

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder andere Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. – Fotokopien für den persönlichen und sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopien hergestellt werden.

Reinold Funke

ERSTUNTERRICHT NACH DER METHODE *LESEN DURCH SCHREIBEN* UND ERGEBNISSE SCHRIFTSPRACHLICHEN LERNENS – EINE METAANALYTISCHE BESTANDSAUFNAHME

1 Fragestellung

Das von dem Schweizer Lehrer Jürgen Reichen konzipierte Verfahren *Lesen durch Schreiben* (Reichen 1988) ist eine der in der Gegenwart am meisten diskutierten Methoden des Erstunterrichts im Lesen und Schreiben. Die Kinder werden angeleitet, Wörter zu schreiben, indem sie sie entsprechend ihrer phonetischen Wahrnehmung schrittweise in Laute zerlegen. Jeden dieser Laute ordnen sie einem Bild zu, welches sie in einer Anlauttabelle aufsuchen (z. B. Bild eines Rades für den Laut [r]). Neben dem Bild findet sich ein zu dem Laut passender Buchstabe (z. B. Buchstabe <R> neben dem Bild des Rades). Die Kinder übertragen diesen und gehen zum nächsten Laut über. Stück für Stück entsteht so das komplette Schriftwort, ohne dass fremde Hilfe notwendig ist. Was die Kinder schreiben, ergibt sich aus ihrer jeweiligen kommunikativen Absicht, so dass das Schreiben von Anfang an in einen sinnvollen Zusammenhang eingebettet ist. Das Lesen wird in dieser Methode gar nicht unterrichtet. Nach Reichen stellt sich die Lesefähigkeit im Laufe der Zeit von selbst ein.

Der Methode *Lesen durch Schreiben* liegt der Gedanke zugrunde, dass schriftsprachliches Können nur in einem selbstgesteuerten Lernprozess erworben werden kann. In der Fachdiskussion ist sie umstritten. Den Kindern stellt sich ihre Tätigkeit beim Verschriften mit der Anlauttabelle als phonetische Wiedergabe von Wörtern dar. Die Schrift gibt jedoch Wörter als sprachlich interpretierte Einheiten wieder, deren phonologische, morphologische und syntaktische Eigenschaften sie kennzeichnet. Um flüssig lesen und orthographisch schreiben zu können, muss man erfassen, wie sie das tut. Auf dem Weg dahin, so die Kritik, bleiben die Kinder in der Methode *Lesen durch Schreiben* sich selbst überlassen. Für lernschwache Kinder wird die Eignung von *Lesen durch Schreiben* aus diesem Grund bezweifelt, und zwar sowohl im Blick auf den Erwerb des Lesens (Valtin 1998) wie den des Rechtschreibens (Röber-Siekmeyer/Spiekermann 2000). Die Kritik ist im Kern motiviert durch die Annahme, dass Menschen, wenn sie lernen, sprechsprachliche und schriftsprachliche Einheiten in Beziehung zu setzen, angeregt werden, ihr sprachliches Wissen umzubauen und zu vertiefen. Der Schriftspracherwerb wird deshalb als eine Sprachlernerfahrung betrachtet. Mit welcher Qualität diese stattfindet, ist, so wird angenommen, insbesondere für zweitsprachliche Kinder von Bedeutung (Bredel 2012).

In der vorliegenden Arbeit werden empirische Studien ausgewertet, welche Daten zu den Lernergebnissen in Klassen enthalten, die nach der Methode *Lesen durch Schreiben* unterrichtet wurden, und welche sie mit denen in Klassen vergleichen, die Erstunterricht nach einer Fibel erhalten haben. In Forschungsberichten zu *Lesen durch Schreiben* gelangen Metze (2008) und Schröder-Lenzen (2009) zu einem skeptischen Fazit, was die Lernerfolge angeht. Den Hintergrund der Berichte von Brügelmann/Brinkmann (2012) und Hanke (2005) bildet der Gedanke, dass die Frage nach der Eignung von *Lesen durch Schreiben* als Methode falsch gestellt sei, da nicht Methoden als solche für den Erfolg des Erstunterrichts ausschlaggebend seien, sondern die Qualität des Unterrichts, in dem sie angewendet werden. Im Folgenden werden metaanalytische Verfahren genutzt, um den Überblick über die empirischen Daten zu systematisieren. Das Ziel ist, zu klären, was sich aus den Daten hinsichtlich der Lernwirksamkeit von *Lesen durch Schreiben* entnehmen lässt. *Lernwirksamkeit* muss von *Lernergebnis* unterschieden werden. Wenn man, wie es im Titel dieser Arbeit der Fall ist, von den bei einer Methode erreichten (Lern-)Ergebnissen spricht, sind beobachtete Lernleistungen gemeint, ohne dass das mit einer Festlegung darüber verbunden ist, ob sie der Nutzung der Methode zuzurechnen sind oder nicht. Wenn man dagegen von der Lernwirksamkeit einer Methode spricht, bringt das zum Ausdruck, dass Lernergebnisse auf den Einsatz der Methode zurückgeführt werden. Es liegt auf der Hand, dass man aus den in *Lesen durch Schreiben*-Klassen im Vergleich zu in Fibelklassen erreichten Lernergebnissen alleine nicht auf die Lernwirksamkeit einer der beiden Methoden schließen kann, da es möglich ist, dass die Klassen mit jeweils unterschiedlichen Lernvoraussetzungen gestartet sind. Es geht also in der vorliegenden Bestandsaufnahme nicht darum, die Nullhypothese zu prüfen, die Lernwirksamkeit von *Lesen durch Schreiben* und von Fibelunterricht sei gleich, um auf diesem Weg einen Vorteil einer der beiden Methoden aufzuzeigen. Stattdessen soll überprüft werden, in welchem Umfang den Daten zu den Lernergebnissen in *Lesen durch Schreiben*-Klassen Aussagekraft im Hinblick auf die Lernwirksamkeit der Methode zukommt.

Zu den zentralen Erkenntnissen der aktuellen Unterrichtsforschung gehört, dass man alleine aus der Kenntnis, welche Unterrichtsmethode angewandt wurde, Lernerfolge nur in geringem Umfang voraussagen kann. Als für die Lernwirksamkeit maßgebliches Merkmal wird stattdessen die Unterrichtsqualität angenommen (Helmke 2009). Aktuelle Arbeiten stützen diesen Ansatz mit Blick auf den Erstunterricht (Hanke 2005; Poerschke 1999). Wenn man fachdidaktisch begründete Zugänge zum Verständnis der Wirksamkeit von Unterricht sucht, bereitet er jedoch, so gut er auch belegt ist, Schwierigkeiten. Ein Aspekt, unter dem man Unterrichtsmethoden betrachten kann, ist, dass sie Werkzeuge sind. So wie man die Qualität einer Reparatur nicht aufgrund der Werkzeuge beurteilen kann, die der (die) Handwerker(in) gebraucht hat, kann man die Lernerfolge bei Schülerinnen und Schülern nicht aufgrund der Methoden voraussagen, die die Lehrkraft eingesetzt hat. Trotzdem wird ein(e) Handwerker(in) Wert darauf legen, über eine gut sortierte Werkzeugkiste zu verfügen, und sich bei jedem Arbeitsgang Gedanken über das geeignete Werkzeug machen. Für Lehrkräfte wie für Fachdidaktiker(innen), die Unterricht aus deren Per-

spektive betrachten, gibt es deshalb kaum eine Alternative zur Auseinandersetzung über Methodenfragen, obwohl unstrittig ist, dass Unterrichtsmethoden als solche Lernergebnisse nur begrenzt voraussagen vermögen.

2 Methode

2.1 Datenbasis

Während die publizierten Studien zu *Lesen durch Schreiben* durch die oben aufgeführten, in der Literatur enthaltenen Überblicke erschlossen werden, ist es schwierig, nicht publizierte Studien zuverlässig zu erfassen. Recherchen in deutschen und Schweizer Literaturdatenbanken (<http://www.fachportal-paedagogik.de>, <http://www.dnb.de>, <http://edudoc.ch>) führten auf zwei empirische Dissertationen, die *Lesen durch Schreiben*-Klassen einbeziehen (Friedrich 2010; Sander 2006). Recherchen in Datenbanken mit laufenden Forschungsprojekten (<http://www.gesis.org>, <http://www.forschungsportal.ch>, <http://www.skbf-csre.ch>) erbrachten keinen weiteren Treffer.¹

In der vorliegenden Arbeit werden empirische Erhebungen in den Klassenstufen 1-4 berücksichtigt, die im Erstunterricht nach *Lesen durch Schreiben* unterrichteten Klassen mit Fibeln unterrichtete Klassen gegenüberstellen und quantitativ auswertbare Angaben zu Lernergebnissen im schriftsprachlichen Bereich enthalten. Diese Erhebungen, die im Literaturverzeichnis durch einen Asterisk (*) gekennzeichnet sind, umfassen 16 im Zeitraum 1985 bis 2010 fertiggestellte Studien mit 21 unabhängigen Stichproben. Die meisten Daten liegen zum Rechtschreiben vor, wenige zum Lesen, nur sporadisch solche zu textuellen Merkmalen frei konzipierter Schülertexte. Da die Analyse in explorativer, nicht in konfirmatorischer Absicht erfolgt, werden keine Stichproben von vorneherein wegen methodologischer Merkmale der Studie, der sie entstammen, aus der Datenbasis ausgeschlossen. Die einzigen Einschlusskriterien sind, dass die Studien sich auf Klassenstufe 1-4 beziehen sollen und die in den untersuchten Klassen praktizierte Methode von den Autor(inn)en selbst als *Lesen durch Schreiben* klassifiziert werden soll. Wenn die Methode zwar nach der Beschreibung, die von ihr gegeben wird, dem Ansatz *Lesen durch Schreiben* zu ähneln scheint, aber von den Autor(innen) anders bezeichnet wird, wird die Studie nicht aufgenommen (so bei Einsiedler, Frank et al. 2002; vgl. Kirschhock 2003). Soweit die Methode als *Lesen durch Schreiben* beschrieben wird, aber keine Daten zu Klassenstufe 1-4 vorliegen, bleibt sie ebenfalls unberücksichtigt (so bei Albert et al. 2009). Einbezogen werden schließlich keine Daten, die sich auf Messzeitpunkte vor dem Ende der Klassenstufe 1 beziehen.

1 Maja Wiprächtiger-Geppert sei für ihren Hinweis auf die Schweizer Datenbanken gedankt.

2.2 Statistisches Modell

Wie bereits angesprochen, lassen sich gegen den Versuch, aus den vorhandenen Studien auf die Lernwirksamkeit von *Lesen durch Schreiben* schließen zu wollen, methodologische Einwände erheben. Ihre Erörterung im Einzelnen erübrigt sich, da sie sich bündeln lassen in der Feststellung, dass die Studien nicht experimenteller Natur sind. Die meisten sind im Sinne der Terminologie von Borenstein et al. (2009) als ‚observational‘ einzuordnen. Das heißt, es werden Klassen, in denen ohnehin nach *Lesen durch Schreiben* bzw. nach einer Fibel Erstunterricht erteilt worden war, aufgesucht, und ihr Lernstand wird erhoben. Von einem experimentellen Vorgehen könnte man sprechen, wenn einzelne Schülerinnen und Schüler (randomisiertes Experiment) oder ganze Klassen (Quasi-Experiment) von außen einer der Methoden *Lesen durch Schreiben* oder Fibel zugeordnet werden und ihr Lernerfolg in vorher festgelegten Bereichen untersucht wird. Das ist bei keiner der genutzten Studien der Fall. Einige dieser Studien (Diem 1990; Hüttis-Graff 1998; May 2001; Schröder-Lenzen/Mücke 2005) entstanden im Rahmen von Schulversuchen, bei denen es aber um etwas anderes ging als um den Vergleich zwischen *Lesen durch Schreiben*- und Fibelklassen. In zwei weiteren Studien war es das Kernziel, *Lesen durch Schreiben*- und Fibelklassen hinsichtlich ihrer Lernergebnisse zu vergleichen, und die Erhebung wurde methodisch von vorneherein entsprechend angelegt (Poerschke 1999; Weinhold 2006). Welcher der beiden Methoden eine Klasse zugeordnet wurde, ergab sich aber daraus, was die Lehrkräfte ohnehin vorhatten. Es wurde nicht als experimentelle Bedingung festgelegt.

Die gängigen metaanalytischen Verfahren setzen voraus, dass für einen Leistungsbe- reich, etwa das Rechtschreiben, Messwerte aus mehreren unabhängigen Stichproben vorliegen. Jede Stichprobe umfasst Versuchs- und Kontrollprobanden. Das Ausmaß, in dem sich Versuchs- und Kontrollprobanden in dem Leistungsbereich unterscheiden, wird durch die Effektstärke d beziffert. Sie wird üblicherweise wie folgt berechnet:

$$d_{HO} = (m_{VG} - m_{KG}) / s_{pooled}$$

In dieser Formel steht d_{HO} für die Effektstärke (nach Hedges/Olkin 1985), m_{VG} und m_{KG} stehen für die Mittelwerte der Versuchs- bzw. Kontrollprobanden und s_{pooled} steht für die ‚gepoolte‘ Streuung (die als mittlere Streuung von Versuchs- und Kontrollgruppe betrachtet werden kann). Die Effektstärke bezieht sich ausschließlich auf den Lernstand zum Messzeitpunkt. Unterschiedliche Lernausgangsstände berücksichtigt sie nicht. Eine Effektstärke $d > 0$ zeigt an, dass die Lernergebnisse bei den Versuchsprobanden im Durchschnitt besser waren als bei den Kontrollprobanden, eine Effektstärke $d < 0$ zeigt das Gegenteil an. Für die Aussagekraft, die eine Effektstärkenschatzung d hat, ist ihre Varianz V_d maßgeblich (bzw. deren Wurzel, der Standardfehler s_d). Sie stellt ein Maß der Präzision dar, mit der der ‚Effekt‘ der Versuchsbedingung gegenüber der Kontrollbedingung anhand von d bestimmt werden

kann. Ihre Höhe ist stark von der Zahl der Versuchs- und Kontrollprobanden abhängig, die zur Stichprobe gehören.

Daraus entsteht nun ein Problem bei der Auswertung von Daten aus nicht-experimentellen Studien. Dass eine Stichprobe umso mehr Information hinsichtlich der Effektstärke enthält, je mehr Probanden der Stichprobe angehören, trifft insoweit zu, wie jede(r) Proband(in) einen von den übrigen Probanden unabhängigen Beitrag zur Bestimmung der Effektstärke leistet. Wenn Stichproben zusammengestellt werden, indem ganze Klassen der Versuchs- oder der Kontrollbedingung zugeordnet werden, ist diese Voraussetzung nur bedingt erfüllt. Denn was jeder einzelne Proband zur Einschätzung der Höhe von d beiträgt, ist dann nur zu einem Teil neu. Zu einem anderen Teil reproduziert es etwas, was man bereits von den anderen Probanden der gleichen Klasse weiß. In klassenweise zusammengestellten Stichproben ist jeder einzelne Proband somit ‚weniger informativ‘ als in Zufallsstichproben. Das bedeutet, dass die Varianz der Effektstärke in Ersteren nicht im gleichen Ausmaß mit der Stichprobengröße sinkt wie in Letzteren. Das muss bei der Analyse berücksichtigt werden, damit man den Daten nicht eine Aussagekraft zuschreibt, die ihnen abgeht, und so zu ungedeckten Schlussfolgerungen verleitet wird.

Die vorliegende Arbeit folgt einem von Hedges (2009) beschriebenen Vorgehen, bei dem an Stelle von d_{HO} das Effektmaß d_T verwendet wird. Bei der Berechnung von d_T wird d_{HO} unter Berücksichtigung der klassenweisen Abhängigkeit der Messwerte korrigiert. Die Korrektur schlägt sich in einer erhöhten Varianzschätzung nieder, während die Effektstärke selbst nur wenig beeinflusst wird.

Die für die Berechnung von d_T wesentliche Größe ist die Intraklassenkorrelation ρ . Sie kann als Maß dafür interpretiert werden, in welchem Umfang Effekte der Versuchs- und der Kontrollbedingung auf Unterschiede zurückzuführen sind, die zwischen Klassen bestehen, oder auf Unterschiede, die zwischen Probanden bestehen. ρ nimmt Werte zwischen 0 und 1 an, wobei ein Wert von 0 bedeuten würde, dass Unterschiede voll und ganz der Probandenebene zuzurechnen sind, während ein Wert von 1 bedeuten würde, dass sie voll und ganz der Klassenebene zuzurechnen sind. Hinweise auf die Höhe von ρ enthalten die hier ausgewerteten Studien in der Regel nicht. Nach Mücke (2006) errechnet sich für die Würzburger Leise Leseprobe (vgl. Küspert/Scheider 1998) in der Berliner Stichprobe von Schröder-Lenzen/Mücke (2005) in Klassenstufe 1 eine Intraklassenkorrelation von 0,16 und in Klassenstufe 2/3 von 0,25. May (2001, S. 79) macht Angaben zu den Streuungen der in seiner Stichprobe bei der Hamburger Schreibprobe (vgl. May 1995) erreichten Graphemtreffer auf Probanden- und Klassenebene. Daraus lassen sich Schätzungen für ρ errechnen, deren Median über die Klassenstufen 1-4 hinweg bei 0,23 liegt. Aus weiteren Quellen (nachgewiesen in einer Datendokumentation zur vorliegenden Studie unter <http://www.ph-heidelberg.de/wp/funke/Publikationen.htm>) können Angaben zur Intraklassenkorrelation für unterschiedliche schriftsprachliche Tests, Klassenstufen und Regionen entnommen werden. Der Median der gefundenen Werte liegt bei 0,18 (Rechtschreiben) bzw. 0,15 (Lesen), wobei die höchsten Werte nahe 0,30 liegen und aus Stichproben stammen, die mit frei konstruierten statt normierten Aufgabenstellungen getestet wurden. Hedges/Hedberg (2007) errechnen für bei amerikani-

schen Schülerinnen und Schülern eingesetzte Lesetests eine mittlere Intraklassenkorrelation von 0,22. In den von ihnen ausgewerteten Daten ist die Intraklassenkorrelation in großstädtischen Regionen höher als in ländlichen. Das könnte daraus resultieren, dass sich in großstädtischen Regionen Schulen mit mehr oder weniger privilegierten Einzugsbereichen ausbilden, und könnte auch für deutsche Verhältnisse zutreffen. Für die folgenden Berechnungen wird als Default der Intraklassenkorrelation $\rho = 0,20$ angenommen. Für May (2001) und Schröder-Lenzen/Mücke (2005), für die ebenfalls aus einer großstädtischen Region stammenden Daten von May (1994) und Hüttis-Graff (1998) sowie für Stichproben, in denen nicht normierte Messinstrumente eingesetzt wurden, wird abweichend davon $\rho = 0,25$ angesetzt. Die Berechnungen wurden entsprechend der Empfehlung von Hedges (2009) überprüft, indem sie mit anderen Defaults ($\rho = 0,15$, $\rho = 0,25$ und $\rho = 0,30$) wiederholt wurden. Dabei ergaben sich keine bedeutsamen Unterschiede in den Resultaten.

2.3 Berechnung von Effektstärken

Während die für die Berechnung von d_T benötigten Angaben über Mittelwerte meistens aus den ausgewerteten Studien zu entnehmen sind, fehlen wiederholt Angaben zur Streuung in den verglichenen *Lesen durch Schreiben*- und Fibelklassen. In einigen Fällen konnte die Streuung in der Gesamtstichprobe der Studie entnommen oder aus Testhandbüchern geschätzt werden; einmal (Herff 1993) musste sie aus einer Vergleichsuntersuchung imputiert werden. In anderen Fällen ließen sich die Effektstärke und ihre Varianz mit Hilfe der Probit-Methode schätzen (vgl. Sánchez-Meca et al. 2003). Soweit Effektstärken anhand der Originaldaten nur auf Klassenebene berechnet werden konnten, wurden sie unter Nutzung der von Hedges (2009, S. 341 u. 345) angegebenen Formeln in Effektstärken auf Probandenebene transformiert. In einem Fall (Fay 2010) beruhen die Angaben zur Effektstärke auf einer Reanalyse der in der Studie enthaltenen Originaldaten. Einzelheiten zum Vorgehen bei der Berechnung von Effektstärken und ihren Varianzen sowie zu deren numerischer Höhe enthält die genannte Datendokumentation.

Die ermittelten Effektstärken wurden für Ende Klassenstufe 1 einerseits und Klassenstufe 2-4 andererseits zusammengefasst. Bei längsschnittlichen Studien, die für ein- und dieselbe Stichprobe Ergebnisse aus mehreren Jahrgängen der Klassenstufen 2-4 enthalten, wurden die für die einzelnen Jahrgänge errechneten Effektstärken und ihre Varianzen aggregiert (vgl. Borenstein et al. 2009, 227 f.). Die für die Aggregation der Varianzen benötigten Korrelationen der Leistungen in folgenden Schuljahren wurden einheitlich mit 0,75 für unmittelbar aufeinander folgende Klassenstufen und 0,65 für mittelbar aufeinander folgende Klassenstufen angesetzt. Das entspricht sporadischen Angaben in den Studien (Einsiedler et al. 2002; May 2001) sowie den Handbüchern der verwendeten Tests (Marx 1998; May 1995) und ist vereinbar mit Empfehlungen in der Literatur (Borenstein et al. 2009). Bei der Interpretation der Ergebnisse sollte man im Auge behalten, dass weder die Effektstärken-Schätzungen für Klassenstufe 1 und 2-4 noch die für Lesen und Rechtschreiben stets voneinander unabhängig sind, da sie z. T. an den gleichen Stichproben gewonnen wurden.

Für die metaanalytische Prüfung wurde ein Modell fester Effekte zugrunde gelegt, weil im Sinne der gegebenen Fragestellung nicht vorausgesetzt wird, dass die in die Auswertung eingehenden *Lesen durch Schreiben*- und Fibelklassen zufällig gewählte Repräsentanten zugrunde liegender Grundgesamtheiten sind. Zur Kontrolle der metaanalytischen Berechnungen sowie zur grafischen Darstellung ihrer Ergebnisse wurde das frei zugängliche Paket *metafor* der Programmierumgebung R genutzt (Viechtbauer 2010).

3.1 Rechtschreiben

Die zu den Rechtschreibleistungen in Klassenstufe 1 vorliegenden Effektstärken sind in Abbildung 1 wiedergegeben. Die Stichproben werden darin nach Erstautor(in) der Studie bezeichnet, bei Bedarf ergänzt um eine Angabe zur Klassenstufe oder zum Erscheinungsjahr. Neben den Effektstärken (hervorgehobene Punkte) sind deren Standardfehler (Balken) ersichtlich. Die Größe des eine Effektstärke d repräsentierenden Punktes zeigt das Gewicht an, das die Stichprobe bei der Berechnung der mittleren Effektstärke hat. Es ist von dem Standardfehler s_d abhängig. Hier wie im Folgenden zeigen negative Werte von d einen Effekt zuungunsten der *Lesen durch Schreiben*-Klassen im Vergleich zu den Fibelklassen an, positive einen Effekt zugunsten der *Lesen durch Schreiben*-Klassen.

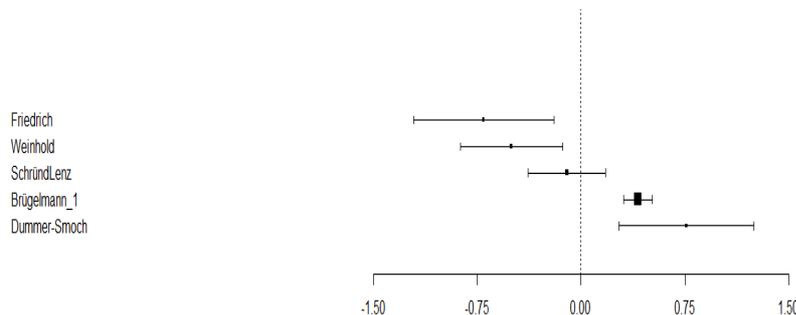


Abb. 1: Effektstärken zum Rechtschreiben in Klassenstufe 1.

Mittlere gewichtete Effektstärke $d = 0,28$; $s_d = 0,09$; $p < 0,01$;

Homogenitätsstatistik $Q(4) = 12,06$; $p = 0,02$.

Für die Rechtschreibleistungen am Ende der Klasse 1 errechnet sich eine mittlere gewichtete Effektstärke von $d_T = 0,28$ mit Standardfehler $s_{dT} = 0,09$. Sie weicht signifikant von 0 ab ($p < 0,01$). Man könnte auf den ersten Blick versucht sein, daraus zu schließen, dass der Vergleich zugunsten von *Lesen durch Schreiben* ausfällt. Jedoch ist, wie die im Kommentar zur Abbildung ausgewiesene Homogenitätsstatistik Q zeigt, kein einheitlicher Effekt gegeben ($Q(4) = 12,06$; $p = 0,02$). Die in den verschiedenen Stichproben gefundenen Effekte weichen so stark voneinander ab, dass man sie nicht als zufällige Varianten einer gemeinsamen Wirkgröße von ein- und

derselben Ausprägung auffassen kann. Die mittlere gewichtete Effektstärke ist damit inhaltlich nicht interpretierbar. Sie spiegelt vermutlich etwas anderes als die Wirkungen, die ein Erstunterricht nach *Lesen durch Schreiben* im ersten Schuljahr auf die Rechtschreibung hat.

In zwei der fünf Stichproben (Friedrich 2010; Weinhold 2006) wurde sichergestellt, dass die *Lesen durch Schreiben*- und die Fibelklassen mit gleichen kognitiven Eingangsvoraussetzungen in den Anfangsunterricht hineingegangen sind. Sie weisen von den anderen Stichproben deutlich abweichende, zuungunsten von Lesen durch Schreiben ausfallende Effektstärken auf. Um eine solche Beobachtung systematisch bewerten zu können, sollten jedoch insgesamt mindestens zehn Stichproben gegeben sein (Borenstein et al. 2009, S. 188).

Eine in dieser Größenordnung liegende Zahl von Stichproben liegt zur Rechtschreibung in Klassenstufe 2-4 vor (vgl. Abb. 2). Hier errechnet sich eine durchschnittliche gewichtete Effektstärke von $d_T = -0,26$ ($s_{dT} = 0,05$; $p < 0,01$). Die *Lesen durch Schreiben*-Klassen erreichten damit in Klassenstufe 2-4 signifikant schlechtere Rechtschreibleistungen als die Fibelklassen. Die angegebene durchschnittliche Effektstärke kann als Kennwert betrachtet werden, der alle Stichproben repräsentiert, da die Effektstärken sich als homogen darstellen ($Q(14) = 14,01$; $p = 0,45$).

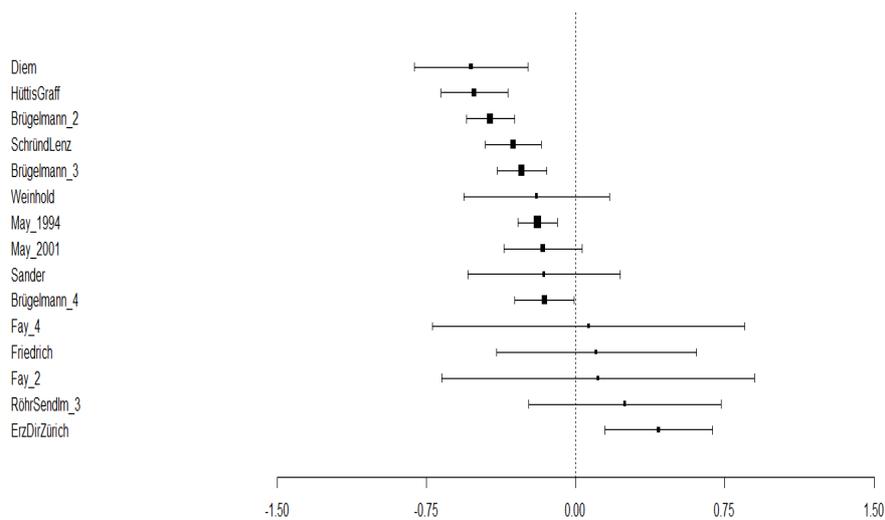


Abb. 2: Effektstärken zum Rechtschreiben in Klassenstufe 2-4.

Mittlere gewichtete Effektstärke $d = -0,26$; $s_d = 0,05$; $p < 0,01$;

Homogenitätsstatistik $Q(14) = 14,01$; $p = 0,45$.

Die unauffällige Homogenitätsstatistik ergibt sich allerdings auch aus den teilweise hohen Standardfehlern der Effektstärken, die ihrerseits aus der Notwendigkeit resultieren, die Intraklassenkorrelation zu berücksichtigen. Die Abbildung 2 zeigt an-

sonsten große Unterschiede, die im Einzelnen schwer zu erklären sind. Eine Stichprobe (Erziehungsdirektion des Kantons Zürich 1985) fällt aus dem Rahmen, da in ihr im Gegensatz zu allen anderen Stichproben die *Lesen durch Schreiben*-Klassen deutlich besser abschneiden als die Fibelklassen. Sie liegt am Rande der gesamten Stichprobenverteilung ($z = 2,45$). Über die Gründe lässt sich nur spekulieren. Fasst man eine metaanalytische Auswertung als Versuch der indirekten Messung eines Methodeneffekts auf, so würde es Sinn machen, diese Stichprobe als statistischen Ausreißer zu eliminieren (Rustenbach 2003). Wenn man das tut, setzt man voraus, dass es grundsätzlich möglich ist, aus den vorliegenden Effektstärken einen Effekt der Methode *Lesen durch Schreiben* zu ermitteln. Bei der hier gegebenen Fragestellung steht jedoch diese Voraussetzung selbst zur Diskussion. In dieser Situation ist das Auftreten von Ausreißer-Werten als solches der interessierende Befund, so dass eine Eliminierung nicht angezeigt ist. Bemerkenswert ist noch, dass eine andere Stichprobe, die aus demselben Zeitraum und ebenfalls aus der Schweiz stammt (Diem 1990), ein genau gegenteiliges Ergebnis zeigt.

Die in Abbildung 2 wiedergegebenen Ergebnisse legen nahe, dass die *Lesen durch Schreiben*-Klassen in den Klassenstufen 2-4 signifikant schlechtere Rechtschreibleistungen erreichen als die Fibelklassen. Ein weniger eindeutiges Bild ergibt sich aber, wenn man sich auf Stichproben beschränkt, bei denen die Eingangsvoraussetzungen kontrolliert wurden, entweder indem anhand von normierten Tests sichergestellt wurde, dass die verglichenen *Lesen durch Schreiben*- und Fibelkinder über gleiche kognitive Fähigkeiten oder gleiche Lernausgangslagen verfügen (Friedrich 2010; Röhr-Sendlmeier et al. 2007; Weinhold 2006), oder indem Effektstärken für Residuen einer Regression der Lernergebnisse auf den Ausgangsstand ermittelt wurden (Fay 2010; May 2001). Die mittlere gewichtete Effektstärke beträgt für diese Stichproben $d_T = -0,09$ ($s_{dT} = 0,15$; $p = 0,57$), ist also nicht signifikant von Null verschieden. Homogenität ist gegeben ($Q(5) = 0,92$; $p = 0,97$), und das stimmt auch mit dem überein, was der Augenschein nahe legt. Die Verringerung der ermittelten durchschnittlichen Effektstärke von $-0,26$ auf $-0,09$ ist eindrucksvoll. Jedoch erreicht der Unterschied der Effektstärken aus Stichproben mit und ohne Kontrolle der Eingangsvoraussetzungen wegen der hohen Standardfehler keine Signifikanz ($z = 1,15$; $p = 0,25$).

3 Ergebnisse

3.2 Lesen

Ergebnisse zum Lesen werden in den Primärstudien durchgängig unterschieden in solche, die die elementare Lesefertigkeit (Dekodieren) betreffen, und solche, die das Leseverstehen betreffen. Für die Zwecke der vorliegenden Arbeit werden diese Ergebnisse in einem einzigen Lese-Kennwert pro Stichprobe zusammengefasst. Der Grund ist, dass es bei den herangezogenen Stichproben oft schwierig ist, zu entscheiden, in welchem Umfang dokumentierte Leseleistungen dem Decodieren oder dem Leseverstehen zuzuordnen sind (so insbesondere bei dem wiederholt ein-

gesetzten Test *Knuspels Leseaufgaben*; Marx 1998). Die Dekodierfähigkeit steht zudem zum Leseverstehen in einer Beziehung, die sich im Laufe der Grundschulzeit stark ändert (Hoover/Gough 1990; Gough et al. 1996). Bei jüngeren und leeschwachen Kindern ist das Leseverstehen weitgehend durch sie determiniert. Bei älteren und lesestärkeren Kindern ist es weitgehend von ihr unabhängig. Im Fazit wurde auf den Versuch, die in den ausgewerteten Studien gemessenen Leseleistungen dem Dekodieren oder dem Leseverstehen zuzuordnen, verzichtet, und beides wurde in einem einzigen Kennwert verbunden.

Bei den am Ende von Klassenstufe 1 erreichten Leseleistungen beträgt die gewichtete mittlere Effektstärke $d_T = -0,27$ zuungunsten von *Lesen durch Schreiben*. Sie weicht signifikant von Null ab ($s_{dT} = 0,14$; $p = 0,04$; vgl. Abb. 3). Die Effektstärken sind homogen ($Q(3) = 1,00$; $p = 0,80$). Die Studien, in denen die kognitiven Eingangsvoraussetzungen kontrolliert wurden (Friedrich 2010; Poerschke 1999), erbringen im Mittelfeld liegende Ergebnisse.

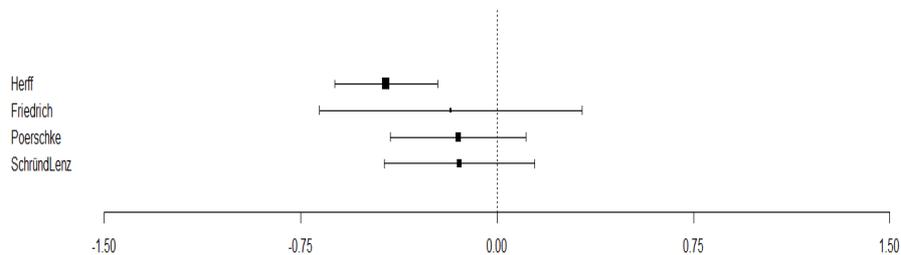


Abb. 3: Effektstärken zum Lesen in Klassenstufe 1.

Mittlere gewichtete Effektstärke $d = -0,28$; $s_d = 0,14$; $p = 0,04$;

Homogenitätsstatistik $Q(3) = 1,00$; $p = 0,80$.

Wie die Abbildung 3 allerdings zeigt, ist die Effektstärke stark bestimmt durch das Ergebnis der Studie von Herff (1993), bei der der für die Berechnung von d benötigte Streuungswert sich nur auf indirektem Weg schätzen ließ. Setzt man höhere, immer noch plausible Schätzungen dieses Streuungswerts an, so würde die aus allen Stichproben gegen Ende der Klasse 1 für das Lesen berechnete Effektstärke nicht mehr signifikant von Null verschieden sein. Die Homogenität der Effekte bliebe jedoch erhalten. Das Gesamtbild, das die Studien zeigen, ist also einheitlich, aber hinsichtlich der Frage, ob man wirklich von einem signifikanten Rückstand der *Lesen durch Schreiben*-Klassen sprechen kann, nicht zwingend eindeutig.

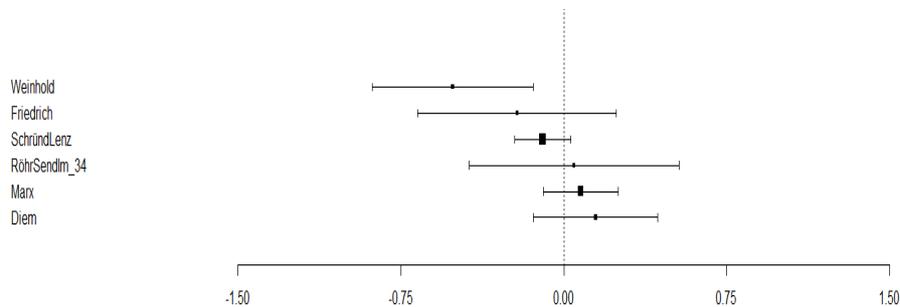


Abb. 4: Effektstärken zum Lesen in Klassenstufe 2-4.

Mittlere gewichtete Effektstärke $d = -0,05$; $s_d = 0,09$; $p = 0,60$;
 Homogenitätsstatistik $Q(5) = 2,75$; $p = 0,74$.

Die in Abbildung 4 wiedergegebenen, zur Leseleistung in Klassenstufe 2-4 ermittelten Effekte sind homogen ($Q(5) = 2,75$; $p = 0,74$) und es errechnet sich eine mittlere gewichtete Effektstärke von $d_T = -0,05$ ($s_{dT} = 0,09$; $p = 0,60$). Zwischen *Lesen durch Schreiben*-Klassen und Fibelklassen sind demnach im Lesen keine Unterschiede feststellbar. Von den beiden Studien mit Kontrolle der kognitiven Eingangsvoraussetzungen (Friedrich 2010; Weinhold 2006) weist die zweite eine deutlich unter den anderen liegende Effektstärke auf; die erste liegt im unauffälligen Bereich. Zudem erscheint auch bei anderen Stichproben aus Abbildung 4 eine systematische Begünstigung der *Lesen durch Schreiben*-Klassen durch bessere Eingangsvoraussetzungen wenig wahrscheinlich, so etwa bei der Stichprobe von Marx (2000), die aus einer Normierungsuntersuchung stammt.

3.3 Kommunikative Nutzung des Schreibens

Da die Kinder bei *Lesen durch Schreiben* das Schreiben von Anfang an als eine sinnvolle kommunikative Tätigkeit kennen lernen, liegt die Vermutung nahe, dass sie ein positiveres Verhältnis zum Schreiben entwickeln als Fibelkinder. Ob das zutrifft, scheint jedoch nie systematisch untersucht worden zu sein. Eine der ausgewerteten Studien (May 1994) schloss eine Erhebung zur Textqualität in Klassenstufe 4 ein, ohne aber die Ergebnisse für *Lesen durch Schreiben*- und Fibelklassen getrennt auszuweisen. In einigen Studien (vgl. Datendokumentation im Internet) finden sich Angaben zur Länge freier Schülertexte, gemessen an der Zahl der Wörter. Die mittlere gewichtete Effektstärke beträgt in Klassenstufe 1 $d_T = 0,32$, in den Klassenstufen 2-4 aber $d_T = -0,16$. Die Texte waren also in der Klassenstufe 1 bei den untersuchten *Lesen durch Schreiben*-Klassen länger als bei den Fibelklassen, nicht jedoch in der Klassenstufe 2-4. Da die Länge freier Texte in hohem Maß aufgaben- und situationsabhängig sein dürfte und eine Schätzung der Intraklassenkorrelation für sie spekulativ bleibt, wird keine Prüfstatistik berechnet.

3.4 Besondere Schülergruppen

Es wird angenommen, dass Schülerinnen und Schüler mit ungünstigen Lernvoraussetzungen in offenen Formen des Erstleseunterrichts geringere Lernerfolge erreichen als in Unterrichtsformen mit explizit instruierendem Charakter (Stahl/Miller 1989). Daten dazu enthalten die Studien von Poerschke (1999), der den Lernstand im Lesen Ende Klasse 1 erfasste, und von May (2001), der den Lernfortschritt im Rechtschreiben bis Klassenstufe 4 erhob. Für Poerschkes Stichprobe beträgt die Effektstärke in der Gruppe der Probanden mit unterdurchschnittlichen kognitiven Voraussetzungen $d_T = -0,95$ gegenüber $d_T = -0,15$ in der Gesamtgruppe. Aus den bei May angegebenen Korrelationen errechnet sich für lese-rechtschreibschwache Schülerinnen und Schüler eine Effektstärke von $d_T = -0,29$ gegenüber $d_T = -0,16$ in der Gesamtgruppe. Der Unterschied ist im Fall der Stichprobe von Poerschke signifikant.

Denkbar ist, dass Effekte sich in der Gruppe der zweitsprachlichen Schülerinnen und Schüler in ähnlicher Weise zuungunsten von *Lesen durch Schreiben* verschieben. Studien, die Daten dazu enthalten (Erziehungsdirektion des Kantons Zürich 1985; Friedrich 2010; Schründer-Lenzen/Mücke 2005 für Rechtschreiben; Friedrich 2010; Herff 1993; Marx 2000; Schründer-Lenzen/Mücke 2005 für Lesen), scheinen für diese Vermutung zu sprechen. Die folgende Tabelle 1 gibt einen Überblick über die aus diesen Studien berechneten mittleren gewichteten Effektstärken.

Tab. 1: Mittlere gewichtete Effektstärken für Schülerinnen und Schüler mit Deutsch als Erstsprache und für Schülerinnen und Schüler mit nichtdeutscher Erstsprache

| | Rechtschreibung | | Lesen | |
|-------------------------------|-----------------|---------|-------|---------|
| | Kl. 1 | Kl. 2-4 | Kl. 1 | Kl. 2-4 |
| Deutsch ist Erstsprache | 0.31 | 0.08 | 0.08 | -0.02 |
| Deutsch ist nicht Erstsprache | -0.33 | 0.10 | -0.72 | -0.18 |

Aus dem Rahmen fällt die für Rechtschreibung in Klassenstufe 2-4 errechnete Effektstärke bei Nicht-Erstsprachlern. Sie ist stark durch die Studie der Erziehungsdirektion Zürich (1985) bestimmt, deren Ergebnisse hier (wie insgesamt) eine Sonderstellung einnehmen. Ansonsten zeigt die Tabelle eine Auswahl von Stichproben, in denen die erstsprachlichen Schülerinnen und Schüler der *Lesen durch Schreiben*-Klassen im Wesentlichen die gleichen schriftsprachliche Leistungen erbringen wie die der Fibelklassen, während die zweitsprachlichen Schülerinnen und Schüler in denselben Klassen schlechtere Ergebnisse erreichen.

Es ist allerdings fraglich, ob die Schülerinnen und Schüler, die hier als ‚zweitsprachlich‘ zusammengefasst sind, sehr viel miteinander gemein haben. Zu ihnen gehören Probanden aus älteren Studien, die dort als ‚ausländisch‘ bezeichnet werden (Erziehungsdirektion des Kantons Zürich 1985; Herff 1993). Dieses Merkmal muss sich (insbesondere bei der Züricher Stichprobe) mit dem Merkmal ‚zweitsprachlich‘

nicht decken. Auch Probanden aus neueren Studien, die als Schülerinnen und Schüler mit Migrationshintergrund bezeichnet werden, bilden kaum eine homogene Gruppe. Wegen der anzunehmenden Heterogenität der in Tabelle 1 als nicht-erstsprachlich geführten Gruppen wird keine Prüfstatistik berechnet. Sie könnte eine Präzision vortäuschen, die nicht besteht.

3.5 Moderatoren

Den Einfluss von Moderatoren, d. h. Merkmalen der Stichproben, auf die gefundenen Effektstärken systematisch zu untersuchen, würde die Datenbasis überstrapazieren. Die folgenden Überlegungen zielen nicht darauf, inhaltliche Erkenntnisse zu gewinnen, sondern forschungsmethodologische Fragen herauszustellen. Im Blick darauf wurden die Stichproben nach drei Kriterien klassifiziert, die die Validität der Schätzung einer Effektstärke für sie betreffen. Zu diesen Kriterien gehören die bereits angesprochene Kontrolle der kognitiven Lernvoraussetzungen (Kriterium *Eingangsvoraussetzungen*), die Erhebung von Lernergebnissen mittels normierter Erhebungsverfahren (Kriterium *Messinstrument*) und die Zuverlässigkeit der Zuordnung von Klassen zu *Lesen durch Schreiben* oder Fibelunterricht.² Es sei hervorgehoben, dass die drei Kriterien die Validität der hier vorgenommenen Berechnung einer Effektstärke aus der jeweiligen Stichprobe betreffen, nicht aber unbedingt Rückschlüsse zulassen auf die Validität der Studien selbst, aus denen die Daten der Stichprobe entnommen wurden. Die Studien zielen z. T. gar nicht auf eine Evaluation von *Lesen durch Schreiben*. Wenn sie Stichproben enthalten, zu denen eine nur wenig valide Effektstärkenschätzung möglich ist, können sie dennoch mit Blick auf ihre eigentliche Fragestellung valide sein.

Für die im Blick auf die Rechtschreibung in Klassenstufe 2-4 berechneten Effekte wurde oben bereits festgestellt, dass die Effektstärke geringer ist in Stichproben, bei denen die Eingangsvoraussetzungen kontrolliert wurden, als in anderen Stichproben. Der Unterschied zwischen beiden Gruppen von Stichproben ist jedoch nicht signifikant. Gleiches lässt sich mit Blick auf die beiden anderen Kriterien (*Messinstrument* und *Zuordnungsvalidität*) sagen. Die jeweils niedrigere gewichtete mittlere Effektstärke findet sich in den Stichproben, in denen eine Effektstärkenschätzung mit größerer Validität möglich ist. Das sind die Stichproben, bei denen als Messinstrument ein genormter Test verwendet wurde, bzw. die Stichproben, bei denen die Zuordnung von Klassen zu *Lesen durch Schreiben* oder Fibelunterricht besser gesichert ist. Signifikanz erreichen die Unterschiede jedoch nicht, auch nicht bei gleichzeitiger Berücksichtigung aller drei Validitätskriterien.

Fasst man die in Klasse 1-4 zum Rechtschreiben und zum Lesen ermittelten Effekte stichprobenweise zu jeweils einem Effekt zusammen, der als globales Maß für den

2 Kriterium *Zuordnungsvalidität*; zu den genauen Definitionen der drei Kriterien sowie den Zuordnungen der einzelnen Stichproben zu Ausprägungsgraden der Kriterien vgl. Datendokumentation

Unterschied der schriftsprachlichen Lernergebnisse in *Lesen durch Schreiben*- und Fibelklassen interpretiert werden kann, so erhält man insgesamt 21 Effektstärken. Bemerkenswert ist, dass diese Effektstärken sich in jedem Fall als untereinander inhomogen darstellen, wenn nur die Stichproben mit einer nach einem der drei Kriterien weniger validen Schätzung berücksichtigt werden, dagegen als homogen, wenn nur die Stichproben mit jeweils eher valider Schätzung einbezogen sind. Die Validität der Schätzung der Effektstärke in einer Stichprobe könnte also eine Rolle dafür spielen, in welchem Umfang die Schätzung durch Zufallseinflüsse bestimmt wird.

4 Diskussion

Im Folgenden werden die Ergebnisse zunächst unter dem Aspekt diskutiert, inwieweit die Befunde zu den Lernergebnissen bei *Lesen durch Schreiben* zu einem interpretierbaren, theoretisch konsistenten Bild seiner Lernwirksamkeit führen könnten. Im Anschluss wird die für die vorliegende Bestandsaufnahme zentrale Frage angesprochen, welche forschungsmethodologischen Konsequenzen sich ergeben.

Zu einem inhaltlich konsistenten Bild scheinen die Befunde zum Lesen zu führen. Während die Leistungen der *Lesen durch Schreiben*-Klassen gegen Ende der Klassenstufe 1 unter denen der Fibelklassen liegen, weichen sie in Klassenstufe 2-4 nicht erkennbar von denen der Fibelklassen ab. Soweit diese Ergebnisse einen Rückschluss auf die realen Lesefähigkeiten zulassen, könnten sie als Hinweis auf einen vorübergehenden Rückstand der *Lesen durch Schreiben*-Kinder aufgefasst werden, der sich daraus ergibt, dass bei ihnen der Leselernprozess in Klasse 1 nicht forciert wird. Die Hypothese Reichens, Lesen könne gelernt werden, ohne dass es gelehrt wird, scheint dagegen auf längere Sicht Bestand zu haben.

Im Rechtschreiben scheinen die untersuchten *Lesen durch Schreiben*-Klassen auf den ersten Blick gegen Ende der Klassenstufe 1 signifikant bessere Leistungen zu erreichen als die Fibelklassen. Diese Feststellung muss überraschen angesichts des Umstands, dass ein systematischer Unterricht in Rechtschreibung im Erstunterricht nach *Lesen durch Schreiben* gar nicht vorgesehen ist. Jedoch sind die gefundenen Effekte nicht homogen.

In diesem Zusammenhang ist aufschlussreich, dass bei zwei längsschnittlich verfolgten Stichproben, in denen die *Lesen durch Schreiben*-Klassen gegen Ende Klasse 1 schlechter abschnitten als die Fibelklassen, gegen Mitte der Klasse 1 noch eine Überlegenheit in der Rechtschreibung bestanden hatte (Schründer-Lenzen/Mücke 2005 und Weinhold 2006; ähnlich Kirschhock 2003 für Klassen mit „entwicklungsorientiertem Unterricht“). In beiden Studien wurde zur Erfassung der Rechtschreibleistung die Zahl der Graphemtreffer nach HSP (May 2002) herangezogen (bei Kirschhock war es eine frei gestellte Schreibaufgabe). Es ist ohne weiteres plausibel, dass Kinder, die rein phonetisch orientiert schreiben, in einem frühen Stadium eine höhere Zahl von richtigen Buchstaben pro Wort erreichen als Kinder, die sich über die orthographisch richtige Schreibung Gedanken machen. Denkbar ist darüber hinaus, dass diese Kinder bei Tests seltener als andere einen Schreibversuch verweigern. Die auf diese Weise zustande kommende frühe Rechtschreibleistung hat

jedoch möglicherweise nur eine geringe prognostische Aussagekraft für die spätere Rechtschreibung. Die Überlegenheit, die die *Lesen durch Schreiben*-Klassen mancher Stichproben im Rechtschreiben noch gegen Ende der Klassenstufe 1 zeigten, könnte ein reales und durchaus charakteristisches Merkmal dieser Klassen widerspiegeln, aber nicht eine Rechtschreibleistung in dem Sinne, wie sie in späteren Klassenstufen und in genormten Testverfahren verstanden wird. Bei der eigentlichen Rechtschreibleistung dagegen könnte es so sein, dass sie sich in der Mitte der Klassenstufe 1 noch nicht so konsolidiert hat, dass sie zuverlässig messbar ist. Jedenfalls scheint die Art der gestellten Schreibaufgabe das Ergebnis einer Messung zu diesem Zeitpunkt nachhaltig zu beeinflussen. In der Studie von Brügelmann et al. (1994) korrelieren die Rechtschreibleistungen im Diktat und in einer freien Schreibaufgabe in Klassenstufe 1 deutlich weniger als in den Klassenstufen 2-4.

Der beim Rechtschreiben in den Klassenstufen 2-4 zentrale Befund ist, dass die in den *Lesen durch Schreiben*-Klassen erreichten Lernergebnisse signifikant unter denen der Fibelklassen liegen, dass die Unterlegenheit jedoch nicht mehr erkennbar ist, wenn man sich auf Stichproben beschränkt, in denen eine Erklärung unterschiedlicher Lernergebnisse durch unterschiedliche kognitive Lernausgangslagen ausgeschlossen werden kann. Dieser Befund ist mehrdeutig. Wenn zwischen den Rechtschreibleistungen von *Lesen durch Schreiben*-Klassen und Fibelklassen mit gleichen kognitiven Eingangsvoraussetzungen kein signifikanter Unterschied besteht, bedeutet das ja nicht zwingend, dass die Leistungen gleich sind. Auch dass die Effekte, soweit sie sich bestätigen würden, im Sinne gängiger Klassifikationsschemata (vgl. etwa Bortz/Döring 2006) als gering einzustufen sein würden, besagt nicht viel. Denn es geht um langfristige Auswirkungen eines Unterrichtsansatzes, der in Klasse 1 praktiziert wurde, auf Leistungen bis hin zur Klassenstufe 4. Solche Auswirkungen anhand der gleichen Maßstäbe zu beurteilen, wie man sie für kurzfristige Effekte nutzt, wäre wenig überzeugend. Im Ergebnis muss man dennoch sagen, dass eine schlechtere Rechtschreibleistung von *Lesen durch Schreiben*-Klassen gegenüber Fibelklassen in Klassenstufe 2-4 bei gleichen Eingangsvoraussetzungen derzeit nicht belegt ist.

Gegen diese Schlussfolgerung liegt folgender Einwand nahe: Falls Methodeneffekte in Klassenstufe 2-4 in der Wirklichkeit nicht bestehen, so müssten sie sich, wenn sie in den Daten auftreten, zufällig verteilen. Es müssten, mit anderen Worten, Unterschiede im Rechtschreiben mal zugunsten von *Lesen durch Schreiben*, mal zugunsten von Fibeln ausfallen. Die tatsächlich gefundenen Ergebnisse legen jedoch ein anderes Bild nahe. Wie Abbildung 2 zeigt, geht der Vergleich der Rechtschreibleistungen in den Klassenstufen 2-4 bei zehn Stichproben zuungunsten von *Lesen durch Schreiben* aus und nur bei fünf Stichproben zu dessen Gunsten.

Dieser Einwand ist jedoch nicht zwingend, da sämtliche vorliegenden Untersuchungen observational sind. Das bedeutet, dass sie Klassen untersuchen, deren Lehrkräfte sich selbst einer der Unterrichtsmethoden *Lesen durch Schreiben* oder *Fibel* zugeordnet hatten. Lehrkräfte, die eine Entscheidung für *Lesen durch Schreiben* treffen, könnten sich jedoch von sonstigen Lehrkräften zusätzlich in anderen Punkten unterscheiden. In den Untersuchungen von May (1994; 2001) gaben sie seltener als Fi-

bellehrkräfte an, dass Rechtschreibsicherheit für sie ein wichtiges Unterrichtsziel sei, und legten stattdessen mehr Wert darauf, dass die Kinder Freude am Schreiben gewinnen. Denkbar ist auch, dass *Lesen durch Schreiben*-Lehrkräfte ihrem schulischen Umfeld nach keine Zufallsauswahl darstellen. So konnte Sander (2006) für ihre im rheinischen Raum durchgeführte Untersuchung *Lesen durch Schreiben*-Klassen nur in öffentlichen Gemeinschaftsschulen, nicht in konfessionellen Schulen finden. Aus diesen Gründen muss man mit der Möglichkeit rechnen, dass die *Lesen durch Schreiben*-Klassen der untersuchten Stichproben keine (im statistischen Sinne) ‚zufälligen‘ Realisierungen der Methode darstellen, sondern dass zwischen ihnen übergreifende Gemeinsamkeiten bestehen, die den gehaltenen Unterricht gleichsinnig beeinflussen.

Befunde aus einzelnen Studien legen nahe, dass *Lesen durch Schreiben* für Schülerinnen und Schüler mit ungünstigen Lernvoraussetzungen, möglicherweise auch für zweitsprachliche Schülerinnen und Schüler, keine optimalen Lernwege bereitstellt. Die Resultate sind jedoch nicht einheitlich, so dass weiterer Forschungsbedarf besteht.

Obwohl zur Rechtschreibung in *Lesen durch Schreiben*-Klassen deutlich mehr Studien existieren als zum Lesen, bleiben im Fazit gerade im Blick auf die Rechtschreibung mehr Fragen offen als im Blick auf das Lesen. Das heißt nicht, dass es unmöglich ist, aus den Daten inhaltliche Aufschlüsse zu gewinnen, die als Mosaiksteine in ein insgesamt theoretisch konsistentes Bild eingehen können. *Lesen durch Schreiben* vermittelt den Kindern ein phonographisches Ausgangsmodell vom Funktionieren der Schrift. Das heißt, es legt die Vorstellung nahe, die Schrift beruhe auf einer Zugum-Zug-Zuordnung von gesprochenen Lauten zu Buchstaben. Eine solche Vorstellung stellt eine Simplifizierung dar, könnte jedoch hinreichend sein, um einen Einstieg in das Lesen zu finden, der einen selbsttragenden Lernprozess in Gang setzt (*self-teaching hypothesis*, vgl. Share 1995). Wenn Kinder ein unbekanntes Wort rekodierend zu erlesen versuchen, so vermögen sie aus dem Ergebnis selbständig und unmittelbar eine Information darüber zu gewinnen, ob ihr Versuch erfolgreich war. Auf diese Situation stellt die *self-teaching hypothesis* Shares ab. Sie ist alles andere als klar formuliert, besagt aber im Kern, dass das schrittweise rekodierende Erlesen eines unbekanntes Wortes den Erwerb der Fähigkeit fördert, das Wort fortan anhand orthographischer Merkmale zu erkennen. In diesem Sinne wird sie durch empirische Untersuchungen zum frühen Lesen gestützt (Cunningham 2006; Share 1999), mit der Einschränkung, dass *self-teaching* beim Lesen neben erfolgreichem Dekodieren möglicherweise zusätzlich orthographische Fähigkeiten voraussetzt (Connors et al. 2011; Cunningham et al. 2002). Die *self-teaching hypothesis* steht in enger Beziehung zum Modell der Entwicklung des Sichtwortlesens nach Ehri (2005), das die Entwicklung des Worterkennens beim Lesen nicht als Wechsel von einer alphabetischen zu einer orthographischen Lesestrategie versteht, sondern als zunehmende Entfaltung und Ausdifferenzierung des alphabetischen Lesens. Beide Modelle beziehen sich auf den Erwerb des orthographischen Wissens, welches zum Lesen (*word identification*) benötigt wird. Sie können nicht unbesehen auf das orthographische Wissen übertragen werden, das zum Rechtschreiben (*spelling*) benötigt

wird. Wenn Kinder sich mit einem unbekanntem Wort auseinandersetzen, indem sie es rekodierend erlesen, ist der Effekt auf Rechtschreiben geringer als der auf Lesen (Conrad 2007). Umgekehrt ist der Effekt auf Rechtschreiben größer, wenn sie sich mit einem unbekanntem Wort auseinandersetzen, indem sie es schreiben, als indem sie es lesen (Shahar-Yames/Share 2008).

Das wichtigste Fazit, das sich insgesamt ziehen lässt, ist, dass die Aussagekraft der vorliegenden empirischen Studien zu *Lesen durch Schreiben* durch methodologische Beschränkungen begrenzt wird. Dabei lassen sich drei Punkte hervorheben. Erstens: Das durchgehend observationale Vorgehen bei der Evaluation der Lernergebnisse in *Lesen durch Schreiben*-Klassen könnte die Befunde dadurch beeinflusst haben, dass die untersuchten *Lesen durch Schreiben*-Klassen sich außer durch die Erstunterrichtsmethode auch durch Merkmale der Lehrkräfte von Fibelklassen unterscheiden. Zweitens: Wie wiederholt auftretende inhomogene Statistiken und Ausreißer-Werte nahelegen, könnte eine eingeschränkte Validität der Ermittlung der Lernergebnisse in *Lesen durch Schreiben*- und Fibelklassen das Zustandekommen von Zufallsbefunden befördert haben. Drittens: In den vorliegenden Studien werden Rechtschreibleistungen evaluiert, in einem geringeren Umfang auch Leseleistungen und peripher der Umfang geschriebener Texte. Keine vorliegende Studie unternimmt es aber, *Lesen durch Schreiben* systematisch an seinem eigenen Anspruch zu messen, der darin besteht, dass Kinder, die selbstgesteuert lesen und schreiben lernen, ein positiveres Verhältnis zur Nutzung der Schriftsprache gewinnen als andere.

Abschließend sollen die methodologischen Beschränkungen, denen die vorliegende Bestandsaufnahme ihrerseits unterliegt, noch einmal zusammenfassend aufgeführt werden. Erstens wurden alle aufgefundenen Studien zu *Lesen durch Schreiben* in die Auswertung einbezogen, ohne eine Auswahl anhand von Ausschlusskriterien vorzunehmen. Das praktizierte nicht-selektive Vorgehen bedeutet, dass man aus den Ergebnissen der vorliegenden Bestandsaufnahme Rückschlüsse auf die Forschungslage ziehen kann, nicht aber unbesehen auf die Lernwirksamkeit von *Lesen durch Schreiben* schließen sollte. Zweitens wurden Effektstärken zum Teil auf der Grundlage von externen Angaben zu Stichprobenstreuungen geschätzt. Eine Überprüfung mittels Metaregression zeigt allerdings, dass die Stichproben, in denen so vorgegangen wurde, zu keinen von den anderen signifikant verschiedenen Effektstärkenschätzungen führen ($z = 0,37$; $p = 0,71$). Gleichwohl bleibt in einem Fall (Herff 1993) eine Unsicherheit, die aus der Imputation einer Stichprobenstreuung resultiert. Drittens war schwer sicherzustellen, dass alle einschlägigen Studien berücksichtigt werden, da die bibliographische Erschließung unzureichend ist. Ein zusätzliches, bei Metanalysen generell bestehendes Risiko ergibt sich daraus, dass nach einer häufig geübten, dem Erkenntnisfortschritt aber wenig zuträglichen Praxis Autor(inn)en nicht signifikante Ergebnisse unpubliziert lassen oder Zeitschriften sie nicht veröffentlichen. Für observationale Studien wie die ausgewerteten könnte das allerdings weniger zu befürchten sein als für experimentelle, da diese Studien nicht unbedingt in konfirmatorischer Absicht durchgeführt werden. Anzumerken bleibt schließlich, dass man nicht nur bei *Lesen durch Schreiben*-, sondern auch bei Fibelklassen kaum unterstellen kann, dass die Lernergebnisse stets untereinander homogen und insbe-

sondere von der Unterrichtsqualität unabhängig sind. Wie bereits zum Ausdruck gebracht wurde, bedeutet die unternommene Auswertung von Lernergebnissen bei *Lesen durch Schreiben* nicht, dass die größere Bedeutung der Unterrichtsqualität gegenüber der Unterrichtsmethode in Frage gestellt wird. Wenn Methoden selbst die Lernwirksamkeit des Erstunterrichts auch nur in engen Grenzen bestimmen, so ist die Diskussion über sie für Fachdidaktiker(innen) wie für Lehrkräfte gleichwohl unausweichlich – unter anderem, da auf dem Weg über sie Qualitätsfragen zugänglich werden.

Literatur

- Albert, R., Roder, A., Rokitzki, C. & Teepker, F. (2009). Alphabetisierung von erwachsenen Einwanderern. Methodisches Vorgehen bei der Evaluation von Lehrmethoden. *Report. Zeitschrift für Weiterbildungsforschung* 32, Heft 4. S. 43-54.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009). *Introduction to meta-analysis*. New York: Wiley.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation*. 4. Aufl. Heidelberg: Springer.
- Bredel, U. (2012). (Verdeckte) Probleme beim Orthographieerwerb des Deutschen in mehrsprachigen Klassenzimmern. In W. Griebhaber & Z. Kalkan (Hrsg.). *Orthographie- und Schriftspracherwerb bei mehrsprachigen Kindern*. Freiburg: Fillibach. S. 125-142.
- Brügelmann, H. & Brinkmann, E. (2012). Freies Schreiben im Anfangsunterricht? Eine kritische Übersicht über Befunde der Forschung. *leseforum.ch* 2/2012. S. 1-22. www.leseforum.ch/bruegelmann_brinkmann_2012_2.cfm
- * Brügelmann, H., Hengartner, E. & Reichen, J. (1994). Richtig schreiben durch freies Schreiben? In H. Brügelmann & S. Richter (Hrsg.). *Wie wir recht schreiben lernen*. Lengwil: Libelle. S. 135-148.
- Connors, F. A., Loveall, S. J., Moore, M. S., Hume, L. E. & Maddox, C. D. (2011). An individual differences analysis of the self-teaching hypothesis. *Journal of Experimental Child Psychology* 108. S. 402-410.
- Conrad, N. J. (2007). From reading to spelling and spelling to reading: Transfer goes both ways. *Journal of Educational Psychology* 100. S. 869-878.
- Cunningham, A. E. (2006). Accounting for children's orthographic learning while reading text: Do children self-teach? *Journal of Experimental Child Psychology* 95. S. 56-77.
- Cunningham, A. E., Perry, K. E., Stanovich, K. & Share, D. (2002). Orthographic learning during reading: examining the role of self-teaching. *Journal of Experimental Child Psychology* 82. S. 185-199.
- * Diem, M. (1990). *Evaluation eines Legasthenie-Prophylaxe-Kurses für Lehrerinnen*. Dissertation Universität Zürich.
- * Dummer-Smoch, L. (1995). Vergleichsuntersuchungen zum Erfolg im Leselernprozess. In W. Niemeyer (Hrsg.). *Kommunikation und Lese-Rechtschreibschwäche*. Bochum: Winkler. S. 135-138.

- Ehri, L. C. (2005). Development of sight word reading: phases and findings. In M. J. Snowling & C. Hulme (Hrsg.). *The science of reading*. Malden, MA: Blackwell. S. 135-154.
- Einsiedler, W., Frank, A., Kirschhock, E.-M., Martschinke, S. & Treinies, G. (2002). Der Einfluss verschiedener Unterrichtsmethoden auf die phonologische Bewusstheit sowie auf Lese- und Rechtschreibleistungen im 1. Schuljahr. *Psychologie in Erziehung und Unterricht* 49. S. 194-209.
- * Erziehungsdirektion des Kantons Zürich (1985). Erstlesemethode und Rechtschreibkompetenz. In J. Reichen. *Lesen durch Schreiben*. Heft 1. 3. Aufl. Zürich: Sabe 1988. S. 58-59.
- * Fay, J. (2010). *Die Entwicklung der Rechtschreibkompetenz beim Textschreiben*. Frankfurt a. M.: Peter Lang.
- * Friedrich, K. (2010). *Unterrichtskonzept und Schriftspracherwerb*. Dissertation Pädagogische Hochschule Heidelberg.
- Gough, P. B., Hoover, W. A. & Peterson, C. L. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Hrsg.). *Reading comprehension difficulties*. Mahwah, NJ: Erlbaum. S. 1-13.
- Hanke, P. (2005). *Öffnung des Unterrichts in der Grundschule*. Münster: Waxmann.
- Hedges, L. V. (2009). Effect sizes in nested designs. In H. Cooper, L. V. Hedges & J. C. Valentine (Hrsg.). *The handbook of research synthesis and meta-analysis*. 2nd edition. New York, NY: Russell Sage Foundation. S. 337-356.
- Hedges, L. V. & Hedberg, E. C. (2007). Intraclass correlations for planning group randomized experiments in rural education. *Journal of Research in Rural Education* 22. S. 1-15.
- Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität*. 2. Aufl. Seelze: Klett Kallmeyer.
- * Herff, I. M. (1993). *Die Gestaltung des Leselernprozesses als elementare Aufgabe der Grundschule – neuere Entwicklungen und gegenwärtige Situation an den Grundschulen des Regierungsbezirks Köln – ein Beitrag zur pädagogischen Tatsachenforschung*. Dissertation Universität zu Köln.
- Hoover, W. A. & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal* 2. S. 127-160.
- * Hütts-Graff, P. (1998). *Rechtschreiblernen und Unterricht: Der Blick auf die Klassen*. In C. Osburg (Hrsg.). *Textschreiben, Rechtschreiben, Alphabetisierung*. Baltmannsweiler: Schneider. S. 44-71
- Kirschhock, E.-M. (2003). *Die Entwicklung schriftsprachlicher Kompetenzen im ersten Schuljahr*. Dissertation Friedrich-Alexander-Universität Erlangen-Nürnberg.
- Küspert, P. & Schneider, W. (1998). *Würzburger Leise Leseprobe (WLLP)*. Göttingen: Hogrefe.
- Marx, H. (1998). *Knuspels Leseaufgaben (KNUSPEL-L)*. Göttingen: Hogrefe.
- * Marx, H. (2000). Einfluss von Lehrmethoden und Hörverstehen auf das Lesenlernen im Grundschulalter. In R. Duit & C. von Rhöneck (Hrsg.). *Ergebnisse fachdidaktischer und psychologischer Lehr-Lernforschung*. Kiel: IPN. S. 191-231.

- * May, P. (1994). Rechtschreibfähigkeit und Unterricht. Rechtschreibleistungen Hamburger Schüler/innen im 4. Schuljahr in Zusammenhang mit Merkmalen schriftsprachlichen Unterrichts. Hamburg: Universität Hamburg.
- May, P. (1995). Hamburger Schreibprobe zur Erfassung der grundlegenden Schreibstrategien. Hamburg: vpm.
- * May, P. (2001). Lernförderlicher Unterricht. Band 1. Frankfurt a. M.: Pater Lang.
- May, P. (2002). Hamburger Schreibprobe zur Erfassung der grundlegenden Schreibstrategien. Neustandardisierung 2001. Hamburg: vpm.
- Metze, W. (2008). Lernwegorientierter Schriftspracherwerb im Spiegel der Empirie und des Schulalltags.
http://www.wilfriedmetze.de/Vortrag_Zurich_31.5.08.pdf
- Mücke, S. (2006). Vorhersagestabilität von Kontextbedingungen auf die basalen Leseleistungen. In A. Schröder-Lenzen (Hrsg.). Risikofelder kindlicher Entwicklung. Berlin: Springer. S. 87-108.
- * Poerschke, J. (1999). Anfangsunterricht und Lesefähigkeit. Münster: Waxmann.
- Reichen, J (1988). Lesen durch Schreiben. 3. Aufl. Zürich: sabe.
- Röber-Siekmeyer, C. & Spiekermann, H. (2000). Die Ignorierung der Linguistik in der Theorie und Praxis des Schriftspracherwerbs. Zeitschrift für Pädagogik 46. S. 753-771.
- * Röhr-Sendlmeier, U. M., Wagner, H. & Götze, I. (2007). Die Auswirkungen unterschiedlicher Didaktiken und elterlicher Anregungen auf den Orthographieerwerb im Grundschulalter. *Bildung und Erziehung* 60. S. 357-375.
- Rustenbach, S. J. (2003). Metaanalyse. Bern: Huber.
- Sánchez-Meca, J., Marín-Martínez, F. & Chacón-Moscoso, S. (2003). Effects-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods* 8. S. 448-467.
- * Sander, E. (2006). Rechtschreibprobleme von Schülern am Ende der Grundschulzeit. Dissertation Universität zu Köln.
- Schröder-Lenzen, A. (2009). Schriftspracherwerb und Unterricht. 3. Aufl. Wiesbaden: VS.
- * Schröder-Lenzen, A. & Mücke, S. (2005). Mit oder ohne Fibel – was ist der Königsweg für die multilinguale Klasse? In H. Bartnitzky & A. Speck-Hamdan (Hrsg.). Deutsch als Zweitsprache lernen. Frankfurt a. M.: Grundschulverband. S. 210-222.
- Shahar-Yames, D. & Share, D. L. (2008). Spelling as a self-teaching mechanism in orthographic learning. *Journal of Research in Reading* 31. S. 22-39.
- Share, J. (1995). Phonological recoding and self-teaching: *sine qua non* of reading acquisition. *Cognition* 55. S. 151-218.
- Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology* 72. S. 95-129
- Stahl, S. A. & Miller, P. D. (1989). Whole language and language experience approaches for beginning reading: A quantitative research synthesis. *Review of Educational Research* 59. S. 87-116.
- Valtin, R. (1998). Der ‚neue‘ Methodenstreit oder: (Was) können wir aus der amerikanischen Leseforschung lernen? In H. Balhorn, H. Bartnitzky, & A. Speck-Hamdan (Hrsg.). Schatzkiste Sprache 1. Von den Wegen der Kinder in die Schrift. Frankfurt a. M.: Arbeitskreis Grundschule. S. 14-46.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* 36, 1-48.

* Weinhold, S. (2006). Entwicklungsverläufe im Lesen- und Schreibenlernen in Abhängigkeit didaktischer Konzepte. In S. Weinhold (Hrsg.). *Schriftspracherwerb empirisch*. Baltmannsweiler: Schneider. S. 120-151.

Anschrift des Verfassers:

Prof. Dr. Reinold Funke, Pädagogische Hochschule Heidelberg, Im Neuenheimer Feld 561, D-69120 Heidelberg

funke@ph-heidelberg.de