

Erfassung des Lernfortschritts durch curriculumbasierte Messung

von Karl Josef Klauer

In diesem Beitrag wird über eine Familie neuerer diagnostischer Verfahren berichtet, die in den Vereinigten Staaten von Amerika in sonderpädagogischem Kontext entwickelt worden ist und die die Förderung des Lernens insbesondere von Kindern mit Lernschwächen substantiell verbessern soll. Insofern gibt es also gewisse Parallelen zur deutschen Förderdiagnostik (z.B. Kornmann, 1982), die allerdings an der Diskrepanz zwischen Wollen und Können gescheitert ist (Schlee, 1985). Ein Hauptvorteil der neuen Tests besteht darin, dass (a) der Lernverlauf über eine ungewöhnlich lange Zeitspanne abgebildet wird, und (b) die Tests streng auf die Lehr-Lern-Ziele ausgerichtet sind. Die Verfahren lassen frühzeitig erkennen, ob der Lehr-Lern-Prozess so fortschreitet wie erwartet oder ob entsprechende Interventionen eingeleitet werden sollten, deren Erfolg sich dann ebenfalls im Lernverlauf abbilden würde.

Schlüsselwörter: Lernfortschrittmessung, Diagnostik des Lernverlaufs

Bildungspolitischer Hintergrund in den USA

Bevor die Details näher dargestellt werden, erscheint es sinnvoll, zunächst einen kurzen Blick auf die bildungspolitische Situation in den USA zu werfen. Entwickelt wurden die Methoden zur Messung des Lernfortschritts unabhängig von den bildungspolitischen Tendenzen in den Vereinigten Staaten, doch wird der Einsatz dieser neuen Verfahren forciert durch die dortigen schulpolitischen Entwicklungen der letzten Jahrzehnte. Die schon ältere *Accountability*-Politik hält die Schulen für den Lernfortschritt der Kinder direkt verantwortlich. Früher waren die Schulen wie bei uns nur verpflichtet, das vorgesehene Bildungsangebot bereitzustellen, und es lag dann an den Kindern (und den Eltern), inwieweit

sie von dem Angebot Gebrauch machten. Inzwischen aber wird von den Schulen gefordert, nicht nur das Lernangebot bereitzustellen, sondern den Lernerfolg der Schülerinnen und Schüler zu garantieren. Die Schulen tragen also Verantwortung für das Lernen der Kinder, und das bleibt für die Schulen nicht folgenlos. Verschärft wird diese Situation durch das *No Child Left Behind*-Gesetz aus dem Jahre 2001. Danach sind die Schulen gefordert, dafür zu sorgen, dass 95% der Schülerinnen und Schüler die innerhalb des jeweiligen Bundesstaates geltenden Normen im Lesen und in der Mathematik erreichen. Bis zum Schuljahr 2013–2014 müssen sogar 100% der Kinder die Norm erfüllen. Schulen, die zweimal hintereinander das Leistungsniveau nicht schaffen, werden als „verbesserungsbedürftig“ eingestuft. Mit diesem Status sind schwerwiegende Sanktionen ver-

bunden, und über Schulen, die für längere Zeit darin verbleiben, hängt sogar das Damoklesschwert der Kündigung des gesamten Lehrpersonals (Deno, 2003a). Besonders stark sind natürlich jene Schulen betroffen, die sich in sozialen Brennpunkten befinden oder einen hohen Anteil von Minoritäten aufweisen (Kim & Sunderman, 2005). Die Stoßrichtung dieser gesetzlichen Vorgaben ist, wie der Name des Gesetzes andeutet, gerade schwache Kinder besonders stark und so zu fördern, dass sie nicht mehr zurückbleiben. Von daher sind Schulen zu beraten, insbesondere den Lernfortschritt leistungsschwacher Kinder systematisch zu überwachen, um nötigenfalls frühzeitig die Förderung zu intensivieren oder besser anzupassen.

Student Progress Monitoring by Curriculum-based Measurement

This article reviews diagnostic tools developed in the United States by a group of researchers in special education. The tools aim to improve both teaching and learning mainly in the field of learning disabilities. There are certain parallels to the German „Förderdiagnostik“ (e.g. Kornmann, 1982) which failed because of a disconnection between goals and realizations (Schlee, 1985). The main advantages of curriculum-based measurement which these tools provide are (a) documentation of the learning process over a long period of time and (b) direct measurement of instructional objectives. This enables discovery of whether or not learning does continue as expected or whether appropriate interventions should be started, the effects of which could also be demonstrated.

Keywords: Student progress monitoring, curriculum-based measurement

Was ist mit curriculumbasierter Messung gemeint?

Der Grundgedanke zur Realisierung der curriculumbasierten oder Lernfortschrittmessung lässt sich wie folgt präzisieren. In regelmäßigen Abständen werden repräsentative Stichproben von Aufgaben zu ein und demselben Lehrziel vorgelegt. Durch die wiederholten Leistungsmessungen ist dokumentierbar, wie sich die Leistung in dem fraglichen Bereich längerfristig entwickelt. Die Messungen dienen der zügigen Feststellung des Lehr-Lern-Erfolgs und können daher frühzeitig signalisieren, wenn der Fortschritt unbefriedigend und daher eine Modifikation der didaktischen Intervention angezeigt ist. Deren Erfolg lässt sich dann ebenfalls rasch ermitteln.

Wie die so genannte Förderdiagnostik dienen Verfahren der Lernfortschrittmessung nicht der Klassifikation von Schülern, nicht der Einteilung von Schülern nach Niveaugruppen, nicht der Differenzierung zwischen Schülern nach ihrer Leistungsfähigkeit oder der Zuweisung einzelner zu einer speziellen Förderung. Sie dienen also weder der Selektion, noch der Klassifikation oder der Platzierung. Es geht auch nicht primär darum festzustellen, ob ein gegebenes Lehrziel erreicht worden ist oder nicht, wie dies bei lehrzielorientierten Tests der Fall ist. Vielmehr geht es um die Erfassung der Leistungsveränderungen eines Schülers oder einer Klasse in der Zeit, um die *Dokumentation des Lernfortschritts im Verlauf der Zeit*. Die Lernfortschrittmessung dient insbesondere dazu, frühzeitig pädagogisch umsteuern zu können, wenn Gefahr droht, dass nicht so gut gelernt wird, wie das sein sollte.

Diese Messverfahren sind *formativ* statt *summativ*. Die Unterscheidung stammt von Scriven (1967), und sie wurde publik gemacht durch das verbreitete Handbuch von Bloom, Hastings und Madaus (1971). Summativ ist eine Evaluation, wenn der Lernerfolg am Ende des Lernprozesses ermittelt werden soll. Typische Beispiele sind Schulleistungstests oder auch lehrzielorientierte Tests. Letztere dienen der Feststellung, ob beziehungs-

weise wie gut ein Lehrziel erreicht worden ist. Formativ ist dagegen eine Evaluation, wenn sie den Lernfortschritt laufend dokumentiert. Hier geht es also um die Erfassung und Abbildung des *Lernverlaufs*. Ein entscheidender Vorteil formativer Messung besteht darin, dass durch Änderung der Instruktion frühzeitig eingegriffen werden kann, sollte der Fortschritt nicht so sein wie erwartet.

Die Messung des Lernfortschritts wurde an der Universität von Minnesota in Minneapolis von Stanley Deno und einigen seiner Doktoranden in sonderpädagogischem Kontext seit 1972 entwickelt. Verfahren dieser Art beschränkten sich ursprünglich auf die Erfassung von Fortschritten behinderter Kinder im Lesen, Schreiben und Rechnen. Inzwischen wird diese Diagnostik auch im Primarschul- wie im Sekundarschulbereich eingesetzt und natürlich auch in anderen Fächern (Deno, 1985; 2003a; 2003b). Einen guten Überblick über die Entwicklung bietet Fuchs (2004). Für diese Methoden zur Lernfortschrittmessung hat Deno die Bezeichnung „*curriculum-based measurement*“ (CBM) eingeführt. Der Name soll einerseits darauf verweisen, dass die Beherrschung der im *aktuellen Unterricht tatsächlich* vermittelten Inhalte erfasst werden soll (daher „*curriculum-based*“), andererseits aber auch zum Ausdruck bringen, dass es um wissenschaftlich begründete Messverfahren im testpsychologischen Sinne geht und nicht um irgendwelche Beurteilungs- oder Einschätzverfahren (daher „*measurement*“ statt „*assessment*“).

Zur Technik der Lernfortschrittmessung

Das Vorgehen erinnert an verhaltensanalytische Techniken, wie sie etwa in der Einzelfall-Forschung in Gebrauch sind (vgl. Barlow & Hersen, 1984; Julius, Schlosser & Goetze, 2000; Petermann, 1992; Wember, 1994; 1998). In festgelegten Zeitabständen werden Verhaltensstichproben gezogen zu jeweils derselben Leistung. Konkret werden in bestimmten Abständen Stichproben von Aufgaben zur Bearbeitung vorgelegt, die zwar immer neu sein müssen, aber prin-

zipiell die gleiche Leistung fordern. Die Aufgaben müssen durch dieselben Operationen lösbar und insofern auch gleich schwer sein. Außerdem sollen die einzelnen Aufgabenstichproben nur wenig Zeit in Anspruch nehmen, schon damit die Leistungsmessung sehr häufig wiederholt werden kann und sie dann nicht zuviel Zeit wegnimmt, die dem Unterricht abgibt. Daher müssen die Aufgaben auch relativ leicht sein und sich ebenso rasch wie objektiv auswerten lassen.

Im Zentrum der Lernfortschrittmessung steht die Entwicklung von Aufgabenstichproben, die jeweils erfassen, wie sich die Leistung bei ein und demselben Lehrziel entwickelt. Wie Fuchs (2004) darlegte, gibt es hierzu zwei Wege.

Der erste Weg

Man sucht einen Typ von Aufgaben, der für die geforderte Leistung repräsentativ ist, also möglichst alle die Teiloperationen miterfasst, die am Jahresende gekannt sein sollen. Stichproben solcher Aufgaben müssen robust mit anderen relevanten Leistungen korrelieren, wenn sie geeignet sein sollen. Für das Lesen hat man beispielsweise sehr gute Erfahrungen gemacht mit kurzen Passagen gleich schwerer Texte, die man eine Minute lang laut lesen lässt. Angenommen, man hat einen längeren Text – etwa ein Buch – zur Hand, das sprachlich dem Leistungsniveau des Kindes oder der Klasse entspricht. Man schlägt dann irgendwo auf und lässt das (oder jedes) Kind für genau eine Minute laut lesen, um zu registrieren, wie viele Wörter es in der Zeit richtig und wie viele es falsch liest. Die Anzahl richtig gelesener Wörter pro Minute (WpM) korreliert sehr gut mit den Ergebnissen von Lesetests, insbesondere auch mit dem Leseverständnis, obgleich dies hier nicht direkt geprüft wird. Die Flüssigkeit des lauten Lesens ist also ein geeigneter Indikator für die Leseleistung insgesamt (Deno, Mirkin & Chiang, 1982).

Weitere Möglichkeiten demonstrierten Fuchs und Fuchs (1993). Diese Autoren dokumentierten die Leistungen von über

3000 Grundschulkindern von Klasse 1 bis Klasse 6 im Lesen, in der Rechtschreibung und in der Mathematik, und zwar zum Teil jede Woche, zum Teil auch nur monatlich. Zur Erfassung der Leseleistung setzten die Autoren zwei Verfahren ein, (a) das laute Lesen innerhalb einer Minute und (b) einen Lückentext, der am PC offeriert wurde. Das laute Lesen wurde im Einzelversuch geprüft, wobei ein immer anderer, aber in der Schwierigkeit vergleichbarer Text zugrunde gelegt wurde. Im zweiten Versuchsjahr wurde der PC eingesetzt. Hier erhielt jedes Kind für 2,5 Minuten einen 400 Wörter langen Text vorgelesen. Dabei handelte es sich ebenfalls um immer neue Texte, die aber demselben Buch entnommen waren. Nur der erste Satz des Textes wurde lückenlos angeboten. Aber ab dem zweiten Satz war jedes siebte Wort gestrichen, und es standen drei Auswahlantworten zur Verfügung. Die Auswahlantworten waren möglichst gleich lang. Ansonsten waren die beiden Distraktoren aber weder optisch noch akustisch dem Zielwort ähnlich. Die Kinder brauchten nur zwei Tasten zu bedienen, die Leertaste und die Enter-taste. Eine jüngere Variante des Lückentests sieht so aus (Shin, Deno & Espin, 2000, 166).

My mother always likes to go home. She was born on a nice (farm/ big/ soon) in a valley. Her father started (home/ the/ sat) farm before she was born. When (red/she/ told) was a little girl they lived (to/ fun/ in) a very old log house on (call/ date/ the) farm. One day her father said (that/ little/ size) they were going to build a (try/ new/ with) house. That made my mother, her (before/ longer/ sister) and her brothers very excited. ...

Der Fortschritt in der Rechtschreibung wird oft mittels Wortdiktionen erfasst. Die Lehrkraft diktiert beispielsweise 20 Wörter im Abstand von 7 Sekunden, so dass das Diktat nach zwei Minuten und 20 Sekunden beendet ist. Die Wortlisten wurden in der erwählten Arbeit von Fuchs und Fuchs (1993) einer Art Grundwortschatz jeweils per Zufall entnommen, allerdings in zwei Varianten. In der ersten Variante wurden nur solche

Wörter diktiert, die im Grundwortschatz dem jeweiligen Schuljahr zugeordnet waren. Die zweite Variante bestand aus Stichproben, die unabhängig von der Zuordnung zu Schuljahren gezogen wurden und daher insbesondere für die jüngeren Kinder sehr viel schwerer waren.

Der zweite Weg

Bei diesem Vorgehen stellt man zunächst fest, welche Leistungen in einem Fach am Ende des Schuljahres erreicht werden sollen, um dann in einem detaillierten Plan festzulegen, um welche Teilmengen von Aufgaben es sich handelt. Im zweiten Schritt ist zu entscheiden, wie viele Aufgaben für jede der Teilmengen eingesetzt werden. Das Verfahren entspricht der Erstellung von kontextvaliden Tests, einer Technik, die im Arbeitskreis des Verfassers etwas differenzierter entwickelt worden ist (Klauer, 1987, 12–57). Im Mathematikunterricht der Grundschule seien beispielsweise ab der zweiten Klasse Aufgaben zur Addition, Subtraktion (beide mit und ohne Überschreitung der Zehner), zur Multiplikation und Division gefordert, aber auch das Rechnen mit Größen. Sind die Teilmengen der geforderten Aufgaben genau definiert, so kann man für die Stichproben anschließend fixieren, mit wie vielen Aufgaben jede der Teilmengen in jedem Test vertreten sein soll. Auf diese Weise erhält man lehrzielvalide Stichproben. Das Vorgehen sei an Tabelle 1 erläutert.

Der Test von Tabelle 1 realisiert das Prinzip, das Fuchs, Fuchs, Hamlett und Stecker (1991) schon eingesetzt hatten. Aus Copyrightgründen wurde hier nur analog verfahren, wobei die mathematischen „Diagnose- und Förderblätter 2“ für das zweite Schuljahr verwendet wurden (Klauer, 1994). Die Teilziele und damit auch die Teilmengen von Aufgaben sind dort – etwas vereinfacht – so definiert:

- Addition 1: Vier Aufgabenformen ($a + b = ?$, $a + ? = c$, $a + ? = c$; Zehner; $a + ? = 100$) ohne Überschreiten der Zehner.
- Addition 2: Die gleichen Aufgabenformen mit Überschreiten eines Zehners.

- Subtraktion 1: Zwei Aufgabenformen ($c - b = a$, $c - ? = a$) ohne Unterschreiten eines Zehners.
- Subtraktion 2: Die gleichen Aufgabenformen mit Unterschreiten eines Zehners.
- Multiplikation: Drei Aufgabenformen ($a \cdot b = c$, $a \cdot ? = c$, $c = ? \cdot b$).
- Division mit nur einer Aufgabenform ($c : b = ?$).
- Rechnen mit Größen: Vier Standardformen Addition, Subtraktion, Multiplikation, Division kombiniert mit einer von acht Größen.

Somit resultieren sieben Teilmengen, aus denen per Zufall jeweils vier Aufgaben für Tabelle 1 gebildet worden sind. Ersetzt man die Zahlen durch entsprechende andere, so erhält man weitere Paralleltests. Das ergibt dann jedes Mal 28 Testaufgaben nur mit jeweils anderen Zahlen, die jedoch dieselben Bedingungen erfüllen. Daher fordern solcherart konstruierte Tests stets die gleich Leistung, Fuchs, Fuchs, Hamlett und Stecker (1991) hatten auf ähnliche Weise einen Mathematiktest für die fünfte Klasse mit 25 Aufgaben entwickelt, und nach Ersetzung der Zahlen durch entsprechende andere konnten sie so viele Paralleltests herstellen, wie sie brauchten.

Dokumentation des Lernfortschritts

Tests dieser Art sind also relativ schnell durchzuführen. Für den Test von Tabelle 1 wird man höchstens 10–15 Minuten vorsehen, zumal die Kinder viele der Aufgaben erst im Laufe der Zeit zu lösen lernen. Wird nun regelmäßig – täglich oder wöchentlich – ein solcher Test gegeben mit immer anderen, aber vergleichbar schweren Aufgaben, so lässt sich die Leistungsentwicklung anschaulich darstellen.

Abbildung 1 zeigt die Entwicklung im Lesen eines Schülers über ein ganzes Schuljahr, genauer über 36 Wochen. Der Schüler las einen immer anderen Lesebuchtext für genau eine Minute, wobei die Anzahl richtig und falsch gelesener Wörter festgehalten wurde. Das geschah in der Regel einmal pro Woche. Zu-

Tab. 1: Mathematiktest für das zweite Schuljahr

Addition 1	23 + 5 =	41 + _ = 78	57 + 3 =	66 + _ = 100
Addition 2	85 + 8 =	58 + _ = 82	25 + 18 =	79 + _ = 100
Subtraktion 1	17 - 6 =	88 - 55 =	34 - _ = 31	98 - _ = 54
Subtraktion 2	43 - 26 =	91 - 12 =	63 - _ = 47	35 - _ = 9
Multiplikation	4 · 6 =	7 · 8 =	5 · _ = 35	40 = _ · 10
Division	18 : 3 =	42 : 6 =	35 : 7 =	64 : 8 =
Größen	30€ + 24€ =	76 cm - 17 cm =	7 · 4 Wochen =	36 Stück : 4 =

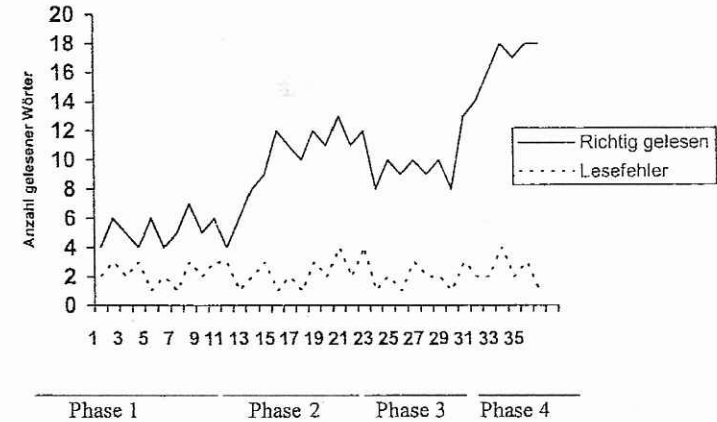


Abb. 1: Leistungsentwicklung eines Kindes über 36 Wochen und vier Phasen (in Anlehnung an Deno, 2003b; vgl. auch Wember, 1999, S. 121)

nächst fällt auf, dass sich die Lesefehler kaum verändern, sondern auf einem relativ niedrigen Niveau bewegen. Anders dagegen die Anzahl richtig gelesener Wörter. Bis zu den ersten zwölf Messungen tut sich da wenig. Erst danach setzt eine deutliche Besserung ein, die sich mit Unterbrechungen bis zum Ende des Schuljahres fortsetzt.

Interessant sind die vier Phasen, die unten markiert sind. Jeder dieser Abschnitte ist durch ein bestimmtes didaktisches Vorgehen der Lehrerin gekennzeichnet.

Die ersten zwölf Wochen gibt sie ihren regulären Unterricht, der aber keine erkennbaren Fortschritte bringt. Sie versucht es dann in einem längeren Abschnitt mit individuell zugeschnittenen Hilfen, auf die das Kind auch deutlich positiv reagiert. Allerdings muss sie ab der 23. Woche erkennen, dass ein deutlicher Rückschlag erfolgt. Nun versucht es die Lehrerin mit einer neuen Kombination von Lernhilfen – und diesmal klappt es. Die Leseleistung steigt nun kontinuierlich an.

Abbildung 1 verdeutlicht eine besonders interessante Einsatzmöglichkeit der Messung des Lernfortschritts, nämlich die Feststellung, wie ein Kind (oder eine ganze Klasse) auf eine bestimmte Fördermaßnahme reagiert, ob es darauf anspricht oder nicht. Vielfach versuchen Lehrkräfte zwar, mit einem neuen Anlauf das Lernen der Kinder zu verbessern, aber wenn der Fortschritt der Lernenden nicht wirklich objektiv dokumentiert wird, so kann man nicht sicher sein, eine gute Entscheidung gefällt zu haben. Möglicherweise hätte die Lehrerin, deren

Lehrerfolg in Abbildung 1 wiedergegeben ist, nicht wiederholt versucht, ihr Vorgehen zu modifizieren, wenn sie keine Rückmeldung gehabt hätte.

In einem späteren Artikel schlugen Fuchs und Fuchs (1998) sogar vor, die Lernfortschrittsmessung zur Diagnostik von Kindern mit einer *Learning Disability* (nicht gleich definiert wie unsere Lernbehinderung) in einem vierschriftigen Verfahren vorzunehmen, das in Abbildung 2 illustriert ist.

- In Phase 1 und 2 wird geprüft, ob das fragliche Kind tatsächlich weniger leistet als die anderen und, falls ja, ob die anderen Kinder in der Klasse denn besser gefördert werden. Beide Bedingungen sind gemäß Phase 1 erfüllt.
- In Phase 3 sollen nun die Möglichkeiten der Förderung innerhalb der regulären Klasse verbessert werden. Im Beispiel von Abbildung 2 wird zu Beginn von Phase 3 ein Motivierungsvertrag geschlossen, der zwar Belohnungen in Aussicht stellt, aber keine erkennbaren Verbesserungen bringt. Im Rahmen dieser Phase bietet die Lehrkraft dem Kind leichtere Lesetexte, die jedoch auch keine Wirkungen zeigen.
- In Phase 4 testet ein Sonderpädagoge, ob das Kind auf sonderpädagogische Förderung positiv anspricht oder nicht. Im konkreten Fall konzentrierte er sich auf zwei spezielle Fehlerarten des Kindes, das offensichtlich

darauf gut auspricht. Immerhin hat es am Ende das Leseniveau erreicht, das die Klassenkameraden im Durchschnitt schon erheblich früher geschafft hatten. Jedenfalls wurde so nachgewiesen, dass das Kind unter der speziellen sonderpädagogischen Förderung wesentlich weiter kommt.

Für den Mathematikunterricht lassen sich solche Leistungsverläufe ebenfalls darstellen, und zwar für einzelne Kinder wie für die Mittelwerte ganzer Klassen. Man zählt dann bei jeder Testung die Anzahl richtig gelöster Aufgaben und erhält analoge Verläufe wie die dargestellten. Die Autoren der curriculumbasierten Messung ziehen es allerdings vor auszuzählen, wie viele Ziffern der Lösung richtig hingeschrieben wurden. Sie wollen auf diese Weise Teilleistungen honorieren, eine zweifellos lobenswerte Absicht, die in der Umsetzung jedoch nicht unproblematisch ist (vgl. Michelsen & Cordes, 2005).

Ein anderer Modus der Lernerfolgsmessung bietet sich an, wenn man über Ausgangs- und Zieldaten verfügt, beispielsweise über das Leistungsniveau, das zu Beginn des Schuljahres vorhanden ist, und über das Leistungsniveau, das am Ende des Schuljahres vorhanden sein soll. In Abbildung 3 geht es um die Leseleistung eines Kindes, das zu Schuljahresbeginn deutlich unterhalb des Leistungsniveaus der Klasse liegt. Die mittlere Leistungsentwicklung der Klasse ist bekannt dank früherer Erhebungen in

sonst vergleichbaren Klasse. Nach Abbildung 3 kann man nun ermitteln, wie der Leistungsfortschritt von Monat zu Monat sein müsste, damit das Kind eine Chance hat, auf das mittlere Leistungsniveau der Klasse aufzuschließen. Man kann also quantitative Zwischenziele festlegen, die es gestatten, frühzeitig zu erkennen, ob man mit der Förderung des Kindes auf dem rechten Weg ist.

Das Beispiel von Abbildung 3 zeigt die mittleren Leseleistungen eines Kindes, wobei die wöchentlichen Werte der Übersichtlichkeit wegen für jeden Monat gemittelt wurden. Die Abbildung verdeutlicht zweierlei: (1) Dass der Lernfortschritt des lesechwachen Kindes bei diesem Ziel deutlich größer sein müsste als der seiner lesetüchtigeren Mitschüler und (2) wie sich die Leistungen des Kindes in den ersten fünf Monaten tatsächlich entwickeln. Das erste Vierteljahr verläuft erfreulich im Sinne des remedialen Lernens. Allerdings bringen die letzten acht Wochen, also der vierte und fünfte Monat keinerlei Fortschritt, und das ist ein ernst zu nehmendes Signal, um sich Gedanken darüber zu machen, ob die Förderung noch angemessen ist oder ob andere Einflüsse die Leistungsentwicklung beeinträchtigen. In jedem Fall müsste der Anstieg des Leistungsniveaus in den verbleibenden Monaten wesentlich *steiler* werden, um die Vorgabe noch erreichen zu können.

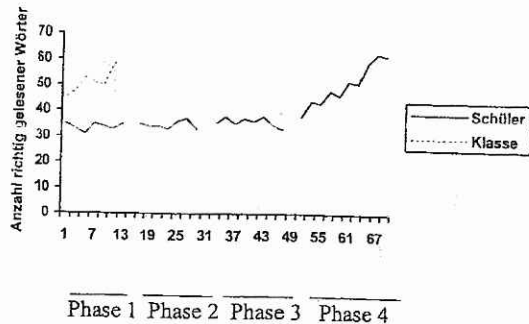


Abb. 2: Datenerhebung zur Diagnostik einer *Learning Disability* im Lesen (in Anlehnung an Fuchs & Fuchs, 1998)

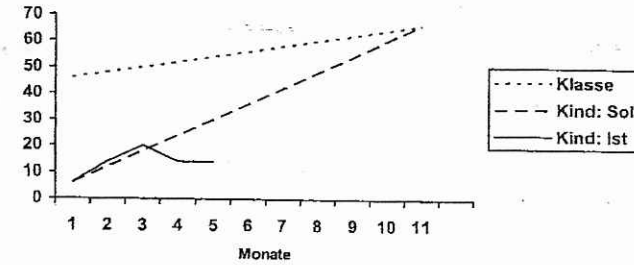


Abb. 3: Wie sich ein deutlich unterdurchschnittlich leistendes Kind verbessern müsste, um am Schuljahresende den erwarteten Leistungsstand der Klasse einstellen zu können und wie sich seine Leistungen in den ersten fünf Monaten tatsächlich verändern

	Sep	Okt	Nov	Dez	Jan	Feb	Mär	Apr	Mai	Jun	Jul
Addition 1	□	■	■	■	■	■	■	■	■	■	■
Addition 2	□	■	■	■	■	■	■	■	■	■	■
Subtraktion 1	□	□	■	■	■	■	■	■	■	■	■
Subtraktion 2	□	■	■	■	■	■	■	■	■	■	■
Multiplikation	□	□	■	■	■	■	■	■	■	■	■
Division	□	□	■	■	■	■	■	■	■	■	■
Größen	□	■	■	■	■	■	■	■	■	■	■

(□ keine, ■ 1, ■ 2, ■ 3, ■ 4 Aufgaben gelöst)

Abb. 4: Leistungsentwicklung eines Kindes im zweiten Schuljahr bei den sieben Teilmengen von Matheaufgaben in Anlehnung an Fuchs et al. (1991)

Im Fall der Mathematik und in manchen anderen Fällen hat man es gemäß Weg 2 mit definierten, klar unterscheidbaren Teilmengen zu tun. Abbildung 4 zeigt die Leistungsentwicklung eines Kindes in Mathematik im Laufe des zweiten Schuljahres, allerdings in einer anderen

Darstellung. Hier steht die Entwicklung bei den sieben Teilfähigkeiten von Tabelle 1 im Vordergrund. Je stärker ein Feld geschwärzt ist, desto mehr der Aufgaben hat das Kind am jeweiligen Termin richtig gelöst. Man sieht, dass die Addition natürlicher Zahlen ganz gut

beherrscht wird, und das gilt auch für die Subtraktion ohne Unterschreiten, die Multiplikation und das Rechnen mit Größen. Dagegen bestehen noch deutliche Unsicherheiten bei der Subtraktion, wenn Zehner zu unterschreiten sind, und bei der Division. Eine Darstellung wie in

Abbildung 4 informiert also über die Leistungsentwicklung bei den einzelnen *Teilleistungen*. Hieraus lassen sich Schlüsse ziehen, in welchen Bereichen besonders gearbeitet werden muss. Auf dieser Grundlage kann man gezielte Interventionen planen an Stelle undifferenzierter und gleichmäßiger Übung aller Rechenoperationen.

In einer umfangreichen Untersuchung in den dritten, vierten und fünften Klassen einer Grundschule konnten VanDerHeyden und Burns (2005) mit ähnlichen Verfahren nachweisen, wie sich die Rechenleistungen von Januar bis April substantiell und mit beachtlichen Effekstärken verbesserten.

Lernkurven und Steigungskoeffizienten

Wie oben bereits erwähnt dokumentierten Fuchs und Fuchs (1993) die Leistungen in Mathematik, im Lesen und in der Rechtschreibung von über 3 000 Grundschulkindern der Klassen 1 bis Klasse 6. Dabei konnten sie für diese drei Bereiche *durchschnittliche Anstiege* pro Woche oder Monat ermitteln, die im Einzelfall zur Beurteilung zugrunde gelegt werden können: Bleibt ein Kind deutlich unter dem durchschnittlichen mittleren Anstieg, so ist das ein Warnsignal und Veranlassung, der Sache nachzugehen. Mittlere Leistungsanstiege informieren also darüber, welche Zuwächse in einem gegebenen Zeitintervall zu erwarten sind.

Für jedes Kind wurden von Fuchs und Fuchs (1993) die aktuellen Daten dokumentiert, darüber hinaus aber auch die linearen Regressionen und die Regressionskoeffizienten errechnet, Verfahrensweisen, die mit statistischen Programmen zu den Routineoperationen gehören. In der Lernerfolgsmessung verknüpft die Regressionslinie den jeweiligen Leistungsstand mit dem Zeitpunkt der Messung. Im linearen Fall ist die Steigung an jeder Stelle der Zeitachse gleich groß, und Lernende können gemäß ihren Steigungskoeffizienten verglichen werden. Je steiler der Anstieg, desto mehr hat ein Kind in der Zeiteinheit gelernt (s. Abb. 5).

Nun hat sich herausgestellt, dass nicht die Daten aller Kinder optimal durch eine *gerade* Regressionslinie repräsentiert werden. Die Leistungsentwicklung von bis zu einem Drittel der Kinder wurde bei Fuchs und Fuchs (1993) besser durch eine Kurve wie in Abbildung 6 repräsentiert. Statistisch gesehen enthält dieser Typ neben der linearen eine quadratische Komponente, die zu einer negativen Beschleunigung führt. Bei einer negativen Beschleunigung werden die Zuwächse immer kleiner: Anfangs gibt es relativ starke Fortschritte, aber danach schwächt sich der Trend ab, die Leistung verbessert sich immer weniger, was ja auch aus Abbildung 6 hervorgeht. In diesem Fall wird noch weiter geübt, auch wenn kein sichtbarer Erfolg mehr erkennbar ist. Man spricht dann von *Overlearning*, das die Reaktionssicherheit und -schnelligkeit deutlich erhöht und lernschwachen Kindern häufig auch Freude macht.

Testgütekriterien

Die Tests zur Lernfortschrittsmessung lassen sich mit einem hohen Maß von *Objektivität* auswerten. In der Regel geht es um Entscheidungen über richtig und falsch bei einfachen Aufgaben. Die Interrater-Übereinstimmung liegt deutlich über 90% der Aufgaben. Beispielsweise wurden in einer Untersuchung im Kindergarten Lernfortschrittsmessungen durchgeführt, um die Lesereife der Kinder zu testen. Bei sechs Testverfahren mit 45 bis 68 Paralleltests und 15 bis 18 Auswertern fand man Übereinstimmungen der Beurteiler in 98 bis 100% der Fälle (VanDerHeyden, Wir, Naquin & Noell, 2001).

Die Tests zeichnen sich ferner durch eine hohe *Reliabilität* aus. Bei den so vielfach wiederholten Testungen liegt es nahe, Retestreliabilitäten zu berechnen. Streng genommen handelt es sich dabei sogar um Retests mit Paralleltests statt mit identischen Tests, und dennoch korrelieren deren Ergebnisse bemerkenswert hoch miteinander. Deno et al. (1982) fanden bei leistungstüchtigen wie bei Kindern mit *Learning Disabilities* Retestreliabilitäten in einem Zeitraum von

fünf Wochen, die im Bereich von $0,73 \leq r_{tt} \leq 0,91$ lagen, wobei die meisten der Koeffizienten größer als 0,80 waren. Vergleichbare Werte fand auch Marston (1989). Bei Kindergartenkindern brachten sechs Testverfahren Paralleltest-Reliabilitäten zwischen 0,58 und 0,84, wobei nur zwei kleiner als 0,70 ausfielen (VanDerHeyden et al., 2001).

Die Validität der Verfahren erweist sich als besonders erfreulich. Für das einmündige laute Lesen beispielsweise wurden mehrfach hohe Korrelationen mit den Ergebnissen standardisierter Lesetests gefunden. In einer differenzierten Studie konnten Riley-Heller, Kelly-Vance und Shriver (2005) darüber hinaus zeigen, dass es keinen signifikanten Unterschied macht, ob die Kinder Texte aus dem Lesebuch der Klasse oder aus standardisierten Tests laut lesen. Die Ergebnisse der wiederholten Tests stehen ferner in engem Zusammenhang mit dem Alter und dem Schuljahr der Kinder und korrelieren hoch mit der Lehrernote für das Lesen (Good & Jefferson, 1998). Analoges gilt für andere Maße des Lernfortschritts, insbesondere auch in der Mathematik (Fuchs, Fuchs & Maxwell, 1988; Good, Simmons & Kameenui, 2001). Die Korrelation zwischen solchen Maßen und standardisierten Schulleistungstests liegen zwischen 0,65 und 0,85 (Deno et al., 2002; Good et al., 2001; Stecker & Fuchs, 2000; vgl. auch VanDerHeyden et al., 2001). Die Validität der wiederholten Leistungsstichproben, wie sie für die Lernfortschrittsmessung typisch sind, ist demnach mehr als befriedigend.

Insgesamt wurde eine relativ große Zahl empirischer Studien durchgeführt, um die Qualität der Lernfortschrittsmessung anhand ihrer Testgütekriterien einschätzen zu können. Alles in allem kann man feststellen, dass sich die Verfahren unter teststatistischen Gesichtspunkten bewähren und daher ernst zu nehmen sind.

Einsatzmöglichkeiten

Deno (2003a) listet eine stattliche Reihe von vierzehn Möglichkeiten auf, zu welchen Zwecken solche Messverfahren eingesetzt werden können. Die wichtigs-

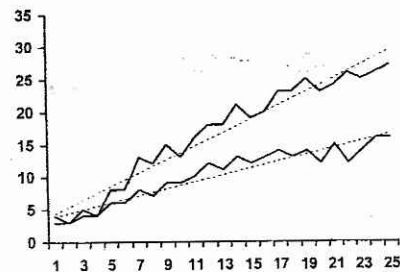


Abb. 5: Unterschiedlich steile Leistungsentwicklung zweier Schüler vom gleichen Ausgangspunkt aus

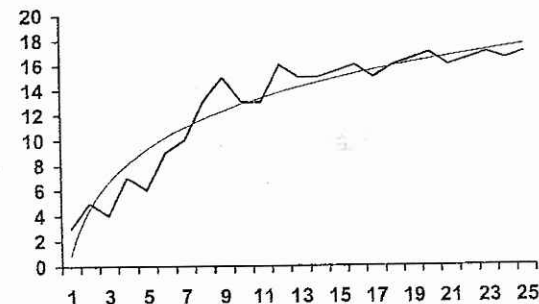


Abb. 6: Kurvilineare Leistungsentwicklung über 26 Wochen (logarithmischer Trend)

ten seien hier kurz vorgestellt.

- *Verbesserung der Instruktionsverfahren für einzelne Lernende oder ganze Klassen.* Die Lernfortschrittsmessung macht darauf aufmerksam, wenn der Prozess zu entgleisen droht. Bei komplexen Lehrzielen zeigt sie überdies, bei welchen Teilzielen besondere Schwächen verblieben sind, wo also anzusetzen ist.
- *Vorhersage der Leistung bei wichtigen Kriterien.* Da die Ergebnisse der Lernfortschrittsmessung hoch mit Kriterien wie etwa Zeugnisnoten, standardisierten Tests usw. korrelieren, eignen sie sich zur Vorhersage späteren Lernens. Das gilt auch für die Einschätzung des Erfolgs einer Klasse bei überregionalen Leistungs-

vergleichen, die in Zukunft auch bei uns eine größere Rolle spielen werden.

- *Entwicklung von Lernnormen.* Wie oben gezeigt wurde, lassen sich mit Hilfe der Verfahren quantitative Normen gewinnen, die Lernende einhalten müssen, wenn ein längerfristiges Ziel, wie etwa die Erreichung des mittleren Leistungsniveaus der Klasse geschaffert werden soll.
- *Erleichterung der Kommunikation mit den Lernenden und den Eltern.* Während die Ergebnisse eines standardisierten Intelligenz- oder Schulleistungstests schwer zu vermitteln sind, ist das Ergebnis der Lernfortschrittsmessung, wie es etwa in den Abbildungen 1 bis 6 deutlich wird, leicht verständlich und wohl jeder-

mann einsichtig.

- *Früherkennung von Kindern, die Gefahr laufen, das Ziel nicht zu erreichen.* Hier geht es insbesondere darum, prophylaktisch einem sonderpädagogischen Förderbedarf entgegen zu wirken, oder, falls dies nicht zum Erfolg führt, für alle Seiten frustrierende weitere Versuche abzuschließen und sonderpädagogische Förderung in Anspruch zu nehmen.

Vielleicht sind nicht alle diese Einsatzmöglichkeiten gleich wichtig. Es dürfte aber deutlich geworden sein, dass mit der curriculumbasierten Messung pädagogisch bedeutsame Informationen gewonnen werden, die insbesondere für remediales Lehren und Lernen bedeutsam sein dürften.

Diskussion

Zur Namensgebung

Die Bezeichnung „*curriculum-based measurement*“, die von Deno vor dem Hintergrund der amerikanischen Diskussionen eingeführt wurde und nach wie vor präferiert wird, lässt sich gut ins Deutsche als *curriculumbasierte Messung* übertragen. Allerdings klingt das relativ blass, und es löst auch keine klare Vorstellung darüber aus, was eigentlich gemeint sei. Denos Erläuterung, der Ausdruck „*curriculumbasiert*“ solle die Verfahren abheben von überregionalen Schulleistungstests, die sich nicht auf das im Unterricht hier und heute Behandelte beziehen, trifft schon länger nicht mehr zu. Beispielsweise konnten Riley-Heller et al. (2005) zeigen, dass es keinen Unterschied macht, ob man die Kinder Klassentexte lesen lässt oder Texte aus standardisierten Lesetests verwendet. Und wenn man Tests gemäß dem Weg 2 erstellt, so werden zwangsläufig Inhalte im Test erscheinen, die im Unterricht noch nicht behandelt worden sind. Im Zusammenhang mit den Tests nach Tabelle 1 wurde das deutlich.

Allerdings ist auch die Bezeichnung „*Student Progress Monitoring*“ in Gebrauch (vgl. zum Beispiel www.student-progress.org/ oder www.progress-monitoring.org/). Diese Bezeichnung lässt sich etwa mit Lernfortschrittsmessung übertragen, um nicht von Überwachung (*monitoring*) zu sprechen. Immerhin scheint der Ausdruck den Kern besser zu treffen. Schließlich geht es um die Dokumentation des Lernfortschritts, wie er sich im Laufe der Zeit darstellt. Einer der Hauptvorteile des Verfahrens besteht doch darin, frühzeitig erkennen zu können, wenn erwartungswidrig wenig gelernt wird. Deswegen ist vielleicht besser von der Lernfortschrittsmessung oder der Lernverlaufsdiagnostik die Rede.

Ziehung von Aufgabenstichproben

Für den Weg 1 von Fuchs (2004) haben sich bislang hauptsächlich das laute Lesen, das Erlesen von Lückentexten und Rechtschreibtests mittels Wortlisten

durchgesetzt. Vielleicht kommen noch einige andere Möglichkeiten hinzu wie etwa die Erfassung bestimmter Rechenfertigkeiten, doch dürfte der Schwerpunkt der Lernfortschrittsmessung auf Weg 2 liegen. Zentral dafür sind Stichproben von Aufgaben, die repräsentativ sind für das, worum es im konkreten Fall geht. Hier scheint manches doch verbesserungsfähig.

Repräsentative Stichproben lassen sich nur dann ziehen, wenn man zuvor weiß, wofür sie repräsentativ sein sollen. Es muss also die Grundmenge genau definiert sein, aus der repräsentative Stichproben zu ziehen sind. Voraussetzung der Lernverlaufsdiagnostik ist demnach die Definition von Lehrzielen durch Grundmengen von Aufgaben, und daran scheint es mitunter bei den amerikanischen Kollegen zu mangeln. Auf diesem Feld haben die CBM-Protagonisten noch Nachholbedarf.

Wie Grundmengen von Aufgaben definiert werden, wurde am Beispiel lehrzielorientierter Tests differenziert erarbeitet, und ebenso, wie aus solchen Mengen repräsentative Stichproben mit Hilfe von Zufallsprozeduren zu ziehen sind (Klauer, 1987). Faktisch werden repräsentative Aufgabenstichproben in der Regel nicht gezogen, sondern erzeugt, hergestellt, und auch hierfür existieren klare Vorgaben. Auf diese Weise erhält man Aufgabenstichproben, die definitionsgemäß lehrzielvalide beziehungsweise kontextvalide sind. Sind die Stichproben wirklich zufällig gezogen oder erzeugt, so kann man gleich schwere Stichproben unterstellen. Manche Schwächen, die dem Konzept der Lernfortschrittsmessung noch anhaften, lassen sich dem Umstand zuschreiben, dass die Aufgabengrundmengen nicht immer richtig definiert und die Stichproben folglich nicht immer repräsentativ sind.

Einzelne Lehrziele sind eng begrenzt, andere Lehrziele sind umfassender, etwa wenn man den gesamten mathematischen Lehrstoff einer Klasse in Betracht zieht, wie dies im Zusammenhang mit Tabelle 1 und Abbildung 4 erläutert wurde. Der Unterschied zwischen Weg 1 und Weg 2 dürfte wesentlich auf die Art der Lehr-

Lern-Ziele zurückzuführen sein. Bei eng begrenzten Lehr-Lern-Zielen handelt es sich in aller Regel um bestimmte *Fertigkeiten*, die sich nach Weg 1 behandeln lassen. Umfassendere Lehr-Lern-Ziele sind insofern komplexer, als sie mehrere solcher Fertigkeiten oder Fertigkeiten und Wissensziele repräsentieren.

Ein wesentlicher Unterschied in der Lernverlaufsdiagnostik von *Fertigkeiten* und *komplexeren* Zielen sollte nicht unbeachtet bleiben. Beide zeichnen sich durch spezifische Vor- und Nachteile aus. Lernfortschrittsmessung bei komplexen Lehr-Lern-Zielen setzt voraus, dass man in dieser Diagnostik schon Aufgaben stellen muss, die noch nicht alle bearbeitet werden können, weil sie noch nicht im Unterricht behandelt worden sind, sondern erst für später im Schuljahr auf dem Plan stehen. Das ist klar ein Nachteil. Man wird dann den Kindern sagen müssen, dass sie noch nicht alle Aufgaben lösen können, dass sie manche erst später kennen lernen werden. Diesen Nachteil muss man in Kauf nehmen, wenn man den Lernverlauf etwa über ein ganzes Schuljahr dokumentieren will. Man handelt sich aber auch einen wichtigen Vorteil ein: Nur auf diese Weise ist es möglich, den Verlauf des Lernfortschritts insgesamt auf einer einheitlichen Skala darstellen zu können.

Bei *eng umgrenzten* Fertigkeiten oder Teilzielen dagegen (wie etwa bei der Groß- und Kleinschreibung oder beim Wissen, welcher Abfall in die gelbe Tonne gehört und welcher nicht) stellt sich als Vorteil dar, dass man Kindern keine Aufgaben vorzulegen braucht, von denen man schon weiß, dass sie die nicht werden beantworten können. Allerdings handelt man sich dabei einen anderen Nachteil ein: Wie sich die Vermittlung der verschiedenen Fertigkeiten und Teilziele im ganzen Schuljahr entwickelt, lässt sich nicht auf einer einheitlichen Skala darstellen. Stattdessen erhält man mehrere unzusammenhängende, nicht miteinander vergleichbare Testergebnisse für jedes der Teilziele und ihrer Entwicklung.

Zusammenfassende Wertung

Als wesentlich neu sind zwei Aspekte zu beachten, die langfristige Verlaufskontrolle und die strikte Orientierung an Lehr-Lern-Zielen. Die langfristige Perspektive wird nur durch die regelmäßigen Testerhebungen in kurzen Abständen gewährleistet, wobei im strengen Sinne Paralleltests einzusetzen sind. Das ist das wirklich Neue, denn es gibt kein anderes Verfahren, das die Möglichkeit bietet, über punktuelle Testerhebungen hinaus langfristig die Leistungsentwicklung abzubilden. Unter pädagogischem Aspekt ist ein solches Vorgehen besonders vorteilhaft. Hinzu kommt, dass die einzelnen Tests der Idee nach jeweils lehrzielvalide sein müssen, dass also jeder einzelne der vielen Tests das fragliche Lehr-Lern-Ziel gültig zu repräsentieren hat. Handelt es sich um einzelne Fertigkeiten, so dürfte es nicht allzu schwierig sein, diese Forderung zu erfüllen. Aber selbst für den Fall komplexer Lehrziele, die sich aus Teilzielen zusammensetzen, stehen geeignete Technologien zur Verfügung (Klauer, 1987).

Trotz erfreulicher Reliabilitäten und Validitäten der Tests muss man natürlich mit Messfehlern rechnen. Nicht jeder „Zacken“ in einer Kurve wird die wahren Verhältnisse widerspiegeln, nicht wenige dürften schlicht auf Messfehlern beruhen – und natürlich weiß man nicht, welche

Literatur

- Barlow, D. H. & Hersen, M. (1984). *Single case experimental designs. Strategies for studying behavior change*. New York: Pergamon Press.
- Bloom, B. S., Hastings, J. B. & Madaus, G. G. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw Hill.
- Deno, S. L. (1985). *Curriculum-based measurement: The emerging alternative*. *Exceptional Children*, 52, 219–232.
- Deno, S. L. (2003a). *Developments in curriculum-based measurement*. *The Journal of Special Education*, 37, 184–192.
- Deno, S. L. (2003b). *Curriculum-based measures. Development and perspectives*. *Assessment for Effective Intervention*, 28(3–4), 3–12.
- Deno, S. L., Mirkin, P. K. & Chiang, B. (1982). *Identifying a valid measure of reading*. *Exceptional Children*, 49,

das im einzelnen sind. Darüber hinaus ist natürlich auch bei Kindern mit echten Tagesschwankungen zu rechnen, also mit Einbrüchen oder besonders guten Leistungen, die am nächsten Tag schon verschwunden sein können. Tagesschwankungen wie Messfehlern begegnet man am besten durch häufigeres Testen, und das ist in der Lernfortschrittsmessung ja nicht nur möglich, sondern ohnedies erwünscht. Auch diese Überlegungen sprechen dafür, die zeitlichen Abstände zwischen den Testungen möglichst kurz zu halten. Darüber hinaus dürfte es sich anbieten, weniger stark auf kurzfristige Schwankungen als auf längerfristige Trends zu achten.

Bei vielen psychologischen Tests hat man das Problem, dass Testwiederholungen selbst einen Effekt haben, der seinerseits das Messergebnis verfälscht. Wiederholt man beispielsweise einen Intelligenztest, so erzielt man bei der Zweitestung in aller Regel bessere Ergebnisse, nicht weil die Probanden intelligenter geworden wären, sondern weil der Test bekannt geworden ist. Diese Schwierigkeit stellt sich hier nicht: Erstens stellt jede Testvariante neue Aufgaben, und wenn dabei ein Lern- oder Übungserfolg stattfindet, so ist der pädagogisch erwünschte und testpsychologisch unschädlich.

Schließlich sei noch auf einen weiteren

Gesichtspunkt verwiesen. Man darf nicht übersehen, dass die Lernfortschrittskurven zwar Verläufe darstellen, dass daraus aber nicht auf die Ursachen geschlossen werden kann, die zu den Verlaufsänderungen führen. Wenn etwa wie in Abbildung 2 nach einer geänderten Intervention ein deutlicher und nachhaltiger Lernerfolg zu verzeichnen ist, so kann man keineswegs immer sicher sein, dass er speziell dieser Intervention zu danken ist. Im konkreten Fall der Abbildung 2 übernahm ein Sonderpädagoge die Förderung mit einem speziell auf gewisse Lesefehler ausgerichteten Lesetraining: Es ist natürlich denkbar, dass es diese Art von Förderung war, die die Wende brachte. Vielleicht lag es aber auch nur am Lehrerwechsel und der besseren Lehrer-Kind-Interaktion. Man kann also nicht sicher sein, was nun das wirksame Agens war, weil kein experimenteller Versuchsplan zugrunde lag. Es gibt zweifellos geeignete Versuchspläne für die Einzelfallforschung, wie dies eingangs deutlich wurde. Wenn es aber um die Förderung von Kindern und nicht um Forschung geht, so braucht sich die Lehrkraft auch nicht um den Nachweis von Kausalzusammenhängen zu kümmern. Wichtiger ist für sie festzustellen, ob die Lernförderung wirksam war oder nicht, und das lässt sich mit der Lernfortschrittsmessung sehr gut darstellen.

Kene Kasalzusammenfassung

- (1991). The role of skills analysis in curriculum-based measurement in math. *School Psychology Review*, 19, 6–22.
- Fuchs, L. S., Fuchs, D. & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education*, 9, 20–28.
- Good, R. & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In: M. R. Shinn (Ed.). *Advanced Applications of Curriculum-based Measurement* (pp. 61–88). New York: Guilford Press.
- Good, R. H. III, Simmons, D. C. & Kameenui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high stakes outcomes. *Scientific Study of Reading*, 5(3), 257–288.
- Julius, H., Schlosser, R. W. & Goetze, H. (2000). *Kontrollierte Einzelfallstudien*. Göttingen: Hogrefe.
- Kirn, J. S. & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, 34(8), 3–13.
- Klauer, K. J. (1987). *Kriteriumsorientierte Tests*. Göttingen: Hogrefe.
- Klauer, K. J. (1994). *Diagnose- und Förderblätter 2. Reihenfertigkeiten 2. Schuljahr*. Berlin: Cornelsen.
- Kormann, R. (1982). Von der Auslesediagnostik zur Förderdiagnostik: Entwicklungen, Konzepte, Probleme. *Behindertenpädagogik*, 21, 293–301.
- Marston, D. (1989). A curriculum-based approach to assessing academic performance: What it is and why do it. In: M. R. Shinn (Ed.). *Curriculum-Based Measurement. Assessing Special Children* (pp. 19–78). New York: Guilford Press.
- Michelsen, U. A. & Cordes, M. (2005). Ein praktikables Verfahren zur Bestimmung der Wissensquote und der Ratequote bei Mehrfachwahlaufgaben mit Bestantwort. *Diagnostica*, 51, 156–166.
- Petermann, F. (1992). *Einzelfalldiagnose und klinische Praxis*. München: Quintessenz.
- Riley-Heller, N., Kelly-Vance, L. & Shriver, M. (2005). Curriculum-based measurement: Generic vs. curriculum-dependent probes. *Journal of Applied School Psychology*, 21, 141–162.
- Schlee, J. (1985). Kann Diagnostik beim Fördern helfen? *Anmerkungen zu den Ansprüchen der Förderdiagnostik*. *Zeitschrift für Heilpädagogik*, 36, 153–165.
- Scriven, M. (1967). The methodology of evaluation. In: R. Tyler, R. Gagné & M. Scriven (Eds.). *Perspectives of curriculum evaluation*. AERA Monograph No. 1. Chicago: Rand McNally.
- Shin, J., Deno, S. L. & Espin, C. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *The Journal of Special Education*, 34, 164–172.
- Stecker, P. M. & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research and Practice*, 15, 128–134.
- VanDerHeyden, A. M. & Burns, M. K. (2005). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention*, 30 (3), 15–31.
- VanDerHeyden, A. M., Witt, J. C., Naquin, G. & Noell, G. (2001). The reliability and validity of curriculum-based measurement readiness probes for kindergarten students. *School Psychology Review*, 30, 363–382.
- Wember, F. B. (1994). Möglichkeiten und Grenzen der empirischen Evaluation sonderpädagogischer Interventionen in quasi-experimentellen Einzelfallstudien. *Heilpädagogische Forschung*, 20, 99–117.
- Wember, F. B. (1998). Die Bedeutung der Einzelfallforschung bei der Evaluation gemeinsamen Unterrichts. In: M. Greisbach, U. Kullik & E. Souvignier (Hrsg.). *Von der Lernbehindertenpädagogik zur Praxis schulischer Lernförderung* (S. 231–240). Lengerich: Pabst.

Relevante Internetadressen:

www.studentprogress.org

www.progressmonitoring.org

Stichwörter im Glossar:

- Accountability-Politik
- Lernfortschrittsmessung
- Lehrzielorientierte Tests
- Testgütekriterien

Anschrift des Autors:

Prof. Dr. K. J. Klauer

Robert-Stolz-Weg 15

D-42781 Haan

E-mail: klauerk@uniduesseldorf.de