

Tests & Trends Bd. 9 (2010): Frühprognose schulischer Kompetenzen

Marcus Hasselhorn und Wolfgang Schneider (Hrsg.)

Kapitel 6

Fähigkeitsindikatoren Primarschule (FIPS) - Erprobung und Weiterentwicklung einer internationalen computerbasierten Schulanfangsdiagnose

Kerstin Bäuerlein, Andrea Beinicke, Gabriele Faust, Ute Franz, Melanie Jost und Wolfgang Schneider

Zusammenfassung

Die computerbasierte Schulanfangsdiagnose *Fähigkeitsindikatoren Primarschule* (FIPS) basiert auf dem weltweit seit fast 20 Jahren verwendeten und vom *Centre for Evaluation and Monitoring* (CEM) in Durham (UK) entwickelten Programm *Performance Indicators in Primary Schools* (PIPS; z.B. Tymms & Albone, 2002). Dadurch können am Anfang und am Ende der ersten Klasse die Voraussetzungen und Entwicklungen von Erstklässlern in den Bereichen Schriftsprache, Mathematik und Lautbewusstheit sowie in ausgewählten Persönlichkeitsaspekten ökonomisch und alltagstauglich erhoben werden. Im Rahmen einer interdisziplinären pädagogisch-psychologischen Zusammenarbeit (Studie *LEIF, Lernausgangslagen erheben – Individuell fördern*, 2007-2010) wird das englische Originalinstrument für den deutschen Anfangsunterricht evaluiert, überarbeitet und normiert. Gleichzeitig wird die Nutzung diagnostischer Informationen durch die Lehrkräfte zur Differenzierung und individuellen Förderung untersucht. Der Beitrag hat zwei Ziele: (1) Die Entwicklung des Diagnoseverfahrens FIPS und sein derzeitiger Stand werden näher dargestellt. (2) Erste Ergebnisse der LEIF-Studie werden berichtet.

1. Einführung

Von Lehrkräften wird erwartet, dass sie über den Lernstand und die -entwicklungen ihrer Schülerinnen und Schüler im Bilde sind. Dies ist nicht nur Voraussetzung für zutreffende Beurteilungen, sondern auch und noch mehr für die *Steuerung des täglichen Unterrichts*, z.B. die Justierung des Schwierigkeitsniveaus, die Auswahl von Fragestellungen und Aufgaben, für Rückmeldungen und spezifische Hilfen und vor allem für Differenzierungs-

und Individualisierungsmaßnahmen. Die diagnostische Aufgabe bezieht sich sowohl auf Aufgaben als auch Personen, fachliche und überfachliche Fähigkeiten, die ganze Klasse und jedes einzelne Individuum, den gegenwärtigen Stand und die erwartete Entwicklung. Personenbezogen sind sowohl das durchschnittliche Niveau der Klasse als auch die Leistungsbreite und die Rangfolge der einzelnen Schülerinnen und Schüler zu ermitteln (Helmke, 2009, S. 84ff.). Eine treffsichere Einschätzung ist besonders für einerseits leistungsstarke, andererseits leistungsschwache Schülerinnen und Schüler wichtig. Standardisierte Lerndiagnosen am Schuljahresende ermöglichen es darüber hinaus, den Lernstand der eigenen Klasse differenziert mit den zu erwartenden Ergebnissen zu vergleichen und zu bewerten.

Eine hohe Diagnosekompetenz steht dabei im Dienst der *Adaptivität des Unterrichts*, insbesondere im Zusammenhang mit Strukturierungs- und Differenzierungsmaßnahmen. Dadurch soll "Passung" zwischen den Lernvoraussetzungen der Schüler und dem Unterrichtsangebot erreicht werden, was Voraussetzung für Unterrichtserfolg ist. Helmke (2009, S. 94) spricht in diesem Sinn von der „Katalysatorvariable“ Diagnosekompetenz, Schrader (2006, S. 98) davon, dass diagnostische Kompetenz die Wirksamkeit des unterrichtlichen Handelns moderiert.

Der Befund, dass im Rahmen der PISA 2000-Studie in den Hauptschulklassen nur etwa 10 % der schwachen Leser von ihren Lehrerinnen und Lehrern erkannt wurden (Artelt, Stanat, Schneider & Schiefele, 2001, S. 119f.), verstärkte das Interesse für die *diagnostische Kompetenz von Lehrerinnen und Lehrern*. Zu berücksichtigen ist zwar, dass diese Beurteilungsaufgabe besonders schwierig war (z.B. war die Definition von „schwacher Leser“ nicht eindeutig) und von Lehrkräften keine exakten, sondern nur hinreichend genaue Einschätzungen erwartet werden. Aber trotz substanzieller Korrelationen zwischen Lehrerurteil und Testleistungen von $r = .62$ bzw. $r = .69$ kommen in zahlreichen Studien erhebliche Leistungsüber- und -unterschätzungen vor und die Lehrkräfte unterscheiden sich in ihrer diagnostischen Kompetenz interindividuell beträchtlich (Schrader, 2006, S. 97). Die Korrelationen zwischen Persönlichkeitsmerkmalen der Schüler und den darauf bezogenen diagnostischen Lehrerurteilen liegen sehr viel niedriger (vgl. z.B. Spinath, 2005). Üblicherweise sind Studien zur Diagnosekompetenz von Lehrkräften unterrichtsfern angelegt. Die Lehrkräfte geben z.B. ein Urteil ab, welche Aufgaben von ihren Schülerinnen und Schülern bei einer zeitgleichen Bearbeitung gelöst werden und welche nicht bzw. wie

schwer die Aufgaben sind. Die Auswirkungen der Diagnosekompetenz auf den Unterricht und - vermittelt über adaptive Lernangebote - auf den Lernfortschritt der Schülerinnen und Schüler sind schwieriger zu untersuchen, weil dazu sowohl die Leistungsdiagnosen als auch die Unterrichtshandlungen der Lehrkräfte zu erfassen und ein Zusammenhang zwischen beiden herzustellen ist.

Besonders wichtig erscheinen zutreffende *Diagnosen am Schulanfang*. Zum einen unterscheiden sich die Kinder gerade zu diesem Zeitpunkt sehr stark in ihren Lernvoraussetzungen, z.B. im Schriftspracherwerb und Mathematik (vgl. z.B. Richter & Brügelmann, 1992; Martschinke & Kammermeyer, 2003; Füssenich & Löffler, 2005; Moser, Stamm & Hollweber, 2005; Grassmann, 2001; Schneider, Krajewski & Schwenck, 2010). Für mathematische Vorkenntnisse ist nachgewiesen, dass die Lehrkräfte zur *starken Unterschätzung* tendieren und den Schulanfängern zu wenig zutrauen (Hasemann, 2003, S. 29). Zum anderen sollten die Lehrkräfte vor allem in diesen beiden Lernbereichen möglichst rasch alle Kinder ihrer Klasse kennen lernen, um von Anfang an alle gezielt zu fordern und zu fördern und den schwachen Schulanfängern passende Hilfen zu geben. Zeit zum allmählichen Ermitteln der Lernvoraussetzungen ist am Schulanfang nicht vorhanden. Ein alltagstaugliches wissenschaftliches Diagnoseinstrument könnte den Lehrkräften gerade zu diesem Zeitpunkt die Diagnoseaufgabe erleichtern.

Im *mathematischen Bereich* könnten sie unter mehreren formellen Tests wählen, die die Lernvoraussetzungen breit erheben (vgl. z.B. Osnabrücker Test zur Zahlbegriffsentwicklung, van Luit, van de Rijt & Hasemann, 2001; TEDI-MATH, Kaufmann, Nuerk, Graf, Delazer & Willmes, 2009, vgl. auch die Übersicht bei Ennemoser und Krajewski, in diesem Band). Im *Schriftspracherwerb* dominieren demgegenüber informelle Verfahren mit geringer Durchführungs- und Auswertungsobjektivität. So sind z.B. am Schulanfang nur wenige formelle Tests zur Diagnose der phonologischen Bewusstheit einsetzbar (Barth & Gomm, 2004; Martschinke, Kirschhock & Frank, 2001). Zur Diagnose der ebenso relevanten Buchstabenkenntnis (Bowey, 2007) und des metasprachlichen Wissens über die Schrift liegen zahlreiche vorstrukturierte Beobachtungsaufgaben vor, mit denen der subjektive Blick der Kinder auf die Sprache und ihr schon vorhandenes Können vornehmlich in Einzeluntersuchungen rekonstruiert werden sollen: Beispiele sind das "Gezinkte Memory", "Embleme Lesen", das "Leere Blatt", "Notizen machen" und "Buchstaben-Monster" (vgl. Füssenich & Löffler, 2005, S. 34ff; Hanke, 2007, S. 89ff., die meisten Aufgaben stammen

von H. Brügelmann und E. Brinkmann bzw. M. Dehn, die durch Hinweise die Auswertung ihrer Beobachtungshilfen zu objektivieren versucht). Diese Aufgaben sind einfallsreich, haben Spielcharakter und lassen sich in den Unterricht einbauen. Aber die Erhebung ist aufwendig, insbesondere wenn die benötigten Materialien erst herzustellen sind. Die Durchführungs- und Auswertungsobjektivität sind ebenso wie Reliabilität und Validität eingeschränkt und damit ist die Qualität der gewonnenen Information begrenzt. Zudem erheben diese Aufgaben nur jeweils Teile der Lernvoraussetzungen für den Schriftspracherwerb, wobei auf einen für das spätere Leseverständnis prädiktiven Wortschatztest im Allgemeinen sogar verzichtet wird. Was fehlt, ist ein die beiden wichtigsten Lernbereiche des Anfangsunterrichts umfassendes, wissenschaftliche Qualitätsansprüche erfüllendes, differenziertes und trotzdem ökonomisches und von den Kindern gern durchgeführtes normorientiertes Diagnoseverfahren.

Weltweit besteht großes Interesse an der Entwicklung *systematischer Diagnoseinstrumente*, die sowohl für die Erfassung von Lernstand und Lernentwicklung der Schülerinnen und Schüler als auch für die Evaluation und Verbesserung des Unterrichts geeignet sind (Tymms & Wyld, 2004). Im angelsächsischen Sprachraum wird seit fast zwei Jahrzehnten das Diagnoseinstrument *Performance Indicators in Primary Schools* (PIPS) vom Schuleintritt in die Reception Class bis zum Ende der Primarstufe eingesetzt. Auf einer bereits vorliegenden deutschsprachigen Computerversion für den Schulanfang aufbauend wird dieses Verfahren in Bayern seit 2007 interdisziplinär erprobt, an die deutsche Situation angepasst und im Schulforschungsprojekt *LEIF – Lernausgangslagen erheben, individuell fördern* auf seine Wirkungen auf die Unterrichtsgestaltung und den Lernerfolg von Erstklässlerinnen und Erstklässlern untersucht. Beteiligt sind Psychologen (Lehrstuhl für Psychologie IV der Universität Würzburg, Leitung Prof. Dr. W. Schneider, der auch das Gesamtprojekt leitet und für die Auswertung der Schülerdaten verantwortlich war), Grundschulpädagogen (Lehrstuhl für Grundschulpädagogik und -didaktik der Universität Bamberg, Prof. Dr. G. Faust in Zusammenarbeit mit Dr. U. Franz, die auch die Instrumente für die Lehrerinnen und Eltern entwickelten und die Lehrerdaten auswerten) sowie seit 2008 auch Schulpädagogen (Lehrstuhl für Schulpädagogik, Universität Augsburg, Prof. Dr. W. Wiater) und empirische Bildungsforscher (Lehrstuhl für Allgemeine Pädagogik, Erziehungs- und Sozialisationsforschung an der LMU München, Prof. Dr. H. Ditton, in Zusammenarbeit mit der Professorin für Grundschulpädagogik und -didaktik Dr. A. Speck-Hamdan). Im

Folgenden werden sowohl das Ausgangsverfahren PIPS als auch die deutschsprachige Adaption *FIPS (Fähigkeitsindikatoren Primarschule)* und die Anlage und erste Ergebnisse der LEIF-Studie dargestellt.¹

2. Performance Indicators in Primary Schools (PIPS)

Ausgangspunkt war das 1992 in Durham (UK) im *Centre for Evaluation and Monitoring* (CEM) unter der Leitung von Prof. Dr. Peter Tymms gestartete Projekt *Performance Indicators in Primary Schools* (PIPS). In ca. 20-minütigen Einzeltests am PC oder als Paper-Pencil-Test erheben die Lehrkräfte sowohl die Leistungen von Schülerinnen und Schülern als auch Aspekte ihrer Persönlichkeitsentwicklung über die gesamte Grundschulzeit hinweg (in England umfasst diese Zeit das Alter von etwa vier bis elf Jahren). Die Rohdaten werden an das CEM übermittelt, dort jeweils schuljahresaktuell zu standardisierten Daten verarbeitet und in Form von benutzerfreundlichen Übersichtstafeln und Grafiken an die Lehrkräfte zurückgemeldet. Ziel ist es, den Schulen und Lehrkräften qualitativ hochwertige, objektive Informationen über den Stand und die Entwicklung ihrer Schülerinnen und Schüler im nationalen Vergleich zur Verfügung zu stellen. Die kostengünstige zentrale Auswertung der Schülerdaten im CEM erlaubt die Ermittlung aktueller Normwerte für jede Klassenstufe, stellt ein wissenschaftsbasiertes unabhängiges Bildungsmonitoring für den englischen Primarbereich dar und erlaubt die Untersuchung zahlreicher aktueller Fragen der Bildungsforschung.

Mit der computerbasierten Diagnose werden *breit und differenziert* alle Bereiche erfasst, die sich in der entwicklungspsychologischen und grundschuldidaktischen Forschung als *Prädiktoren der Lernentwicklung in der Grundschule* erwiesen haben. Am Schulanfang, d.h. in England beim Eintritt in die „Reception Class“, erfasst das PIPS die Fähigkeit, den eigenen Namen zu schreiben, und enthält Aufgaben zu den Bereichen Wortschatz, Vorstellungen vom Lesen, phonologische Bewusstheit, Buchstabenkenntnis,

¹ Das Staatsministerium für Kultus und Unterricht hat den Modellversuch von Anfang an unterstützt. Frau Ltd. Ministerialrätin Elfriede Ohrnberger und Frau Ministerialrätin Anne Blank sei für ihr Engagement herzlich gedankt. Besonderer Dank geht auch an die Schulkinder, Lehrkräfte, Eltern und externen Testleiter, die an den bisherigen Untersuchungen mitgewirkt haben sowie an die zuständigen Schulverwaltungsbehörden, die die Arbeiten engagiert unterstützten. In Würzburg übernahmen Frau Andrea Beinicke und Frau Christiane Horstmeier die Datenauswertung, in Bamberg steuerte Frau Dipl. Päd. Carmen Archie die Erhebungen ab Schuljahresbeginn 2008/09 und betreute dabei vorbildlich das Feld. Sie war auch an der Auswertung der Daten beteiligt.

Worterkennung, Lesen, Vorstellungen von Mathematik, Zählen und Mengen, Identifikation von Formen, Zahlenkenntnis sowie formelle und formale Rechenaufgaben (Tymms, 1999; Tymms & Merrell, 2009). Darüber hinaus können optional neben dem Kurzzeitgedächtnis auch die Einstellung gegenüber einzelnen Fächern und schulrelevanten Aktivitäten sowie das akademische Selbstkonzept - also zentrale Persönlichkeitsmerkmale - erhoben werden (Tymms & Albone, 2002). Die meisten der insgesamt 14 Aufgabentypen wurden vom CEM speziell für das PIPS-Projekt entwickelt. Die Computerversion erlaubt adaptives Testen, d.h. die Itemauswahl wird an die Fähigkeiten des Kindes angepasst. Da das Programm einerseits sehr weit entwickelte schriftsprachliche und mathematische Lernvoraussetzungen berücksichtigt, andererseits nach mehreren erfolglosen Aufgabendarbietungen zur nächsten Aufgabengruppe übergeht, können sehr unterschiedliche Lernvoraussetzungen erhoben werden, ohne die Kinder zu entmutigen, und das Programm ist sowohl am Schulanfang als auch am Ende der ersten Klasse einsetzbar. Zugleich wird dadurch die Testzeit kurz gehalten. In den höheren Primarklassenstufen werden lehrplanbasierte Aufgaben in Mathematik, Lesen und Naturwissenschaften sowie Tests der allgemeinen verbalen und nonverbalen Fähigkeiten durchgeführt.

Mit einer *Retestreliaibilität* von .98 erwies sich das PIPS-Verfahren als sehr zuverlässig (Tymms & Wylde, 2004). Eine Vorhersage der Leistung in einem Mathematik- und Deutschtest drei Jahre nach der PIPS-Testung beim Schuleintritt ergab eine *Validität* von etwa .7 (Tymms & Wylde, 2004). Eine Korrelation der PIPS-Ergebnisse vom Anfang der ersten Klasse mit den Schulleistungen im Alter von elf Jahren betrug ebenfalls .7 (Tymms & Wylde, 2004; Tymms et al. 2007). Insgesamt konnten die Ergebnisse einer PIPS-Testung bei der Einschulung im Alter von viereinhalb Jahren fast 50 % der Leistungsvarianz am Ende der Grundschulzeit erklären (Tymms et al., 2007; Tymms et al., 2009). Während für die Leseleistung im Alter von sieben Jahren die Buchstabenkenntnis der beste Prädiktor war, kamen die besten Vorhersagen für die Leistung in den Bereichen Lesen, Mathematik und Naturwissenschaften im Alter von elf Jahren aus dem mathematischen Bereich (Tymms et al., 2007; Tymms et al., 2009). Informelle Rechenaufgaben und Zahlenkenntnis erwiesen sich dabei als beste Prädiktoren. Die Identifikation von Buchstaben war ein etwas schwächerer, aber immer noch signifikanter Prädiktor. Auch die Fähigkeit zur Lautanalyse stellte sich als vielversprechender Prognosefaktor heraus (Tymms et al., 2009).

Durch die zweimalige Durchführung, jeweils am Anfang und am Ende eines Schuljahres, können die Schulen erkennen, welche Schülerinnen und Schüler zurückbleiben oder besonders gute Fortschritte machen. Die Berechnung des individuellen Lernfortschrittes erfolgt regressionsanalytisch. Durch den Vergleich einer Schülerin oder eines Schülers mit anderen Schülerinnen und Schülern, die bei der Einschulung eine ähnliche Lernausgangslage aufwiesen, wird mit dem PIPS ein fairer Maßstab für die individuelle Lernentwicklung angelegt. Lehrkräfte und Schulleitungen können dadurch die Effektivität der Arbeit in ihren Klassen bzw. Schulen in Relation zu anderen Klassen und Schulen beurteilen (Tymms & Wylde, 2004).

Eine *Befragung an 500 Schulen*, die das PIPS-Programm einsetzen, ergab, dass die Lehrkräfte die Rückmeldung, die sie vom CEM über den Leistungsstand ihrer Schüler und Klassen erhalten in vielfältiger Weise nutzen können (Tymms & Albone, 2002). Beispielsweise gaben über 93 % der Lehrkräfte an, auf der Basis der PIPS-Ergebnisse die Stärken und Schwächen einzelner Kinder zu identifizieren. 76 % verwendeten die Daten zur Kontrolle des Lernfortschritts der ganzen Klasse. Auf die Frage, für wie wertvoll die Lehrkräfte die PIPS-Rückmeldung hielten, zeigte sich eine fast durchweg positive Beurteilung (s. Tab. 1; Tymms & Albone, 2002). Vor allem Rückmeldungen zu Beginn und am Ende der ersten Klasse werden von Lehrkräften als wertvoll empfunden (Tymms, 1999).

Tabelle 1. Bewertung der PIPS-Rückmeldungen durch die Lehrkräfte .

Bewertung	1	2	3	4	5
Prozent	3 %	6 %	22 %	47 %	20 %

Anmerkung: 1 = nicht wertvoll, 5 = sehr wertvoll

Das PIPS wird *weltweit* in über 4.000 Schulen in acht verschiedenen Sprachräumen und 25 verschiedenen Ländern (z.B. den Niederlanden, England, Australien, Schottland, Wales, Südafrika, Hongkong, Serbien, Luxemburg und Neuseeland) sowie zahlreichen internationalen Schulen auf der ganzen Welt eingesetzt (Tymms et al., 2007; Tymms et al., 2009), wobei jeweils landesweite Normwerte zur Verfügung stehen. Da sich bei der Feststellung des Lernstands und der Beobachtung der Fortschritte sehr ähnliche Profile in allen einbezogenen Ländern zeigen, eignet sich das PIPS-Programm für

grenzüberschreitende Untersuchungen sowie Vergleiche des Lernstandes und des Lernzuwachses zwischen verschiedenen Ländern (Tymms & Wylde, 2004).

3. Lernausgangslagen erheben – individuell fördern (LEIF)

Im Rahmen der LEIF-Studie wurde erprobt, inwieweit sich eine an das englische PIPS-Projekt angelehnte Schulanfangsdiagnose auf den deutschen Schulanfang übertragen lässt. Als Diagnoseinstrument diente eine von der ehemaligen Lehrerin an der Deutschen Schule London Monika Wylde erarbeitete und an dieser Schule bewährte Übersetzung des PIPS-PC-Programms ins Deutsche. Aus dem englischen „PIPS“ war dabei das deutsche „FIPS (Fähigkeitsindikatoren Primarschule)“ geworden. Die LEIF-Studie untersuchte neben der Alltagstauglichkeit für deutsche Verhältnisse, ob und inwieweit die Erstklasslehrkräfte die dadurch erhobenen diagnostischen Daten im Unterricht nutzen. Vor allem interessierte die Rolle der diagnostischen Informationen bei *Differenzierungs- und individuellen Fördermaßnahmen*. Da die englischen Kinder im Alter von vier bis fünf Jahren in die Reception Class eintreten, die deutschen aber erst ca. zwei Jahre später in die Grundschule, wurde außerdem ermittelt, welche Aufgabengruppen, Aufgaben und Items für die Testung der deutschen Schulanfänger adaptiert werden müssen. Im Gegensatz zum englischen PIPS ist das deutsche FIPS bisher auf den Einsatz zu Beginn und am Ende der ersten Klasse beschränkt und liegt nur als Computerversion vor, was adaptives Testen und den Einsatz am Schuljahresanfang und -ende erlaubt sowie die Testzeit kurz hält.

3.1 Stichprobe und Untersuchungsdesign der LEIF-Studie

In den Schuljahren 2007/08 („LEIF 1“) und 2008/09 („LEIF 2“) fanden *zwei Pilotuntersuchungen* statt. An LEIF 1 nahmen knapp 400 Schülerinnen und Schüler aus 20 ersten Klassen an acht Schulen aus den Regionen Würzburg und Bamberg teil. In denselben Regionen wurde ein Jahr später mit den identischen Instrumenten und exakt parallel getaktet LEIF 2 mit knapp 300 Schülerinnen und Schülern aus 15 ersten Klassen an sieben Schulen durchgeführt. Aus organisatorischen Gründen fand die Eingangserhebung von weiteren 270 Schülerinnen und Schülern in den Regionen Augsburg und München zeitlich verzögert statt. Zwar wurden auch an diesen Klassen und Schulen Lehrer- und Elternfragebögen sowie Unterrichtstagebücher eingesetzt, aber die Schülerdaten sind nicht mit den in Mittel- und Unterfranken ermittelten Daten vergleichbar (zur Übersicht über die Stichprobe vgl. Tab. 2). Insgesamt nahmen an LEIF 1 20 Lehrerinnen teil, an LEIF 2 15 Lehrkräfte. Die

Datenerhebungen zu LEIF 2 sind abgeschlossen, die Schülerdaten vom Schuljahresende liegen jedoch noch nicht vor. Die im Abschnitt 3.3 berichteten Ergebnisse beziehen sich deshalb nur auf LEIF 1. Die Schülerdaten dieses ersten Durchgangs bildeten auch die Grundlage für Itemanalysen und entsprechende Überarbeitungen des Programms. Im Schuljahr 2009/10 findet in allen vier beteiligten Regionen eine Normierungsuntersuchung mit dem überarbeiteten FIPS-Programm statt.

Zur Überprüfung der Fragen, ob sich das Programm unter deutschen Bedingungen einsetzen lässt und ob und inwieweit die Lehrkräfte die dadurch gewonnenen diagnostischen Informationen im Unterricht und insbesondere zur individuellen Förderung ihrer Schulanfänger nutzen, wurde jede Klasse einer Schule einer von *drei Versuchsbedingungen* zugewiesen:

- Bei der *Experimentalgruppe 1* (EG 1) führten die Klassenlehrerinnen nach entsprechender Schulung die Testungen in ihrer Klasse selbst durch und lernten somit durch die in der zweiten und dritten Schulwoche durchgeführten Testungen schon kurz nach Schulbeginn den Stand aller Schülerinnen und Schüler ihrer Klasse unmittelbar kennen. Während sie das Computerprogramm durchführten, wurden sie im Unterricht von mobilen Reservisten vertreten. Sobald die Auswertungen aller Schülerdaten vorlagen, was bei LEIF 1 Anfang Januar 2008 der Fall war, erhielten sie noch einmal eine schriftliche Rückmeldung auf Klassen- und Individualebene, auch im Vergleich zu den übrigen an der Studie teilnehmenden Klassen.
- Bei der *Experimentalgruppe 2* (EG 2) wurden die Lehrerinnen Anfang Januar 2008 schriftlich über die Anfangsleistung ihrer gesamten Klasse im Vergleich zu den übrigen Klassen informiert, nicht jedoch über die Lernvoraussetzungen der einzelnen Kinder.
- Bei der *Kontrollgruppe* (KG) erhielten die Lehrkräfte keine Rückmeldung über das Abschneiden ihrer Klasse oder einzelner Schüler.

Zusätzlich wurde die Variable *Testleiter* variiert, indem in der zweiten und dritten Gruppe entsprechend geschulte externe Testleiterinnen und Testleiter die Schülerdaten erhoben.

Variiert wurde also die Testleitung, das Maß an Information, das die Lehrkräfte über ihre Schülerinnen und Schüler erhielten, sowie der Zeitpunkt der Rückmeldung (s. Tab. 2).

Tabelle 2. Versuchsdesign der ersten Pilotierungsphase LEIF 1.

	Versuchsbedingung		
	<i>Experimental- gruppe 1 (EG 1)</i>	<i>Experimental- gruppe 2 (EG 2)</i>	<i>Kontroll- Gruppe (KG)</i>
Testleiter	Lehrkraft selbst	Externer Testleiter	Externer Testleiter
Ergebnis- rückmeldung	Individual- <u>und</u> Klassenebene	Nur Klassen- ebene	Keine Rückmeldung
Zeitpunkt der Ergebnisrückmeldung	Schuljahresanfang Individualebene, Januar 2008 Klassenebene	Januar 2008	----

Den Tabellen 3 und 4 sind die Stichprobengrößen von LEIF 1 und 2 zu entnehmen.

Tabelle 3. Stichprobe der ersten Pilotierungsphase LEIF 1, Schuljahr 2007/08 (Schuljahresende, ohne drop-out).

	EG1	EG2	KG	Gesamt
Würzburg	46	80	58	184
Bamberg	59	69	77	205
Gesamt	105	149	135	389

Tabelle 4. Stichprobe der zweiten Pilotierungsphase LEIF 2, Schuljahr 2008/09 (Schuljahresanfang).

	EG1	EG2	KG	Gesamt
Würzburg	41	45	39	125
Bamberg	61	43	54	158
Augsburg	33	28	33	94
München	61	61	53	175
Gesamt	196	177	179	552

3.2 Instrumente und Fragestellungen der LEIF-Studie

Zusätzlich zur Schülertestung am Schuljahresbeginn und -ende wurden am Schuljahresbeginn ein Lehrer- und ein Elternfragebogen, während des Schuljahrs zweimal ein Unterrichtstagebuch (in LEIF 2 ein weiteres Unterrichtstagebuch vor der ersten klassenbezogenen Rückmeldung) und am Schuljahresende erneut ein Lehrerfragebogen eingesetzt. Außerdem sollen die Mathematik- und Deutschnoten der Kinder am Ende der

zweiten und dritten Klasse sowie ihre Leistungen in den bayernweit einheitlichen Orientierungsarbeiten erfasst werden. Damit soll im weiteren Verlauf die *Vorhersagekraft* des FIPS-Verfahrens für die spätere schulische Laufbahn überprüft werden. Die Instrumente für die Hand der Lehrkräfte erhoben schwerpunktmäßig Informationen zu ihren *Leistungsdiagnosen* und zum *Umgang mit den heterogenen Ausgangslagen* der Schülerinnen und Schüler im Unterricht. Zusätzlich wurde jeweils nach den Gründen für die Bildung von Teilgruppen bzw. die individuelle Förderung einzelner Kinder gefragt.

Lehrerfragebogen I am Schuljahresanfang

Der Lehrerfragebogen erfasste neben Informationen zur Person und zur beruflichen Laufbahn die *unterrichtlichen Rahmenbedingungen*. Getrennt für die Lernbereiche Deutsch und Mathematik wurden die Stundenanzahl einschließlich Förder- und Teilungsunterricht und das übliche Vorgehen bei der *Unterrichtsdifferenzierung* erhoben. In Mathematik wurden die verwendeten Unterrichtswerke, in Deutsch die Ziele der Lehrkraft für das Ende des 1. Schuljahrs und die grundsätzliche Anlage des Schriftspracherwerbsunterrichts erfragt. Außerdem sollten die Lehrkräfte in beiden Lernbereichen den *Stand der Klasse* einschätzen.

Elternfragebogen

Über die Elternfragebögen wurden Informationen zum *sozioökonomischen Hintergrund* des Kindes einschließlich eines möglichen Migrationshintergrunds und zu den Bildungserwartungen der Eltern an das Kind erhoben. Dadurch sollten der durchschnittliche soziale Status und der Migrantenteil der Klasse bei der Aufklärung der Leistungsfortschritte der Klasse einbezogen werden können. Erfragt wurden u.a. der Bildungs- und Berufsabschluss sowie die derzeitige berufliche Tätigkeit beider Eltern, der Reichtum in der Familie ("wealth"), die Geburtsländer beider Eltern und die in der Familie bzw. vom Kind gesprochene Sprache. Zusätzlich wurde auch nach dem Kindergartenbesuch des Kindes und dessen Dauer gefragt.

Unterrichtstagebuch

Alle Unterrichtstagebücher waren gleich aufgebaut. Für die fünf Tage einer Unterrichtswoche sollten die Lehrkräfte getrennt für den Deutsch- und den Mathematikunterricht am Ende des Unterrichtstages *konkrete Fragen zur Unterrichtsgestaltung* beantworten: Dauer des Deutsch- oder Mathematikunterrichts,

Vorkommen von Teilungs- oder Förderunterricht und deren Thema, Vorkommen und Dauer von differenzierten Unterrichtsformen wie Freie Arbeit, Wochenplanunterricht u.ä., Gruppeneinteilungen sowie quantitative und/oder qualitative Differenzierung (z.B. Art der Aufgaben oder besondere Hilfsmittel für einzelne Kinder), schließlich geplante oder ungeplante Arbeit mit einzelnen Kindern und deren Dauer. Die Lehrkräfte sollten jeweils die Kinder nennen, die an den Maßnahmen teilgenommen hatten, und die Ziele der jeweiligen Maßnahmen. Am Ende der Woche wurde außerdem gefragt, aus welchen Quellen die Lehrkraft diagnostische Informationen zur Lernentwicklung ihrer Schülerinnen und Schüler gewinnt und ob sie einzelne Inhalte aufgrund dieser Informationen unterrichtet hat.

Lehrerfragebogen II am Schuljahrsende

Im zweiten Lehrerfragebogen sollten die Lehrkräfte für jeden Lernbereich und getrennt für das erste und zweite Schulhalbjahr noch einmal zusammenfassend ihre *Differenzierungsmaßnahmen* beschreiben. Außerdem sollten sie für beide Lernbereiche jeweils die drei nach ihren *Einschätzungen* leistungsstärksten und leistungsschwächsten Kinder angeben sowie die Kinder, bei denen die stärksten positiven oder negativen Veränderungen während der ersten Klasse stattfanden.

Durch die Instrumente für die Hand der Lehrkräfte sollen folgende *Fragen* beantwortet werden:

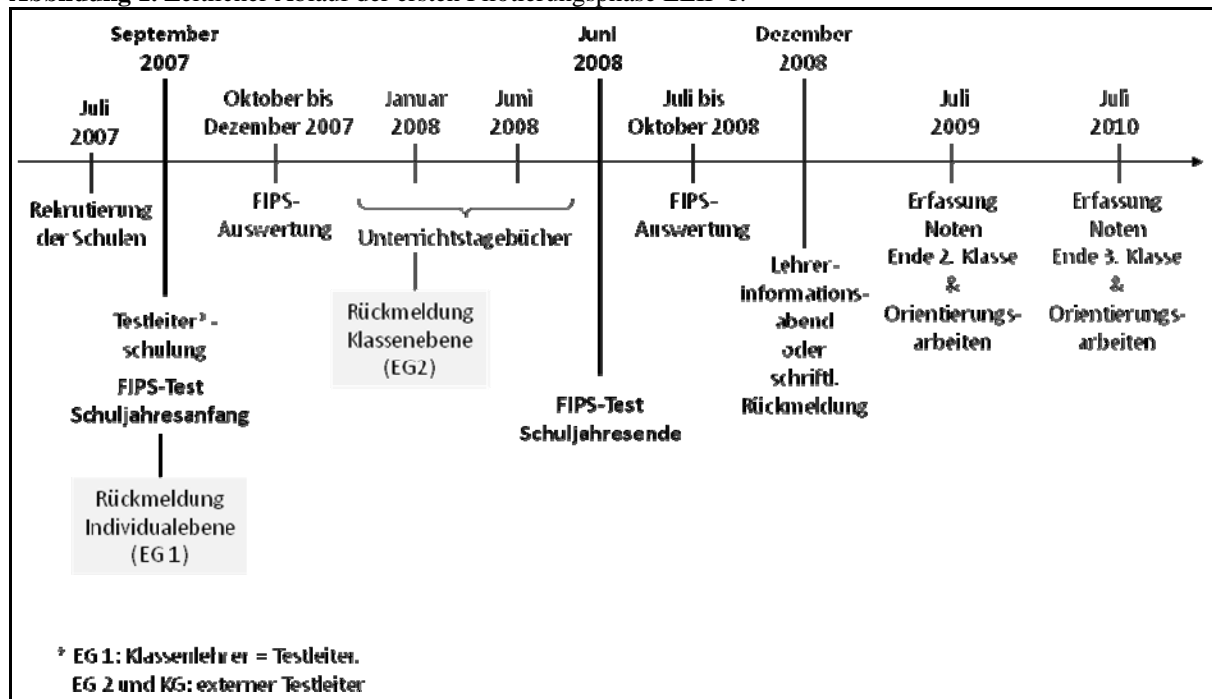
- In welchem Umfang kommen welche Differenzierungs- und individuellen Fördermaßnahmen in den ersten Klassen vor? Werden leistungsstarke und leistungsschwache Kinder im Anfangsunterricht gleichermaßen gefördert?
- Welche diagnostischen Informationen ziehen die Lehrkräfte heran und wie gut ist ihre diagnostische Kompetenz ausgeprägt?
- Welchen Lernzuwachs weisen die Klassen im Mittel und die jeweils leistungsstärksten und leistungsschwächsten Schüler in den Klassen auf?

Bei allen Fragestellungen interessieren insbesondere mögliche Unterschiede zwischen den drei Lehrer- bzw. Klassengruppen (EG 1, EG 2, KG). Die Antworten auf diese Fragen

stehen noch aus, da aufgrund der geringen Lehrerzahl pro Treatment die weitere Auswertung dieser Fragestellungen zumindest bis zum Abschluss von LEIF 2 verschoben werden muss.

Der zeitliche Ablauf der ersten Pilotuntersuchung geht aus Abbildung 1 hervor.

Abbildung 1. Zeitlicher Ablauf der ersten Pilotierungsphase LEIF 1.



3.3 Ergebnisse

3.3.1 Itemanalyse und -überarbeitung

Nach Abschluss der Datenerhebungen wurden alle FIPS-Items genauer analysiert. Dabei zeigten sich im Aufgabenbereich *Lautbewusstheit* bereits zu Beginn der ersten Klasse Deckeneffekte. Im Aufgabenbereich *Lesen* kamen am Jahresende weitere *Deckeneffekte* hinzu. Lediglich die Ergebnisse im Bereich *Mathematik* erwiesen sich als *annähernd normalverteilt* und konnten deshalb weitgehend beibehalten werden. Die linksschiefen Häufigkeitsverteilungen in den Bereichen *Lesen* und *Lautbewusstheit* lassen sich wahrscheinlich auf die mangelnde Anpassung der Items dieser ersten FIPS-Version an den deutschen Anfangsunterricht und die Lernausgangslage der beim Schuleintritt im Vergleich zu den englischen Erstklässlern deutlich älteren deutschen Schulanfänger zurückführen. Auf der Grundlage weiterer Itemanalysen wurden die Aufgaben im Bereich *Lautbewusstheit* noch einmal nahezu vollständig überarbeitet. Auch im Bereich *Lesen* wurden einige

Verbesserungen vorgenommen. In der zweiten Pilotphase im Schuljahr 2008/09 wurden zusätzlich zum computerbasierten FIPS-Programm die neuen Aufgaben in Papierform durchgeführt, um sie auf ihre Tauglichkeit zu überprüfen und die Schwierigkeitsindizes der Items zu bestimmen.

Verbesserungsvorschläge der Testleiter sowie fachlicher Rat von Lehrkräften und Grundschulpädagogen führten zu Veränderungen bei den Instruktionen und von Details der Programmsteuerung, um den Ablauf und die Durchführung noch weiter zu erleichtern. Es wurde darauf geachtet, dass Aufgaben nach aufsteigender Schwierigkeit angeordnet sind. Die *revidierte Fassung des FIPS-Programms* wird im Abschnitt 4 genauer beschrieben.

3.3.1.1 Lautbewusstheit

Bei den Aufgaben zur Lautbewusstheit zeigte sich in der Rasch-Analyse eine nicht zufriedenstellende Personenreliabilität von .32 und eine Itemreliabilität von .98. Außerdem umfasste der Bereich Lautbewusstheit nur 17 Items. Diese waren zudem sehr leicht und konnten nicht ausreichend zwischen den Kindern differenzieren. Aus der Reimaufgabe wurden daher zu leichte und zu wenig trennscharfe Items (mind. 95 % Kinder lösten die Items richtig) entfernt und eine neue Reimaufgabe mit höherem Schwierigkeitsgrad aufgenommen, bei der aus einer Reihe von vier vorgesprochenen Wörtern dasjenige Wort ausgewählt werden soll, das sich nicht mit den übrigen drei Wörtern reimt (in Anlehnung an Bradley & Bryant, 1985). Neu ist außerdem eine Aufgabe zur Lautanalyse, bei der die Kinder mehrere Wörter in einzelne Laute zerlegen sollen (in Anlehnung an Küspert, 1998; Küspert & Schneider, 2006; Roth, 1999). Insgesamt ist dieser Bereich durch *drei Aufgaben* mit jeweils mehreren Items abgedeckt.

3.3.1.2 Lesen

Für den Bereich Lesen ergab die Rasch-Analyse eine Personenreliabilität von .91 und eine Itemreliabilität von .99. Beide Reliabilitätswerte sind hoch, was auf eine große Streubreite der Aufgabenschwierigkeit und eine gute Diskriminationsleistung der Skala Lesen zwischen den Fähigkeiten der Kinder hinweist. Der Mittelwert der Schülerfähigkeit lag etwas über dem Mittelwert der Itemschwierigkeiten. Das spricht dafür, dass die Leseaufgaben tendenziell zu leicht waren. Beim Ausschluss der Aufgabe "Bilderwortschatz" blieben beide Reliabilitäten hoch (Personenreliabilität = .91, Itemreliabilität = .99) und die Mittelwerte der

Schülerfähigkeiten und Itemschwierigkeiten stimmten fast überein, wobei die Itemschwierigkeit etwas höher (aber nicht zu hoch) für die Schülerfähigkeiten war.

Inhalt dieser Aufgabe, mit der der Wortschatz erfasst wird, waren drei Bilder (Landschaft mit Schloss, Küche und Spielzeugladen), wobei die Kinder nacheinander auf mehrere vorgegebene Objekte deuten sollten. Das erste Bild wurde aufgrund der sehr geringen Schwierigkeit eliminiert, zusätzlich wurden auf den beibehaltenen Tafeln "Küche" und "Spielzeugladen", die typische semantische Felder des frühen Lexikons von Kindern darstellen (z.B. Lebensmittel, Haushaltsgegenstände, Spielzeug; Kauschke & Stan, 2004), einzelne zu benennende Objekte gegen schwierigere ausgetauscht. Dabei wurde berücksichtigt, dass Wortlänge, Häufigkeit im Sprachgebrauch und die Verwendung von Ober- vs. Unterbegriffen einen Einfluss darauf haben, ob Kinder ein Wort bereits kennen bzw. benennen können oder nicht (Kauschke & Stan, 2004; Siegmüller, 2005).

Bei der Prüfung der "Vorstellungen vom Lesen", bei der zu dem Bild eines Klassenzimmers beispielsweise die Aufforderung „Zeige mir ein Kind, das schreibt!“ erfolgte, wurde die erste von zwei Teilaufgaben wegen zu geringer Schwierigkeit eliminiert.

Nach der Darbietung von Buchstaben und vor dem Lesen von Wörtern und Sätzen wurde als Zwischenstufe eine Aufgabe zum Lesen von Silben aufgenommen, mit der ermittelt wird, ob die Schulanfänger schon zwei Buchstaben synthetisieren können. Die Items stammen aus dem Test *Wortgewandt und Zahlenstark* von Moser und Berweger (2007) und waren auch in der deutschen Grundschuluntersuchung PERLE erfolgreich verwendet worden (Corvacho & Greb, 2007). Bei einer vollständigen Bearbeitung des Bereichs Lesen sind *zehn Aufgaben* mit jeweils mehreren Items zu bewältigen.

3.3.1.3 Mathematik

Dieser Bereich umfasste Aufgaben zu Vorstellungen von Mathematik, Zählen, Zahlenidentifikation sowie einfache Rechenaufgaben, Formidentifikation und Lösung mathematischer Probleme. Hier ergab die Rasch-Analyse mit einer Personenreliabilität von .86 und einer Itemreliabilität von .98 zufriedenstellende Werte. Die Aufgaben konnten auch am Jahresende noch ausreichend zwischen den Schülern differenzieren. Da aber die Aufgaben "Vorstellungen von Mathematik", "Zählen" und "Formen" schon am Schuljahresanfang sehr leicht waren und von mehr als 95 % der Schülerinnen und Schüler vollständig gelöst wurden, wurde der Mathematikteil um diese Aufgaben gekürzt, so dass der Teil revidiert *vier Aufgaben* umfasst.

3.3.2 Ergebnisse zu den Schülerleistungsdaten

Die Datenauswertung der ersten Pilotphase ergab, dass sich die Lernausgangslage am Schuljahresanfang nicht signifikant zwischen den Versuchsbedingungen und den Standorten unterschied. Das bedeutet, dass alle Klassen, wie notwendig, die *gleichen Ausgangsvoraussetzungen* hatten. Unterschiede zwischen den Versuchsgruppen am Schuljahresende oder im Lernzuwachs können somit möglicherweise – vorbehaltlich einer genaueren Prüfung der Lehrerdaten – auf die unterschiedlichen Treatments zurückgeführt werden.

Es zeigte sich *kein Einfluss der Testleitervariation*, das heißt, es gab keinen signifikanten Unterschied zwischen Klassen, in denen die Lehrkraft selbst testete, und Klassen, in denen externe Testleiter die Kinder durch das FIPS-Programm führten. Die Unabhängigkeit der FIPS-Ergebnisse vom Testleiter spricht für die Objektivität des Verfahrens.

Die weitere Auswertung erfolgte separat für die einzelnen Bereiche Lesen, Mathematik und Lautbewusstheit. Dabei zeigte sich in *Mathematik* der stärkste Leistungszuwachs in der EG 1, ein etwas schwächerer Lernzuwachs in der EG 2 und der geringste Lernzuwachs in der KG. Als Wertzuwachs wurde dabei die mittlere Differenz zwischen den FIPS-Ergebnissen vom Jahresanfang und Jahresende betrachtet. Eine Varianzanalyse ergab, dass der Unterschied zwischen den Gruppen signifikant ist ($F = 4.51, p < .01$). Entsprechende Kontraste zeigten, dass der Leistungszuwachs im Bereich Mathematik in der EG 1 signifikant größer war als der Leistungszuwachs der KG (mittlere Differenz = $-2.09, p < .01$). Das heißt also, dass die Klassen, deren Lehrkräfte unmittelbar nach Schuljahresanfang bei der FIPS-Testung diagnostische Informationen zu den Lernvoraussetzungen ihrer einzelnen Schülerinnen und Schüler erhalten konnten, eine größere Leistungssteigerung in Mathematik erzielten als die Klassen, deren Lehrkräfte keine diagnostischen Informationen erhalten hatten.

Im Bereich *Lesen* zeigten sich allerdings am Schuljahresende keine Leistungsunterschiede zwischen den Gruppen. Jedoch erwies sich die Varianz der FIPS-Ergebnisse in den Klassen der EG 1 als signifikant geringer im Vergleich zur KG ($F = 1.69, p < .01$). In Klassen, in denen die Lehrkräfte die Lernvoraussetzungen ihrer einzelnen Schülerinnen und Schüler im

Schriftspracherwerb mit Hilfe des FIPS-Programms kennenlernten, wurde also am Ende der ersten Klasse ein einheitlicheres Niveau der Lesefähigkeit erreicht als in Klassen, in denen die Lehrkräfte nicht über die Ergebnisse der FIPS-Testungen informiert wurden.

Im Bereich *Lautbewusstheit* zeigten sich weder bezüglich der FIPS-Ergebnisse noch bezüglich der entsprechenden Varianzen Unterschiede zwischen den Gruppen. Dies lässt sich wahrscheinlich auf die mangelnde Passung der Items in der ursprünglichen FIPS-Version zurückführen.

4. Fähigkeitsindikatoren Primarschule (FIPS) – revidierte Version

In die im Jahr 2009 fertiggestellte Endversion des computerbasierten FIPS-Programms wurden die neuen Aufgaben, die in der zweiten Pilotierungsphase (LEIF 2) zusätzlich zum FIPS im Paper-Pencil-Format getestet worden waren und sich dabei bewährt hatten, aufgenommen. Für die Lautanalyse wurden dafür vom CEM entsprechende Bilder erstellt und ergänzt.

4.1 Durchführung

Das deutsche FIPS-Programm existiert bisher nur in computerisierter Fassung. Beim Schuleintritt und (optional) am Ende des ersten Schuljahres soll jedes Kind in einer etwa 15- bis 30-minütigen Einzelsitzung von einer Lehrkraft durch das Programm geführt werden. Die Items werden größtenteils direkt am PC dargeboten, die Instruktionen erfolgen über die Audioausgabe des PCs. Lediglich für die Aufgaben "Schreiben" und "Gemischte Mathematikaufgaben" werden dem Kind zusätzlich Papier und Bleistift sowie für die Aufgabe "Lautanalyse" kleine Plastikplättchen zur Verfügung gestellt. Darüber hinaus sind keine weiteren Arbeitsmittel oder Lernhilfen nötig.

Testleiter und Schüler sitzen gemeinsam vor dem Bildschirm und der Testleiter zeichnet mit der Computermaus die Antworten des Kindes auf. Der PC speichert die Ergebnisse automatisch auf der Festplatte. Das Programm interagiert mit dem Kind und passt die Itemauswahl fortwährend an die Fähigkeiten des Kindes an. Kann ein Kind drei Items in Folge oder insgesamt fünf Items eines Aufgabenbereichs nicht richtig lösen, wird automatisch zum nächsten Aufgabenbereich übergegangen. Analog entfallen Aufgaben ganz, wenn einfachere Aufgaben zuvor nicht gelöst wurden. Auf diese Weise können Kinder mit

sehr unterschiedlichen Lernvoraussetzungen mit demselben Programm getestet werden, ohne dass ein Kind über- oder unterfordert wird.

Am Schuljahresende setzt FIPS automatisch bei jedem Kind mehrere Items vor dem letzten richtig gelösten Item vom Schuljahresanfang ein. Von dort wird wieder mit aufsteigender Aufgabenschwierigkeit fortgefahren, bis entweder alle Aufgaben richtig gelöst sind oder erneut eines der beiden Abbruchkriterien erfüllt ist.

4.2 Aufbau des Diagnoseprogramms, Aufgaben und Erweiterungen

Die aus wiederholten Itemanalysen und Überarbeitungen hervorgegangene revidierte Version des deutschen FIPS-Programms umfasst *17 Aufgaben* und *vier optionale Erweiterungen* (davon zwei für das Kind und zwei zur Bearbeitung durch die Lehrkraft). Die Vorkenntnisse im *Lesen* werden durch zehn Aufgaben überprüft; drei Aufgaben beziehen sich auf die *Lautbewusstheit* und vier auf *Mathematik*. Außerdem können vom Kind eine Aufgabe zur Prüfung des Kurzzeitgedächtnisses und ein Fragebogen zu Persönlichkeitsaspekten bearbeitet werden. In den folgenden Abschnitten werden die einzelnen Aufgaben in der Reihenfolge der Darbietung beschrieben:

Schreiben

In der Einführungsaufgabe am Schuljahresanfang und -ende wird das Kind aufgefordert, seinen vollständigen Vor- und Nachnamen auf ein Blatt Papier zu schreiben, wobei die ausschließliche Verwendung von Großbuchstaben auch die Anforderungen erfüllt. Der Testleiter bzw. die Testleiterin beurteilt die Qualität des Schriftbildes mit einem Mausklick auf das entsprechende Feld einer vorgegebenen Skala von 0 (kann nicht schreiben) bis 5 (sehr klar, Vor- und Nachname).

Bilderwortschatz

Die nächste Aufgabe beginnt mit der Darbietung der Bildtafel einer Küche mit zahlreichen typischen Objekten. Das Kind wird aufgefordert, mit dem Finger auf bestimmte Objekte (z.B. Obst) zu zeigen (s. Abb. 2). Wenn das Kind diese Aufgabe bewältigt hat, wird anschließend das Bild eines Spielzeugladens mit verschiedenen Spielsachen eingeblendet. Auch hier soll das Kind auf einzelne Objekte deuten. Die Aufgabe überprüft den Wortschatz, die Fähigkeit, die angegebenen Objekte auf der Bildtafel zu finden, und die Fähigkeit und

Bereitschaft, Aufgaben auszuführen, und ist dadurch vermutlich prädiktiver als eine reine Wortschatzaufgabe.²

Abbildung 2. Bildschirm bei der Aufgabe *Bilderwortschatz*



Vorstellungen vom Lesen

Bei dieser Aufgabe werden dem Kind zwei Sätze und eine dazu passende Abbildung dargeboten. Es soll mit dem Finger auf bestimmte Bestandteile der Sätze (z.B. den Anfang eines Satzes, einen Großbuchstaben, einen Punkt) zeigen.

Silben nachsprechen

Über die Audioausgabe des PCs dargebotene Silben sollen vom Kind nachgesprochen werden. Danach folgen Silbenkombinationen (Pseudowörter). Hier kann nicht nur festgehalten werden, ob das Kind die Silben richtig oder falsch nachspricht, sondern auch, ob ggf. eine Sprachstörung vorliegt oder das Kind das Vorgesprochene nicht wiederholen möchte.

Reimwörter

Anschließend werden dem Kind immer vier Wörter und - zur Entlastung des Gedächtnisses - die entsprechenden Bilder vorgegeben. Das Kind soll erkennen, welches weitere Wort sich

² Hinweis Prof. Dr. S. Weinert, Universität Bamberg

auf das erste Wort reimt. Kann das Kind die Aufgabe bewältigen, werden vier Wörter (ohne Bilder) dargeboten und das Kind soll umgekehrt entscheiden, welches der Wörter sich nicht mit den anderen reimt.

Lautanalyse

Bei der Lautanalyse werden dem Kind einzelne Wörter mit einer entsprechenden Abbildung dargeboten. Das Kind soll das Wort jeweils in die einzelnen Laute zerlegen und für jeden Laut ein Plastikplättchen auf den Tisch legen. Es werden zunächst Wörter mit zwei Lauten dargeboten, bei Erfolg danach auch Wörter mit drei und vier Lauten.

Buchstaben

Die nächste Aufgabe des Kindes besteht darin, einzelne Groß- und Kleinbuchstaben vorzulesen. Hier zählen sowohl der Laut als auch der Buchstabenname als richtige Lösung. Als Erstes erscheint immer der erste Buchstabe des Vornamens des Kindes. Die Anzahl der darauf folgenden Buchstaben richtet sich nach der Buchstabenkenntnis, über die das Kind bereits verfügt.

Silben lesen

Hat das Kind eine ausreichende Anzahl an Buchstaben richtig erkannt, wird es aufgefordert, eingeblendete Silben aus zwei Buchstaben vorzulesen. Gelingt dies, folgen einfache Silbenkombinationen.

Vertauschte Wörter

Sofern das Kind die ersten beiden Leseaufgaben (Buchstaben und Silben lesen) bewältigen konnte, sollen nun aus eingeblendeten Wörtern und Bildern zusammengehörige Bild-Wort-Paare gebildet werden.

Bild-Wort-Zuordnung

Ähnlich wie bei der vorhergehenden Aufgabe soll das Kind hier aus mehreren Wörtern dasjenige auswählen, das zu einem eingeblendeten Bild passt.

Sätze lesen

Kann das Kind Silben und Wörter lesen, wird es gebeten, mehrere Sätze, die zusammen eine kleine Geschichte ergeben, vorzulesen.

Sprechblasen lesen

Anschließend werden Bilder mit Sprechblasen dargeboten, die das Kind vorlesen soll. Hierbei liegt die Schwierigkeit darin, einen Satz zu lesen, der sich über mehrere Zeilen erstreckt.

Lückentext

Wurden alle vorangehenden Leseaufgaben bewältigt, folgt ein längerer Lückentext. Das Kind soll mehrfach aus drei Antwortmöglichkeiten das in den Zusammenhang passende Wort aussuchen und die Sätze vorlesen. So wird mit einer sehr schwierigen Aufgabe also auch das Leseverständnis auf Satz- und Textebene erfasst.

Rechnen mit Bildern

Der mathematische Bereich beginnt mit leichten Additions- und Subtraktionsaufgaben, die durch Bilder mit der entsprechenden Anzahl von Gegenständen erleichtert werden. Die Instruktion bei dieser Aufgabe lautet z.B.: „Hier sind sechs Eistüten. Wenn du drei wegnimmst, wie viele bleiben übrig?“

Zahlen lesen

Bei der nächsten Aufgabe werden erst ein-, dann zwei- und schließlich dreistellige Zahlen dargeboten, die das Kind vorlesen soll.

Rechnen mit Punkten

Konnte das Kind die Aufgabe Rechnen mit Bildern bewältigen, folgen nun Additions- und Subtraktionsaufgaben mit höherem Abstraktionsgrad, die nur noch durch Punkte in der entsprechenden Anzahl unterstützt werden. Die Instruktion lautet dabei beispielsweise: „Hier sind sieben Punkte. Es werden drei Punkte weggenommen. Wie viele bleiben übrig?“

Gemischte Matheaufgaben

Wenn das Kind auch beim Rechnen mit Punkten eine ausreichende Anzahl an Aufgaben richtig gelöst hat, folgen im erweiterten mathematischen Bereich Aufgaben höheren Schwierigkeitsgrades. Neben Aufgaben mit formalen Rechenzeichen werden zum Beispiel auch Aufgaben zur Flächenidentifikation gestellt oder Zahlenreihen sind fortzusetzen.

Kurzzeitgedächtnis

Die Leistung des Kurzzeitgedächtnisses wird erfasst, indem eine Abfolge farbiger Kreise auf dem Bildschirm dargeboten wird, deren Reihenfolge das Kind anschließend korrekt wiedergeben soll. Bei richtiger Lösung werden die Folgen, die sich das Kind merken muss, immer länger.

Einstellungen

Über die aufgeführten Aufgaben hinaus kann das Kind am Ende des ersten Schuljahres optional einen Fragebogen ausfüllen, der seine Einstellungen gegenüber schulrelevanten Aktivitäten, wie beispielsweise dem Hören von Geschichten, Spielen auf dem Spielplatz oder Zählen abfragt. Dabei antwortet das Kind auf einer dreistufigen Smiley-Skala.

Damit enden die Aufgaben für das Kind. Für die *Hand der Lehrkraft* sind die zwei folgenden optionalen Erweiterungen gedacht:

Verhalten

Am Ende des Schuljahres kann die Lehrkraft zusätzlich einen Fragebogen zum Verhalten der einzelnen Kinder ausfüllen. Dieser Fragebogen enthält alle DSM-IV-Kriterien für eine Aufmerksamkeits-Hyperaktivitäts-Störung. Die Erweiterung wird angeboten, da Zusammenhänge mit der Leistungsentwicklung bestehen (Tymms & Wylde, 2004).

Personale, soziale und emotionale Entwicklung

In einem weiteren Fragebogen kann die Lehrkraft die personale, soziale und emotionale Entwicklung des Kindes beurteilen. Dabei werden Anpassung (Wohlbefinden, Selbstständigkeit), Persönlichkeit (Selbstvertrauen, Konzentration) und soziale Kompetenzen (Beziehung zu Gleichaltrigen, Beziehung zu Erwachsenen, Einhalten von Regeln, kulturelles Bewusstsein, Kommunikation) berücksichtigt.

4.3 Auswertung und Rückmeldungen

Die *Rohdaten der gesamten Klasse und aller Schülerinnen und Schüler* können direkt nach der Erhebung in einer Übersicht aufgerufen werden und stehen für erste Analysen sofort zur Verfügung. Auch die vom Kind stammenden Informationen zum Bereich Einstellungen und die von der Lehrkraft eingegebenen Verhaltensdaten werden in dieser Übersicht dargestellt. Die Ergebnisse der Lehrerbeurteilung der personalen, sozialen und emotionalen Entwicklung lassen sich separat abrufen und werden grafisch veranschaulicht.

Die Rückmeldung der FIPS-Daten erfolgt in unterschiedlicher Form. Die *kognitiven Daten* werden normbezogen als relative Maße in Form von *Normwerten* rückgemeldet, während die *Einstellungen* der einzelnen Schülerinnen und Schüler kriteriumsbezogen lediglich als *Punktzahl* auf einer Dreipunkteskala berichtet und ein Mittelwert über alle Items gebildet wird (Tymms & Albone, 2002).

4.3.1 Kognitive Daten

4.3.1.1 Normwerte

Die Normwerte der kognitiven Daten werden auf den Daten der Normstichprobe basieren, die im Schuljahr 2009/10 erhoben wird. Sie werden als *T-Werte* mit einem Mittelwert von 50 und einer Standardabweichung von 10 ausgedrückt (Tymms & Albone, 2002). Auf eine Skalierung mit einem Mittelwert von 100 und der Standardabweichung 15 wurde bewusst verzichtet, um eine Missinterpretation der Ergebnisse als IQ-Werte zu vermeiden (Tymms & Albone, 2002). Aufgrund der gewählten T-Skalierung liegen die Normwerte von etwa zwei Dritteln der Kinder zwischen 40 und 60. Somit kann aus der Normwerttabelle abgelesen werden, wie sich die Lernvoraussetzungen des Kindes im Vergleich zu den Ergebnissen der Normstichprobe verhalten: Liegt der Normwert über 60, gehört das Kind zu den oberen 16 % der Vergleichsgruppe, liegt er unter 40, gehört es zu den unteren 16 %. Normwerte über 70 und unter 30 sind sehr selten. Nur 2 bis 3 % der Kinder liegen jeweils in diesen Bereichen.

Die Lehrkräfte erhalten als Rückmeldung sowohl ein *genormtes Gesamtergebnis für jeden Schüler* als auch jeweils *genormte Teilergebnisse für die einzelnen Bereiche Lesen, Lautbewusstheit und Mathematik*. Der genormte Gesamtwert berechnet sich dabei nicht aus der Summe der Normwerte für die Teilbereiche. Schüler mit außergewöhnlich hohen

Normwerten in allen drei Bereichen haben einen noch höheren Gesamtnormwert als die Summe der Teilnormwerte. Genauso haben Kinder mit außergewöhnlich niedrigen Normwerten in allen Teilbereichen einen Gesamtnormwert, der noch unterhalb der Summe der Teilnormwerte liegt. Dies kommt dadurch zustande, dass es noch seltener ist, in allen drei Bereichen außergewöhnliche Werte zu erreichen als nur in einem einzigen Bereich.

4.3.1.2 Lernentwicklung

Nach der zweiten Erhebung am Schuljahresende werden neben den Normwerten, die die Kinder am Ende der ersten Klasse erreicht haben, auch regressionsanalytisch ermittelte *Werte des Lernzuwachses* zurückgemeldet. Das heißt, auch die Lernentwicklung der einzelnen Kinder kann mit der anderer Kinder verglichen werden, die bei der Einschulung ähnliche Lernvoraussetzungen hatten. Die sogenannte „Regression zur Mitte“, d. h. die Tendenz außergewöhnlicher Werte, bei einer Wiederholungsmessung näher am Mittelwert zu liegen, wird bei der Berechnung des Lernzuwachses berücksichtigt. Um den Wertzuwachs in einem einzelnen Bereich oder insgesamt zu bestimmen, wird der am Jahresende erzielte Gesamt- bzw. Teilnormwert mit dem Gesamtnormwert vom Jahresanfang verglichen. Es wird immer auf den normierten Gesamtwert und nicht auf die normierten Teilwerte der Jahresanfangserhebung zurück gegriffen, da sich der normierte Gesamtwert als bester Prädiktor für die einzelnen Bereiche und den Gesamtwert am Jahresende erwiesen hat.

4.3.2 Einstellungen

Am Schuljahresende kann auch die Einstellung der Schülerinnen und Schüler gegenüber schulrelevanten Aktivitäten erfasst werden. Für diesen Bereich werden keine Normwerte berechnet. Es wird lediglich *für jedes Kind ein Mittelwert über alle Items* gebildet. Das heißt, es wird berechnet, ob ein Kind im Mittel den schulrelevanten Aktivitäten eher positiv, neutral oder negativ gegenübersteht.

4.3.3 Hinweise auf besonderen Förderbedarf

Erzielen Kinder am Anfang der ersten Klasse *T-Werte über 40* (d.h. bis minus eine Standardabweichung vom T-Wert 50), ist es unwahrscheinlich, dass sie in diesem Aufgabenbereich einen besonderen Förderbedarf haben. Für Kinder mit niedrigeren Werten wird empfohlen, die Lernentwicklung sorgfältig zu beobachten. Vor allem auf Schülerinnen

und Schülern, deren T-Werte zu Beginn der ersten Klasse *unter 30* liegen, sollte besonderes Augenmerk liegen, ebenso auf Kindern, deren Normwerte sich deutlich von den Erwartungen der Lehrkraft unterscheiden. Entscheidungen über einen besonderen Förderbedarf eines einzelnen Kindes in einem oder mehreren Aufgabenbereichen sollten jedoch nie allein aufgrund auffälliger Werte bei der FIPS-Schulanfangsuntersuchung getroffen werden. Die Ergebnisse, die das FIPS-Verfahren liefert, sind nur ein Teil der Informationen über ein Kind. Sie ergänzen die Berichte der Eltern, des Kindergartens, der Klasse oder anderer Quellen und die gezielte Beobachtung der Lehrkraft in strukturierten und offenen Situationen. Am besten erschließt sich besonderer Förderbedarf aus Informationen, die über einen längeren Zeitraum hinweg gesammelt werden. Im Verlauf des Schuljahrs können zudem weitere und gegebenenfalls spezifischere Diagnoseverfahren eingesetzt werden.

Kinder mit sehr hohen Ergebniswerten (*70 und höher*) sind ebenfalls außergewöhnlich. Nur zwei bis drei Prozent der Kinder gehören normalerweise dieser Gruppe an. In einigen Schulen sind solche Ergebnisse nicht so selten, in anderen hingegen können sie als Ausnahme ebenfalls eine größere Herausforderung darstellen. Auch hier geht es hauptsächlich darum, dass die Lehrkraft durch das FIPS-Programm auf die besonders guten Lernvoraussetzungen des Kindes aufmerksam wird und mögliche Unterforderungen vermeiden kann.

4.3.4 Fazit

Die Lehrkräfte können also mit Hilfe der Rückmeldungen, die sie über die FIPS-Ergebnisse erhalten, zum einen erkennen, wie ihre gesamte Klasse im Vergleich zur Normstichprobe zu Beginn und am Ende des ersten Schuljahres steht. Zum anderen wird deutlich, ob die Klasse die auf Grund der Ausgangssituation erwarteten Lernfortschritte gezeigt hat oder dahinter zurückgeblieben ist. Gleiches gilt auch für die Leistungen einzelner Schülerinnen und Schüler, wobei hier nicht nur der Stand und die Entwicklung im Vergleich zur Normstichprobe, sondern auch im Vergleich zu anderen Kindern der eigenen Klasse festgestellt werden kann.

Durch die Berechnung von getrennten Normwerten für die einzelnen Bereiche Lesen, Lautbewusstheit und Mathematik ist es darüber hinaus möglich zu erkennen, ob ein Kind in

einem Bereich besonders gut oder besonders schlecht abschneidet oder ob es beispielsweise in allen Bereichen deutlich über oder unter dem Durchschnitt liegt.

Die Lehrkraft kann somit gleich *zu Beginn der ersten Klasse ihr eigenes diagnostisches Urteil* bezüglich der Lernvoraussetzungen ihrer Schülerinnen und Schüler *mit wissenschaftlich fundierten objektiven und normorientierten Daten abgleichen und ergänzen*. Dadurch erhält sie für ihre Unterrichtsgestaltung und insbesondere die Differenzierungs- und individuellen Fördermaßnahmen wertvolle Informationen. Durch eine zweite Erhebung am Jahresende und eine Rückmeldung über den Lernzuwachs der Kinder kann die Lehrkraft auch die Fortschritte ihrer Schüler sowie den Erfolg ihrer Fördermaßnahmen erkennen.

5. Ausblick

Nach der erfolgreichen Überarbeitung des FIPS-Programms und ersten vielversprechenden Ergebnissen im Rahmen der Pilotphase LEIF 1 stehen nun noch die Ergebnisse der zweiten Pilotphase LEIF 2 aus. Diese sollen weitere Informationen darüber liefern, ob die Lehrkräfte die Rückmeldung über die FIPS-Ergebnisse auf Individualebene für die Unterrichtsgestaltung sinnvoll nutzen können und ob die Schülerinnen und Schüler davon profitieren. Darüber hinaus soll die *schulische Entwicklung* der „FIPS-Kinder“ aus LEIF 1 und 2 *bis zum Ende der dritten Klasse* weiter beobachtet werden.

Im Schuljahr 2009/10 findet im süddeutschen Raum die *Normierung* des im Juli 2009 fertiggestellten revidierten FIPS-Programms statt, um Vergleichswerte für künftige erste Klassen zur Verfügung stellen zu können. Die Schulen werden zum FIPS-Programm ein Manual erhalten, das Normtabellen für die einzelnen Bereiche Lesen, Lautbewusstheit und Mathematik sowie die Gesamtleistung am Schuljahresanfang und -ende enthält. In der Zukunft wird möglicherweise auch eine Online-Auswertung angeboten, bei der die FIPS-Schulen ihre Rohdaten an eine Sammelstelle schicken und die ausgewerteten Ergebnisse zurückgemeldet erhalten. Dies würde eine ständige Aktualisierung der Normwerte ermöglichen.

Die FIPS-Testungen am Schuljahresanfang wurden bisher nur *nach der Einschulung* durchgeführt, und die Lehrkräfte waren durch den mehrtägigen Einsatz von mobilen Reserven vom Unterricht teilweise oder ganz entlastet. Das FIPS-Programm könnte auch

einige Wochen früher *vor der Einschulung* im Rahmen der Kooperation von Kindergarten und Grundschule durchgeführt werden.³ Zwar wäre dies mit zusätzlichem Aufwand für die Lehrkräfte verbunden, aber es entfielen dadurch zumindest am Schulanfang die Abhängigkeit von einer Vertretung im Unterricht. Zudem würde sich dadurch die Möglichkeit ergeben, sich mit den pädagogischen Fachkräften der Kindertagesstätte sowohl über die Lernstände einzelner Kinder als auch über die am Schulanfang nötigen und vorhandenen Vorkenntnisse in den drei Aufgabenbereichen Lesen, Lautbewusstheit und Mathematik auszutauschen. Die gemeinsame Arbeit mit dem FIPS-Programm könnte so zu einem Kernelement der lernbereichsbezogenen Zusammenarbeit von pädagogischen Fachkräften und Lehrerinnen und Lehrern werden und die Anschlussfähigkeit der beiden ersten Bildungsstufen verbessern.

³ Diese Anregung stammt von Frau Prof. Dr. A. Speck-Hamdan, LMU München.

Literatur

- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert (Hrsg.), *Pisa 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69-140). Opladen: Leske und Budrich.
- Barth, K. & Gomm, B. (2004). *Gruppentest zur Früherkennung von Lese- und Rechtschreibschwierigkeiten (PB-LRS)*. München: Reinhardt.
- Bowey, J. A. (2007). Predicting Individual Differences in Learning to Read. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 155-172). Malden: Blackwell.
- Bradley, L. & Bryant, P. (1985). *Rhyme and reason in reading and spelling*. Ann Arbor: The University of Michigan Press.
- Corvacho, I. & Greb, K. (2007). Persönlichkeits- und Lernentwicklung von Grundschulkindern: Zur Anlage des Projekts PERLE. In K. Möller, P. Hanke, C. Beinbrech, K. Hein, T. Kleickmann & R. Schages (Hrsg.), *Qualität von Grundschulunterricht. Entwickeln, erfassen und bewerten*. (Jahrbuch Grundschulforschung; 11, 1. Aufl., S. 313–316). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Füssenich, I. & Löffler, C. (2005). *Schriftspracherwerb. Einschulung, erstes und zweites Schuljahr*. München: E. Reinhardt.
- Grassmann, M. (2001): Kinder wissen viel - 4 Jahre Untersuchungen zu mathematischen Vorkenntnissen von Grundschulkindern. In G. Kaiser (Hrsg.), *Beiträge zum Mathematikunterricht 2001: Vorträge auf der 35. Tagung für Didaktik der Mathematik vom 5. – 9. März in Ludwigsburg* (S. 221-224). Hildesheim: Franzbecker.
- Hanke, P. (2007). *Anfangsunterricht – Leben und Lernen in der Schuleingangsphase* (2. Aufl.). Weinheim: Beltz.
- Hasemann, K. (2003). *Anfangsunterricht Mathematik*. Heidelberg: Spektrum.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität*. Seelze-Velber: Klett.
- Kaufmann, L., Nuerk, H.-C., Graf, M., Delazer, M. & Willmes, K. (2009). *TEDI-MATH: Test zur Erfassung numerisch-rechnerischer Fertigkeiten für 4-8 Jährige*. Zürich: Hans Huber.

- Kauschke, C. & Stan, A. (2004). Semantische Entwicklung am Beispiel kindlicher Fehlbenennungen. *Linguistische Berichte*, 198, 191-219.
- Küspert, P. (1998). *Phonologische Bewußtheit und Schriftspracherwerb: Zu den Effekten vorschulischer Förderung der phonologischen Bewußtheit auf den Erwerb des Lesens und Rechtschreibens*. Frankfurt: Lang.
- Küspert, P. & Schneider, W. (2006). *Hören, lauschen, lernen. Sprachspiele für Kinder im Vorschulalter* (5. Aufl.). Göttingen: Vandenhoeck & Ruprecht.
- Martschinke, S. & Kammermeyer, G. (2003). Jedes Kind ist anders. Jede Klasse ist anders. Ergebnisse aus dem KILIA-Projekt zur Heterogenität im Anfangsunterricht. *Zeitschrift für Erziehungswissenschaft*, 2, 257-275.
- Martschinke, S., Kirschhock, E.-M. & Frank, A. (2001). *Der Rundgang durch Hörhausen. Das Nürnberger Erhebungsverfahren zur phonologischen Bewusstheit*. Donauwörth: Auer.
- Moser, U. & Berweger, S. (2007). *Wortgewandt & zahlenstark*. St. Gallen: Kantonaler Lehrmittelverlag St. Gallen.
- Moser, U., Stamm, M. & Hollenweger, J. (2005). *Für die Schule bereit? Lesen, Wortschatz, Mathematik und soziale Kompetenz beim Schuleintritt*. Aarau: Sauerländer.
- Richter, S. & Brügelmann, H. (1992). Stellenwert schriftsprachlicher Prädiktoren bei der Vorhersage späterer Rechtschreibleistungen. *Psychologie in Erziehung und Unterricht*, 39, 253-263.
- Roth, E. (1999). *Prävention von Lese- und Rechtschreibschwierigkeiten: Evaluation einer vorschulischen Förderung der phonologischen Bewußtheit und der Buchstabenkenntnis*. Frankfurt: Lang.
- Schneider, W., Krajewski, K. & Schwenck, C. (2010). Rechenstörungen: Möglichkeiten der Prävention und Intervention. In W. von Suchodoletz (Hrsg.), *Therapie von Entwicklungsstörungen. Was wirkt wirklich?* (S. 129-152). Göttingen: Hogrefe.
- Schrader, F.-W. (2006). *Diagnostische Kompetenz von Eltern und Lehrern*. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. Aufl., S. 95-100), Weinheim: Beltz.
- Siegmüller, J. (2005). Einflüsse von Frequenz und Erwerbsalter auf das Benennen bei Kindern mit Wortfindungsstörungen. *L.O.G.O.S. interdisziplinär*, 13, 15-20.

- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19, 85-95.
- Tymms, P. (1999). Baseline assessment, value-added and prediction of reading. *Journal of Research in Reading*, 22 (1), 27-36.
- Tymms, P. & Albone, S. (2002). Performance indicators in primary schools. In A. J. Visscher & R. Coe (Eds.), *School improvement through performance feedback* (pp. 191-218). Lisse, Niederlande: Swets & Zeitlinger.
- Tymms, P. & Merrell, C. (2009). On-entry baseline assessment across cultures. In A. Anning, C. Culling & M. Fleers (Eds.), *Early childhood education: Society and culture* (2nd ed., pp. 117-129). London: Sage.
- Tymms, P. & Wylde, M. (2004). Basisprüfverfahren und Dauerbeobachtung in der Grundschule. In G. Faust, M. Götz, H. Hacker & H.-G. Roßbach (Hrsg.), *Anschlussfähige Bildungsprozesse im Elementar- und Primarbereich* (S. 190-203). Bad Heilbrunn: Klinkhardt.
- van Luit, J. E. H., van de Rijt, B. A. M. & Hasemann, K. (2001). *Osnabrücker Test zur Zahlbegriffsentwicklung (OTZ)*. Göttingen: Hogrefe.