

Diagnostik der Lese-Rechtschreibstörung

von Wolfgang Deimel

Zusammenfassung

- Die Beschäftigung mit Schwierigkeiten beim Erlernen der Schriftsprache findet in mehreren Disziplinen - Pädagogik, Psychologie und Medizin – statt. Erwartungsgemäß kommt es dabei zu unterschiedlichen Sichtweisen. Die medizinische (und zum Teil die psychologische) Sichtweise ist dadurch gekennzeichnet, dass hier weniger der Prozess des Lesen- und Schreibenlernens im Mittelpunkt steht. Die Situation ist vielmehr, dass Kinder als Patienten vorgestellt werden, die in der Schule das Lesen und / oder Schreiben nicht hinreichend erlernen konnten. Hieraus erklärt sich der diagnostische Ansatz, d.h. es wird der Ist-Zustand mit geeigneten Mitteln erfasst, und es werden dem entsprechend Behandlungsmaßnahmen vorgeschlagen. Dieser Ansatz ist in gewisser Weise komplementär zu dem „pädagogischen“ Ansatz, wo eher der Prozess des Lernens im Fokus steht. Eine Überlappung zwischen medizinischer und pädagogischer Forschung findet leider zu wenig statt.
- Die Erfassung der Rechtschreibleistung birgt - unabhängig vom Ansatz - eine Reihe methodischer Probleme. Zunächst handelt es sich bei Rechtschreibleistung um ein Kontinuum. Eine Dichotomisierung in z.B. auffällig vs. unauffällig, förderungsbedürftig vs. nicht förderungsbedürftig etc. ist willkürlich und – zumindest bei Kindern - nicht durch Außenkriterien zu validieren. Während es bei Erwachsenen durchaus diskutierbare Ansätze für Außenkriterien gibt (z.B. erreicht durchschnittliches Niveau eines Grundschulers, oder ist in der Lage, bestimmte Gebrauchstexte zu erlesen), fehlt ein solcher Ansatz völlig für Kinder, da deren Rechtschreibleistung sich ja ständig verändert, und der Bezug auf die Klassennorm das einzig objektive Kriterium darstellt. Daraus resultierend gibt es keine allgemein akzeptierte Möglichkeit, eine Prävalenz für Kinder festzustellen.
- Die Diagnosevorschriften der Manuale ICD-10 und DSM-IV sind durch eine Reihe von Diskrepanzen geprägt: Diskrepanz zum IQ, Diskrepanz zwischen Rechtschreibleistung und Lesen, Diskrepanz zur Beschulung und Diskrepanz zu anderen Schulleistungen. Hier gelten bezüglich der Kategorisierung die gleichen Überlegungen, da es sich hier auch um Kontinua handelt. Zudem wären hier Überlegungen zur Reliabilität solcher multiplen Entscheidungen nötig, die es zur Zeit noch nicht mal im Ansatz gibt.
- Die Diskussion darüber, ob für die Diagnose der IQ berücksichtigt werden sollte, nimmt vom Umfang her eine Sonderstellung ein. Aufgrund uneinheitlicher Subgruppendefinitionen und unterschiedlicher untersuchter Variablen lässt sich nur schwer eine klare Linie finden, jedoch spricht eine Mehrheit der Befunde dafür, dass sich ALS (Allgemeine Leistungsschwäche) und SLS (Spezifische Leistungsschwäche) weder hinsichtlich der Ätiologie noch hinsichtlich der Therapierbarkeit bedeutsam unterscheiden. Wohlgermerkt gibt es auch anderslautende Befunde, so dass die Frage zur Zeit als noch nicht entschieden zu betrachten ist. Einige relevante Konzepte wurden auch noch nicht untersucht, zum Beispiel, ob sich die Gruppen hinsichtlich des Trainingserfolgs bei kognitiv anspruchsvollen Verfahren, wie einem Regeltraining, unterscheiden.
- Entscheidend für die Qualität der testpsychologischen Diagnostik sind die Normen der verwendeten Tests. Es ist aus dem klinischen Alltag bekannt, dass Tests rasch veralten, was dazu führt, dass Rechtschreibung und Lesen unterschätzt, und Intelligenz überschätzt wird. Eine Liste aktueller Rechtschreib- und Lesetests findet sich hier: <http://www.kjp.uni-marburg.de/kjp/legast/leg/diagnose.htm>.
- Die Diagnostik von Lese-Rechtschreibschwierigkeiten sollte interdisziplinär durchgeführt werden (Psychologen, Ärzte). Gerade in den frühen Grundschulklassen sollte auch Information der Lehrerinnen über den Stand des Kindes relativ zur Klasse einbezogen werden. Rechtschreib- und Lesetests sind einigermaßen zuverlässig erst ab Ende der zweiten Klasse zu verwenden. Die

meistens verwendete Kategorisierung in auffällig vs. unauffällig sollte sorgfältig abgewogen werden, und statistische Normwerte sollten gerade bei grenzwertigen Ergebnissen nicht sklavisch befolgt werden.

1 Einführung

Es waren Ärzte, die das Phänomen des gestörten Schriftspracherwerbs Ende des 19. Jahrhunderts als erste beschrieben. Einige Namen in diesem Zusammenhang sind Berkhan, Morgan und Hinselwood (Übersicht bei Warnke, 1990). Stephenson (1907) fasste die damals vorhandenen Kasuistiken zusammen und definierte die „Kongenitale Wortblindheit“ als ein Syndrom, welches bereits das heute in der Medizin verwendete Konzept einer Teilleistungsstörung vorwegnahm. Insbesondere wurde herausgestellt, dass einige Menschen das Lesen und Schreiben *trotz* mindestens normaler Begabung nicht angemessen zu lernen vermochten. Damit war der Versuch verbunden, Subgruppen zu definieren; so nahm zum Beispiel Ranschburg 1928 an, dass ‚legasthenische‘ Kinder aufgrund einer speziellen „physiologischen oder auch pathologischen Unreife“ (1928, S. 90) besondere Schwierigkeiten beim Erlernen des Lesens und Schreibens haben, und er grenzte diese Verursachung klar von einer allgemeinen intellektuellen Beeinträchtigung ab.

In den Siebziger Jahren des letzten Jahrhunderts kam von pädagogischer Seite erheblicher Widerspruch gegen dieses Konzept (s. dazu Dummer-Smoch 2001). Man plädierte dafür, bei der Förderung lediglich das Schreib- bzw. Leseniveau zu berücksichtigen. Förderung sollte auf keinen Fall den überdurchschnittlich intelligenten Kindern aus ‚günstigen‘ Elternhäusern vorbehalten bleiben (Dummer-Smoch 2001). Diese Diskussion fand ihren Niederschlag in den Empfehlungen der Kultusministerkonferenz von 1978, in denen nicht mehr von Legasthenie oder Rechtschreibschwäche die Rede war, sondern nur noch von „Schülern mit besonderen Schwierigkeiten beim Erlernen des Lesens und Rechtschreibens“.

Diese Diskussion hält unvermindert an. Auf der einen Seite ist die Lese-Rechtschreibstörung als Krankheit von der WHO anerkannt und kann per ICD-10 (Dilling, Mombour und Schmidt 1991) verschlüsselt werden, auf der anderen Seite wird eine Legasthenietherapie in Deutschland i.d.R. nicht von den Krankenkassen finanziert. Auf der einen Seite gibt es einen großen Markt zur Therapie zahlreicher vermuteter Teilleistungsstörungen, auf der anderen Seite heißt es von pädagogischer Seite „Wir halten LRS in erster Linie für ein pädagogisches Problem“ (Naegele & Valtin 2001, S. 10).

Aktuell werden von pädagogischer Seite zwei Hauptargumente gegen das medizinische (IQ-diskrepanzdefinierte) „Legastheniekonstrukt“ (Scheerer-Neumann 2001) ins Feld geführt. Zum einen wird aus der Tatsache, dass es zur Ursachenforschung durchaus widersprüchliche Befunde gibt, und aus den Befunden zur Ursachenforschung bisher keine wirksame Therapie abgeleitet werden konnte, rigoros geschlossen, dass „alle Annahmen des ‚klassischen‘ kausalen Legasthenie-Konzepts als falsifiziert gelten“ (Valtin 2001, S. 16); zum anderen wird auf Befunde verwiesen, die zeigen, dass sich am IQ gebildete Subgruppen weder therapeutisch noch ätiologisch unterscheiden, und dies als Argument gegen die Verwendung einer IQ-Diskrepanz verstanden. Zum ersten Punkt kann angemerkt werden, dass es zwar widersprüchliche Befunde gibt, dass aber insgesamt die Belege für eine genetische Mitverursachung, sowie für auditive und visuelle Wahrnehmungsstörungen mittlerweile ein beeindruckendes Ausmaß erreicht haben (Schulte-Körne 2001). Die *ausschließliche* Betrachtung von LRS als „pädagogisches Problem“ ist daher auf jeden Fall unangemessen.

Zu dem anderen Punkt, nämlich ob die Berücksichtigung des IQ zur Diagnose einen Sinn ergibt, liegt eine umfangreiche Literatur vor, auf die ich weiter unten eingehen werde.

Sofern in den folgenden Abschnitten englischsprachige Literatur zitiert wird, so bezieht sich diese in der Regel auf Untersuchungen der Leseleistung, wogegen in der deutschsprachigen Literatur vorwiegend Rechtschreibleistung oder eine Kombination aus Lese- und Rechtschreibleistung untersucht wurde und wird. Theoretische Überlegungen zur Messung werde ich anhand von Recht-

schreibleistung erläutern, analog können diese Argumente aber auch auf Leseleistung angewendet werden.

2 Was wird gemessen?

Bei der Messung von Rechtschreibleistung gibt es konzeptuell erfreulicherweise kaum Dissens¹. Die inhaltlich valide und nachvollziehbare Operationalisierung von Rechtschreibleistung ist die Zahl der Richtigschreibungen bezogen auf die Gesamtzahl der Wörter in einem Text, üblicherweise in Form eines Lückendiktats. Das sieht so aus, dass jeweils Sätze diktiert werden, in denen ein Wort fehlt, welches dann verschriftlicht werden muss.

Rechtschreibleistung ist ein kontinuierliches Maß, d.h. die individuelle Ausprägung dieser Variable lässt sich auf einem Kontinuum von „sehr schlecht“ bis „sehr gut“ darstellen. Statistisch handelt es sich dabei um eine Absolutskala, da sie einen natürlichen Nullpunkt besitzt.

Bei der Bewertung von Rechtschreibleistung kommt man aber normalerweise um eine Dichotomisierung nicht herum, da die Messung ja bestimmten Zwecken dient. In der Regel geht es darum, zu entscheiden, ob es bei einer bestimmten individuellen Rechtschreibleistung einen Handlungsbedarf im Sinne einer Förderung gibt. Dieses methodische Problem stellt sich schulübergreifend, auch wenn die resultierende Kategorisierung entweder Förderbedarf oder Diagnose genannt wird. Auch bei epidemiologischer Fragestellung („Wieviel Prozent der Bevölkerung sind rechtschreibgestört?“), also der Ermittlung einer Prävalenzrate, steht man vor der Aufgabe, das vorliegende Kontinuum zu dichotomisieren.

Hierzu muss man zunächst einen Blick auf die Form der Verteilung werfen. Es scheint zunächst plausibel anzunehmen, dass Rechtschreibleistung, wie zahlreiche andere natürliche Maße auch, normalverteilt ist.

Oft werden in Handbüchern von Rechtschreibtests die Verteilungen der Rohwerte mitgeteilt, so dass man sich praktisch einen guten Eindruck über die realen Verteilungen verschaffen kann.

Diese Verteilungen unterscheiden sich in erster Linie dadurch, ob bzw. wie stark die Verteilung abgeschnitten ist, da der Test insgesamt zu leicht ist. Dieses Phänomen entsteht dadurch, dass bei der Testkonstruktion darauf geachtet wird, dass der Test im unteren Leistungsbereich gut differenziert. Ein extremes Beispiel dafür ist der GRT4+ (Birkel 1990, s. Abb. 1).

¹ Es ist hier von der quantitativen Rechtschreibleistung die Rede, d.h. eine einzige Maßzahl beschreibt die individuelle Rechtschreibleistung. Ansätze zur differenzierten Erfassung von Rechtschreibleistung mit Hilfe von Fehlerkategorien sind zahlreich, aber nicht empirisch belegt. Auch die sogenannten „klassischen Legasthenerfehler“, wie z.B. Verwechslungen von b und d, entbehren jeglicher empirischen Fundierung.

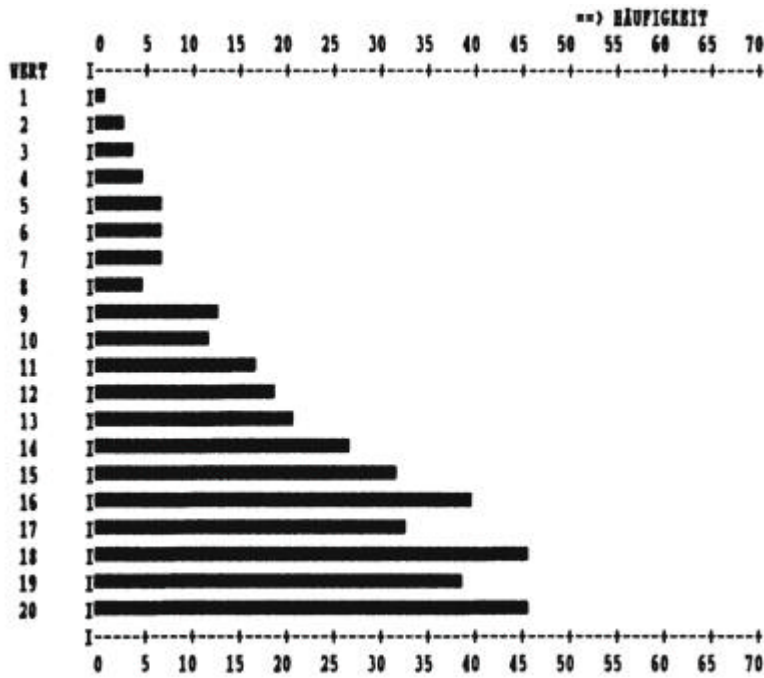


Abb. 1: Rohwertverteilung des GRT4+, Kurzform A (aus Birkel 1990)

Dieser Test dürfte in der Tat vollkommen ungeeignet sein, um Leistungen in der oberen Hälfte des zu erwartenden Kontinuums zu differenzieren. Wie man sieht, gibt es am unteren Rand einen kleinen Buckel, so als wäre eine kleine Verteilung mit sehr niedrigem Mittelwert der anderen Verteilung aufmoduliert. Dies könnte man eventuell im Sinne der „Hump“-These verstehen (s.u.), jedoch findet sich dieses Phänomen nur beim GRT4+; bei allen anderen der gebräuchlichen Tests, bei denen Rohwertverteilungen mitgeteilt werden, gibt es dieses Phänomen nicht.

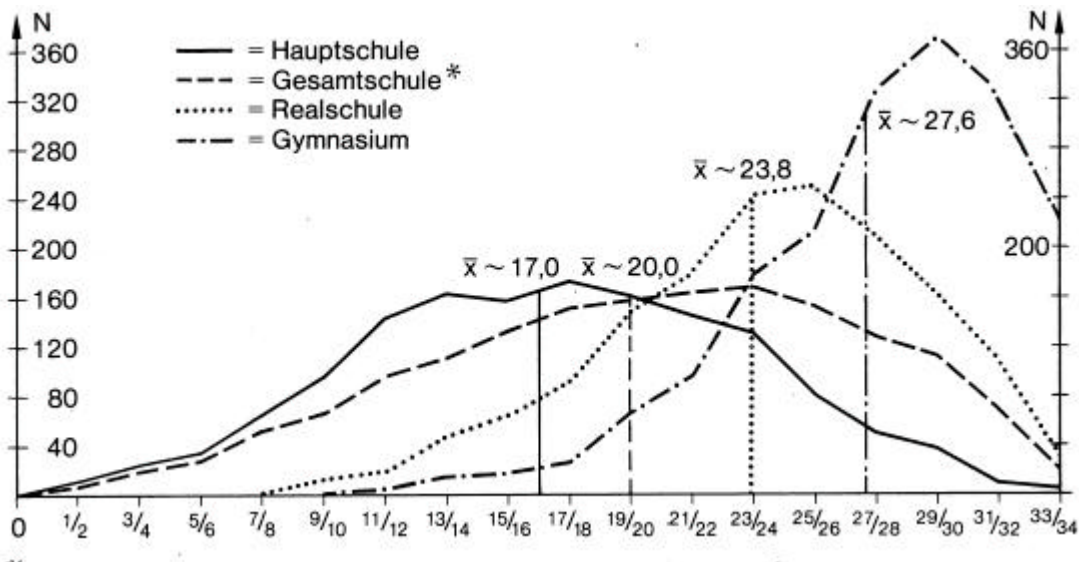


Abb. 2: Rohwertverteilung des WRT4/5 in Abhängigkeit von der Schulform (aus Rathenow 1979)

Da viele der Tests mehr als nur einen Messzeitpunkt und/oder mehrere Schulformen mit dem gleichen Material abdecken, ist oft der Grad der Schiefe der Verteilungen davon abhängig. Dies illustriert eine Abbildung aus dem Handbuch des WRT4/5 (Abb. 2, Rathenow 1979). Während die Verteilung der Hauptschüler symmetrisch ist, ist die Verteilung der Gymnasiasten deutlich abgeknippt.

Man kann also aufgrund der empirischen Befunde davon ausgehen, dass die Verteilungen der Rechtschreib-Rohwerte eingipfelig sind.

Zur Erstellung von Testnormen wird dann jeweils eine Flächentransformation durchgeführt, d.h. die vorliegenden Daten werden per Transformation in die genaue Form einer Normalverteilung gezwungen. Dieses Vorgehen erscheint gerechtfertigt, man sollte sich lediglich darüber im Klaren sein, dass bei manchen Tests die Differenzierung im oberen Leistungsbereich faktisch nicht gegeben ist.

Für die Leseleistung wurde vermutet, dass es am unteren Ende der Verteilung eine signifikante Abweichung von dieser angenommenen Normalverteilung in Form eines Buckels gibt. (Isle-of-Wight-Studie, Rutter & Yule 1975). Diese Buckel-Theorie (engl.: hump) wurde für die Leseleistung aufgestellt (Rutter & Yule 1975), und ließ sich empirisch nicht bestätigen (Shaywitz et al. 1992, Shaywitz, Fletcher und Shaywitz 1995, Rodgers 1983, Aaron 1997)².

Dieses Ergebnis ist von Bedeutung, da die Existenz eines Buckels implizieren würde, dass es schlechte Leseleistung als qualitativ abgrenzbares Syndrom gibt (s.a. Shaywitz, Fletcher und Shaywitz 1994). Da das nicht so zu sein scheint, haben wir es mit einer Größe zu tun, die ab einer nicht a-priori festliegenden Ausprägung als auffällig gelten kann.³

Die Frage nach dem kritischen Wert, bei dessen Unter- oder Überschreitung eine zu behandelnde Störung vorliegt, ist also weder statistisch noch inhaltlich zu beantworten. Wenn Leseleistung sich in Form einer Normalverteilung abbildet, so gibt es keinen a-priori begründbaren Grenzwert („Cut-off point“). Man kommt also nicht mehr über den Grenzwert zur Prävalenz, sondern es ist gerade umgekehrt: man muss erst eine Prävalenz festlegen, und gelangt dann zum Grenzwert. Analoges gilt für die Rechtschreibleistung. Dieses Problem wäre nur durch Außenkriterien zu umgehen, was aber für Kinder kaum möglich ist. Ein denkbare Außenkriterium für Erwachsene ist z.B., ob sie das Lese- und Schreibniveau von durchschnittlichen Grundschulern erreichen (s.a. Strehlow und Haffner 2002).

3 Die Diskussion um den IQ

Bereits 1895 postulierte der schottische Augenarzt Hinshelwood zwei Untergruppen bei den schlechten Lesern – solche, die in anderen kognitiven Bereichen gut abschnitten und also nur im Lesen versagten (Wortblindheit), und solche, für die das nicht galt.

Dieses Phänomen des spezifischen Versagens wurde als „under-achievement“ bezeichnet (Thorndike 1963). „Under-achiever“ sind also Menschen, die es trotz geeigneter Beschulung nicht schaffen, eine IQ-entsprechende Rechtschreibung zu entwickeln. Man spricht hier auch von der „erwartungswidrigen Minderleistung“.⁴ Der IQ wird also bei diesem Ansatz als das Maß für das Potential betrachtet, das ein Mensch für die Entwicklung von Fertigkeiten wie Lesen und Schreiben hat (Siegel 1992). Es handelt sich konzeptuell beim IQ aber nicht etwa um eine definierte Obergrenze für die Ausbildung von Fertigkeiten, wie oft fälschlicherweise kritisiert wurde.

Die statistische Grundlage dieses Modells ist die Korrelation zwischen Fähigkeit (IQ) und Fertigkeit (Rechtschreibleistung). Mittels Regression kann man die Rechtschreibleistung auf Grund des IQ vorhersagen (Evans 1990).

2 Es gab dazu auch eine statistisch-methodische Diskussion, auf die hier nicht näher eingegangen werden soll. S.a. Van der Wissel & Zegers 1985, sowie die Leserbriefdiskussion über den Artikel von Shaywitz et al. 1992 im New England Journal of Medicine 327/4.

3 Eine analoge Größe aus dem medizinischen Bereich ist Blutdruck – es gibt Richtwerte der WHO, aber eine ‚natürliche‘ Grenze für zu behandelnden Hochdruck gibt es nicht.

4 „Erwartungswidrig“ bezieht sich hier auf die Leistung des Individuums. Bezogen auf eine Gruppe ist es natürlich erwartungskonform, dass es eine bestimmte Anzahl von Personen mit erwartungswidriger Minderleistung gibt. Die fehlende Unterscheidung zwischen diesen beiden Aspekten führte oft zu der Behauptung, es gebe bei nicht-perfekten Zusammenhängen keine erwartungswidrigen Abweichungen (z.B. Marx et al. 2001).

Da die Korrelation zwischen IQ und Rechtschreibleistung nicht 1 ist, gibt es in der Praxis Abweichungen von dieser Vorhersage: manche Menschen schreiben besser, als es auf Grund des IQ zu erwarten ist, manche schreiben schlechter.

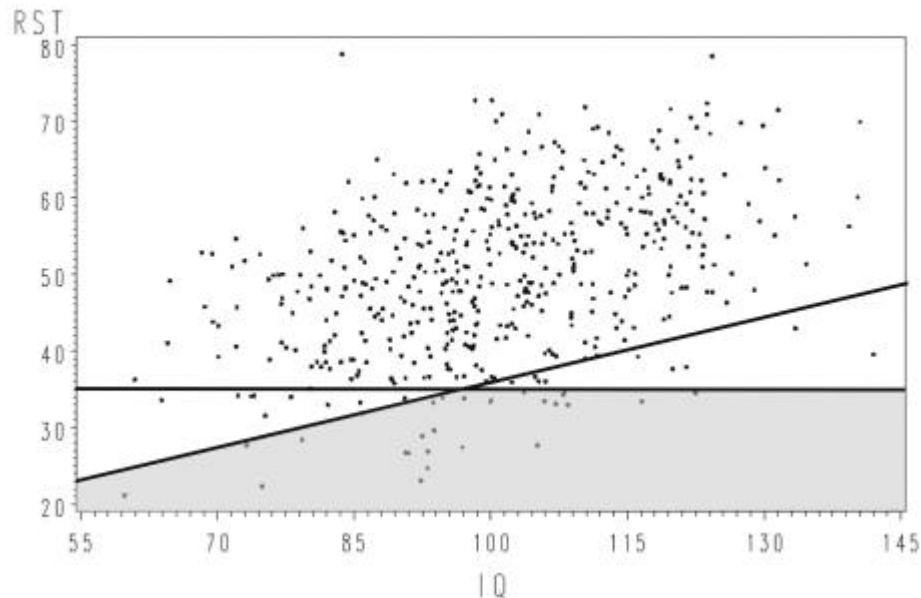


Abb. 3: Zufalls-Punktvolke zweier Variablen mit $r=0,4$. Die waagerechte Linie markiert das Kriterium *Rechtschreibleistung < T-Wert 35*, die diagonale Linie markiert das Kriterium *1,5 Standardabweichungen zwischen tatsächlicher und aufgrund des IQ erwarteter Rechtschreibleistung*.

Abbildung 3 zeigt exemplarisch eine per Computer generierte Punktvolke zweier normalverteilter Variablen mit Korrelation 0,4. Der Einfachheit halber wurden die Achsenbeschriftungen so gewählt, dass die Abszisse in IQ-Werten und die Ordinate in T-Werten (Rechtschreibtest) skaliert ist. Der unter der parallel zur Abszisse verlaufenden Linie liegende Teil der Wolke ist gekennzeichnet durch einen RST-Wert < 35 ($= 1,5$ Standardabweichungen unter RST-Erwartungswert), und der unter der schrägen Linie liegende Teil der Wolke ist gekennzeichnet durch eine Diskrepanz von $> 1,5$ Standardabweichungen zwischen IQ-Wert und erwarteter Rechtschreibleistung (Regression), bezogen auf die Residuen der Regression.

Die inhaltliche Grundlage dieses Diskrepanzmodells lässt sich in 3 Annahmen fassen (Solity 1996): (1) Kinder mit Rechtschreibproblemen bilden zwei distinkte Gruppen, die sich mit Hilfe des IQ definieren lassen; (2) IQ hängt kausal mit der Rechtschreibung zusammen; und (3) IQ repräsentiert das Potential zum Erlernen der Rechtschreibung.

Bereits sehr früh etablierte sich diese Sichtweise, und es wurde angenommen, dass es für diese beiden Gruppen eine unterschiedliche Ätiologie gibt – zunächst allerdings ohne empirische Fundierung (Stanovich 1994)

Rutter & Yule (1975) etablierten dann endgültig die Einbeziehung des IQ bei der Diagnose von Lesestörungen (Stevenson 1988). Sie hatten mit den Daten der Isle-of-Wight-Studie empirisch zwei Untergruppen gebildet, die sich nicht hinsichtlich ihrer Leseleistung, wohl aber hinsichtlich ihres IQ unterschieden, und hinsichtlich einer Reihe von Variablen verglichen. Die eine Gruppe war gekennzeichnet durch eine Diskrepanz der Leseleistung zum IQ (Spezifische Leistungsschwäche, SLS), die andere durch eine schlechte Leseleistung ohne IQ-Diskrepanz (Allgemeine Leistungsschwäche, ALS). Die Gruppe der spezifisch schlechten Leser (SLS) wies seltener Hirnschäden, neurologische und motorische Auffälligkeiten auf, der Anteil an Jungen war größer, und die Therapierbarkeit schlechter als in der ALS-Gruppe. Damit gab es einen empirischen Hinweis darauf, dass die Unterscheidung der beiden Gruppen tatsächlich gerechtfertigt ist. Eine weitere bekannte Studie (Shaywitz

et al. 1992) kommt allerdings zu dem Schluss, dass sich die beiden Gruppen hinsichtlich sprachlicher Fähigkeiten, visueller Wahrnehmung und Händigkeit *nicht* unterscheiden. Ganz offensichtlich ist es also auch entscheidend, hinsichtlich *welcher* Eigenschaften die beiden Gruppen verglichen werden. Bei der Vielzahl von möglichen Variablen beeinflusst das Studiendesign also ganz erheblich die Ergebnisse. Neuere Studien, die Gruppenunterschiede finden, beschreiben die IQ-diskrepanten schlechten Leser als multipel auffällig in kognitiven Testbatterien (Meyer 2000) und in Kleinhirn-Funktionstests (Fawcett, Nicolson and Maclagan 2001).

Genetische Studien beschäftigten sich ebenfalls mit der Rolle des IQ. Aufgrund ungünstiger Stichprobenselektionen sind deren Ergebnisse aber oft nicht sehr aufschlussreich (Wadsworth et al. 2000). In einer neueren Arbeit kommen Wadsworth et al. (2000) zu der Einschätzung, dass die Heritabilität von Leseschwäche linear vom IQ abhängt, und für Kinder mit hohem IQ größer ist als für solche mit niedrigem IQ.⁵ Andere Arbeiten fanden keine Hinweise auf unterschiedliche genetische Ätiologie (Pennington et al. 1992) oder auf Unterschiede bei der familiären Häufung (Schulte-Körne et al. 1996).

Übersichtsarbeiten, die sich mit der Validität des IQ-Diskrepanzkriteriums beschäftigen, sind sich hinsichtlich der Schlussfolgerungen nicht einig. Aaron (1997) vertritt die Auffassung, dass es keine Hinweise auf Unterschiede zwischen ALS und SLS gibt. In ihrer ausgezeichneten Übersicht zu diesem Thema kommt Meyer (2000) zu der Schlussfolgerung, dass die Frage als ungelöst zu betrachten ist. Es muss allerdings der Vollständigkeit halber angemerkt werden, dass die Zahl der Studien überwiegt, die keine Unterschiede zwischen den Gruppen finden. Bezüglich der phonologischen Bewusstheit, die als eine Hauptursache der Leseschwäche betrachtet wird, ist die Zahl der Negativbefunde beeindruckend (Siegel 1992, Aaron 1997).

Marx et al. (2001) schließen aus der Befundlage treffend, dass „eine Unterteilung leseschwacher Kinder in abgrenzbare Ätiologien möglich zu sein scheint. Fraglich bleibt jedoch, inwieweit die Diskrepanzdefinition einen Beitrag zur Bestimmung dieser Gruppen leistet.“ (S. 87)

Diese Diskussion wurde praktisch ausschließlich für den englischen Sprachraum und für die Lese- störung geführt. Einige der kritisierten Aspekte der Diskrepanzdefinition treffen so auf den deutschsprachigen Raum allerdings nicht zu: es wurde kritisiert, dass die im Englischen üblicherweise verwendeten sprachgebundenen Intelligenztests (WISC-R) in einer wechselseitig kausalen Beziehung zur Leseleistung stehen (Aaron 1997, Meyer 2000). Das bedeutet, dass bessere Leser auch tendenziell besser in Intelligenztests abschneiden, *weil* sie besser lesen (Matthäus-Effekt, Meyer 2000). Für die in Deutschland meistens verwendeten sprachfreien Tests (CFT) dürfte das kaum zutreffen. Stanovich & Siegel 1994 (Übersicht bei Meyer 2000) finden außerdem, dass es keine substantielle Korrelation zwischen Lesen und IQ gibt, und die Beziehung zwischen beiden nicht über die gesamte Skala linear verläuft. Dies gilt im deutschsprachigen Raum für Rechtschreibleistung und IQ nicht, hier finden sich konsistent lineare Zusammenhänge um 0,4 (Schulte-Körne 2001).

Trotz dieser Unterschiede gibt es jedoch auch für deutschsprachige Kinder keinen empirischen Beleg dafür, dass sich die Gruppen spezifisch und allgemein Lese-Rechtschreibschwacher ätiologisch bedeutsam unterscheiden (Klicpera & Gasteiger-Klicpera 1993, Marx et al. 2001). Marx et al. (2001) fanden lediglich, dass in der Gruppe der spezifischen LRS der Anteil der Jungen höher war. Zur Frage der unterschiedlich erfolgreichen Förderbarkeit liegt ein deutschsprachiger, und zwar negativer, Befund vor (Weber, Marx und Schneider 2001).

Es gibt einen weiteren, sehr praktischen Aspekt der Diskussion. Wenn ein Kind in einer psychologischen oder kinderpsychiatrischen Praxis oder Klinik wegen Rechtschreibproblemen vorgestellt wird, so sind in der Regel mindestens die Eltern davon überzeugt, dass die Probleme des Kindes

5 Die Autoren räumen in der Diskussion der Arbeit allerdings ein, dass die Eltern von Kindern mit höherem IQ eine längere und bessere Ausbildung hatten und sich auch in weiteren Variablen (Anzahl der Bücher im Haushalt etc.) von den Eltern von Kindern mit niedrigem IQ unterscheiden. Von daher wäre es denkbar, dass die Umwelteinflüsse bei Kindern mit höherem IQ sowohl homogener als auch günstiger sind, was das Ergebnis der höheren Heritabilität bereits erklären würde.

nicht innerschulisch zu lösen sind. An diesem Punkt ist es selbstverständlich bedeutsam, ob es sich um ein spezifisches Versagen in einem Bereich, oder um ein allgemeines Schulversagen des Kindes handelt. Während im ersteren Fall eine spezielle Förderung in Frage kommt, ist bei letzterem Fall zu überlegen, ob das Kind allgemein leistungsmäßig überfordert ist, oder ob es emotionale oder sonstige Gründe für das Versagen gibt. Unabhängig von etwaigen wissenschaftlichen Überlegungen hat hier die Unterscheidung von spezifischem und allgemeinem Leistungsversagen eine ganz praktische Bedeutung.

Schließlich seien noch zwei methodische Aspekte angemerkt. Das „entweder – oder“ der Diskussion täuscht eine Unterscheidbarkeit vor, die so nicht existiert. In Abbildung 3 ist ersichtlich, dass die Gruppe der Messwertpaare, die das IQ-Diskrepanzkriterium und gleichzeitig das einfache Rechtschreib-Kriterium erfüllen (grauer Bereich), größer ist als die Summe derer, die entweder das eine oder das andere Kriterium erfüllen. Mehr als 2/3 (69%) der nach Diskrepanzmodell Betroffenen sind gleichzeitig nach einfachem Rechtschreib-Kriterium betroffen, und ebenfalls 69% der nach einfachem Rechtschreib-Kriterium Betroffenen sind gleichzeitig nach Diskrepanzmodell betroffen. Leider ist die Vergleichbarkeit der Befunde zu Unterschieden zwischen SLS und ALS erheblich eingeschränkt, weil die Abgrenzung der Gruppen unterschiedlich vorgenommen wurde. In der Arbeit von Rutter & Yule wurden z.B. *alle*, die das SLS-Regressionskriterium erfüllten verglichen mit jenen, die *nur* das ALS-Kriterium, aber nicht das SLS-(Regressions-)Kriterium erfüllten. In der Arbeit von Marx et al. (2001) wurden demgegenüber *alle*, die das ALS-Kriterium erfüllten aufgeteilt in zwei Gruppen, nämlich die mit $IQ < 85$ und die mit $IQ > 85$, wobei letztere dann zusätzlich noch das Regressionskriterium erfüllen mussten. Diese beiden Gruppeneinteilungen von Rutter & Yule (1975) und Marx et al. (2001) sind derart unterschiedlich, dass gleiche Ergebnisse gar nicht unbedingt zu erwarten gewesen wären. Die Arbeiten zu dieser Fragestellung sind also in zweierlei Hinsicht schwer vergleichbar: die jeweils untersuchten Variablen und die Gruppeneinteilungen sind unterschiedlich.

4 Praktische Probleme der Messung: Normierung

Jede Messung erhält ihre Wertigkeit durch den Vergleich mit einer relevanten Vergleichsgruppe. Ob eine bestimmte Rechtschreibleistung gut oder schlecht ist, kann man dem Rohwert nicht ansehen. Daher ist es von entscheidender Bedeutung, dass das Verfahren an einer hinreichend großen und repräsentativen Stichprobe normiert wurde. Normen werden bei Rechtschreibtests sinnvollerweise nicht für Altersgruppen, sondern für Klassenstufen angegeben. Dies setzt allerdings voraus, dass alle Klassen auch nach der gleichen Methode unterrichtet werden, was ganz zweifellos nicht gegeben ist. Dies wirkt sich vor allem in den ersten beiden Klassen aus, was eine Diagnose außerordentlich erschwert.

Ein weiterer sehr wichtiger Punkt ist, dass Normen veralten. Von Intelligenztests ist bekannt, dass die Normen im Laufe der Zeit zu leicht werden (d.h. Zwanzigjährige vor 10 Jahren waren im Bewältigen der gleichen IQ-Aufgaben schlechter als Zwanzigjährige von heute). Dies ist als „Flynn-Effekt“ bekannt geworden (Rodgers 1998). Der genau gegenteilige Effekt wurde bei Schulleistungstests beobachtet (Rodgers 1998), d.h. das allgemeine Rechtschreibniveau sinkt. Dazu kommt noch, dass die Items selbst veralten, d.h. die Items von älteren Tests wie dem WRT6+ sind zu einem Teil heute ungebräuchlich, und demzufolge werden sie auch seltener gekannt. Für eine deutsche Stichprobe konnte belegt werden, dass veraltete Normen für die Rechtschreibleistung zu einer Unterschätzung, und für die Intelligenz zu einer Überschätzung der tatsächlichen Werte führen (Strehlow und Haffner 2002, Zerahn-Hartung et al. 2002).

Schließlich gibt es noch ein wenig bekanntes Phänomen bei der Zusammensetzung von Normstichproben, was bei der gleichzeitigen Verwendung von IQ-Tests und Rechtschreibtests relevant ist: für die Normierung von Rechtschreibtests werden z.B. in der Primarstufe nur Grundschulen, aber keine Sonderschulen herangezogen. Bei der Normierung von IQ-Tests werden die Sonderschüler aber

berücksichtigt. Dies führt dazu, dass – bezogen auf die Normgruppe des IQ-Tests – die Normen des Rechtschreibtests etwas zu schwer sind, da ja ein Teil der schlechteren Schreiber (Sonderschüler) nicht berücksichtigt wurden.

Schließlich gibt es noch das Problem mit den Therapierten – falls eine Therapie auch nur teilweise erfolgreich war, so fällt die Leistung eines Kindes eventuell nicht mehr in den Bereich, der zu einer Diagnose LRS führt. Die Frage, wie lange eine Therapie ggf. durchzuführen ist, lässt sich also ebenfalls nicht auf testdiagnostischem Wege beantworten.

Diese Effekte sind bei der Interpretation von Testwerten angemessen zu berücksichtigen.

5 Diagnostische Kriterien

DSM-IV (Saß, Wittchen und Zaudig 1996) und ICD-10 (Dilling, Mombour und Schmidt 1991) formulieren für Schulleistungsstörungen im Kern die selben Kriterien:

1. die jeweilige Schulleistung (wie in einem standardisierten Test gemessen) ist kleiner, als aufgrund von IQ, Alter und Beschulung zu erwarten ist.
2. es liegt eine klinisch relevante Beeinträchtigung bzw. ein Leidensdruck vor.
3. Ausschluss: unangemessene Beschulung, periphere Seh- und Hörstörungen, neurologische und emotionale Störungen, soweit sie ursächlich für die Schulleistungsproblematik sind.

Die Unterschiede zwischen DSM-IV und ICD-10 fallen marginal aus. Die Kriterien im DSM-IV sind insgesamt knapper gefasst.

- Im DSM-IV finden neurologische und emotionale Erkrankungen als Ausschlusskriterien keine Erwähnung.
- Das DSM-IV fordert eine Diskrepanz von >2 Standardabweichungen, mindestens jedoch 1 Standardabweichung (im begründeten Einzelfall) zwischen Schulleistung und Intelligenz. Im ICD-10 wird die geforderte „bedeutsame“ Diskrepanz in den klinischen Kriterien nicht quantifiziert, sondern es wird lediglich in den Forschungskriterien eine Diskrepanz von 2 Standardabweichungen gefordert.
- Im ICD-10 wird ausdrücklich eine Diskrepanz gefordert, die mit dem Regressionsansatz berechnet wurde. Das DSM-IV erwähnt den Regressionsansatz nicht explizit, sondern spricht nur von Leistungen, die kleiner als „erwartet“ sind.

Schließlich unterscheiden sich auch noch die Kategorien. Das ICD-10 unterscheidet Lese- und Rechtschreibstörung (F 81.0) und Isolierte Rechtschreibstörung (F 81.1). Eine – theoretisch denkbare – Isolierte Lesestörung ist nicht vorgesehen. Das DSM-IV sieht als Kategorien die Lesestörung (315.00) und die Störung des schriftlichen Ausdrucks (315.2) vor, wobei beide lt. DSM-IV oft zusammen auftreten.

In der Praxis sind die beiden Kategorien (jeweils aus ICD-10 oder DSM-IV) nicht methodisch zufriedenstellend abgrenzbar. Dazu müsste z.B. genau definiert werden, wie stark beeinträchtigt die Leseleistung zusätzlich zur isolierten Rechtschreibstörung lt. ICD-10 sein müsste, um die Diagnose Lese-Rechtschreibstörung zu stellen. Eine solche Operationalisierung existiert nicht.

Wichtig ist ebenfalls die Abgrenzung zur Rechenstörung. Die Diagnose Rechenstörung (F81.2) darf laut ICD-10 *nicht* zusammen mit entweder isolierter Rechtschreibstörung (F81.1) oder Leserechtschreibstörung (F81.0) gegeben werden. Stattdessen ist in diesem Fall von einer „kombinierten Störung schulischer Fertigkeiten“ (F81.3) auszugehen. Dabei handelt es sich allerdings um eine schlecht definierte Restkategorie. Man kann aber vermuten, dass bei gleichzeitigem Versagen im Lesen, Schreiben und Rechnen nicht mehr von einer „Teilleistungsstörung“ ausgegangen werden kann, sondern es sich eher um ein weitreichendes allgemeines Schulversagen handelt.

6 Praktisches Vorgehen zur Diagnostik

Ausgehend von den Kriterien der ICD-10 und DSM-IV sind drei Bereiche zu berücksichtigen: Leistungsmessung, Beeinträchtigung und Ausschlusskriterien. Die praktischen Teile der Untersuchung sind dabei Eingangsgespräch und Anamnese, Testdiagnostik und medizinische Untersuchung.

Leistungsmessung. Die Werte von standardisierten Rechtschreib- und Lesetests sind erfahrungsgemäß erst ab etwa Ende der 2. Klasse einigermaßen zuverlässig. Davor ist die Variation durch unterschiedliche Unterrichtsmethoden noch zu groß.

Bei Rechtschreib- und Lesetests sollten schulformübergreifende Klassennormwerte verwendet werden.

In einem Artikel aus unserer Arbeitsgruppe (Schulte-Körne, Deimel und Remschmidt 2001) wurde ein Vorschlag für eine Operationalisierung der ICD-10-Kriterien entwickelt. Es handelt sich um ein kombiniertes Kriterium aus Rechtschreib-Prozentrang < 16 und regressionsbasierter IQ-Diskrepanz von 1,5 Standardabweichungen. Die benötigte Tabelle ist im Internet zugänglich unter <http://www.kjp.uni-marburg.de/kjp/legast/leg/diagnose.htm>.

Ein zweiter Aspekt der Leistungserfassung ist die Abgrenzung zwischen spezifischem und allgemeinem Schulversagen. In der Primarstufe heißt das praktisch, dass die Leistungen in Deutsch und Rechnen verglichen werden. Ist das Kind im Rechnen deutlich besser als im Schreiben und/oder Lesen, so ist das als Hinweis auf eine Teilleistungsproblematik zu werten. Hierbei ist allerdings zu berücksichtigen, dass Kinder durch die ständigen Frustrationen in einem Bereich eine zunehmende allgemeine Schulunlust bzw. emotional bedingtes allgemeines Schulversagen entwickeln können. Eine Abgrenzung durch den Vergleich verschiedener Fächer ist also u.U. ab einem bestimmten Alter nicht mehr möglich. Ebenfalls erschwert wird die Leistungsmessung durch bereits durchgeführte Fördermaßnahmen, die ja einen positiven Einfluss auf die Leistungen haben sollen. Das könnte dazu führen, dass ein gut gefördertes Kind die Kriterien für eine Teilleistungsstörung gerade nicht mehr erfüllt, obwohl eine Fortführung der Förderung natürlich sinnvoll ist.

Diese beiden Beispiele zeigen, dass Testergebnisse nicht immer buchstabengetreu „geglaubt“ werden können. Vielmehr ist eine Integration aller Befunde erforderlich, wofür eine umfangreiche Erfahrung erforderlich ist. Das gleiche gilt für grenzwertige Befunde, d.h. solche, wo ein Kind die Kriterien für eine Teilleistungsstörung nur sehr knapp erreicht oder verfehlt. Tatsächliche Leistungsschwankungen und weitere Einflüsse wie Messfehler und ‚Tagesform‘ lassen eine anhand eines solchen grenzwertigen Resultats vorgenommene Kategorisierung sehr unsicher werden. Auch hier ist die Integration aller Informationen sehr wichtig, um zu einer insgesamt stimmigen Entscheidung zu kommen.

Beeinträchtigung bzw. Leidensdruck. Dabei handelt es sich um zwei verschiedene Dinge. Der individuelle Leidensdruck des Kindes wird über eine psychologische Untersuchung festgestellt. Hierzu gehören eine gründliche Exploration und die Durchführung geeigneter Testverfahren zur Emotio-

naldiagnostik. Das Ausmaß der Beeinträchtigung kann nur indirekt über die Befragung von Kind, Eltern und Lehrerinnen festgestellt werden.⁶

Ausschlusskriterien. Dabei handelt es sich größtenteils um medizinische Kriterien. Zur Feststellung ist also eine medizinische Untersuchung nötig. Der Ausschluss von emotionalen ursächlich verantwortlichen Erkrankungen ist im Rahmen einer psychiatrischen oder psychologischen Untersuchung zu prüfen. Die Angemessenheit der Beschulung kann am besten durch Exploration der Eltern festgestellt werden.

Die Ergebnisse der Informationserhebung in diesen Bereichen sind dann zu integrieren, und es kann ggf. über eine geeignete Behandlung (Legastheniebehandlung und/oder Psychotherapie) entschieden werden. Es handelt sich auf jeden Fall um eine fachübergreifende Fragestellung, die in der Regel nicht von einer Person allein bearbeitet werden kann. Umfangreiche Kenntnisse und Erfahrung sind natürlich Voraussetzung für eine valide Diagnose.

Die *Frühdiagnose im Kindergartenalter* stellt einen Sonderfall der Diagnostik dar. Mit dem Bielefelder Screening zur Frühdiagnose (BISC, Jansen et al. 1999) ist es lt. Handbuch möglich, mit einer guten Zuverlässigkeit bereits im späten Kindergartenalter die Kinder zu identifizieren, die später Probleme beim Schriftspracherwerb bekommen werden.

Das BISC besteht aus neun Untertests, von denen vier dem Bereich Phonologie zugeordnet werden können. Drei weitere Untertests befassen sich mit Abfragen des Langzeitgedächtnisses, ein Untertest überprüft das Kurzzeitgedächtnis und ein Untertest erfasst die visuelle Aufmerksamkeit.

Die Ergebnisse zur Vorhersage wurden allerdings bisher nicht repliziert, und das Verfahren dürfte auch in der Praxis am ehesten durch flächendeckende Vorsorgeuntersuchungen anwendbar sein, für die es zur Zeit keinen organisatorischen Rahmen gibt.

Literatur

- Aaron PG (1997). The impending demise of the discrepancy formula. Review of Educational Research 67(4), 461-502.
- Birkel P (1990). Grundwortschatz Rechtschreibtest für 4. und 5. Klassen (GRT4+). Beltz, Weinheim und Basel.
- Dilling H, Mombour W, Schmidt MH (1991). Internationale Klassifikation psychischer Störungen, ICD-10. Huber, Bern.
- Dummer-Smoch L (2001). Frühe Diskrepanzdiagnose oder Screening der „Phonologischen Bewusstheit“. In Schulte-Körne G (Hrsg): Legasthenie: erkennen, verstehen, fördern. Beiträge zum 13. Fachkongress des Bundesverbandes Legasthenie 1999. Winkler, Bochum.
- Evans LD (1990). A conceptual overview of the Regression Discrepancy Model for evaluating severe discrepancy between IQ and achievement scores. Journal of Learning Disabilities 23(7), 406-412.
- Fawcett AJ, Nicolson RI, Maclagan F (2001). Cerebellar tests differentiate between groups of poor readers with and without IQ discrepancy. Journal of Learning Disabilities 34(2), 119-135.
- Jansen H, Mannhaupt G, Marx H, Skowronek H (1999). Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC). Hogrefe, Göttingen.
- Klicpera C, Gasteiger-Klicpera B (1993). Lesen und Schreiben. Entwicklung und Schwierigkeiten. Huber, Bern.
- Marx P, Weber JM, Schneider W (2001). Legasthenie versus allgemeine Lese-Rechtschreibschwäche: Ein Vergleich der Leistungen in der phonologischen und visuellen Informationsverarbeitung. Zeitschrift für Pädagogische Psychologie 15(2), 85-98.
- Meyer MS (2000). The ability-achievement discrepancy: Does it contribute to an understanding of learning disabilities? Educational Psychology Review 12(3), 315-337.
- Naegele IM, Valtin R (2001). LRS – Legasthenie in den Klassen 1 – 10. Handbuch der Lese-Rechtschreibschwierigkeiten Band 1. Beltz, Weinheim.
- Pennington BF, Gilger JW, Olson RK, DeFries JC (1992). The external validity of age- versus IQ-discrepancy definitions of reading disability: lessons from a twin study. Journal of Learning Disabilities 25(9), 562-73.

6 Es könnte z.B. sein, dass ein Kind aus einer allgemein schlechten Klasse zwar ein schlechtes Ergebnis in einem Rechtschreibtest erzielt, das Kind in der Klasse aber wegen des allgemein schlechten Niveaus nicht weiter auffällt. Hier ist vermutlich kein zusätzlicher Förderbedarf vorhanden.

- Ranschburg P (1928). Die Lese- und Rechtschreibstörungen des Kindesalters. Carl Marhold, Halle.
- Rathenow P (1979). Westermann Rechtschreibtest 4/5. Westermann, Braunschweig.
- Rodgers B (1983). The identification and prevalence of specific reading retardation. British Journal of Educational Psychology 53(3), 369-373.
- Rodgers JL (1998). A critique of the Flynn Effect: Massive IQ gains, methodological artifacts, or both? Intelligence 26(4), 337-356.
- Rutter M, Yule W (1975). The concept of specific reading retardation. Journal of Child Psychology and Psychiatry 16(3), 181-97.
- Saß H, Wittchen HU, Zaudig M (1996). Diagnostisches und Statistisches Manual Psychischer Störungen DSM-IV. Hogrefe, Göttingen.
- Scheerer-Neumann G (2001). LRS und Legasthenie: Eine Bestandsaufnahme. In Naegele IM & Valtin R: LRS – Legasthenie in den Klassen 1 – 10. Handbuch der Lese-Rechtschreibschwierigkeiten Band 1. Beltz, Weinheim.
- Schulte-Körne G. (2001). Legasthenie und Sprachwahrnehmung. Waxmann, Münster.
- Schulte-Körne G, Deimel W, Müller K, Gutenbrunner C, Remschmidt H (1996). Familial aggregation of spelling disability. Journal of Child Psychology and Psychiatry 37, 817-822.
- Schulte-Körne G, Deimel W, Remschmidt H (2001). Zur Diagnostik der Lese-Rechtschreibstörung. Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie 29 (2), 113-116.
- Shaywitz BA, Fletcher JM, Shaywitz SE (1995). Defining and classifying learning disabilities and attention-deficit/hyperactivity disorder. Journal of Child Neurology 10 Suppl 1, 50-57.
- Shaywitz SE, Escobar MD, Shaywitz BA, Fletcher JM, Makuch R (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. New England Journal of Medicine 326(3), 145-150.
- Shaywitz SE, Fletcher JM, Shaywitz BA (1994). A new conceptual model for dyslexia. In Capute AJ, Accardo PJ, Shapiro BK (Hrsg): Learning disabilities spectrum: ADD, ADHD, and LD. York Press, Maryland, 1-15.
- Siegel LS (1992). An evaluation of the discrepancy definition of dyslexia. Journal of Learning Disabilities 25(10), 618-629.
- Solity J (1996). Discrepancy definitions of dyslexia: An assessment-through-teaching perspective (a response to Ashton). Educational Psychology in Practice 12(3), 141-151.
- Stanovich KE (1994). Annotation: does dyslexia exist? Journal of Child Psychology and Psychiatry 35(4), 579-595.
- Stanovich KE, Siegel LS (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. Journal of Educational Psychology 86(1), 24-53.
- Stephenson S (1907). Six cases of congenital word-blindness affecting three generations of one family. Ophthalmoscope 5, 482-484.
- Stevenson J (1988). Which aspects of reading ability show a "hump" in their distribution? Applied Cognitive Psychology 2(1), 77-85.
- Strehlow U, Haffner J (2002). Definitionsmöglichkeiten und sich daraus ergebende Häufigkeit der umschriebenen Lese- bzw. Rechtschreibstörung – theoretische Überlegungen und empirische Befunde an einer repräsentativen Stichprobe junger Erwachsener. Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie 30 (2), 113-126.
- Thorndike RL (1963). The concepts of over- and under-achievement. Bureau of Publications, Teachers College, Columbia University, New York.
- Valtin R (2001). Von der klassischen Legasthenie zur LRS – notwendige Klarstellungen. In Naegele IM und Valtin R: LRS – Legasthenie in den Klassen 1 – 10. Handbuch der Lese-Rechtschreibschwierigkeiten Band 2. Beltz, Weinheim.
- Van der Wissel A, Zegers FE (1985). Reading Retardation Revisited. British Journal of Developmental Psychology 3, 3-9.
- Wadsworth SJ, Olson RK, Pennington BF, DeFries JC (2000). Differential genetic etiology of reading disability as a function of IQ. Journal of Learning Disabilities 33(2), 192-199.
- Warnke A (1990). Legasthenie und Hirnfunktion. Hans Huber, Bern.
- Weber JM, Marx P, Schneider W (2002). Profitieren Legastheniker und allgemein lese-rechtschreibschwache Kinder in unterschiedlichem Ausmass von einem Rechtschreibtraining. Psychologie in Erziehung und Unterricht 49(1), 56-70.
- Zerahn-Hartung C, Strehlow U, Haffner J, Pfüller U, Parzer P, Resch F (2002). Normverschiebungen bei Rechtschreibleistungen und sprachfreier Intelligenz. Praxis der Kinderpsychologie und Kinderpsychiatrie 51, 281-297.