Ricken, Gabi; Fritz, Annemarie; Balzer, Lars

Mathematik und Rechnen - Test zur Erfassung von Konzepten im Vorschulalter (MARKO-D). Ein Beispiel für einen niveauorientierten Ansatz

Empirische Sonderpädagogik 3 (2011) 3, S. 256-271

Empirische Sonderpädagogik

ISSN 1869-4845 3. Jahrgang 2011 | Heft 3

Schwerpunktthema:
Trends in der sonderpädagogischen Diagnostik
Gastherausgeberin: Gabi Ricken

Pabst Science Publishers

Empirische Sonderpädagogik, 2011, Nr. 3, S. 256-271

Mathematik und Rechnen – Test zur Erfassung von Konzepten im Vorschulalter (MARKO-D) – ein Beispiel für einen niveauorientierten Ansatz

Gabi Ricken[1], Annemarie Fritz[2], Lars Balzer[3]

[1]Universität Hamburg; [2]Universität Duisburg-Essen; [3]Eidgenössisches Hochschulinstitut für Berufsbildung (EHB) Zollikofen, Schweiz

Eine prozessorientierte Diagnostik erfordert eine theoretische Dimension, anhand derer Veränderungen beschrieben und interpretiert werden können. Für den Test zur Erfassung von Konzepten im Vorschulalter, MARKO- D, wurde eine solche Dimension bzw. Skala aus der Analyse theoretischer Aussagen und empirischer Befunde für die Entwicklung mathematischer Konzepte im Vorschulalter abgeleitet. Die fünf zentralen Konzepte sind: Zählzahl, mentaler Zahlenstrahl, Kardinalzahl, Teil-Teil-Ganzes-Konzept und Relationalzahl. Mit der empirischen Prüfung gelang der Nachweis der Gültigkeit des Modells (eindimensionales Raschmodell). Damit steht ein Testkonzept zur Verfügung, das sowohl den Vergleich mit der Sozialnorm als auch insbesondere eine Einordnung eines Kindes in einen Entwicklungsverlauf und damit ein individuelles Bezugssystem erlaubt.

Schlüsselwörter: mathematische Konzepte, prozessorientierte Diagnostik, Kompetenzdiagnostik, Rechenstörungen, Entwicklungsmodell

Math and Calculation – A Test for Diagnosing Concepts at Pre-school Age – An Example of a Level-oriented Approach

Process-oriented diagnostics require a theoretical framework which allows to describe and interpret individual competence changes. For MARKO-test, a corresponding dimension/scale of mathematical achievement in preschool age was developed on the basis of theoretical assumptions and empirical data. Five essential concepts are: numbers as counting sequence, ordinal number line, cardinal understanding, part – part – whole and concept of congruent intervals. There is empirical evidence for the validity of the model, using a unidimensional Rasch model. Therefore, a concept of testing is available which on the one hand allows to compare individual data with a social norm and on the other hand is usable to make valid statements about individual changes and the current development status of a child.

Key words: mathematical concepts, process-oriented diagnostics, specific learning disabilities, dyscalculia, competencies, developmental model

Grundidee der Testkonstruktion

Die aktuelle Entwicklung in der pädagogischen Diagnostik zeigt eine Veränderung der Aufgabenstellung von einer Zustandserhebung zu einer Verlaufserfassung. Damit werden diagnostische Verfahren benötigt, die für Veränderungen sensibel sind: Veränderungen, die einerseits durch quantitative Zuwächse und andererseits durch Weiterentwicklungen von Wissensinhalten entstehen.

Im Folgenden wird mit dem Test MARKO-D ein Beispiel für einen Ansatz vorgestellt, der die Abbildung qualitativer Veränderungen erlaubt. Inhaltlich betrachtet werden mathematische Konzepte erfasst, die sich bereits im Vorschul- und frühen Grundschulalter entwickeln. Die Auswahl dieses Altersbereichs lässt sich mit der besonderen prognostischen Bedeutung der Entwicklung mathematischer Kompetenzen für das Schulalter begründen (Landerl & Kaufmann, 2008). Kinder mit besonders guten oder schlechten Voraussetzungen könnten so vor Schulbeginn diagnostiziert werden, um bereits zu so frühen Zeitpunkten ihre Entwicklung angemessen zu unterstützen.

Üblicherweise werden zwei Prinzipien genutzt, um Entwicklungsstände zu erfassen: eine Auswahl von Aufgaben, die Curriculumsanforderungen repräsentiert (z.B. DEMAT 4 – Gölitz, Roick & Hasselhorn, 2006) oder eine Auswahl von sogenannten Basiskompetenzen (Eggenberger Rechentests 1+, Schaupp, Holz & Lenart, 2007). Tests, die so konstruiert werden, enthalten meist mehrere Untertests. Durch die Bestimmung der richtigen Lösungen pro Untertest werden die individuellen Werte (richtig gelöste Aufgaben) mit der Verteilung der Werte in der sozialen Normgruppe verglichen. Insofern sich Untertests zu Faktoren bilden, sind Stärken und Schwächen von Kindern auch für Untertestgruppen (Faktoren) ausweisbar. Die Frage, die sich daraufhin stellt, ist die nach der inhaltlichen Interpretation. Was bedeutet eine quantitativ geringere Ausprägung, eine geringe Punktzahl in einem Untertest? Im Grunde kann damit ein Bereich oder ein Aufgabentyp benannt werden, bei dem ein Kind nur wenige richtige Lösungen erzeugt, also über kein ausreichend anwendbares Wissen verfügt, um in vergleichbarer Weise zu seiner Altersgruppe oder einem Kriterium die Anforderungen zu bewältigen. Bei Kindern, die Schwierigkeiten in der Entwicklung zeigen, entsteht eine Liste von Aufgaben, die nicht gut genug bewältigt werden.

Damit ist der Bereich inhaltlich umrissen, an dem in einem Förderprozess gearbeitet werden sollte. Noch nicht beantwortet ist, wie der Aufbau von Wissen nun erfolgen kann, in welcher Reihenfolge Aufgaben bearbeitet werden müssen, um Kompetenzen oder Konzepte zu entwickeln.

Mit dem MARKO-D wird ein Ansatz vorgestellt, der von einer Ordnung im Aufbau mathematischen Wissens ausgeht (**kompetenzorientierter Ansatz**). Ausgewählt wird als Wissensbereich das deklarative Wissen, speziell mathematische Konzepte, die sich im Vorschulalter entwickeln. Diese Wahl ist damit zu begründen, dass Konzepte eine entscheidende Rolle spielen, wenn mathematische Sachverhalte verstanden werden sollen. Damit wird der Aspekt der Beherrschung, insbesondere der Schnelligkeit der Beherrschung von Rechenoperationen, der als eine Bedingung für Rechenstörungen diskutiert wird (Haffner et al, 2005), hier explizit nicht weiterverfolgt.

Die Konstruktion des MARKO-D erfordert zwei Arbeitsschritte: Erstens sind Konzepte auszuwählen, die für alle Kinder eine zentrale Bedeutung insofern haben, als dass sie erworben werden müssen, um den nächsten Entwicklungsschritt zu vollziehen. Zweitens ist eine Reihenfolge der Konzepte auf der Basis theoretischer und empirischer Befunde sowie auf der Grundlage einer empirischen Überprüfung herzustellen. Für den zweiten Arbeitsschritt wird mit einem eindimensionalen Raschmodell geprüft, ob sich auf der Basis der Lösungswahrscheinlichkeiten der Items eine Skala bilden lässt, die mit den theoretischen Vorhersagen übereinstimmt. Für Items, die hinsichtlich ihrer Schwierigkeit eng beieinander liegen, ist bei Modellgültigkeit davon auszugehen, dass sie je ein Konzept repräsentieren. Die Itemgruppen sollten dann die angenommene Abfolge der Konzepte in der Entwicklung abbilden. Damit läge eine Skala mit Abschnitten, mit unterscheidbaren Entwicklungsniveaus vor.

Ein so konstruierter Test erlaubt insgesamt Aussagen der folgenden Art: Durch die Anzahl gelöster Items (Rohwertsumme) liegt ein Kind z.B. in Niveau III. Das bedeutet, das Kind entwickelt gerade das Konzept dieses Niveaus. Die Items der darunter liegenden Niveaus I und II werden bewältigt, was als Vorhandensein der entsprechenden Konzepte zu interpretieren ist. Als Nächstes ist die Bewältigung der Items des darüber liegenden Niveaus und damit die Entwicklung dieses Konzepts zu erwarten.

Die *Grundidee* der Entwicklung des Testkonzeptes für MARKO-D ist wie folgt zusammenzufassen: Es werden zentrale Konzepte der frühen mathematischen Entwicklung ausgewählt, die Annahmen zu ihrer Ordnung hinsichtlich der Aufeinanderfolge in ihrer Entwicklung werden empirisch und auf der Basis des eindimensionalen Raschmodells geprüft. Die erhaltene Skala ermöglicht eine Platzierung eines Kindes und damit die Aussage, welche Entwicklungsschritte ein Kind bewältigt hat, was es gegenwärtig entwickelt und was in der Zukunft entstehen wird. Dieser Stand kann mit anderen Kindern im Sinne der Sozialnorm abgeglichen werden. Von besonderem Wert ist jedoch, dass der Entwicklungsstand bei einer wiederholten Testung durch eine Veränderung des Platzes auf der Skala beschrieben werden kann. Das kann sich als quantitativer Zuwachs innerhalb eines Niveaus und als qualitativer Zuwachs durch den Wechsel ins nächste Niveau ausdrücken.

Zu benennen sind nunmehr die ausgewählten Konzepte, zu beschreiben ist die empirische Prüfung und darzustellen ist exemplarisch die Testumsetzung. Erste Ergebnisse hinsichtlich der Beobachtung von Veränderungen bei Kindern entlang dieser Skala liegen einschließlich von Interventionsbedingungen vor und werden an späterer Stelle publiziert.

Wesentliche Konzepte der mathematischen Entwicklung im Vorschulalter

Empirische Untersuchungen haben in den letzten Jahren gezeigt, zu welchen Leistungen Säuglinge und Kleinkinder bereits in der Lage sind. Mengenveränderungen und Mengenunterschiede werden, wenn sie deutlich genug sind, sowohl bei kleinen als auch bei großen Mengen bemerkt (Butterworth, 2005). Kinder sind von Geburt an auf den Umgang mit Numerositäten durch einen „Zahlensinn" (number sense) vorbereitet.

Wesentliche Schritte im Vorschulalter sind die Entstehung einer festen Reihenfolge von Zahlworten, die einen mentalen Zahlenstrahl bilden, das Aus- und Abzählen von Mengen erlauben, indem Zahlworte und Zählobjekte einander zugeordnet werden. Zunächst steht das letzte Zahlwort für das zuletzt gezählte Objekt. Der Entwicklung des Zählens wird insgesamt eine große Bedeutung beigemessen. Kinder erlangen im Vorschulalter erste Einsichten in die Vermehrung, Verminderung und Teilbarkeit von Mengen, erste Einsichten hinsichtlich der Bedeutung der Anzahlen von Mengen als wesentliche Merkmale. Zahlen bezeichnen Mengen mit spezifischen Anzahlen, Mengen mit gleichen Anzahlen von Elementen sind gleich groß und natürliche Zahlen setzen sich aus anderen Zahlen (Mengen) zusammen. Zahlen drücken auch Abstände zwischen den natürlichen Zahlen aus. Ziffern und Zahlworte werden ineinander transkodierbar (Kaufmann & Nuerk, 2007). Mit dem nachfolgend beschriebenen Modell, das dem MARKO-D Test zugrunde liegt, werden diese Konzepte präzisiert und im Sinne von Niveaus aufeinander bezogen.

Da das Ziel in der Konstruktion eines Testverfahrens für Kinder ab 4 Jahren bestand, wurden die ganz früh entstehenden Fertigkeiten nicht einbezogen. Im Folgenden werden die für den Test gesetzten Entwicklungsniveaus expliziert.

Fünf Entwicklungsniveaus

Niveau I: Zählzahl

Zahlen werden zunächst nur als Wortreihe gelernt. Ganz allmählich wird die Bedeutung von Zahlen als Zähl- und später als Kardinalzahl konstruiert. Zunächst werden kleine Mengen aus- und abgezählt, indem jedem Objekt ein Zahlwort zugeordnet wird. Die Frage danach, wie viele Elemente die Menge enthält, wird mit dem letzten Zahlwort beantwortet. Aus den Befunden von Le Corre et al. (2006) und Wynn (1990, 1992) lässt sich schlussfolgern, dass sich das Aus- und Abzählen kleiner Mengen von Zahlwort zu Zahlwort entwickelt: Zuerst wird zuverlässig nur ein Objekt, später werden zwei Objekte, danach drei und schließlich vier Objekte ausgezählt. Wenn vier Objekte sicher ab- und ausgezählt werden können, ist das Zählprinzip erworben.

Niveau II: Repräsentation eines mentalen Zahlenstrahls

In einer nächsten Phase wird das Wissen über Ordnungen von Zahlen differenzierter. Zahlworte sind geordnet und werden allmählich "größer". Es wird angenommen, dass Kinder Zahlenrepräsentationen in der Form eines mentalen Zahlenstrahls repräsentieren. Jede Zahl hat eine Position auf diesem Zahlenstrahl, ohne dass Abstände quantifiziert werden. Zahlen vor oder nach einer Zahl können benannt werden (Vorgänger- und Nachfolgerzahlen). Zahlen, die in der Reihe später kommen, sind „größer", Vorgängerzahlen „kleiner". Mit diesem Wissen können Zahlen über ihre Position in der Zahlwortreihe miteinander verglichen werden. Da Kinder das Vermehren und Vermindern von Mengen verstehen, sind Additionsaufgaben (a + b = ?) numerisch präzise zu lösen. Dies tun die Kinder, indem sie Mengen zusammenschieben und die Zahlwortreihe jeweils bei 1 beginnend den Objekten zuordnen. Mit dieser Kompetenz können Rechenoperationen als abstrakte (Zähl-) Handlungen vollzogen werden, wobei alle (Teil-) Mengen einzeln ausgezählt werden müssen.

Niveau III: Kardinalität und Zerlegbarkeit

Haben die Kinder gelernt, die Mächtigkeit von Mengen aus- und abzuzählen sowie die Gesamtmenge mit dem letzten Zahlwort zu benennen (last-word-rule, Fuson, 1988), heißt das noch nicht, dass sie verstanden haben, dass das letzte Zahlwort für alle Elemente der Menge steht, unabhängig davon, von wo (order irrelevance-principle) und was (Repräsentation) ausgezählt wird. Das Konzept kardinaler Einheiten, die durch bestimmte Zahlworte benannt werden, entwickelt sich erst auf dem Niveau III.

Das Kardinalzahlkonzept ermöglicht Additionen und Subtraktionen, bei denen eine Anzahl von Elementen als Teilmenge betrachtet wird. Das bedeutet, die Bearbeitung der Aufgaben erfordert nicht mehr das sukzessive Auszählen aller Elemente oder das Abzählen an den Fingern, jeweils bei eins beginnend, sondern erfolgt durch das Bilden einer Gesamtmenge aus zwei Teilmengen bzw. das Herstellen von zwei Teilmengen aus einer Gesamtmenge. Bei Additionsaufgaben zählen Kinder von der ersten Teilmenge aus weiter.

Die Zahlwortreihe wird als Sequenz größer werdender kardinaler Einheiten verstanden. Das bedeutet, Größenvergleiche zwischen zwei Zahlen finden nun nicht mehr auf der Ebene des Rangplatzes in der Zahlwortreihe statt (die 5 ist größer als die 4, da sie später in der Zahlwortreihe auftaucht), sondern auf der Ebene des Vergleichs der Mächtigkeit zweier Mengen (die 5 ist größer als die 4, da sie mehr Elemente enthält).

Niveau IV: Enthaltensein

Im nächsten Schritt erfolgt eine Differenzierung des Wissens über Verhältnisse zwischen Mengen. Auf der Basis des Prinzips der Klasseninklusion (Piaget, 1964) entwickelt sich die Einsicht, dass Zahlen andere Zahlen enthalten. Jede Zahl der Zahlenreihe enthält alle vorangegangenen Zahlen. Wird nun eine Teilmenge aus einer Zahl herausgelöst (die Zahl 7 aus der Zahl 12), kann sie als Teilmenge in Beziehung zu ihrer Gesamtmenge betrachtet werden.

Auf früheren Entwicklungsniveaus zerlegen Kinder Mengen bereits handelnd und fügen Teilmengen wieder zu einer Gesamtmenge zusammen. Zerlegen und Zusammenfügen sind hier noch sequentielle Prozesse, in deren Abfolge je neue Mengen hergestellt und für sich ausgezählt werden. Dass die Teilmengen Teile der Gesamtmenge sind, in der sie **enthalten** sind, wird aber eben erst auf dem Niveau IV verstanden. Die Kenntnis zweier Mengen reicht folglich aus, um die dritte zu bestimmen. Mit diesem Verständnis können Sachaufgaben zur Addition und Subtraktion gelöst werden, bei denen nach der Endmenge, einer Austausch- oder der Ausgangsmenge gefragt wird. Nach Riley et al. (1983) können ca. 50% der Erstklässler diese Aufgaben lösen.

Niveau V: Relationalität

Mit der Erkenntnis, dass Zahlworte für zusammengesetzte („united“, Fuson, 1992) Mächtigkeiten stehen und die Zahlwortreihe eine Sequenz aufeinander folgender Mächtigkeiten darstellt, ist noch nicht verstanden, wie genau sich die Mächtigkeiten der Mengen in der Zahlfolge unterscheiden. Kinder erkennen größer-als/kleiner-als-Beziehungen zwischen Mengen. Aber erst mit der Erweiterung dieses Wissens um ordinale Relationen und die Kenntnis der Mächtigkeit der Zahlen entwickeln Kinder ein Verständnis dafür, dass jedes nachfolgende Zahlwort sich von seinem Vorgänger um die Mächtigkeit + 1 unterscheidet.

Das bedeutet, dass die Intervalle zwischen aufeinander folgenden Zahlen gleich groß, nämlich eins sind und damit eine Art Maßstab zur Verfügung steht, um zwei Mengen exakt miteinander zu vergleichen. So werden Differenzen zwischen Mengen unterschiedlicher Größe bestimm- und vergleichbar (die Differenz zwischen 7 und 9 Elementen sowie 43 und 45 Elementen ist je 2). Zahlen stehen in diesem Sinne folglich auch für die Zählschritte zwischen zwei Zahlen. Damit kommt ihnen eine neue Bedeutung zu, nämlich die Abstände bzw. Relationen zwischen anderen Zahlen zu bezeichnen (Stern, 2003). Berechnungen von Aufgaben, die unabhängig vom Nullpunkt das Addieren oder Subtrahieren um eine bestimmte Zahl erfordern, werden nun lösbar. Dass der Erwerb des relationalen Zahlbegriffs deutlich schwieriger und für die meisten Kinder erst ab dem 2. Schuljahr zu erwarten ist, zeigen wiederum die Studien von Riley et al. (1983). Erstklässlern gelang die Bewältigung entsprechender Textaufgaben zur Bestimmung der Differenzmenge (Selina hat 8 Murmeln. Fritz hat 5 Murmeln. Wie viele Murmeln hat Selina mehr?) nur zu 33% bzw. 28% und zur Bestimmung der Referenzmenge (Selina hat 9 Murmeln. Sie hat 4 Murmeln mehr als Fritz.

Wie viele Murmeln hat Fritz?) gar nur zu 11% bzw. 22%.

Empirische Prüfung

Die Operationalisierungen dieser fünf Niveaus wurden in mehreren Studien entwickelt und geprüft. Mit ca. 3000 Kindern wurden verschiedene Itemversionen und -zusammensetzungen erprobt. Dabei wurden Korrekturen hinsichtlich der Präzision der Aufgabenstellung, der Bewertung und der Niveauzuordnung der Items erforderlich. Im Wesentlichen konnte das Modell in seinem Aufbau repliziert werden (Ricken, Fritz & Balzer, im Druck).

Für die Konstruktion des Tests ist die Prüfung von Annahmen über den Zusammenhang zwischen der durch den Test bestimmbaren Testleistung und der eigentlich interessierenden, aber nicht direkt beobachtbaren latenten Personenfähigkeit (hier: erworbene Konzepte) relevant. Die empirische Modellprüfung ist auf der Grundlage der Item-Response-Theory (IRT) und dem dichotomen Raschmodell möglich (z. B. Rasch, 1960; Rost, 2004).

Eine zentrale Voraussetzung für die Gültigkeit dieses Raschmodells ist die Raschhomogenität. Gilt diese, besteht ein Zusammenhang zwischen der Itemschwierigkeit und der Personenfähigkeit und beide sind auf derselben latenten Dimension gemeinsam darstellbar. Ist die Itemschwierigkeit niedriger als die Personenfähigkeit ist es wahrscheinlicher, dass das Item gelöst wird, und umgekehrt. Je höher die Differenz zwischen Itemschwierigkeit und Personenfähigkeit, desto wahrscheinlicher ist das Lösen bzw. Nicht-Lösen eines Items.

Zur Prüfung der Modellgültigkeit stehen diverse Prüfstatistiken zur Verfügung (Linacre, 2002). Mit der Infit-Statistik vergleicht man die tatsächlich beobachteten Lösungshäufigkeiten mit den auf Grundlage des angenommenen Modells berechneten Lösungswahrscheinlichkeiten. Diese Prüfung erfolgt mit dem Gesamtantwortmuster im Datensatz. Die Outfit-Statistik ist hingegen sensitiv gegen Ausreißerpersonen, die unerwartet richtig bei eher schweren Items (z. B. Raten) bzw. unerwartet falsch bei eher leichten Items (z. B. Leichtsinnsfehler) antworten. Beide Werte sind Maße für die Passung der Daten zum Modell und seinen Annahmen. Nach Linarce (2002) sind schlechte Outfit-Werte für das Modell weniger bedeutsam als auffällige Infit-Werte (vgl. auch Adams & Wu, 2002).

Im Weiteren sind Grenzwerte für die Prüfstatistiken festzulegen. Ein durchschnittlicher MNSQ (= standardisierter Fitwert) in der Nähe von 1 (Einheit: logit) entspricht einem guten Modellfit, also einer guten Passung zwischen Daten und Modell. Ein zu hoher MNSQ-Wert spricht für eine zu niedrige Trennschärfe, ein zu niedriger für redundante Items im Test. Als Grenzwerte für MNSQ definieren Wright & Stone (1979) den Bereich 1±0.5. Die strengste Vorgabe liefern Wright & Linacre (1994) mit 1±0.2; sie sprechen aber auch Werten zwischen 1±0.3 Modellgültigkeit zu.

Nachfolgend wird das Ergebnis der empirischen Modellprüfung unter Verwendung des Statistikprogramms WINSTEPS 3.66 auf der Basis der Normierungsdaten zu MARKO-D dargestellt. Der Normierungsdatensatz besteht aus Daten von 1095 Kindern im Alter zwischen 48 und 87 Monaten (M = 64.6; SD = 7.2). 567 Kinder (51.8%) sind Jungen und 528 (48.2%) sind Mädchen. Die Normdaten wurden von Februar 2009 bis Februar 2010 erhoben und über Kindergärten rekrutiert, da 91.2% der Kinder im Alter von 3 bis 6 Jahren in Kindereinrichtungen (Kindertagesstätten, Kitas) betreut werden (Statistisches Bundesamt, 2010, S. 232). Weitere Stichprobenmerkmale sind im Testmanual (Ricken, Fritz & Balzer, im Druck) dargestellt.

Die Modellprüfung ergab für alle 55 Items einen MNSQ Infit im Bereich von 1±0.3; für 53 Items gilt sogar das strengste

Kriterium 1±0.2. Als Schwellenwerte für die logit-Konvergenz-Einstellungen wurde sowohl für die Item- als auch für die Personenschätzung (RCONV und LCONV) das eher strenge .00001 festgelegt. Diese Schwellenwerte stellen die obere Grenze für logit-Veränderungen dar. Erst wenn diese unterschritten werden, ist eine hinreichende Konvergenz zum Modell erreicht und es benötigt keine weiteren Iterationen mehr. Das Modell konvergiert mit diesen Einstellungen in 150 Iterationen.

Die Anforderungen an ein eindimensionales Raschmodell können damit als erfüllt angesehen werden. Bringt man Personen und Items in die für das eindimensionale Raschmodell typische Darstellung, so resultiert die Abbildung 1.

Auf der gemeinsamen intervallskalierten Fähigkeits-/Schwierigkeitsskala mit einer Bandbreite von -5 bis +6 sind im „Person/Item MAP" jeweils links die Personen (das Zeichen '#' repräsentiert 6 Personen; das Zeichen '.' repräsentiert 1-5 Personen) und rechts die direkt benannten Items abgetragen. Je höher die Position einer Person, desto höher ihre Fähigkeit – und je höher die Position eines Items, desto schwieriger ist es. Je höher die Position einer Person im Vergleich zu einem Item, desto höher ist die Wahrscheinlichkeit, dass die Person das Item richtig löst. Liegen Person und Item auf gleicher Höhe, liegt die Wahrscheinlichkeit einer richtigen Lösung bei 50%. Die auf der Y-Achse angegebenen Intervalle werden in logit ausgegeben. M, S und T (links der Trennlinie für die Personen und rechts der Trennlinie für die Items) repräsentieren den Mittelwert sowie 1 bzw. 2 Standardabweichungen der entsprechenden Verteilung. Per Konvention wird der Itemmittelwert auf 0 logit fixiert.

Die „Person/Item MAP" zeigt, dass es mit Hilfe der eingesetzten Items gelingt, die Dimension der Konzepte in ihrer ganzen Breite abzudecken. Es gibt kein Kind ohne richtige Lösung und lediglich 1 Kind mit ausschließlich korrekten Lösungen. Somit existieren weder ein Decken- noch ein Bodeneffekt.

Die Grenzen zwischen den Niveaus werden gebildet, indem die Items theoriegeleitet dem jeweiligen Niveau, für das sie konstruiert worden waren, zugeordnet werden. In der Abbildung 1 ist der Wechsel von einem Niveau zum nächst höheren mit einer Trennlinie gekennzeichnet. Empirisch ist die Grenzziehung mit Bezug zu den Item-Schätzfehlern zu kontrollieren. Bei einem mittleren Item-Schätzfehler von 0.10 logit ergibt sich ein 95%-Konfidenzintervall von ± 0.2 logit um die jeweilige Itemschwierigkeit. Das bedeutet, dass Items, die mehr als 1 Zeile voneinander entfernt sind (1 Zeile entspricht 0.2 logit), zuverlässig voneinander trennbar sind. Damit gibt es lediglich zwischen Niveau 2 und 3 geringfügige Überlappungen der Konfidenzintervalle im Grenzbereich der Niveaus. Eine weitere Statistik des Raschmodells bestätigt diesen Befund. Der sogenannte Separation-Index gibt an, wie viele Niveaugrenzen im Modell empirisch belegbar sind. Mit einem Wert von 3.26 werden 3-4 Niveaugrenzen, also 4-5 verschiedene Niveaus nahe gelegt. Eine starke theoretische Untermauerung unterstützt einen solchen empirischen Befund. Tabelle 1 enthält die Itemkennwerte für die Normierungsstichprobe.

Dargestellt sind die Schwierigkeit der Items in logit und das resultierende Kompetenzniveau. Count gibt an, wie viele Kinder der Normierungsstichprobe ein Item bearbeitet haben. Score gibt an, wie viele Kinder das Item richtig gelöst haben. Infit ist das Maß dafür, wie gut ein Item dem eindimensionalen Raschmodell folgt. Mit dem Schätzfehler kann pro Item ein Konfidenzintervall bestimmt werden, dessen untere (KI unten) und obere (KI oben) Grenzen bei einer angenommenen Irrtumswahrscheinlichkeit von 95% jeweils angegeben sind.

Insgesamt können damit die Modellierung der unterschiedlichen Konzepte und die Abfolge in Form aufeinanderfolgender Ni-

```
                       PERSONS - MAP - ITEMS
                            <more>|<rare>
              6                 . +
                                  |
                                  |
                                  |
                                . |
              5                   +
                                  |
                                  |
                                . |
                                . |
              4                 . +T
                                  |
                                . |
                                . |     46
                                  |     26    33
Niveau 5      3                .# +     22
                                . T|    24    32
                               .# |
                            .#### |     20    34    35
                            .#### |
              2                ## +S    31
                             .### |     29    39
                          .###### |     38    44    45
Niveau 4                  .###### S|
                           .##### |     43    47
              1         .######## +     30    49
Niveau 3                  .###### |     18    55
                         ######## |     17    54     7
                        ######### |     12    14
                        .######## |     16    13    15    10
              0        .######### M+M    2     5    28    36    37
                       .######### |      4    11
                     .########### |     48
Niveau 2                 .####### |
                    .############ |      3
             -1          .####### +
                         .####### |
                        .######## |
                        .######## S|
                             #### |     53
             -2               .## +S     6
                             .### |     41    50
                              ### |      9    51
Niveau 1                       ## |     42    52
                              .## |      8
             -3                .# T+     1    27
                                . |     25    40
                                . |
                                . |     19    23
                                . |     21
             -4                   +T
                                  |
                                . |
                                  |
                                  |
             -5                   +
                            <less>|<frequ>
         EACH '#' IS 6.
```

Abb. 1: Rasch-Skala mit Niveaugrenzen, Personen und Items

Tab. 1: Itemkennwerte für die Normierungsstichprobe

Item-nummer	Schwie-rigkeit	Niveau	Count	Score	Infit	Schätz-fehler	KI unten	KI oben
46	3.34	5	555	42	1.05	0.18	2.99	3.69
26	3.27	5	555	44	1.02	0.17	2.94	3.60
33	3.18	5	555	47	1.16	0.17	2.85	3.51
22	3.07	5	220	16	0.87	0.28	2.52	3.62
32	2.87	5	555	59	1.22	0.15	2.58	3.16
24	2.85	5	221	19	0.84	0.27	2.32	3.38
34	2.45	5	555	79	1.07	0.14	2.18	2.72
20	2.39	5	555	82	1.04	0.14	2.12	2.66
35	2.32	5	555	86	1.16	0.13	2.07	2.57
31	2.08	5	1095	183	0.99	0.09	1.90	2.26
39	1.85	5	1095	212	0.93	0.09	1.67	2.03
29	1.85	5	1095	212	1.15	0.09	1.67	2.03
38	1.6	4	1095	246	0.9	0.08	1.44	1.76
45	1.58	4	1095	249	0.85	0.08	1.42	1.74
44	1.53	4	1095	256	0.82	0.08	1.37	1.69
47	1.21	4	1095	306	1.03	0.08	1.05	1.37
43	1.13	4	1095	318	0.88	0.08	0.97	1.29
49	0.96	3	1095	348	1.03	0.08	0.80	1.12
30	0.95	3	1095	349	1.2	0.08	0.79	1.11
55	0.83	3	1095	370	0.97	0.08	0.67	0.99
18	0.7	3	1095	393	1.04	0.07	0.56	0.84
54	0.68	3	1095	398	1.03	0.07	0.54	0.82
17	0.57	3	1095	417	0.9	0.07	0.43	0.71
7	0.56	3	1044	394	1.14	0.08	0.40	0.72
14	0.31	3	1095	466	0.83	0.07	0.17	0.45
12	0.31	3	1095	467	1.08	0.07	0.17	0.45
16	0.24	3	1095	480	1.15	0.07	0.10	0.38
13	0.19	3	1095	490	0.91	0.07	0.05	0.33
15	0.18	3	1095	491	0.95	0.07	0.04	0.32
10	0.14	2	1095	500	1.03	0.07	0.00	0.28
5	0.1	2	1095	508	0.88	0.07	-0.04	0.24
37	0.07	2	1095	514	0.87	0.07	-0.07	0.21

Tab. 1: Fortsetzung

Item-nummer	Schwie-rigkeit	Niveau	Count	Score	Infit	Schätz-fehler	KI unten	KI oben
36	0.07	2	1095	514	0.89	0.07	-0.07	0.21
28	0.02	2	1095	524	1.25	0.07	-0.12	0.16
2	-0.04	2	1095	534	0.91	0.07	-0.18	0.10
11	-0.12	2	1094	549	1.13	0.07	-0.26	0.02
4	-0.19	2	1095	565	0.91	0.07	-0.33	-0.05
48	-0.32	2	1095	589	0.96	0.07	-0.46	-0.18
3	-0.73	2	1095	669	0.97	0.07	-0.87	-0.59
53	-1.7	1	1095	837	0.95	0.08	-1.86	-1.54
6	-2.08	1	1095	890	1.16	0.09	-2.26	-1.90
50	-2.22	1	1095	908	1.01	0.09	-2.40	-2.04
41	-2.23	1	1095	909	0.89	0.09	-2.41	-2.05
9	-2.31	1	1050	884	1.02	0.09	-2.49	-2.13
51	-2.41	1	1095	930	1.05	0.09	-2.59	-2.23
52	-2.54	1	1095	945	0.9	0.1	-2.74	-2.34
42	-2.63	1	1095	954	0.85	0.1	-2.83	-2.43
8	-2.72	1	1065	936	1	0.1	-2.92	-2.52
1	-2.93	1	1095	982	0.99	0.11	-3.15	-2.71
27	-3.03	1	1095	990	1.01	0.11	-3.25	-2.81
40	-3.23	1	1095	1005	0.87	0.12	-3.47	-2.99
25	-3.26	1	1095	1007	0.89	0.12	-3.50	-3.02
23	-3.51	1	1095	1023	1.05	0.13	-3.76	-3.26
19	-3.51	1	1095	1023	0.86	0.13	-3.76	-3.26
21	-3.73	1	1095	1035	0.97	0.14	-4.00	-3.46

veaus aus theoretischer und empirischer Perspektive als erfolgreich bezeichnet werden.

Im nächsten Abschnitt wird sodann der Aufbau des Tests beschrieben.

Testaufbau

Der Test besteht aus 55 Items, die Kindergartenkindern zwischen 4;0 und 6;5 Jahren präsentiert werden. Für einen unkomplizierten Zugang zu den Kindern und in die Testsituation wurden die Items in eine Geschichte eingebettet. Zwei Eichhörnchen – Ben und Lisa – zählen, vergleichen und bestimmen Men-

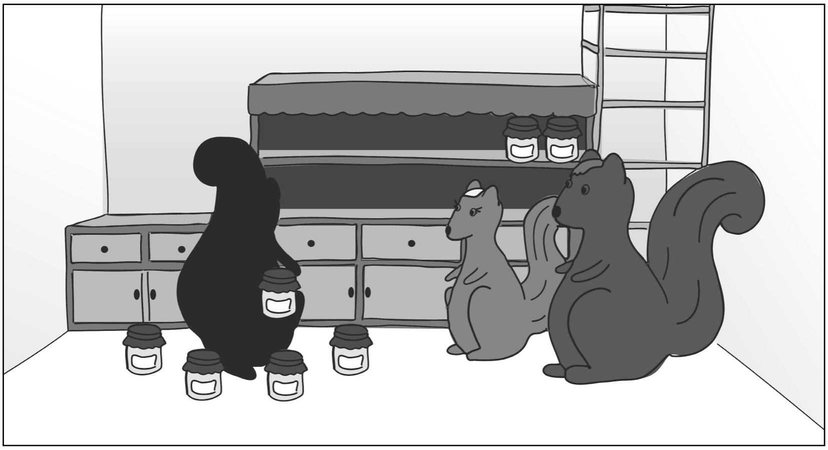


Abb. 2: Beispiel für dekorative Bilder-Instruktion: "Ben und Lisa sind damit beschäftigt, die Vorräte für den Winter zu ordnen. Sie wollen Zahlen an die Gläser schreiben. Lisa macht kleine Knobelaufgaben draus."

gen und Zahlen. Die Rahmenhandlung wird durch dekorative Bilder illustriert (Abb. 2). Das Arbeitsgedächtnis wird durch die Geschichte und die Bilder nicht zusätzlich belastet (Ehlert et al., submitted). Die Kinder benötigen insgesamt etwa 20 bis 30 Minuten, um alle Aufgaben zu bearbeiten.

Die Bilder und die Geschichte dienen zur Vorbereitung der Items. Einige Items werden mit instruktiven Bildern und mit Plättchen präsentiert (Abb. 3). Die Inhalte wurden auf wesentliche Informationen reduziert.

Tabelle 2 enthält Aufgabenbeispiele für die Niveaus und die Angabe der Anzahl der Items, die das jeweilige Niveau repräsentieren.

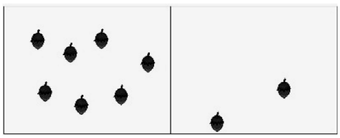
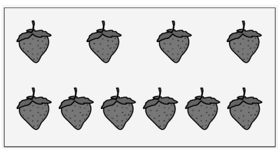
Die Itemreihenfolge entspricht nicht den Niveaus. Vielmehr wechseln sich Items mit unterschiedlichem Schwierigkeitsgrad ab. Dieser Aufbau wurde gewählt, um zu erreichen, dass alle Kinder alle Aufgaben bearbeiten, ohne dass es zu Überforderungssituationen oder Testabbrüchen kommt, wodurch sich die Zuverlässigkeit für die Schätzung des individuellen Entwicklungsniveaus verschlechtern würde.

Auswertungsaspekte

Neben der Einordnung in die Normstichprobe ist in MARKO-D vor allem die *inhaltliche Beschreibung* des erworbenen Entwicklungsstandes wesentlich. Die Zuordnung Rohwertsumme zum Entwicklungsniveau erfolgt auf Basis der Informationen, die in Abbildung 1 abgebildet sind. Da Rohwertsummen und Personenfähigkeiten direkt miteinander korrespondieren, die Personenfähigkeiten im Raschmodell wiederum auf einer gemeinsamen Dimension mit den Itemschwierigkeiten liegen und die Entwicklungsniveaus bezüg-

Abb. 3: Instruktion: "Hier siehst Du 4 Sterne (auf Sterne zeigen), und unter der Wolke (zeigen) sind noch 3 Sterne. Wie viele Sterne sind das zusammen?"

Tab. 2: Beispiele für Operationalisierungen von Items pro Niveau

Niveau	Konzepte	Itemanzahl pro Niveau	Itembeispiele
Niveau I	Zählen	16	Gib mir 5 Plättchen.
			Kannst Du die Beeren für die Eichhörnchen so aufteilen, dass jeder gleich viele hat? Lege die Plättchen unter Ben und Lisa.
Niveau II	Mentaler Zahlenstrahl	10	Wie heißt die Zahl, die zwischen der 5 und der 7 kommt?
			Sein Bruder hat heute Morgen 2 Nüsse gefunden und dann hat ihm der Biber noch 2 geschenkt. Wie viele hat er jetzt? Kannst du mir die Aufgabe mit diesen Plättchen legen?
Niveau III	Kardinalität und Zerlegbarkeit	12	Hier (links) hat Ben hingemalt, wie viele Nüsse er hat. Lege bitte hier (rechtes Kästchen) genau so viele Punkte hin, dass es genau so viele sind wie hier (auf das linke Kästchen zeigen). 
Niveau IV	Enthaltensein	5	Jetzt möchte der Biber 6 Blumen haben - mehr blaue als rote. Gib mir bitte 6 Plättchen, davon sollen mehr blau als rot sein.
Niveau V	Rationaler Zahlbegriff	12	Und jetzt bring mir 8 Blumen. Es sollen 2 mehr blaue als rote sein.
			Ben und Lisa haben Erdbeeren entdeckt. (In welcher Reihe sind weniger?) Und wie viele sind es weniger? 

lich ihrer Lage zu den Items und ihren Schwierigkeiten eindeutig bestimmt sind, kann man diese Verbindung direkt herstellen.

Eine punktgenaue Einschätzung der mathematischen Leistungsfähigkeit auf statistischer Basis ist aufgrund von Fehlereinflüssen nicht möglich. Aus dem Raschmodell resultieren im Unterschied zu Messfehlern in der klassischen Testtheorie Schätzfehler für individuelle Fähigkeitswerte.

Im ersten Schritt wird auf diese Weise das Entwicklungsniveau aufgrund des Gesamtwertes (Rohwertsumme) bestimmt. Dies ist immer dann eine zuverlässige Variante, wenn die Items zu niedrigeren Niveaustufen komplett gelöst wurden. Schwieriger wird die Interpretation, wenn nur einige Items zu niedrigeren, dafür aber einige Items zu höheren Niveaustufen korrekt bearbeitet wurden. Dies kommt aufgrund der Gültigkeit des eindimensionalen Raschmodells nur selten vor, muss aber für eine individuelle Interpretation in Betracht gezogen werden. Damit erfordert eine Niveauzuordnung eine differenzierte Betrachtung der gelösten Items pro Niveau und eine Analyse der Lösungsmuster.

Für die Daten der Normierungsstichprobe wurde die Analyse mit folgenden Ergebnissen durchgeführt. Eine Zuordnung zu einem Niveau schließt nicht gelöste Items ein, die durch andere Faktoren im Testverlauf als durch die Fähigkeiten der Kinder entstanden sind. Die Festlegung einer Anzahl nicht gelöster Items ist aus inhaltlicher Perspektive nicht zu leisten, deshalb wurde der empirische Weg gewählt und auf der Basis der Normstichprobe das Kriterium gesucht, das die besten Verhältnisse zwischen interpretierbaren und nicht interpretierbaren Lösungsmustern erzeugt. Dies ist der Fall für das Kriterium: Das Niveau gilt als bewältigt, wenn 75% der Items richtig gelöst werden.

Nach dieser Regel können von den 1095 Kindern der Normierungsstichprobe 993 eindeutig einem Niveau zugeordnet werden, was einer Quote von 90.7% entspricht. Im Detail befinden sich 75 dieser Kinder auf Niveau 5 (haben also Niveau 4 gemeistert), 50 Kinder auf Niveau 4, 114 Kinder auf Niveau 3, 592 Kinder auf Niveau 2 und 162 Kinder auf Niveau 1.

Interessant ist dann, wie viele Items eines Niveaus von Kindern auf unterschiedlichen Niveaus im Durchschnitt gelöst werden. So haben Kinder auf Niveau 4 durchschnittlich 98% aller Items auf Niveau 1, 90% auf Niveau 2 und 84% auf Niveau 3 gelöst. Auf ihrem Niveau 4 lösen sie durchschnittlich 38% aller Items, auf Niveau 5 nur noch 29%. Diese Häufigkeiten belegen die Annahme, dass jeweils Items unter dem aktuellen Niveau gelöst und in den drüber liegenden Niveaus eher nicht gelöst werden (Abb.4).

In der Normierungsstichprobe stimmt die Auswertung der Rohwertsumme und der Lösungsmuster für 717 Kinder überein. Das be-

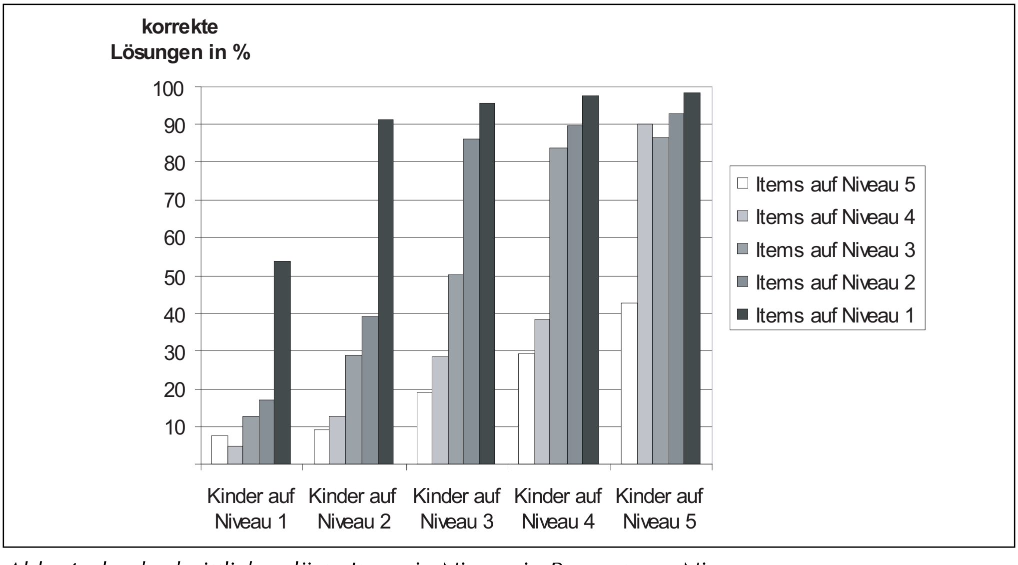


Abb. 4: durchschnittlich gelöste Items je Niveau in Prozent pro Niveau

deutet, dass beide Auswertungen zur gleichen Niveauzuordnung führen.

Für 102 Kinder ist die Niveauzuordnung über die 75%-Regel nicht möglich, womit auch keine Diskrepanz prüfbar ist. Bei 276 Kindern führen beide Methoden der Niveauzuordnung zu Diskrepanzen in der Größenordnung von einem Niveau. Nur für 8 Kinder beträgt der Unterschied mehr als ein Niveau.

Die Interpretation wird in Fällen der Diskrepanz in der folgenden Weise vorgeschlagen:

Fall A: Gesamtwert fällt besser als das Lösungsmuster aus

Kind A erreicht insgesamt 30 Punkte, nach der Rohwertsummen-Auswertung entspricht dies dem Niveau III. Eine Betrachtung des Lösungsmusters zeigt, dass sich die Punkte folgendermaßen verteilen: Auf Niveau I werden 100% der Punkte, auf Niveau II nur 57% und auf Niveau III 75% erreicht. Damit wird das 75%-Kriterium von Niveau II unterschritten. Entgegen der Erwartung löst das Kind allerdings so viele Aufgaben von Niveau III, dass dieses als bewältigt gelten kann (Kriterium erreicht). Eine Zuordnung der Leistung des Kindes zu Niveau III kann jedoch aufgrund der geringen Anzahl richtiger Lösungen von Niveau II nicht sicher erfolgen. Da das Kind aber bereits viele Aufgaben, die ein kardinales Verständnis erfordern, bewältigt, könnte eine Zuordnung zu Niveau II eine Unterschätzung seines Entwicklungsstandes darstellen. Insgesamt ist die Aussage hier durch weitere Beobachtungen zu stützen, da eine eindeutige Interpretation nicht möglich ist.

Fall B: Gesamtwert fällt schlechter aus als das Lösungsmuster

Der Gesamtwert des **Kindes B** liegt mit 28 Punkten auf dem Niveau III. Dieses Kind löst alle 12 Aufgaben aus dem Niveau I, 75% der Aufgaben aus Niveau II und 78% der Aufgaben auf Niveau III. Damit ist das 75%-Kriterium auf allen drei Niveaus erfüllt und es ist davon auszugehen, dass das Kind die Konzepte dieser Entwicklungsniveaus beherrscht und aktuell das Konzept des nächsten Niveaus, das Teile-Ganze-Konzept entwickelt (Niveau IV). In diesem Fall würde eine Zuordnung der Leistung zu Niveau III möglicherweise eine Unterschätzung darstellen. Auch hier müssen weitere Beobachtungen hinzugezogen werden, um zu verifizieren, ob das Kardinalzahlkonzept eventuell nicht schon vorhanden ist.

Fall C & D: Werte in den Randbereichen

Für besonders schwache und besonders starke Leistungen sind folgende Interpretationen vorzunehmen:

Kind C erreicht insgesamt 11 Punkte und wird mit diesem Gesamtwert dem Niveau I zugeordnet. Die Analyse des Lösungsmusters zeigt, dass das Kind 8 Punkte im Niveau I und 3 Punkte im Niveau II erreicht. Für beide Niveaus wird das 75%-Kriterium unterschritten. Hier stimmen die Angaben überein, das Kind befindet sich auf Niveau I, d.h. es ist dabei zu verstehen, dass Zahlen Objekten zugeordnet und mit Hilfe von Zahlen Mengen aus- und abgezählt werden können (Zählzahlkonzept).

Kind D: Mit einer Gesamtpunktzahl von mehr als 40 Punkten wird dem Rohsummenwert zufolge die Leistung dem Niveau V zugeordnet. Zeigt das Lösungsmuster, dass das 75%-Kriterium auf allen Niveaus erfüllt ist, so kann davon ausgegangen werden, dass das Kind auch alle Konzepte bereits verstanden hat. Für dieses Kind tritt somit ein Deckeneffekt ein, da die Zone seiner aktuellen Entwicklung mit dem vorhandenen Modell nicht mehr korrekt angegeben werden kann.

Für weitere Validierungsmerkmale des Verfahrens wird auf das Manual verwiesen, da sie für den hier vorgestellten Konstruktionsansatz nur am Rand eine Rolle spielen.

Fazit

Die Annahmen zur Ordnung der Konzepte bewähren sich in der empirischen Prüfung. Dies zeigt sich insgesamt in einer guten Interpretierbarkeit der Entwicklungsniveaus, die nur für wenige Kinder zu nicht eindeutigen Ergebnissen führt. Zwischenbefunde, die in aktuellen Studien validiert werden, legen die Annahme nahe, dass sich individuelle Entwicklungsprozesse als Verschiebung auf der Skala darstellen lassen und für rechenschwache Kinder ein längeres Verweilen auf unteren Niveaus anzunehmen ist. Diese Daten sind durch umfangreichere Längsschnittdaten zu vervollständigen. Des Weiteren scheinen gezielte Fördermaßnahmen zu gezielten Veränderungen von Konzepten zu führen (MARKO-T: Gerlach, Fritz & Leutner, im Druck).

Literatur

Adams, R.J. & Wu, R. (Eds.) (2002). PISA 2000 technical report. Paris: OECD.

Butterworth, B. (2005). The development of arithmetical abilities. Journal of Child Psychology and Psychiatry, 46 (1), 3-18.

Ehlert, A., Fritz, A., Ricken, G. & Leutner, D. (submitted). Besser rechnen mit Bildern? Zum Einfluss dekorativer Bilder auf die Rechenleistung.

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

Fuson, K.C. (1992). Research on learning and teaching addition and subtraction of whole numbers. In G. Leinhardt, R. Putnam & R.A. Hattrup (Eds.), Analysis of arithmetic for mathematics teaching (pp. 53-187). Hillsdale, NJ: Erlbaum.

Gerlach, M., Fritz, A. & Leutner, D. (i. Dr.). MARKO-T. Mathematik und Rechnen – Training zur Förderung von Konzepten im Vorschul- und frühen Grundschulalter. Göttingen: Hogrefe.

Gölitz, D., Roick, T. & Hasselhorn, M. (2006). DEMAT 4 Deutscher Mathematiktest für vierte Klassen. Göttingen: Hogrefe.

Haffner, J., Baro, K., Parzer, P. & Resch, F. (2005). Heidelberger Rechentest – HRT 1–4. Göttingen: Hogrefe.

Kaufmann, L. & Nuerk, H.C. (2007). Zahlverarbeitung: typische und atypische Entwicklungsverläufe. In L. Kaufmann, L., H.C. Nuerk, K. Konrad & K. Willmes (Hrsg.), Kognitive Entwicklungsneuropsychologie (S. 383-398). Göttingen: Hogrefe.

Landerl, K. & Kaufmann, L. (2008). Dyskalkulie: Modelle, Diagnostik, Intervention. München: Reinhardt.

Le Corre, M., van de Walle, G., Brannon, E.M. & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. Cognitive Psychology, 52, 130-169.

Linacre, J.M. (2002). What do Infit and Outfit, Mean-square and Standardized mean? Rasch Measurement Transactions, 16 (2), 878.

Piaget, J. (1964). Rechenunterricht und Zahlbegriff. Die Entwicklung des kindlichen Zahlbegriffes und ihre Bedeutung für den Rechenunterricht. Braunschweig: Westermann.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Nielsen & Lydiche.

Resnick, L B. (1983). A developmental theory of number understanding. In H. Ginsburg (Ed.), The development of mathematical thinking (pp. 109-151). New York, NY: Academic Press.

Ricken, G. & Fritz, A. (2009). Diagnostik mathematischer Kompetenzen, In A. Fritz, G. Ricken & S. Schmidt (Hrsg.), Handbuch Rechenschwäche. Lernwege, Schwierigkeiten und Hilfen, 2. überarbeitetet Aufl. (S. 292-321). Weinheim: Beltz.

Ricken, G., Fritz, A. & Balzer, L. (im Druck). MARKO-D – Mathematik und Rechnen – Test zu Erfassung von Konzepten im Vorschulalter. Göttingen: Hogrefe.

Riley, M.S., Greeno, J.G. & Heller, J.H. (1983). Development of children´s problem-solving ability in arithmetic. In: H.P. Ginsburg (Ed.), The development of mathematical thinking (pp. 153-196). New York: Academic Press.

Rost, J. (2004). Lehrbuch Testtheorie – Testkonstruktion (2. Auflage). Göttingen: Hans Huber.

Schaupp, H., Holzer, N. & Lenart, F. (2007). Eggenberger Rechentest 1+. Göttingen: Hogrefe.

Statistisches Bundesamt (2009). 30% der betreuten Kinder von 3 bis 5 Jahren in Ganztagsbetreuung. Pressemitteilung Nr.010 vom 09.01.2009.

Stern, E. (2003). Früh übt sich – Neuere Ergebnisse aus der LOGIK-Studie zum Lösen mathematischer Textaufgaben. In A. Fritz, G. Ricken & S. Schmidt (Hrsg.), Handbuch Rechenschwäche (S. 116-130). Weinheim: Beltz.

Wright, B.D. & Linacre, J.M. (1994). Reasonable mean-square fit values. Rasch Measurement Transactions, 8 (3), 370.

Wright, B.D. & Stone, M.H. (1979). Best Test Design. Chicago: MESA Press.

Wynn, K. (1990). Children´s understanding of counting. Cognition, 36, 155-193.

Anschriften der Autoren

Prof. Dr. Gabi Ricken
Fakultät EPB der Universität Hamburg
Sedanstr. 19
20146 Hamburg
gabriele.ricken@uni-hamburg.de

Prof. Dr. Annemarie Fritz
FB Bildungswissenschaften der
Universität Duisburg-Essen
Campus Essen
Universitätsstr. 1
45117 Essen
fritz-stratmann@uni-due.de

Dr. Lars Balzer
Eidgenössisches Hochschulinstitut für
Berufsbildung (EHB)
Kirchlindachstr. 79
CH-3052 Zollikofen
Schweiz
evaluation@lars-balzer.info