angesprochen. Das konzipierte Inventar soll hierzu einen Beitrag leisten, indem es Lehrenden frühzeitig signalisiert, wenn Schwierigkeiten beim Erlernen arithmetischer Unterrichtsinhalte bei einzelnen Kindern auftreten. Es eröffnet somit die Möglichkeit, auf Grundlage der erhobenen Lernstände früh und gezielt Fördermaßnahmen zu initiieren bzw. über den Vergleich der Lernstände zu den einzelnen MZP Rückmeldungen über die Wirksamkeit eingeleiteter Fördermaßnahmen zu erhalten. Im Vergleich zu rein qualitativ ausgerichteten Verfahren, die in Einzeltestungen durchgeführt werden müssen, eignet sich das Inventar „Rechenfische" v.a. für Lehrkräfte, die fachfremd eingesetzt werden, noch sehr unerfahren sind, mit der Klassensituation als solche bereits stark gefordert sind und/oder Einzeltestungen im Rahmen ihres Unterrichts nicht ermöglichen können. Bezüglich der Möglichkeit einer detaillierten Fehler- und Strategieanalyse ist jedoch auch der Nutzen qualitativ orientierter Verfahren herauszustellen. Im Rahmen von Einzeltestungen können die Strategien der Kinder wesentlich genauer beobachtet und analysiert werden, als dies im Rahmen eines Gruppentests möglich ist. Nachteil jener Verfahren und damit wieder ein Vorteil des Inventars „Rechenfische" ist hingegen, dass Vergleiche der Lernstände nicht aussagekräftig sind, wenn Verfahren nicht standardisiert durchgeführt werden und ihre testtheoretische Güte nicht gewährleistet ist. Das Inventar „Rechenfische" wurde in Hinblick auf seine testtheoretische Güte geprüft und bietet die Möglichkeit, die normorientierte Bezugsgröße zur Einschätzung des Kenntnisstandes heranzuziehen, um einen umfassenden Einblick in den Leistungsstand eines Kindes zu erhalten. In Hinblick auf den Vergleich mit einer normorientierten Bezugsgröße muss jedoch einschränkend angemerkt werden, dass die vorliegenden Daten ausschließlich in Mecklenburg-Vorpommern erhoben wurden. Es ist gleichwohl zu vermuten, dass die Ergebnisse mit einer repräsentativen Stichprobe nicht erheblich anders als die hier berichteten ausfallen. Ein weiterer Einsatz des Inventars „Rechenfische" in einer nach repräsentativen Aspekten ausgewählten Eichstichprobe wäre trotzdem sinnvoll. Er könnte Zweifel an der Generalisierung der Ergebnisse ausräumen.

Ein weiterer Forschungskomplex, der sich anschließt, umfasst die folgende Fragestellung: Ist das Inventar „Rechenfische" eine effektive Möglichkeit zur Prävention von Lernschwierigkeiten in Mathematik oder sind hierfür weitere zusätzliche Maßnahmen wie bspw. die Beratung von Lehrenden bei der Interpretation der Ergebnisse, ein zusätzlicher Einsatz qualitativ orientierter Verfahren oder das Durchführen in sich geschlossener Förderprogramme notwendig? Hier wäre auch zu klären, ob sich das Inventar mit Förderprogrammen im Sinne eines „whole-in-one"-Paketes, wie es Hartke, Diehl und Vrban (2008) anregen, kombinieren lässt.

Literatur

Aster, v., M. (2005). Wie kommen die Zahlen in den Kopf? In M. v. Aster & J.H. Lorenz (Hrsg.), Rechenstörungen bei Kindern. Neurowissenschaft, Psychologie, Pädagogik (S. 13-33). Göttingen: Vandenhoeck & Ruprecht.

Aunola, K., Leskinen, E., Lerkkanen, M.-K. & Nurmi, J.E. (2004). Developmental Dynamics of Math Performance From Preschool to Grade 2. Journal of Educational Psychology, 96, 699-713.

Balzer, L., Fritz, A., Ricken, G. & Jäger, R.S. (2007). Der Rechenschwäche auf der Spur – eine Re-Analyse von Mathematik-Leistungsdaten eines kompletten Schülerjahrgangs der achten Klassenstufe in Rheinland-Pfalz. Psychologie in Erziehung und Unterricht, 54, 177-190.

Baroody, A. (2006). Why Children Have Difficulties Mastering the Basic Number Combinations and How to Help Them. Teaching children mathematics, 1, 22-31.

Bortz, J. & Döring, N. (2006). Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler (4. Aufl.). Heidelberg: Springer.

Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Hrsg.), Psychologie des Unterrichts und der Schule. Enzyklopädie der Psychologie. Serie I Pädagogische Psychologie (S. 177-214). Göttingen: Hogrefe.

Carpenter, T.P., Franke, M.L. & Levi, L. (2003). Thinking Mathematically Integrating Arithmetic and Algrebra in Elementary School. Portsmouth: Heinemann.

Dehaene, S. (1992). Varieties of numerical abilities. Cognition, 44, 1-42.

Dehaene, S. (1999). Der Zahlensinn oder Warum wir rechnen können. Basel: Birkhäuser Verlag.

Diehl, K. & Hartke, B. (2007). Curriculumnahe Lernfortschrittsmessung. Sonderpädagogik, 37, 195-211.

Diehl, K., Hartke, B. & Knopp, E. (2009). Curriculum-Based Measurement & Leerlingsonderwijsvolgsysteem – Konzepte zur theoriegeleiteten Lernfortschrittsmessung im Anfangsunterricht Deutsch und Mathematik? Zeitschrift für Heilpädagogik, 4, 122-130.

Fritz, A., Ricken, G. & Gerlach, M. (2007). Kalkulie. Handreichung zur Durchführung der Diagnose. Berlin: Cornelsen.

Fuson, K.C. (1982). An Analysis of the Counting-On Solution Procedure in Addition. In T.P. Carpenter, J.M. Moser & T.A. Romberg (Hrsg.), Addition and Subtraction: a Cognitive Perspective (S. 67-81). Hillsdale: Erlbaum.

Fuson, K.C. (1988). Children´s counting and concepts of number. New York: Springer.

Gaupp, N., Zoelch, Ch. & Schumann-Hengsteler, R. (2004). Defizite numerischer Basiskompetenzen bei rechenschwachen Kindern der 3. und 4. Klassenstufe. Zeitschrift für Pädagogische Psychologie, 18, 31-42.

Geary, D.C., Hoard, M.K., Byrd-Craven, J. & DeSoto, M.C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. Journal of Experimental Child Psychology, 88, 121-151.

Gerster, H.-D. & Schultz, R. (2004). Schwierigkeiten beim Erwerb mathematischer Konzepte im Anfangsunterricht. Bericht zum Forschungsprojekt Rechenschwäche – Erkennen, Beheben, Vorbeugen. Online verfügbar unter: www.freidok.uni-freiburg.de/volltexte/1397/-21k. Zugriff am 26.7.2006.

Grassmann, M., Klunter, M., Köhler, E., Mirwald, E., Raudies, M. & Thiel, O. (2002). Mathematische Kompetenzen von Schulanfängern. Teil 1: Kinderleistungen – Lehrererwartungen. Potsdamer Studien zur Grundschulforschung, 30. Universität Potsdam.

Grassmann, M., Klunter, M., Köhler, E., Mirwald, E., Raudies, M. & Thiel, O. (2003). Mathematische Kompetenzen von Schulanfängern. Teil 2: Was können Kinder am Ende der Klasse 1? Potsdamer Studien zur Grundschulforschung, 31. Universität Potsdam.

Grube, D. (2006). Entwicklung des Rechnens im Grundschulalter. Basale Fertigkeiten, Wissensabruf und Arbeitsgedächtniseinflüsse. Münster: Waxmann.

Grube, D. & Hasselhorn, M. (2006). Längsschnittliche Analysen zur Lese-, Rechtschreib- und Mathematikleistung im Grundschulalter: zur Rolle von Vorwissen, Intelligenz, phonologischem Arbeitsgedächtnis und phonologischer Bewusstheit. In I. Hosenfeld & F.-W. Schrader (Hrsg.), Schulische Leistung. Grundlagen, Bedingungen, Perspektiven (S. 87-105). Münster: Waxmann.

Hartke, B., Diehl, K. & Vrban, R. (2008). Planungshilfen zur schulischen Prävention – Früherkennung und Intervention bei Lern- und Verhaltensproblemen. In J. Borchert, B. Hartke & P. Jogschies (Hrsg.), Frühe Förderung entwicklungsauffälliger Kinder und Jugendlicher (S. 218-234). Stuttgart: Kohlhammer.

Hofer, M. (1986). Sozialpsychologie erzieherischen Handelns. Göttingen: Hogrefe.

Horstkemper, M. (2006). Fördern heißt diagnostizieren. Diagnostische Kompetenz als Basis für pädagogisches Handeln. In G. Becker, M. Horstkemper, E. Risse, L. Stäudel, R. Werning & F. Winter (Hrsg.), Friedrich Jahresheft: Stärken entdecken – Können entwickeln (S. 4-5). Seelze: Verlag Erhard Friedrich.

Jordan, N.C., Hanich, L.B. & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. Journal Experimental Child Psychology, 85, 103-119.

Kawthar, K.A. (2005). Reliabilitäts- und Validitätsuntersuchungen zum neuen kognitiven Fähigkeitstest für die Primarstufe (KFT 1-2 R) unter besonderer Berücksichtigung von Kindern mit Hochbegabungen und Lernbehinderungen. Unveröffentlichte Dissertation. Rostock: Universität.

Kawthar, K.A. & Perleth, C. (2005). Kognitiver Fähigkeitstest für die Primarstufe (KFT 1-2 R). Versuchsversion. Rostock.

Klauer, K.J. (2006). Erfassung des Lernfortschritts durch curriculum-basierte Messung. Heilpädagogische Forschung, 32, 16-26.

Koch, K. & Ellinger, S. (2007). Flexible Schuleingangsphase für Kinder mit sonderpädagogischem Förderbedarf. Eine kritische Bilanz zur Effektivität von Diagnose- und Förderklassen. Zeitschrift für Heilpädagogik, 3, 82-90.

Koch, K., Hartke, B. & Blumenthal, Y. (2008). Die Lernausgangslage von Kindern mit besonderem Förderbedarf in Grundschulklassen 1 und Diagnoseförderklassen. Rostock: Universität, Institut für Sonderpädagogische Entwicklungsförderung und Rehabilitation.

Koch, K. & Knopp, E. (2010). Pädagogische Förderung bei Rechenschwierigkeiten. In K. Novotny (Hrsg.), Pädagogik in der Sonderpädagogik (S. 41-69). Hohengehren: Schneider.

Knopp, E. (2010). Theoretische Grundlagen, Konzeption und Güte des Inventars „Rechenfische“. Ein Verfahren zur Dokumentation von Fortschritten beim Erlernen arithmetischer Kenntnisse im Anfangsunterricht Mathematik. München: Dr. Hut.

Krajewski, K. (2003). Vorhersage von Rechenschwäche in der Grundschule. Hamburg: Kovac.

Krajewski, K., Küspert, P., Schneider, W. & Visé, M. (2002). DEMAT 1+. Deutscher Mathe-

matiktest für erste Klassen. Göttingen: Hogrefe.
Krajewski, K. & Schneider, W. (2006). Mathematische Vorläuferfertigkeiten im Vorschulalter und ihre Vorhersagekraft für die Mathematikleistungen bis zum Ende der Grundschulzeit. Psychologie und Unterricht, 53, 246-262.
Krohne, H.W. & Hock, M. (2006). Psychologische Diagnostik: Grundlagen und Anwendungsfelder. Stuttgart: Kohlhammer.
Küspert, P. & Schneider, W. (1998). Würzburger Leise Leseprobe (WLLP). Ein Gruppentest für die Grundschule. Göttingen: Hogrefe.
Luit, v., J.E.H., Rijt, v.d., B.A.M. & Hasemann, K. (2001). Osnabrücker Test zur Zahlbegriffsentwicklung. Göttingen: Hogrefe.
Mazzocco, M.M.M. & Thompson, R.E. (2005). Kindergarten Predictors of Math Learning Disability. Learning Disabilities Research & Practice, 20, 142-155.
Mittelberg, A. (2004). Rechenschwächen in der Hauptschule – Eine Studie zu den Rechenleistungen in den Klassen 7 und 8. Dissertation an der Universität Hannover. Online verfügbar unter: http://deposit.d-nb.de/cgi-bin/dokserv?idN=972557490&dok_var=d1&dok_ext=pdf&filename=972557490.pdf. Zugriff am 7.7.2007.
Moser Opitz, E. (2006). Assessments, Förderplanung, Förderdiagnostik – messen und/oder fördern? Schweizerische Zeitschrift für Heilpädagogik, 9, 5-11.
Moser Opitz, E. (2007). Rechenschwäche/Dyskalkulie. Theoretische Klärungen und empirische Studien an betroffenen Schülerinnen und Schülern. Bern: Verlag Paul Haupt.
Ostad, S.A. (1997). Developmental differences in addition strategies: a comparison of mathematically disabled and mathematically normal children. British Journal of Educational Psychology, 67, 345-357.
Padberg, F. (2004). Didaktik der Arithmetik für die Lehrerausbildung und Lehrerfortbildung (3.Aufl.). Heidelberg: Spektrum.
Radatz, H., Schipper, W., Ebeling, A. & Dröge, R. (1996). Handbuch für den Mathematikunterricht. Hannover: Schroedel.
Resnick, L.B. (1989). Developing Mathematical Knowledge. American Psychologist, 44, 162-169.
Rinkens, H.-D. (1997). Arithmetische Fähigkeiten am Schulanfang. Universität Paderborn. Online verfügbar unter: www.rinkens-hd.de/_data/AritFaeh.pdf. Zugriff am 4.5.2006.
Roick, T. (2006). Kognitive Determinanten unterschiedlicher Rechenleistungen in der Primarstufe: Arbeitsgedächtnismerkmale und Aufmerksamkeitsaspekte. Göttingen: Georg-August-Universität.
Rost, J. (2004). Lehrbuch Testtheorie – Testkonstruktion (2.Aufl.). Bern: Hans Huber.
Scherer, P. (2007). Elementare Rechenoperationen. In J. Walter & F.B. Wember (Hrsg.), Sonderpädagogik des Lernens (S. 590-605). Göttingen: Hogrefe.
Schrader, F.-W. (2006). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost (Hrsg.), Handwörterbuch Psychologie (3. Aufl., S. 95-100). Weinheim: Beltz.
Siegler, R., Deloache, J. & Eisenberg, N. (2006). How Children Develop (2. Aufl.). New York: Worth Publishers.
Stamm, M. (2005). Zwischen Exzellenz und Versagen. Frühleser und Frührechnerinnen werden erwachsen. Zürich: Rüegger.
Stern, E. (1998). Die Entwicklung des mathematischen Verständnisses im Kindesalter. Lengerich: Pabst Science Publishers.
Stern, E. (2003). Früh übt sich – Neue Ergebnisse aus der LOGIK-Studie zum Lösen mathematischer Textaufgaben. In A. Fritz, G. Ricken & S. Schmidt (Hrsg.), Rechenschwäche. Lernwege, Schwierigkeiten und Hilfen bei Dyskalkulie (S. 116-130). Weinheim: Beltz.
Strathmann, A.M. & Klauer, K.J. (2008). Diagnostik des Lernverlaufs. Eine Pilotstudie am Beispiel der Entwicklung der Rechtschreibkompetenz. Sonderpädagogik, 38, 5-24.
Torbeyns, J., Verschaffel, L. & Ghesquière, P. (2002). Strategic competence: Applying Siegler´s theoretical and methodological framework to the domain of simple addition. European Journal of Psychology of Education, 17, 275-291.
Torbeyns, J., Verschaffel, L. & Ghesquière, P. (2004). Strategic aspects of simple addition and subtraction: the influence of mathematical ability. Learning and Instruction, 14, 177-195.

Walter, J. (2008). Adaptiver Unterricht erneut betrachtet: Über die Notwendigkeit systematischer formativer Evaluation von Lehr- und Lernprozessen und die daraus resultierende Diagnostik und Neudefinition von Lernstörungen nach dem RTI-Paradigma. Zeitschrift für Heilpädagogik, 6, 202-215.

Wartha (2009). Rechenstörungen in der Sekundarstufe: die Bedeutung des Übergangs von der Grundschule zur weiterführenden Schule. In A. Heinze, M. Grüßing (Hrsg.), Mathematiklernen vom Kindergarten bis zum Studium. Kontinuität und Kohärenz als Herausforderung für den Mathematikunterricht (S. 157-180). Münster: Waxmann.

Weberschock, U. & Grube, D. (2006). Zur Spezifität von Einflüssen der Arbeitsgedächtniskapazität und des arithmetischen Faktenwissens auf Rechenleistungen von Viertklässlern. Psychologie in Erziehung und Unterricht, 53, 291-302.

Werning, R. (2006). Lern- und Entwicklungsprozesse fördern – Pädagogische Beobachtung im Alltag. In G. Becker, M. Horstkemper, E. Risse, L. Stäudel, R. Werning & F. Winter (Hrsg.), Friedrich Jahresheft: Stärken entdecken – Können entwickeln (S. 11-15). Seelze: Friedrich.

Anschriften der Autoren:

DR. EVA KNOPP
Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik
Olshausenstr. 62
24098 Kiel
knopp@ipn.uni-kiel.de

PROF. DR. BODO HARTKE
Institut für Sonderpädagogische Entwicklungsförderung und Rehabilitation
Universität Rostock
Philosophische Fakultät
August-Bebel-Str. 28
18051 Rostock
bodo.hartke@uni-rostock.de