

**Einordnung der Ergebnisse der „Bonner Studie“¹
zur Wirkung verschiedener Ansätze des Lese- und Schreibunterrichts
auf die Entwicklung der Rechtschreibleistung
im Verlauf der Grundschulzeit**

- Kurzfassung² (Stand: 8.5.2020) -

Bei der Publikation handelt sich um eine Dissertation, die als Qualifikationsleistung zwar anders zu bewerten ist als die Studie eines Forschungsinstituts, aber da die Untersuchung von der Betreuerin schon vorweg als „Bonner Studie“ für die öffentliche Diskussion instrumentalisiert wurde, ist eine intensivere Auseinandersetzung mit ihrer begrenzten Aussagekraft notwendig.

Die erst jetzt vorgelegte ausführliche Darstellung bestätigt die im Herbst 2018 vorgetragenen Vorbehalte gegenüber dem forschungsmethodischen Status der Ergebnisse. Damit bleiben die Einschränkungen der Aussagekraft der Studie und des Geltungsanspruchs der Folgerungen, die von der Betreuerin in ihren Vorveröffentlichungen gezogen wurden, bestehen. Das zentrale Problem der Studie ist die Unklarheit des Gegenstandes: Schon konzeptuell werden die verglichenen Ansätze nicht sauber definiert und unterschiedlich gruppiert. Darüber hinaus bleiben der Unterricht selbst, dem die Ergebnisse zugerechnet werden sollen, und sein Umfeld eine „black box“. Im Einzelnen beeinträchtigen die folgenden Schwächen die Aussagekraft der Studie - teils schon für die Bonner Stichprobe selbst, vor allem aber für eine Verallgemeinerung der Daten:

1. Auf der **konzeptionellen** Ebene unterscheidet der Autor fünf Ansätze: traditionelle und moderne Fibel, „Lesen durch Schreiben“, Spracherfahrungsansatz und Rechtschreibwerkstatt. In der empirischen **Untersuchung** verglichen werden drei Ansätze: Fibel (wobei die beiden verschiedenen Lehrgänge nicht unterschieden werden), „Lesen durch Schreiben“

¹ Kuhl, T. (2020): Rechtschreibung in der Grundschule. Eine empirische Untersuchung der Auswirkungen verschiedener Unterrichtsmethoden. Springer: Wiesbaden.

² Die Langfassung kann angefordert werden über: hans.bruegelmann@uni-siegen.de

und Rechtschreibwerkstatt. Die **Folgerungen** kontrastieren verallgemeinernd nur noch zwei Ansätze: „Fibel-Didaktik“ und „offene Methoden“. Nicht deckungsgleich ist deshalb das Verhältnis zwischen dem, was *analysiert*, dem, was konkret *untersucht* wurde, und dem, was in den zu oft verallgemeinernden Folgerungen *bewertet* wurde, z. B. Fibelunterricht als „direkte Instruktion“, „Lesen durch Schreiben“ und Rechtschreibwerkstatt als „selbstgesteuertes Lernen“.

2. Die Darstellung der Ansätze folgt bereits auf der **Konzept-Ebene** unterschiedlichen Logiken: Während bei den Fibel-Lehrgängen zwischen traditionellen und modernen Ansätzen unterschieden wird, stammt die letzte zu „Lesen durch Schreiben“ zitierte Publikation aus dem Jahr 2001, die Weiterentwicklung dieses Ansatzes in den letzten 20 Jahren wird nicht berücksichtigt. Bei den fibellosen Ansätzen wird zu Recht zwischen „Lesen durch Schreiben“, Rechtschreibwerkstatt und Spracherfahrungsansatz unterschieden (wobei letzterer allerdings nicht in die empirische Untersuchung eingeht), während die Tobi-Fibel und die Bausteine-Fibel trotz ihrer konzeptionellen Unterschiede als *ein* Ansatz zusammengefasst und gemeinsam bewertet werden. Zudem werden die Urteile noch darüber hinaus verallgemeinert, als ob mit der Studie eine Bewertung „der“ Fibellehrgänge insgesamt möglich wäre, obwohl diese ein didaktisch und methodisch ähnlich breites Spektrum an Varianten umfassen wie die sog. „offenen“ Ansätze.

3. Da der **Unterricht** nicht beobachtet wurde, die Lehrer*innen auch nicht ersatzweise zu ihrem konkreten Vorgehen befragt worden sind, ist anhand der Daten nicht nachvollziehbar, wie weit sich die Ansätze, die auf der Konzept-Ebene einander gegenübergestellt wurden, und der Unterricht tatsächlich decken, so dass die erhobenen Effekte einer „black box“ zugerechnet werden müssen. Das gilt schon für den Anfangsunterricht, aber erst recht für den Rechtschreibunterricht in den Jahren danach. Über ihn wurden nicht einmal auf der Konzept-Ebene Informationen erhoben, die nötig wären, um tragfähige Aussagen über die Ursachen von Effekten machen zu können.

4. Bei den verglichenen Kindern handelt es sich weder um Zufalls- noch um gezielt (nach inhaltlichen Kriterien) ausgewählte **Stichproben**. Ihre soziale Zusammensetzung ist unklar und die Vergleichbarkeit der **Lernvoraussetzungen** und der **Lernbedingungen** der Kinder für die verschiedenen Ansätze nicht zureichend gesichert. Vor allem im Längsschnitt hätte

der Sozialstatus als Ko-Variate einbezogen werden müssen, wenn die Kinder schon nicht über die Ansätze hinweg parallelisiert werden konnten. Das wäre in einer Stadt wie Bonn besonders wichtig gewesen, wo die (in die Studie einbezogene) andere Muttersprache sowohl mit einem akademischen als auch mit einem schulfernen Milieu assoziiert sein kann, also nicht für eine bestimmte soziale Herkunft steht.

5. Die **Auswertung** der Daten ist **unvollständig**. Schon Tabellen zur Größe der verschiedenen Stichproben fehlen, die man sich – soweit überhaupt möglich – selbst zusammenstellen muss. Außerdem fehlen für die verschiedenen Teilstudien Übersichten über Grunddaten der einzelnen Ansätze wie Muttersprache, Geschlecht, Sozialstatus. Für die Rechtschreibleistung im Längsschnitt, aber auch für die Lese- und Schreibbefragung sind die Rohdaten nicht ausgewiesen. Inhaltlich fehlt eine Auswertung auf Klassen- statt nur auf Schüler*innen-Ebene. So kann der Einfluss des Faktors Lehrer*in nicht eingeschätzt werden, der nach anderen Studien oft zu einer großen Streuung der Effekte innerhalb der Ansätze führt. Auch kann nicht nach Bedingungen für erfolgreichen Unterricht mit „Lesen durch Schreiben“ bzw. der Rechtschreibwerkstatt und Gründen für schwächere Leistungen von einzelnen Fibel-Klassen gesucht werden.

6. Die Ergebnisse werden nur intern verglichen, **nicht** aber auf die **Normen der Hamburger Schreibprobe** bezogen. Nach diesem bundesweit repräsentativen Maßstab entsprechen „Lesen durch Schreiben“-Kinder in Kuhls Studie schon zum Ende des Anfangsunterrichts und über die Grundschulzeit hinweg der bundesdeutschen Norm. Sogar der Anteil besonders leistungsschwacher Schüler*innen ist niedriger als in der Bonner Gesamtstichprobe – besonders auffällig in Klasse 1 und 2. Die „Ausreißer“ bilden die (zusammengefassten) Fibel-Klassen (nach oben) und die Rechtschreibwerkstatt-Klassen (nach unten). Insofern wäre zu fragen, welche besonderen Bedingungen in diesen beiden Fällen zu den erwartungswidrigen Ergebnissen geführt haben (spezifische Stärken/ Schwächen der eingesetzten Methoden; Zusammensetzung der Schülerschaft; Engagement/ Kompetenz der Lehrer*innen).

7. Für die Bewertung der Leistungen bleibt die Wahl des **Maßstabs** (richtige Grapheme) begründungsbedürftig. So wurden die unterschiedlichen Ansprüche an das Schreibniveau Ende Klasse 1/2 (alphabetische Strategie) und die Entfaltung dieser Basiskompetenz durch

orthographische und morphematische Strategien (Klasse 2-4) bei der Auswertung nicht berücksichtigt.

8. Die äußerst knappe Darstellung des Forschungsstands in Kap. 1.10 beschränkt sich auf deutschsprachige Studien und dabei zusätzlich auf vergleichende Evaluationen. In der lediglich tabellarischen Übersicht werden sie zudem unvollständig und durch fehlende Kommentierung zum Teil unzutreffend vorgestellt. Bei der Einordnung der Ergebnisse wurden der deutschsprachige und der internationale **Forschungsstand** nur **einseitig** berücksichtigt. Weder die Grundlagenforschung zum Schriftspracherwerb noch neuere Experimentalstudien werden diskutiert, die die Bedeutung des lautorientierten Schreibens in der Anfangsphase für die spätere Rechtschreibentwicklung belegen.

9. Zum Schluss: Wer sich ein Bild von den Stärken und Schwächen der in dieser Studie untersuchten Ansätze machen will, sollte den Hinweis des Autors ernst nehmen, dass die Rechtschreibleistung in einem Wörter- und Satz-Diktat nur ein **Teil-Aspekt schulischer Ziele** ist; die inhaltliche, sprachliche und orthographische Qualität freier Texte, die Entwicklung der Lesefähigkeit, aber auch übergreifende Kompetenzen wie die Förderung der Selbstständigkeit der Schüler*innen wurden in der Studie (bewusst) nicht erfasst.

Es bleibt festzuhalten: Die Anforderungen an eine Dissertationsstudie, deren Logistik bei Kuhl durchaus ein beachtliches Niveau erreicht hat, sind andere als an besser ausgestattete Untersuchungen, die mit Hilfe von Drittmitteln durchgeführt werden können. Aber entsprechend vorsichtig hätten die Folgerungen aus den unter diesen eingeschränkten Bedingungen gewonnenen Daten sein müssen.

Eine empirische Grundlage für das teilweise in den Medien geforderte und von einigen Ministerien auch umgesetzte Verbot von „Lesen durch Schreiben“ oder generell eines lautorientierten Schreibens im Anfangsunterricht liefert auch diese Untersuchung nicht.

Vielmehr stellt sich die Frage, warum in dieser Studie zwei Methoden, die beide mit dem lautorientierten Schreiben beginnen, so unterschiedlich abgeschnitten haben. Diese unterschiedlichen Ergebnisse machen noch einmal deutlich, dass es nicht um das OB geht, sondern auf das WIE ankommt, das leider in dieser Studie nicht untersucht wurde.

**Einordnung der Ergebnisse der „Bonner Studie“³
zur Wirkung verschiedener Ansätze des Lese- und Schreibunterrichts
auf die Entwicklung der Rechtschreibleistung
im Verlauf der Grundschulzeit**

- Langfassung⁴ (Stand: 8.5.2020) -

Im Herbst 2018 hat die Vorabveröffentlichung von Ergebnissen aus der sog. „Bonner Studie“ (Kuhl/ Röhr-Sendlmeier 2018a+b) in den Medien für einen erheblichen Aufruhr gesorgt und auch drastische bildungspolitische Folgen gehabt, obwohl wesentliche Informationen fehlten. Kritiker*innen bemängelten deshalb die unmittelbar folgenden politischen Reaktionen bis hin zu Methoden-Verboten, da weder die Grunddaten noch das methodische Vorgehen nachvollziehbar dargestellt waren, so dass eine fachliche Auseinandersetzung mit den Folgerungen aus der Studie unmöglich war⁵. Es hat anderthalb Jahre gebraucht, bis jetzt die Grunddaten und Informationen über die Untersuchungsbedingungen zugänglich sind - allerdings immer noch nicht vollständig. Bei der Publikation handelt sich um eine sehr knapp gehaltene Dissertation, die vermutlich keinen breiten Leserkreis findet, aber da sie von der Betreuerin vorweg für die öffentliche Diskussion instrumentalisiert wurde, ist eine intensivere Auseinandersetzung mit ihrer begrenzten Aussagekraft notwendig.

Als wissenschaftlicher Mitarbeiter, der für seine Dissertation eine derart umfangreiche Studie mit 2.800 (Querschnitt) plus 284 Kindern (Längsschnitt) aus insgesamt 12 Bonner Schulen und mit Unterstützung von 28 Qualifikationsarbeiten organisiert, hat Tobias Kuhl logistisch eine beachtliche Leistung erbracht. Auch inhaltlich werfen seine Diskussion der

³ Kuhl, T. (2020): Rechtschreibung in der Grundschule. Eine empirische Untersuchung der Auswirkungen verschiedener Unterrichtsmethoden. Springer: Wiesbaden.

⁴ Für hilfreiche Anmerkungen zu Vorfassungen dieses Gutachtens danke ich Axel Backhaus, Erika Brinkmann, Ursula Carle und Gerheid Scheerer-Neumann.

⁵ Vgl. etwa die Kritik des Grundschulverbands unter <https://grundschulverband.de/pressemitteilung-recht-schreibunterricht/> [Abruf: 20.4.2020]

Ansätze und der gewonnenen Befunde bedenkenswerte Fragen zu den didaktischen Konzeptionen bzw. ihren Umsetzungen auf. Aber sie geben wegen der im Folgenden diskutierten methodischen Schwächen keinerlei tragfähige Antworten auf diese Fragen, bieten vor allem keine methodisch gesicherte Grundlage für Verallgemeinerungen und rechtfertigen damit vor allem nicht die 2018 geforderten und teilweise auch umgesetzten politischen Maßnahmen. Bei einer sorgfältigeren Planung, forschungsmethodisch saubereren Anlage und besseren Ausstattung hätte der Erkenntnisgewinn der Studie deutlich höher sein können. Dabei überrascht die Begründung für einen Verzicht auf Drittmittel, dass dieser die Untersuchung unabhängiger gemacht habe (143), wenn man an eine mögliche Finanzierung durch die Deutsche Forschungsgemeinschaft oder durch gemeinnützige Stiftungen denkt. Aber selbst wenn man die Studie akzeptiert, wie sie vorliegt, ist entgegen der 2018 so hoch gepushten öffentlichen Diskussion das wichtigste Ergebnis: Die Kinder, die in den untersuchten Klassen nach „Lesen durch Schreiben“ unterrichtet wurden, erreichen im Mittel dieselben Leistungen wie der bundesdeutsche Durchschnitt. Erklärungsbedürftig sind also nicht die Ergebnisse der LdS-Klassen, entsprechen sie doch völlig der Normerwartung, sondern das Abschneiden der beiden anderen Gruppen: Warum schneiden die Kinder aus den Tobi- und Bausteine-Klassen in dieser Studie deutlich besser ab als der bundesweite Durchschnitt von anderen Fibeln und alternativen Konzepten? Warum schneiden andererseits die mit der Rechtschreibwerkstatt unterrichteten Kinder deutlich schlechter ab als die beiden anderen Gruppen, obwohl dieses Programm als einziges von den untersuchten Ansätzen Rechtschreibung ausdrücklich und systematisch für die ganze Grundschulzeit zum Gegenstand hat? Gründe lassen sich auf verschiedenen Ebenen suchen. Hier sind vor allem zu nennen

1. unterschiedliche Qualität der didaktischen Konzeptionen,- deren mehr oder weniger gelungene Umsetzung in die konkreten Lehrwerke/ Materialien;
2. unterschiedliche Kompetenz bzw. unterschiedliches Engagement der beteiligten Lehrer*innen,
3. unterschiedliche Zusammensetzung der untersuchten Lerngruppen (denn diese sind weder nach Zufall noch nach inhaltlichen Auswahlkriterien einbezogen worden).

Leider lassen sich diese Erklärungen anhand der verfügbaren Daten nicht beurteilen. Kuhls Fokussierung auf Ursachenebene 1 ist vor diesem Hintergrund voreilig.

Die vielfältigen und zum Teil sehr detaillierten Anmerkungen in den folgenden Ausführungen der Zusammenfassung sind nur teilweise als Kritik an der Studie oder gar als Vorwurf an den Autor zu verstehen. In wesentlichen Teilen beziehen sie sich auf grundsätzliche methodische Schwierigkeiten einer Evaluation unter Alltagsbedingungen. Solche Studien hatte ich selbst mir früher einmal einfacher vorgestellt, als wir schon vor 40 Jahren vergleichende Evaluationen von Didaktiken des Anfangsunterricht im Lesen und Schreiben gefordert haben (vgl. Brügelmann/ Söhnen 1981a+b), wo sich bereits eine Reihe der im Folgenden begründeten Anforderungen an eine aussagekräftige Evaluation von Lehrwerken/ Methoden finden. Die kritisierten Schwächen zu verstehen ist wichtig, um den beschränkten Geltungsbereich von Urteilen über die untersuchten Ansätze zu erkennen und die Grenzen für verallgemeinernde Forderungen an Bildungspolitik und Schulpraxis zu beachten.

1. Auf der **konzeptionellen** Ebene unterscheidet der Autor fünf Ansätze: traditionelle und moderne Fibel, „Lesen durch Schreiben“, Spracherfahrungsansatz und Rechtschreibwerkstatt. In der empirischen **Untersuchung** verglichen werden drei Ansätze: Fibel (wobei die beiden verschiedenen Lehrgänge nicht unterschieden werden), „Lesen durch Schreiben“ und Rechtschreibwerkstatt. Die **Folgerungen** kontrastieren verallgemeinernd nur noch zwei Ansätze: „Fibel-Didaktik“ und „offene Methoden“.

Nicht deckungsgleich ist deshalb das Verhältnis zwischen dem, was *analysiert*, dem, was konkret *untersucht* wurde, und dem, was in den zu oft verallgemeinernden Folgerungen *bewertet* wurde, z. B. Fibelunterricht als „direkte Instruktion“, „Lesen durch Schreiben“ und Rechtschreibwerkstatt als „selbstgesteuertes Lernen“.

Fehlende Stringenz auf den verschiedenen Darstellungsebenen erschwert eine Beurteilung der Ergebnisse und ihrer Deutung bzw. Folgerungen. Schon, dass so verschiedene Ansätze wie „Lesen durch Schreiben“, Rechtschreibwerkstatt und Spracherfahrungsansatz einerseits und unterschiedliche Fibel Lehrgänge andererseits pauschal gegenübergestellt werden, wird den Differenzierungen nicht gerecht (vgl. dazu im Einzelnen Scheerer-Neumann 2020). Die auch in der Studie selbst stark voneinander abweichenden Ergebnisse von „Lesen durch Schreiben“ und Rechtschreibwerkstatt hätten den Autor zur Vorsicht gegenüber Allgemeinurteilen über „offene Ansätze“ und „Fibellehrgänge“ mahnen müssen.

Die Diffusität der Definition wird schon in den unterschiedlichen Gegenüberstellungen deutlich: So ist einmal von zwei Ansätzen (lehrgangsorientiert vs. offener Unterricht bzw. Fibel vs. Spracherfahrungsansatz) die Rede (1, 148), dann von drei Ansätzen, indem der offene Unterricht weiter ausdifferenziert wird in „Lesen durch Schreiben“ und Spracherfahrungsansatz (2, 67) und schließlich von vier, indem der Spracherfahrungsansatz (41ff.) nicht nur von „Lesen durch Schreiben“ (27ff.), sondern zusätzlich von der Rechtschreibwerkstatt (44ff.) unterschieden wird, auch wenn die Beziehung zwischen diesen drei Ansätzen an anderer Stelle wieder sehr verschwimmt (IX). Für den Fibel-Ansatz wiederum stellt der Autor nur den Tobi-Lehrgang vor (53ff.), der nach unserer Rekonstruktion der Stichproben (s. unten 4.) im Längsschnitt gar nicht untersucht wird, während die Konzeption der dort eingesetzten Bausteine-Fibel mit ihren Besonderheiten nicht vorgestellt oder mit ihren Besonderheiten von Tobi abgegrenzt wird.

Auch inhaltlich bleibt die Position des Autors selbst unklar. Einerseits kritisiert er aus linguistischer Sicht, die Kinder in der Anfangsphase lautorientiert schreiben zu lassen:

„Nach Bredel et al. (2017) ist Schrift keine Abbildung der deutschen Lautung. Sie ist eher eine Abbildung von Grammatik. Werden Kinder animiert, zu schreiben, wie sie hören und sprechen, wird ihnen der Weg in diese alles entscheidende Einsicht von Anfang an erschwert oder sogar versperrt.“ (53)

Hinzu kommt das verbreitete Missverständnis, Lernen erfolge als Addition von unveränderlichen Kompetenz-Bausteinen:

„Werden Privatschreibungen der Kinder zu Beginn des Schreiblernprozesses toleriert und sogar positiv verstärkt, prägen sie sich diese fehlerhaften Schreibweisen nach der Lerntheorie von Pavlov (1927) ein.“ (135)

Lerntheoretisch hängt Kuhl also einer behavioristischen Position an, wie seine Verweise auf Pavlov und Skinner zeigen (135). Die kognitive Wende in der Psychologie und konstruktivistische Sichtweisen des Lernens werden nicht diskutiert. Konkret bezogen auf das Lesen- und Schreibenlernen interpretieren diese Theorien Fehler - wie beim Spracherwerb - als Vorformen einer schrittweisen Annäherung an die Konvention (vgl. beispielsweise. Downing 1979; Frith 1980; Henderson/ Beers 1980; Temple u. a. 1982; Brügelmann 1984; Valtin/ Naegele 1984; Scheerer-Neumann u. a. 1986; Brinkmann 1997).

Erstaunlicherweise befürwortet auch Kuhl selbst an anderer Stelle eine solche methodische Stufung (140):

„Aus Sicht des Autors der vorliegenden Studie sind Anlauttabellen geeignet, den Schülern das grundlegende Laut-Buchstaben-Prinzip zu vermitteln. Zeitnah sollten die Kinder jedoch auch lernen, dass es sich bei der geschriebenen Sprache mitnichten um eine Lautschrift handelt. Vielmehr sollte die Auf-

merksamkeit der Kinder auf die Vielfalt der Phonem-Graphem-Korrespondenzen gelenkt werden. Sobald die Kinder das Grundprinzip des Laut-Buchstaben-Prinzips begriffen und alle Buchstaben gelernt haben, sollte deshalb vom Einsatz der Anlauttabellen abgesehen werden.“

Dies ist ausdrücklich die Position des von Kuhl kritisierten Spracherfahrungsansatzes (vgl. Brügelmann 1992; Brinkmann 2015; 2018), und sie findet sich in ähnlicher, wenn auch jeweils etwas anders akzentuierter Weise für die Rechtschreibwerkstatt bei Sommer-Stumpenhorst (2015, 112ff.) und für „Lesen durch Schreiben“ bei Wiemer/ Hüttenberger (2015, 140f.).

Kuhl sieht durchaus, dass umgekehrt auch „moderne Fibeln“ Anlauttabellen nutzen (43). So auch die in die Untersuchung einbezogenen Fibellehrgänge, aber sie gewichten beide das lautorientierte Schreiben in der Anfangsphase ebenfalls unterschiedlich. Einerseits bietet auch die Tobi-Fibel eine Buchstabentabelle bzw. (inzwischen) ein „Lautposter“⁶ an, erwartet die „vollständige Verschriftung“ aber nur für „lauttreue Wörter“ (Metze 2016, 13f.). In der BAUSTEINE-Fibel ist das lautorientierte Schreiben in einem sog. „Vorkurs“ zur individuellen Förderung grundlegender Fähigkeiten wichtig: „Er trainiert den Umgang mit der Anlauttabelle und bietet zahlreiche Übungen zur phonologischen Bewusstheit.“⁷

Die beschriebene Konstellation der didaktisch-methodischen Überlappung erlaubt also kein polares *entweder - oder*. Dann aber wäre es wichtig, differenzierter zu schauen, wann, in welchem Umfang und unter welchen Bedingungen die Kinder im konkreten Unterricht mit einer Buchstaben- oder Anlauttabelle schreiben. Die von Kuhl selbst zitierte IGLU-Studie zeigt dazu ein sehr vielfältiges Bild der Unterrichtsprofile in deutschen Grundschulen (40):

Tabelle 3. Frequenz des Einsatzes der Anlauttabelle im Unterricht (Bremerich-Vos & Wendt, 2019).

	%	S.E.
jede Deutschstunde	48.3	4.4
Hälfte der Deutschstunden	32.8	3.5
weniger als die Hälfte der Deutschstunden	15.2	2.9
nicht beantwortet oder ungültig	3.7	1.6

Anmerkung. Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

⁶ <https://www.cornelsen.de/reihen/tobi-360000820000/neubearbeitung-2016-360000820014> [Abruf am 23.4.2029]

⁷ <https://www.westermann.de/reihe/BDFIBEL14/BAUSTEINE-Fibel-Ausgabe-2014#Konzept> [Abruf am 23.4.2020]

Und auch der Umgang mit orthographischen Fehlern im Anfangsunterricht unterscheidet sich stark von Lehrperson zu Lehrperson (40):

Tabelle 4. Umgang mit Rechtschreibfehlern (Bremerich-Vos & Wendt, 2019).

	%	S.E.
Von Anfang an sollte bei allen Schülern korrigiert werden.	38.1	3.9
Im ersten Schuljahr sollte bei allen Schülern nicht korrigiert werden.	16.7	3.5
Im ersten und zweiten Schuljahr sollte bei allen Schülern nicht korrigiert werden.	0.9	0.8
Von Anfang an sollte nur bei Rechtschreibstarken korrigiert werden.	11.7	2.7
Im ersten Schuljahr sollte nur bei Rechtschreibschwachen korrigiert werden.	5.1	1.9
Im ersten und zweiten Schuljahr sollte nur bei Rechtschreibschwachen korrigiert werden.	0.7	0.7
Es sollte nur bei rechtschreibstarken Schülern korrigiert werden.	0.9	0.9
Von Anfang an sollte nur bei starken Rechtschreibern, im zweiten Schuljahr auch bei rechtschreibschwachen Schülern korrigiert werden.	20.2	3.3
keine Angabe	5.6	2.0

Anmerkung. Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

Diese Daten machen deutlich, dass man genau hinschauen muss, welches konkrete Profil der untersuchte Unterricht hat, um die beobachteten Unterschiede in den Rechtschreibleistungen auf bestimmte Bedingungen des Unterrichts beziehen zu können.

2. Die Darstellung der Ansätze folgt bereits auf der **Konzept-Ebene** unterschiedlichen Logiken: Während bei den Fibel-Lehrgängen zwischen traditionellen und modernen Ansätzen unterschieden wird, stammt die letzte zu „Lesen durch Schreiben“ zitierte Publikation aus dem Jahr 2001, die Weiterentwicklung dieses Ansatzes in den letzten 20 Jahren wird nicht berücksichtigt. Bei den fibellosen Ansätzen wird zu Recht zwischen „Lesen durch Schreiben“, Rechtschreibwerkstatt und Spracherfahrungsansatz unterschieden (wobei letzterer allerdings nicht in die empirische Untersuchung eingeht), während die Tobi-Fibel und die Bausteine-Fibel trotz ihrer konzeptionellen Unterschiede als ein Ansatz zusammengefasst und gemeinsam bewertet werden. Zudem werden die Urteile noch darüber hinaus verallgemeinert, als ob mit der Studie eine Bewertung „der“ Fibellehrgänge insgesamt möglich wäre, obwohl diese ein didaktisch und methodisch ähnlich breites Spektrum an Varianten umfassen wie die sog. „offenen“ Ansätze.

Aber auch die explizite Beschreibung der Ansätze ist unzulänglich.

Von den beiden in die Untersuchung einbezogenen Lehrgängen wird „exemplarisch“ die Tobi-Fibel vorgestellt (53ff.), obwohl im Längsschnitt nur die Bausteine-Fibel genutzt wird. „Fibeln“ gibt es sehr unterschiedliche⁸. Das beginnt mit den inhaltlichen Schwerpunkten, ob sie eher mit ganzen Wörtern oder konsequent von einzelnen Buchstabe-Laut-Beziehungen ausgehen oder ob sie die Silbe fokussieren, wobei zusätzlich silbenanalytische Ansätze zu unterscheiden sind von eher unterrichtsmethodisch begründeten Konzepten wie FRESCH. Für die Einschätzung der Effekte wäre z. B. bedeutsam zu wissen, ob die jeweils eingesetzte Version der Tobi-Fibel schon mit einer Silbengliederung arbeitet oder nicht. Unterschiede zwischen Tobi und Bausteine werden gar nicht angesprochen und recht-schreibbezogene Aufgaben oder Aktivitäten werden für beide erst recht nicht benannt. Während Kuhl bei den Fibeln immerhin zwischen traditionellen Ansätzen (24ff.) und modernen Ansätzen (53ff.) unterscheidet (zu denen er die Tobi-Fibel zählt), stammt bei „Lesen durch Schreiben“ die letzte zitierte Publikation von Reichen aus dem Jahr 2001, obwohl eine Zusammenfassung des aktuellen Standes bei Wiemer/ Hüttenberger (2015, speziell zum Rechtschreiben: 140f.) leicht zugänglich ist. Denn „Lesen durch Schreiben“ heute ist nicht mehr „Lesen durch Schreiben“, wie es in den von Kuhl zitierten Schriften von Reichen (1992, zum Teil auch noch 2001) dargestellt und von Kuhl wie folgt zusammengefasst wird:

„Es wird explizit auf Übungen sowie Korrekturen von Fehlschreibungen durch die Lehrkraft oder Eltern verzichtet, damit die Kinder sich frei und individuell entwickeln können.“ (IX)

Diese Zusammenfassung entspricht nicht dem aktuellen Stand von „Lesen durch Schreiben“, das auch schon Reichen weiter ausdifferenziert hat (vgl. die ausgewogene Darstellung bei Scheerer-Neumann 2020). Wer sich die vom Heinevetter-Verlag angebotenen Materialien anschaut, sieht außerdem, dass es gerade im Blick auf die Rechtschreibung seit rund 20 Jahren ergänzende Software und Übungsmaterialien wie „Erstes Verschriften“ (orthographische Einstellung), „Little Genius“, „Big Profi“ und eine ganze Heftserie mit 7 Heften mit Übungen zur Rechtschreibung und Grammatik gibt.

Anders als bei „Lesen durch Schreiben“ berücksichtigt Kuhl bei der Rechtschreibwerkstatt auch neuere Veröffentlichungen von Sommer-Stumpfenhorst (2004-2015). Allerdings

⁸ Vgl. z. B. den Überblick unter: https://www.ph-ludwigsburg.de/fileadmin/subsites/2b-spze-t-01/user_files/Service/Sprachbuchgutachten_V8.pdf [Abruf: 2.5.2020]

reduziert Kuhl diesen Ansatz im Wesentlichen auf Abschreibübungen (48f.), deren mangelnde Effektivität er anschließend als unzureichend für die Entwicklung einer umfassenden Rechtschreibkompetenz kritisiert (50f.). Sommer-Stumpfenhorst (2015, 110) selbst beschreibt seinen Ansatz dagegen umfassender:

„Die Übungen zum Rechtschreibenlernen lassen sich auf fünf grundlegende Methoden zurückführen:

- Sprache nutzen und weiterentwickeln (Lautübungen, deutliches Sprechen, Silbensprechen, Dehnsprechen)
- Abschreiben bzw. Partner- oder Selbstdiktat (Orientierung an Rechtschreibmodellen)
- Gruppieren (Wörter mit gleichen Rechtschreibphänomenen sammeln und sortieren)
- Reflektieren (über das Rechtschreiben mit anderen diskutieren und nachdenken: Was weiß ich schon? Was kann ich mir nicht erklären?)
- Texte korrigieren (Techniken: Texte rückwärts bearbeiten, Umorientierung der Aufmerksamkeit, Kontrollwörter bilden, (Eigen-)Regeln anwenden, Wörterbuch verwenden).“

Zum Spracherfahrungsansatz wird zwar auf die didaktische Landkarte mit ihren verschiedenen Lernfeldern Bezug genommen (42), die dazugehörigen rechtschreibbezogenen Aufgaben und Aktivitäten spielen in der Darstellung des Ansatzes von Kuhl aber keine Rolle, obwohl es dazu eine eigenständige Buchpublikation gibt (Brügelmann/ Richter 1994, u. a. mit „20 Thesen zum Rechtschreibunterricht“) sowie konkrete methodische Vorschläge (u. a. Brinkmann 2015; 2018) und auch ausgearbeitete Materialien gibt (Brinkmann/ Brügelmann 1993/2010; Brinkmann u. a. 2008/2019).

Im Übrigen wird dieser Ansatz zwar ausführlich diskutiert (41ff.), aber weder im Längs- noch im Querschnitt empirisch untersucht. Im Schlussteil wird er trotzdem (kritisch) bewertet - entweder durch Gleichsetzung mit der Rechtschreibwerkstatt (138) oder als verallgemeinerter Gegenpol zum ebenfalls pauschalisierten Fibelunterricht (148). Hier rächt sich die bereits unter (1.) kritisierte unklare Begrifflichkeit.

Dies gilt auch für Konzepte wie „selbstgesteuertes Lernen“, die in den drei Ansätzen durchaus unterschiedlich verstanden werden, was die Formen der Arbeitsorganisation, die Öffnung unterschiedlicher Lernwege bzw. die Akzeptanz individueller Vor- und Zwischenformen der Verschriftung von Sprache betrifft. Insgesamt verkürzt Kuhl dieses Prinzip für die diskutierten Positionen unzulässig, wenn er seine Umsetzung so darstellt, als handle es sich um isoliert vor sich hinarbeitende Kinder, etwa zu „Lesen durch Schreiben“:

„Wie auch beim Spracherfahrungsansatz wird davon ausgegangen, dass die Kinder nach und nach die korrekte Rechtschreibung eigenständig durch Ausprobieren und eigene Erkenntnis erlernen.“ (135)

Das klingt so, als würden die Kinder mit dem *Gegenstand* und den *Aufgaben* allein gelassen. Im *Gegenteil* sind die von ihm (136) geforderten dialogischen Formen geradezu konstitutiv für Konzepte wie den *Spracherfahrungsansatz*: die *Lehrperson*, aber auch die anderen *Kinder* als *anregende, modellierende, erklärende, kritische Gegenüber*, mit denen *Erfahrungen* ausgetauscht, aber auch die *Bedingungen der gemeinsamen Arbeit* ausgehandelt werden. Zwar gibt es *Fehl-Interpretationen* bzw. *Umsetzungen* dieser Konzepte in der *Praxis*, die die *Kinder* sich selbst überlassen. Aber *Kuhl* selbst beschreibt auch für die *Rechtschreibwerkstatt* als *wesentliches Merkmal* die *gemeinsame Planung* als *Grundlage* der *individuellen Arbeit* (44). Und inhaltlich geht es nicht allein darum, *OB* etwas erklärt wird, sondern *WAS* und *WIE* erklärt wird; die *Frage* ist insofern auch nicht, *OB* auf *Fehler* reagiert wird, sondern *WIE*, was also „*Korrektur*“ konkret bedeutet. Insbesondere zum *Spracherfahrungsansatz* gibt es sehr konkrete *methodische Vorschläge* für einen *konstruktiven, weil lernförderlichen Umgang* mit *Fehlern* (Brügelmann 1992; Brinkmann 2018), beispielsweise

- Übersetzung von *Schreibversuchen* der *Kinder* in *Erwachsenenschrift*;
- dort *Markierung* der bereits richtig geschriebenen *Buchstaben* durch das *Kind*;
- *Übertragung* häufig gebrauchter *Wörter* in eine *Übungskartei*.

Die *bildungspolitische* und *schulpraktische Relevanz* der *Ergebnisse* dieses *Methodenvergleichs* sind insofern fraglich: Die *Konzeptionen* werden entweder zu *pauschal* oder *unzutreffend* beschrieben; *Aktualisierungen* werden nicht einbezogen (vor allem bei „*Lesen durch Schreiben*“); und der *Spracherfahrungsansatz* wird - *zusätzlich zu seiner unzutreffenden Darstellung* - zwar auf der *Konzeptebene* kritisiert, aber dann gar nicht untersucht.

Hinzu kommt das *umstandslose Springen* zwischen den drei *Bedeutungen* von „*Spracherfahrungsansatz*“ als

- das *spezifische Konzept* von *Brinkmann/ Brügelmann* (vor allem 41ff.),
- dieses plus *Rechtschreibwerkstatt* (67f., 137f.),
- beide plus „*Lesen durch Schreiben*“ (148).

Insofern kann man die empirischen Daten und Folgerungen im Schlusskapitel nicht stringent auf die Ansätze beziehen. Dabei ist ausdrücklich anzumerken, dass deren oben kritisierten Verkürzungen in der Praxis in der Tat vorkommen (wie auch bei Fibeln die Konzepttreue und die Erfolge breit streuen). Es gibt aber keine Daten zu ihrer Verbreitung, und wie sich im Folgenden zeigen wird, hat der Autor auch für die untersuchte Stichprobe nicht geklärt, ob sie hier tatsächlich eine Rolle spielen.

3. Da der **Unterricht** nicht beobachtet wurde, die Lehrer*innen auch nicht ersatzweise zu ihrem konkreten Vorgehen befragt worden sind, ist anhand der Daten nicht nachvollziehbar, wie weit sich die Ansätze, die auf der Konzept-Ebene einander gegenübergestellt wurden, und der Unterricht tatsächlich decken, so dass die erhobenen Effekte einer „black box“ zugerechnet werden müssen. Das gilt schon für den Anfangsunterricht, aber erst recht für den Rechtschreibunterricht in den Jahren danach. Über ihn wurden nicht einmal auf der Konzept-Ebene Informationen erhoben, die nötig wären, um tragfähige Aussagen über die Ursachen von Effekten machen zu können.

Die Studie geht von zwei Fehlannahmen aus:

1. Wenn verschiedene Lehrer*innen oder gar Schulen (72) sich demselben Konzept zuordnen, dann machen sie auch im Unterricht dasselbe - mit nur „graduellen“ (71, 146) Unterschieden.
2. Wenn eine Lehrperson/ Schule sich im Anfangsunterricht einem bestimmten Konzept zuordnet, dann ist damit auch entschieden, was im Rechtschreibunterricht in Klasse 2 bis 4 passiert, denn es

„...haben sich die Lehrer oder der Rektor für ein Leitmedium entschieden, welches in sämtlichen Klassen einer Schule eingesetzt wurde und den ‚Kurs‘ des Unterrichts maßgeblich bestimmt hat.“ (146)

Diese Annahmen sind nicht belegt und aufgrund der Befunde aus der Schul- und Unterrichtsforschung (s. die Beispiel unten) sehr skeptisch zu betrachten.

So folgt von den untersuchten Ansätzen nur die Rechtschreibwerkstatt einer systematischen Konzeption für den Rechtschreibunterricht von Klasse 1 bis 4.

„Lesen durch Schreiben“ ist - wie der Name schon sagt - ein Konzept für die Förderung des Lesenlernens und bezieht sich auf den Anfangsunterricht. Was Lehrer*innen daneben

bzw. danach tun, unterscheidet sich von Klasse zu Klasse erheblich. Wie unter (2.) erwähnt hat schon Reichen selbst seine Materialien vor Jahren u. a. um eine Lernsoftware für das Üben eines Grundwortschatzes ergänzt. In der Studie wird nicht erfasst, welche Varianten in den untersuchten Klassen realisiert wurden. Zudem gibt es große Unterschiede in der Art, wie die Kinder Rückmeldung zu ihren lautorientierten Schreibversuchen erhalten. Wie „Lesen durch Schreiben“ sind auch die Fibeln didaktische Konzepte für den Anfangsunterricht, die schon in der ersten Klasse auf der Unterrichtsebene unterschiedlich ausdifferenziert und ergänzt werden, für die aber vor allem ab Klasse 2 jeweils vielfältige Anschlussoptionen bestehen. Welche in den untersuchten Klassen genutzt wurden, hat die Studie nicht erfasst. Ob zum Beispiel neben bzw. nach der grundlegenden Einführung in die Phonem-Graphem-Korrespondenz ein Grundwortschatz eingeübt wurde, welche Regeln auf welche Weise vermittelt wurden, bleibt ebenso offen wie der Umfang, dem – auch in den Lehrgängen vorgesehenen Schreiben mit einer Anlauttabelle – tatsächlich Raum gegeben wurde.

Gegen eine 1:1-Umsetzung von methodischen Vorgaben sprechen Befunde der Evaluationsforschung zu den Wandlungen des Curriculums „auf seinem Weg von Autor zum Kind“ (Brügelmann 1978) aufgrund der unvermeidlichen, wegen der unterschiedlichen Bedingungen auch notwendigen „Rekontextualisierungen“ durch die auf verschiedenen Ebenen Beteiligten (Fend 2008, 32⁹), insbesondere durch die Lehrer*innen bei der Auslegung im Unterricht (Brügelmann/ Söhnen 1981a+b).

Wenn die Umsetzung weder durch Unterrichtsbeobachtungen noch zumindest durch eine Befragung der Lehrer*innen konkreter bestimmt wird und wir nichts über die in den aufsteigenden Klassen verwendeten Lehrmittel oder weitere Veränderungen des Unterrichts, z. B. durch den oft nach Klasse 2 praktizierten Lehrerwechsel wissen, muss man die in der Studie erhobenen Effekte *black boxes* zurechnen.

⁹ „Der handlungstheoretische Ausgangspunkt, konkretisiert in der Theorie der Rekontextualisierung, bedeutet verallgemeinert ausgedrückt Folgendes: Auf jeder Handlungsebene, auf jener der Politik, der Verwaltung, der Schulführung, der Lehrpersonen und der Schülerschaft sind die Vorgaben auf der jeweils übergeordneten Ebene Umwelten des Handelns, die eine Umsetzung an die jeweils neuen, ebenenspezifischen Besonderheiten der Umwelten des Handelns erfordern ...“

Dass der Anfangsunterricht die weitere Rechtschreibentwicklung nur sehr begrenzt bestimmt, wird auch deutlich, wenn man die Leistungen Ende der ersten Klasse mit den Leistungen Ende dritter Klasse vergleicht: Die Korrelationen betragen bei Kuhl für die Beziehungen zwischen den Terminen Ende Klasse 1 bis 3:

1 E	2 E	3 E	r =
x	→ x		.45**
	x	→ x	.41**
x	→	x	.25**

Angesichts der Kinderzahl (N= 249-263) sind die Zusammenhänge zwar statistisch signifikant, also als nicht zufällig zu betrachten. Aber ihre praktische Bedeutung ist mit einer Aufklärung der späteren Rechtschreibleistung zu nur 15-20% über ein Schuljahr hinweg und sogar nur 6% bis Ende Klasse 3 eher gering. Leider fehlen Daten zur Entwicklung einzelner Klassen, so dass nicht gesagt werden kann, welche Anteile der Veränderungen (der Rechtschreibleistungen) den Besonderheiten einzelner Kinder zuzurechnen sind und welche der Klassen- und damit der Unterrichtsebene.

Es handelt sich hier nicht nur um theoretisch begründete Einwände. So berichtet Lorenz (2017, 63) aus ihrer Studie, dass in 10 von 11 Klassen, in denen nach Auskunft der Lehrpersonen „Lesen durch Schreiben“ eine „wichtige Rolle“ spielte, zugleich eine Fibel eingesetzt wurde. Ihre Studie zeigt, dass man schon durch eine einfache Befragung der Lehrer*innen ein differenzierteres Bild von der Umsetzung eines Konzepts erhalten kann. Anhand von eigenen Daten aus einer noch laufenden Untersuchung lässt sich ebenfalls eine große Variationsbreite der Umsetzungsformen empirisch belegen, im Kasten illustriert mit Beispielen zu zwei der auch in der Bonner Studie eingesetzten Lehrwerke (vgl. Brügelmann/Brinkmann 2020, Kap. 5).

Zwei Klassen derselben Schule, die beide die Rechtschreibwerkstatt als Leitmedium angaben, wurden Ende Klasse 2 zu ihrem Rechtschreibunterricht befragt.

In Klasse 14 war im Anfangsunterricht nur gelegentlich mit der Anlauttabelle geschrieben worden. Im zweiten Jahrgang haben die Kinder aber mehrmals im Monat freie Texte geschrieben und diese auch individuell überarbeitet. Außerdem haben die Kinder mehrmals pro Woche - allgemein häufige und eigene schwierige Wörter geübt,

- besondere Rechtschreibphänomene und Rechtschreibstrategien bearbeitet und
- sowohl Silben geschwungen als auch die Vokalquantität untersucht sowie gelegentlich das Nachschlagen in Wörterbüchern praktiziert.

In Klasse 15 wurde im Rahmen der Rechtschreibwerkstatt im Anfangsunterricht dagegen mehrmals pro Woche mit der Anlauttabelle geschrieben.

In der zweiten Jahrgangsstufe wurden - wie in der Parallelklasse - allgemein häufige Wörter mehrmals pro Woche geübt, eigene schwierige Wörter aber nur mehrmals im Monat. Letzteres galt auch für das Silbenschwingen. Und nur gelegentlich haben die Kinder die Vokalquantität untersucht und eigene Text überarbeitet.

Dasselbe Lehrwerk - aber zwei völlig unterschiedliche Unterrichtsprofile.

Aus zwei Klassen einer anderen Schule, die beide nach ihrer Angabe im Lehrerfragebogen mit der Tobi-Fibel gearbeitet haben, erhielten wir Ende 1. Klasse über den Lehrerfragebogen folgende Profile für ihren Unterricht:

In Klasse 53 haben die Kinder neben dem Lehrgang häufig selbstständig mit der Anlauttabelle geschrieben und ihre Texte mehrmals im Monat überarbeitet. Außerdem wurden mehrmals pro Woche Wörter nach Silben geschwungen und mehrmals pro Monat Übungen zu Rechtschreibphänomenen durchgeführt.

Auch die Kinder in Klasse 54 haben neben dem Lehrgang häufig selbstständig mit der Anlauttabelle geschrieben, aber ihre Texte nur gelegentlich überarbeitet. Daneben gab es ebenfalls nur gelegentlich Übungen zu ausgewählten Rechtschreibphänomenen und zum Silbenschwingen.

Der Unterricht in beiden Klassen hat sich also ebenfalls in den didaktischen Schwerpunkten und im methodischen Vorgehen unterschieden, obwohl dasselbe Lehrwerk benutzt wurde.

Eine weitere Klasse 52, ebenfalls an dieser Schule, hat das Spektrum noch stärker erweitert, indem sie das Tobi-Arbeitsheft mit Materialien der Fibel Flex+Flora kombiniert hat.

Hätte man in dieser Schule - im Sinne des Kuhl'schen Längsschnitts - auch die Klassen 2 und 3 einbezogen, wäre das Spektrum noch breiter geworden. Hier gab es schon auf der Ebene der Lehrwerke/ Materialien vier unterschiedliche Profile:

Klasse 56 arbeitete mit Flex+Flora und der Rechtschreibleiter, Klasse 55 zusätzlich mit einem Modellwortschatz. Klasse 57 nutzte neben Flex+Flora und der Rechtschreibleiter zusätzlich die FRESCH-Methode, die auch in Klasse 58 verwendet wurde, wobei aber statt der Rechtschreibleiter das Programm Cesar Schreiben zum Einsatz kam. Entsprechend vielfältig war dann auch das Spektrum konkreter Aktivitäten bzw. deren unterschiedliche Gewichtung.

Zurück zur Kuhl-Studie: Auffällig sind die Mittelwertunterschiede zwischen „Lesen durch Schreiben“ und der Rechtschreibwerkstatt schon Ende erster Klasse, obwohl beide die Kinder anfangs lautorientiert schreiben lassen. Auch die großen Standardabweichungen in

den Rechtschreibleistungen *innerhalb* der Ansätze - gerade bei der Rechtschreibwerkstatt und bei „Lesen durch Schreiben“ (103, 107) - können zwar einerseits bedeuten, dass die Streuungen innerhalb dieser Klassen methodenbedingt größer sind als bei den beiden Fibeln. Sie könnten aber auch Ergebnis größerer Unterschiede in der Umsetzung der Konzepte zwischen den Klassen sein. Dazu liegen leider keine Daten vor. Insofern lässt sich auch die unterrichtspraktisch zentrale Frage nicht beantworten, welche Merkmale ein Unterricht oder die Klassen haben, in denen die Rechtschreibwerkstatt bzw. „Lesen durch Schreiben“ besonders erfolgreich waren, oder umgekehrt, wo der Fibel-Unterricht weniger erfolgreich war. Auch zur Entwicklung einzelner Klassen, z. B. bei Veränderungen wie einem Lehrerwechsel (etwa beim Übergang in höhere Klassenstufen) gibt es keine Informationen.

4. Es handelt sich bei den verglichenen Kindern weder um Zufalls- noch um gezielt (nach inhaltlichen Kriterien) ausgewählte **Stichproben**. Ihre soziale Zusammensetzung ist unklar und die Vergleichbarkeit der **Lernvoraussetzungen** und der **Lernbedingungen** der Kinder für die verschiedenen Ansätze nicht zureichend gesichert. Vor allem im Längsschnitt hätte der Sozialstatus als Ko-Variate einbezogen werden müssen, wenn die Kinder schon nicht über die Ansätze hinweg parallelisiert werden konnten. Das wäre in einer Stadt wie Bonn besonders wichtig gewesen, wo die (in die Studie einbezogene) andere Muttersprache sowohl mit einem akademischen als auch mit einem schulfernen Milieu assoziiert sein kann, also nicht für eine bestimmte soziale Herkunft steht.

Kuhl verweist zu Recht auf die Einschränkungen der Befunde aus angefallenen Stichproben in früheren Untersuchungen (63ff.). Er fordert stattdessen (quasi-)experimentelle Untersuchungsdesigns - verlangt sie aber nur für die Einführung neuer Methoden (57, 141f.). Das ist eine überraschende Verschiebung der Beweislast. Denn neue Methoden wurden und werden ja deshalb entwickelt, weil die etablierten nicht erfolgreich waren bzw. sind. Ein Indikator für die fehlenden Erfolge des Fibel-Ansatzes sind die hohe Quoten funktionaler Analphabeten für alle Jahrgänge nach dem 2. Weltkrieg, wie sie seit längerem in retrospektiven Studien nachgewiesen wurden (OECD 1995; Grotlüschen/ Riekmann 2012; Grotlüschen u. a. 2019). Interessanterweise zeigen diese Studien, dass die jüngeren Jahrgänge (im Lesen) besser und in der Rechtschreibung nicht schlechter abschneiden als die älteren,

dass also die oft als „bewährt“ apostrophierten Methoden früherer Zeiten durchaus ihre Probleme hatten.

Bleibt die Frage, warum Kuhl selbst mit einer Gelegenheitsstichprobe arbeitet (146). Einen Grund nennt er selbst (72), der aber für Evaluationen unter Alltagsbedingungen grundsätzlich gilt und dann auch anderen Studien nicht zum Vorwurf gemacht werden kann:

„In einer Laborstudie mit speziell für diese Studie geschulten Lehrern mit einer vollumfänglichen Unterrichtsüberwachung durch geschultes Fachpersonal wäre es vielleicht annähernd möglich, eine Unterrichtsmethode streng nach Vorschrift anzuwenden und diese Anwendung auch zu kontrollieren. Allerdings würde eine derart künstlich angelegte Untersuchung die externe Validität der Ergebnisse stark einschränken und somit wenig praxisrelevante Erkenntnisse liefern. Stattdessen war es das Ziel dieser Studie, die Auswirkungen des regulären Unterrichtsalltags realistisch abzubilden. Deshalb wurde bei der Konzeption der vorliegenden Studie besonderer Wert daraufgelegt, möglichst wenig in den Unterrichtsalltag einzugreifen.“

Damit aber erbringt die Bonner Studie keinen Fortschritt gegenüber früheren Untersuchungen, zumal sie die Möglichkeit vergibt, die eigenen Ergebnisse mit Befunden aus verfügbaren experimentellen Studien zum lautorientierten Schreiben im Anfangsunterricht abzugleichen (s. unten 8.).

Aber auch innerhalb der Studie selbst ist es schwierig, die Befunde aus verschiedenen Quellen systematisch aufeinander zu beziehen. Denn für das Verhältnis der gewonnenen Daten zueinander ergibt sich ein besonderes Problem daraus, dass Kuhl jeweils mit verschiedenen Stichproben von Kindern arbeitet:

- In der *Querschnittsstudie* gibt es für jeden Jahrgang zu jedem Termin Gruppen, deren Anzahl zwischen den didaktischen Ansätzen, aber auch innerhalb dieser Ansätze von Jahrgang zu Jahrgang stark schwankt (107).
- Die *Längsschnittstudie* ist aus diesen Querschnitten nicht nach inhaltlichen Kriterien, sondern allein aufgrund der Verfügbarkeit von Daten aus der Vorerhebung und mindestens vier der fünf HSP-Testtermine erwachsen (88).
- Noch einmal anders zusammengesetzt sind die Gruppen, aus denen Daten über die *Befragung* der Kinder zu Motivation und Verhalten erhoben wurden. Der Rücklauf schwankt zwischen 26,8 und 75,9%, wobei sich die Anteile sehr ungleich auf die didaktischen Ansätze verteilen und zusätzlich ihre Zuordnung zu den einzelnen Jahrgangsstufen unklar ist (126f.).

Wie die folgenden Übersichten zeigen, handelt es sich - je nach Ansatz und Test/ Fragebogen - um sehr unterschiedlich große und stark wechselnde Stichproben, die für den Querschnitt zusätzlich von Jahr zu Jahr bzw. von Testtermin zu Testtermin erheblich schwanken:

Absolute Werte N =	Lese-Fragebogen (Kl. 2-4)	Schreib-Fragebogen (Kl 1-4)	Eltern-Fragebogen	Längsschnitt	Querschnitt	Gesamt
N-Gesamt	1.451	1.957	??	284	2.800	3.084
RSW	880	1.184	??	121	1.768	1.889
LdS	388	300	??	79	432	511
Fibeln	183	491	??	84	600	684

% N bezogen auf Gesamt-N	Lese-Fragebogen (Kl. 2-4)	Schreib-Fragebogen (Kl 1-4)	Längsschnitt	Querschnitt	Gesamt
N-Gesamt	47,0 %	63,5 %	9,2 %	91,7 %	3.084
RSW	46,6 %	62,7 %	6,4 %	93,6 %	1.889
LdS	75,9 %	58,7 %	15,4 %	84,5 %	511
Fibeln	26,8 %	71,8 %	12,3 %	87,7 %	684

Zumindest für den Längsschnitt versucht Kuhl die Schwächen seiner Gelegenheitsstichprobe auszugleichen, indem er ausgewählte Lernvoraussetzungen der Kinder erhebt. Immerhin erfasst er mit der phonologischen Bewusstheit und der Buchstabenkenntnis sowie der Muttersprache für die Rechtschreibentwicklung bedeutsame Voraussetzungen der Kinder. Danach schneiden die „Lesen durch Schreiben“-Klassen sowohl in der phonologischen Bewusstheit als auch in der Buchstabenkenntnis um 4-5 Punkte besser ab als die beiden anderen Gruppen (97). In der Tat kann das so sein. Allerdings ist zu bedenken, dass der Test nicht vor Schulbeginn, sondern im Laufe des „ersten Monat[s]“ durchgeführt wurde. In „Lesen durch Schreiben“- Klassen hängen aber meist von Anfang an große Anlauttabellen sichtbar an der Wand, und außerdem lernen die Kinder schon in den ersten Wochen, Wörter lautorientiert zu verschriften, so dass es durchaus sein kann, dass sie im Test mehr Buchstaben benennen und dass sie geübter in der Lautanalyse und -synthese sind.

Zudem erlauben die individuellen Vorläuferfertigkeiten - phonologische Bewusstheit und Buchstabenkenntnis - nur mit Einschränkungen eine (nachträgliche) Parallelisierung der bereichsspezifischen Voraussetzungen. Denn die von Kuhl (76, 100) als Beleg angeführte Korrelation von .30 bis .40** mit den Rechtschreibleistungen Ende erster und zweiter Klasse belegen eben nicht, dass die spätere Rechtschreib-Leistung gut vorhergesagt wird, erklären sie doch nur 9-16%, also maximal ein Sechstel der Leistungsvarianz später, d. h. es sind deutliche Rangsprünge zu erwarten. Auch hier wäre eine nach didaktischen Ansätzen getrennte Berechnung angebracht gewesen, um zu sehen, ob die eine oder andere Methode die Leistungsentwicklung der Kinder mehr bzw. weniger stark von den Voraussetzungen abkoppelt.

Hinzu kommt ein weiteres Problem bei der individuellen Verrechnung der Voraussetzungen. Der Einfluss der Unterschiede lässt sich nämlich nur begrenzt durch Kovarianzanalysen herausrechnen, weil er doppelt wirkt: unmittelbar als individuelle unterschiedliche Voraussetzung (102), aber zusätzlich über die Zusammensetzung der Klasse als Lernökologie, die ebenfalls einen Einfluss auf die individuelle Lernentwicklung hat.

Insofern sind die Übersicht (97) über die Vorkenntnisse

Tabelle 21. Deskriptive Statistiken der Vorkenntnisse.

	N	PB i.w.S. (A)		PB i.e.S. (B)		PB ge- samt (A + B)		Buchsta- benkennt- nis	
		M	SD	M	SD	M	SD	M	SD
Rechtschreib- werkstatt	121	16.9	3.2	11.6	15.2	28.5	7.4	18.3	7.1
Lesen durch Schreiben	79	17.3	3.3	15.7	5.3	33.0	7.3	22.2	3.9
Fibel	84	16.6	3.8	11.3	5.1	27.9	8.0	17.8	6.9

Anmerkung. PB = Phonologische Bewusstheit; i.e.S. = im engeren Sinn; i.w.S. = im weiteren Sinn.

und deren Interpretation durch den Autor (99):

„Zusammenfassend lässt sich festhalten, dass in der vorliegenden Stichprobe die „Lesen durch Schreiben“-Kinder gegenüber den anderen Kindern höhere Vorkenntnisse aufwiesen und somit bessere Startbedingungen hatten.“

mit Vorsicht zu betrachten, zumal auch für die Vorkenntnisse keine Prozenträge bzw. T-Werte genannt werden, die eine Einschätzung relativ zu den Eichstichproben erlaubten, was für die Generalisierbarkeit der Ergebnisse wichtig wäre. Die Stichproben können zudem auch nicht im Blick auf ihre demographische Zusammensetzung eingeschätzt werden, da die wenigen Werte in Tab. 8 (74) nicht nach den drei Teiluntersuchungen aufgeschlüsselt werden. Über den Elternfragebogen sollten angeblich „sozio-demografische Daten von den Erziehungsberechtigten“ (74) erhoben worden. Die gewählten Merkmale sind allerdings nicht aussagekräftig genug, um die Vergleichbarkeit der sozialen Herkunft überprüfen zu können (was dann in Tab. 8 auch nur für die Familiensprache geschieht - allerdings ohne Informationen zum Rücklauf und zu seiner Verteilung auf die Teilstichproben). Sozialstatistische Grunddaten wären insbesondere deshalb wichtig, weil ein fremdsprachlicher Hintergrund in einer Stadt wie Bonn - anders als in den typischen Ballungszentren - je nach Einzugsbereich der Schule auch mit einem akademischen Hintergrund und einem höheren sozio-ökonomischen Status einhergehen kann. In diese Richtung weist auch der Befund, dass die Familiensprache in der Längsschnittstudie, anders als in anderen Untersuchungen, keinen signifikanten Einfluss auf die Rechtschreibleistung hatte (103).

Angesichts der ungeklärten Unterschiede in den Voraussetzungen und im sozialen Hintergrund lassen sich die Befunde aus gerade einmal zwei Schulen pro Ansatz (bzw. auch nur vier in den Querschnitten) auch wegen ihren ungeklärten Besonderheiten nicht verallgemeinern.

5. Die **Auswertung** der Daten ist **unvollständig**. Schon Tabellen zur Größe der verschiedenen Stichproben fehlen, die man sich - soweit überhaupt möglich - selbst zusammenstellen muss. Außerdem fehlen für die verschiedenen Teilstudien Übersichten über Grunddaten der einzelnen Ansätze wie Muttersprache, Geschlecht, Sozialstatus. Für die Rechtschreibleistung im Längsschnitt, aber auch für die Lese- und Schreibbefragung sind die Rohdaten nicht ausgewiesen. Inhaltlich fehlt eine Auswertung auf Klassen- statt nur auf Schüler*innen-Ebene. So kann der Einfluss des Faktors Lehrer*in nicht eingeschätzt werden, der nach anderen Studien oft zu einer großen Streuung der Effekte innerhalb der Ansätze führt. Auch kann nicht nach Bedingungen für erfolgreichen Unterricht mit „Lesen durch Schreiben“ bzw. der Rechtschreibwerkstatt und Gründen für schwächere Leistungen von einzelnen Fibel-Klassen gesucht werden.

Als Hauptbefund aus dem Längsschnitt nennt Kuhl (XI):

„Die Fibelkinder erreichten zu allen fünf Messzeitpunkten der Hamburger Schreib-Probe signifikant bessere Rechtschreibleistungen mit überwiegend großem Effekt im Vergleich zu den Gruppen, die das Schreiben mit „Lesen durch Schreiben“ oder mit der Rechtschreibwerkstatt erlernten. Zu keinem Messzeitpunkt unterschieden sich die beiden letztgenannten Didaktikgruppen signifikant in ihren Rechtschreibleistungen.“ (XI)

Aber wie aussagekräftig sind diese Werte? Eine Beurteilung ist aus verschiedenen Gründen schwierig, weil die erforderlichen Daten zum Teil an verschiedenen Orten und in unterschiedlichen Formaten vorliegen, so dass sie nicht leicht aufeinander beziehbar sind, und weil vor allem Grunddaten wie Muttersprache, Geschlecht, Sozialstatus nicht nach Ansätzen und Teilstichproben aufgeschlüsselt sind.

In der folgenden Tabelle sind die standardisierten Testwerte für den Längsschnitt im Vergleich zusammengefasst:

Tabelle 24. Deskriptive Statistiken der Rechtschreibleistungen (z-standardisiert) nach Didaktik im Längsschnitt.

	GT 1.2		GT 2.1		GT 2.2		GT 3.1		GT3.2		N
	M	SD	M	SD	M	SD	M	SD	M	SD	
Recht- schreib- werkstatt	-.36	1.10	-.33	1.09	-.38	1.24	-.34	1.19	-.28	1.04	121
Lesen durch Schreiben	.02	.86	.06	.87	.18	.67	.13	.71	-.11	1.08	79
Fibel	.53	.61	.43	.77	.34	.62	.34	.70	.49	.52	84

Anmerkung. GT 1.2 = Graphemtreffer der Hamburger Schreib-Probe Ende Jahrgangsstufe 1.

In den Querschnitten sind die Unterschiede weniger groß, die Rechtschreibwerkstatt schneidet im Vergleich besser, die Fibel schwächer ab als im Längsschnitt:

Tabelle 27. Deskriptive Statistiken der Rechtschreibleistungen (z-transformierte Werte der Graphemtreffer der Hamburger Schreib-Probe) der Jahrgangsstufen Ende eins bis Mitte drei gegliedert nach Didaktik im Querschnitt.

	GT 1.2			GT 2.1			GT 2.2			GT 3.1		
	M	SD	N	M	SD	N	M	SD	N	M	SD	N
RSW	-.17	1.08	552	-.20	1.13	636	-.16	1.09	571	-.16	1.11	513
LdS	.23	.76	84	.17	.76	194	.12	.97	186	-.04	.94	197
Fibel	.47	.51	152	.38	.62	249	.33	.58	210	.36	.67	245
gesamt			788			1079			967			955

Anmerkung. GT 1.2 = Graphemtreffer der Hamburger Schreib-Probe Ende Jahrgangsstufe 1; RSW = Rechtschreibwerkstatt; LdS = Lesen durch Schreiben.

Tabelle 28. Deskriptive Statistiken der Rechtschreibleistungen (z-transformierte Werte der Graphemtreffer der Hamburger Schreib-Probe) der Jahrgangsstufen Ende drei bis Ende vier gegliedert nach Didaktik im Querschnitt.

	GT 3.2			GT 4.1			GT 4.2		
	M	SD	N	M	SD	N	M	SD	N
RSW	-.20	1.14	546	-.14	1.08	615	-.22	1.14	476
LdS	.16	.83	187	.03	1.09	244	.07	.95	180
Fibel	.31	.62	251	.26	.65	301	.35	.56	261
gesamt			984			1160			917

Anmerkung. GT 3.2 = Graphemtreffer der Hamburger Schreib-Probe Ende Jahrgangsstufe 3; RSW = Rechtschreibwerkstatt; LdS = Lesen durch Schreiben.

Auffällig sind die unterschiedlichen Streuungen in den drei Gruppen; allerdings können sie nicht verlässlich auf die Methode zurückgeführt werden. Beispielsweise wies die Gruppe Rechtschreibwerkstatt schon im Vortest bei der phonologischen Bewusstheit im engeren Sinn eine sehr große Varianz auf (97):

Tabelle 21. Deskriptive Statistiken der Vorkenntnisse.

	N	PB i.w.S. (A)		PB i.e.S. (B)		PB ge- samt (A + B)		Buchsta- benkennt- nis	
		M	SD	M	SD	M	SD	M	SD
Rechtschreib- werkstatt	121	16.9	3.2	11.6	15.2	28.5	7.4	18.3	7.1
Lesen durch Schreiben	79	17.3	3.3	15.7	5.3	33.0	7.3	22.2	3.9
Fibel	84	16.6	3.8	11.3	5.1	27.9	8.0	17.8	6.9

Anmerkung. PB = Phonologische Bewusstheit; i.e.S. = im engeren Sinn; i.w.S. = im weiteren Sinn.

Solche Besonderheiten müssten diskutiert werden. Vor allem können noch so saubere statistische Berechnungen Schwächen des Designs (s. oben 3. und 4.) und der Kriterien (s. unten 6. und 7.) nicht ausgleichen.

Im Längsschnitt gingen immerhin einige Voraussetzungen in die Berechnung der Effekte mit ein, um die Vergleichbarkeit der Gruppen zu stärken (102)

„Als Kovariaten dienten die Ergebnisse in Punkten der Gesamtskala Phonologische Bewusstheit sowie die vorschulische Buchstabenkenntnis (vgl. Kapitel 3.2.2). Darüber hinaus ging auch die Familiensprache der Kinder als Kovariate in die längsschnittlichen Berechnungen ein.“

Leider gibt es keine Daten über Unterschiede zwischen Klassen *innerhalb* der Ansätze. Sie könnten helfen, genauer abzuschätzen, in welchem Umfang die Leistungsdifferenzen auf die Konzepte oder auf ihre Umsetzung durch die Lehrperson bzw. besondere Bedingungen in den Lerngruppen zurückzuführen sind (s. zur Bedeutung dieser Ebene ausführlicher oben 3.). Gerade bei so kleinen Stichproben wie hier, können wenige Ausreißer das Gesamtbild erheblich verändern.

Die pauschale Forderung (139f.) man müsse bei den Kindern von Anfang an die Einhaltung orthographischer Regeln einfordern, verkennt deren Komplexität, die weder über silbenanalytische Modelle noch als Faustregeln wie „Nach Kurzvokal stehen zwei Konsonanten; hörst du nur einen, wird er in der Regel verdoppelt“, vollständig erfasst, geschweige denn Anfänger*innen abverlangt werden kann. Auch in dieser Hinsicht ist eine Stufung der Anforderungen geboten, die sich am individuellen Entwicklungsstand orientieren muss.

6. Die Ergebnisse werden nur intern verglichen, **nicht** aber auf die **Normen der Hamburger Schreibprobe** bezogen. Nach diesem bundesweit repräsentativen Maßstab entsprechen „Lesen durch Schreiben“-Kinder in Kuhls Studie schon zum Ende des Anfangsunterrichts und über die Grundschulzeit hinweg der bundesdeutschen Norm. Sogar der Anteil besonders leistungsschwacher Schüler*innen ist niedriger als in der Bonner Gesamtstichprobe - besonders auffällig in Klasse 1 und 2. Die „Ausreißer“ bilden die (zusammengefassten) Fibel-Klassen (nach oben) und die Rechtschreibwerkstatt-Klassen (nach unten). Insofern wäre zu fragen, welche besonderen Bedingungen in diesen beiden Fällen zu den erwartungswidrigen Ergebnissen geführt haben (spezifische Stärken/ Schwächen der eingesetzten Methoden; Zusammensetzung der Schülerschaft; Engagement/ Kompetenz der Lehrer*innen).

Im methodischen Teil begründet Kuhl die Wahl der Hamburger Schreibprobe als Testinstrument wie folgt:

„Als Maß des Lernstands der Kinder ermöglichen umfangreiche Normtabellen und Vergleichswerte aus dem Jahr 2012 jedoch die Interpretation der erreichten Rohwerte.“ (79f.)

In der eigenen Auswertung nutzt er die Möglichkeit, seine Ergebnisse auf die bundesweiten Referenzwerte zu beziehen aber nicht.

Da für den Längsschnitt die Rohdaten nicht berichtet werden, kann dieser Schritt nur für die Querschnitte, zu denen die folgende Übersicht vorliegt, nachträglich ergänzt werden:

Tabelle 25. Deskriptive Statistiken der Rechtschreibleistungen (Rohwerte der Graphemtreffer der Hamburger Schreib-Probe) der Jahrgangsstufen Ende eins bis Mitte drei gegliedert nach Didaktik im Querschnitt.

	GT 1.2			GT 2.1			GT 2.2			GT 3.1		
	M	SD	N	M	SD	N	M	SD	N	M	SD	N
RSW	46.6	9.3	552	50.6	8.5	636	127.8	17.4	571	169.2	17.6	513
LdS	50.1	6.6	84	53.4	5.8	194	132.2	15.5	186	171.1	15.0	197
Fibel	52.1	4.4	152	55.0	4.6	249	135.6	9.3	210	177.4	10.6	245
gesamt	48.0	8.6	788	52.1	7.5	1079	130.4	15.9	967	171.7	15.9	955

Anmerkung. GT 1.2 = Graphemtreffer der Hamburger Schreib-Probe Ende Jahrgangsstufe 1; RSW = Rechtschreibwerkstatt; LdS = Lesen durch Schreiben.

Tabelle 26. Deskriptive Statistiken der Rechtschreibleistungen (Rohwerte der Graphemtreffer der Hamburger Schreib-Probe) der Jahrgangsstufen Ende drei bis Ende vier gegliedert nach Didaktik im Querschnitt.

	GT 3.2			GT 4.1			GT 4.2		
	M	SD	N	M	SD	N	M	SD	N
RSW	171.5	17.8	546	254.5	25.5	615	254.0	23.9	476
LdS	177.0	13.0	187	256.5	25.9	244	259.9	19.8	180
Fibel	179.4	7.8	251	262.1	15.4	301	265.8	11.7	261
gesamt	174.6	15.6	984	255.9	23.7	1160	258.5	20.9	917

Anmerkung. GT 3.2 = Graphemtreffer der Hamburger Schreib-Probe Ende Jahrgangsstufe 3; RSW = Rechtschreibwerkstatt; LdS = Lesen durch Schreiben.

Vergleicht man diese Werte mit den Repräsentativdaten der Hamburger Schreibprobe (May 2014, 88), so kann man die Ergebnisse an einem externen Maßstab messen und die reale Bedeutung der Unterschiede besser einschätzen, als wenn man nur den internen Vergleich untereinander hat:

Grafemtreffer über die Jahrgangsstufen 1 Ende bis 4 Ende hinweg							
HSP-Zielwert	61	63	148	191	191	277	277
HSP-Mittelwert	49.1	53.8	131.9	174.9	177.7	257.9	260.6
arith. Mittel	1E	2M	2E	3M	3E	4M	4E
RSW	46.6	50.6	127.8	169.2	171.5	254.5	254.0
LdS	50.1	53.4	132.2	171.1	177.0	256.5	259.9
Fibeln	52.1	55.0	135.6	177.4	179.4	262.1	265.8

In diesem Vergleich liegen die nach „Lesen durch Schreiben“ unterrichteten Kinder (selbst am Maßstab orthographischer Richtigkeit gemessen) mal geringfügig über, mal etwas unter dem bundesdeutschen Durchschnitt, wobei die Abweichungen im Durchschnitt -0.7 Graphemtreffer betragen, aber zwischen + 1.0 (Ende Klasse 1!) und - 3.8 Graphemtreffern (Mitte Klasse 3) schwanken. Ende Klasse 4 liegen sie bei minus 0.7 auf 277 mögliche Graphemtreffer. Allerdings wird immer wieder behauptet, dass besonders die leistungsschwachen Schüler*innen bei „Lesen durch Schreiben“ Nachteile zu befürchten hätten. Dankenswerterweise hat Kuhl (123ff.) dazu eine Sonderauswertung durchgeführt, wenn auch nur für die Querschnittstudien. Stimmt diese These, müssten sich von den „Lesen durch Schreiben“-Kindern im unteren Leistungsviertel der Bonner Gesamtstichprobe mehr als 25% finden. Tatsächlich sind es über die verschiedenen Messzeitpunkte hinweg nur 16.0%, 15.1%, 17.4%, 26.9%, 21.4%, 21.3% bzw. 20.0% (124). Das bedeutet: Gerade in den ersten beiden Schuljahren, deren Ergebnisse man noch am ehesten auf das Schreiben mit der Anlauttabelle beziehen kann, haben in der Studie auch die leistungsschwachen Kinder keinen Nachteil. Erstaunlicherweise verliert Kuhl diesen Befund in seinen Folgerungen wieder aus dem Blick, in denen er (132, 134, 136f.) generalisierend eine grundsätzliche Benachteiligung schwächerer Schüler*innen durch „offene“ Lehrmethoden behauptet und nicht auf deren jeweils spezifische methodische Ausgestaltung eingeht.

Für die Fibelklassen betragen die Abweichungen im Durchschnitt + 3.1 Graphemtreffer gegenüber den HSP-Durchschnittswerten, mit einer Streuung von +1.2 bis +5.2 und ebenfalls starken Schwankungen über die Jahrgangsstufen hinweg.

Die Rechtschreibwerkstatt liegt im Durchschnitt 4.5 Graphemtreffer unter dem bundesweiten Durchschnitt, wobei die Abweichungen zwischen -2.5 und -6.6 Graphemtreffern schwanken.

Festzuhalten ist, dass es bei allen drei Gruppen keine altersbezogene Tendenz im Sinne einer systematischen Verbesserung oder Verschlechterung gibt, sondern eine wechselnde Zu- bzw. Abnahme. Diese Sprünge - vermutlich den für die einzelnen Ansätze sehr unterschiedlichen Stichproben zu den einzelnen Terminen geschuldet - erschweren zusätzlich eine didaktisch-methodische Interpretation der Effekte, auch wenn die Abweichungen vom bundesdeutschen Durchschnitt gleichsinnig in dieselbe Richtung weisen.

Vier Befunde aber sind eindeutig - immer beschränkt auf die Stichproben der vorliegenden Studie :

1. Die nach „Lesen durch Schreiben“ unterrichteten Kinder liegen sowohl am Anfang als auch am Ende im Rahmen des bundesdeutschen Durchschnitts, selbst wenn man ihre Leistungen am Kriterium orthographischer Richtigkeit misst.
2. Sie erreichen ihr bestes Ergebnis Ende der ersten Klasse - also zu einer Zeit, da das lautorientierte Schreiben mit der Anlauttabelle dominiert. Das gilt auch für die leistungsschwächsten Schüler*innen.
3. Auch die Rechtschreibwerkstatt erreicht ihr (relativ) bestes Ergebnis zu diesem Zeitpunkt, weist aber im Übrigen deutlich schlechtere Werte als der bundesdeutsche Durchschnitt auf.
4. Ebenfalls auffällig sind die Abweichungen der Fibelklassen - allerdings nach oben.

Erklärungsbedürftig sind also die letzten beiden Ergebnisse - nicht aber die in der Öffentlichkeit kontrovers diskutierten Leistungen in den „Lesen durch Schreiben“-Klassen.

7. Für die Bewertung der Leistungen bleibt die Wahl des **Maßstabs** (richtige Grapheme) begründungsbedürftig. So wurden die unterschiedlichen Ansprüche an das Schreibniveau Ende Klasse 1/2 (alphabetische Strategie) und die Entfaltung dieser Basiskompetenz durch orthographische und morphematische Strategien (Klasse 2-4) bei der Auswertung nicht berücksichtigt.

Da sich sowohl die Fibeln als auch „Lesen durch Schreiben“ auf den Anfangsunterricht beschränken (s. oben 3.) ist ein direkter Vergleich der Wirksamkeit (zumindest mit den unter 2. genannten Einschränkungen) nur für Ende der ersten Klasse, evtl. noch Mitte 2. Klasse möglich. Zu diesem Zeitpunkt ist aber die orthographische Richtigkeit von Wörtern nur ein begrenzt taugliches Maß für die Rechtschreibkompetenz, da die Kinder noch kaum über ein zureichendes Rechtschreibwissen verfügen, um unbekannte Wörter mit zureichender Aus-

sicht auf Erfolg richtig schreiben zu können. Richtigkeit kann als Anspruch allenfalls an bekannte Wörter gestellt werden, wie sie den Fibelkindern aus den Lehrgangsstunden verfügbar sind.

Die KMK-Empfehlungen für die Grundschule und die Lehrpläne der einzelnen Bundesländer formulieren deshalb im Einklang mit der Schriftspracherwerbsforschung als Kriterium für die Schreibleistungen in der Anfangsphase die lautgerechte Verschriftung¹⁰.

Hier hätte das - nicht ausgewertete - Pseudowörterdiktat (87) als zusätzliche Schreibprobe hilfreich sein können, auch um auszuschließen, dass die Testwörter mit Kindern einzelner Klassen geübt sein und damit das Ergebnis verfälschen könnten. Allerdings hätte das vorausgesetzt, dass sich der Autor mit Modellen des Schriftspracherwerbs auseinandersetzt und - wie beispielsweise die von ihm verkürzt zitierte Lorenz (2017) - die Beherrschung der alphabetischen Strategie als wichtigen Schritt in der Rechtschreibentwicklung anerkennt. Dieser Hinweis macht deutlich, wie wichtig es ist, Leistungen entwicklungsbezogen zu interpretieren, d. h. auf Anforderungen zu beziehen, die den verschiedenen Phasen der Rechtschreibentwicklung angemessen sind.

Insofern relativiert sich auch Kuhls Hinweis auf die Studie mit freien Texten von Lorenz (2017), die er (148) als Stütze für seine Befunde interpretiert. Bei der Aufgabe handelte es sich um einen Brief, den die Kinder schreiben sollten, und in der Tat machten Kinder umso weniger Rechtschreibfehler, je weniger der Unterricht nach Auskunft der Lehrer*innen durch „Lesen durch Schreiben“ bestimmt war. Bei Kontrolle anderer Einflussfaktoren erklärten die Methodenunterschiede allerdings nur wenige Prozent der Varianz in den Rechtschreibleistungen (a.a.O., 136f.). Zu dem Zeitpunkt (Ende des zweiten Drittels der ersten Klasse) ist eine Richtigschreibung außerdem - wie oben erläutert - nur für Wörter zu erwarten, die den Kindern bekannt waren. Unbekannte Wörter sollten sie lautbezogen konstruieren können - ihre Richtigschreibung ist aber nur dann zu erwarten, wenn sie

¹⁰ So heißt es in den Empfehlungen der Kultusministerkonferenz (2015, 14): „Beim Schriftspracherwerb ist das lautorientierte Schreiben ein Entwicklungsschritt auf dem Weg zum normgerechten Schreiben. Das Kind wird ausgehend von seinen lautorientierten Verschriftungen von Anfang an systematisch an das orthografisch korrekte Schreiben herangeführt.“ (Download unter: <https://t1p.de/KMK-2015> [Abruf: 22.4.2020]) - und im Lehrplan Deutsch für die Grundschulen in NRW (2008, 26): „Grundlage für erfolgreiches Schreibenlernen ist die phonologische Bewusstheit, d. h. die Einsicht in die Lautstruktur der Sprache sowie in die Laut-Buchstaben-Entsprechung der Alphabetschrift. (...) Auf der Grundlage der Laut-Buchstaben-Zuordnung erwerben sie Einsichten in die Besonderheiten der deutschen Rechtschreibung.“ (Download unter: https://www.schulentwicklung.nrw.de/lehrplaene/upload/klp_gs/LP_GS_2008.pdf [Abruf: 22.4.2020])

aus Basisgraphemen bestehen (RAT, WUT, TOMATE usw.). In dieser Hinsicht stellt Lorenz aber fest, dass der Anspruch der lautgerechten Darstellung der Wörter erheblich besser erfüllt wird, je stärker der Unterricht durch „Lesen durch Schreiben“ bestimmt war (a.a.O., 141). Dieser Befund erfordert nach Ansicht von Lorenz eigentlich eine Überprüfung der Hypothese, dass nicht nach „Lesen durch Schreiben“ unterrichtete Kinder eher Lehwörter verwendet haben, während die LdS-Kinder in ihren Texten auch häufiger unbekannte Wörter verwendet haben (a.a.O., 140, 143). Allerdings: So plausibel diese Vermutung auch sein mag, überprüft werden kann sie anhand der vorliegenden Daten auch bei Lorenz nicht. Aber sie zeigt, dass sich Studien nicht nur mit der Feinstruktur des Unterrichts sehr viel genauer beschäftigen müssen, sondern auch mit den Produkten der Kinder, wenn sie Differenzen zwischen den bisherigen Untersuchungen aufklären wollen.

8. Die äußerst knappe Darstellung des Forschungsstands in Kap. 1.10 beschränkt sich auf deutschsprachige Studien und dabei zusätzlich auf vergleichende Evaluationen. In der lediglich tabellarischen Übersicht werden sie zudem unvollständig und durch fehlende Kommentierung zum Teil unzutreffend vorgestellt. Bei der Einordnung der Ergebnisse wurden der deutschsprachige und der internationale **Forschungsstand** nur **einseitig** berücksichtigt. Weder die Grundlagenforschung zum Schriftspracherwerb noch neuere Experimentalstudien werden diskutiert, die die Bedeutung des lautorientierten Schreibens in der Anfangsphase für die spätere Rechtschreibentwicklung belegen.

Kuhl versucht, seine Befunde bzw. deren Interpretation abzustützen, indem er behauptet, diese befänden sich in Übereinstimmung mit „vielen Studien aus dem deutschsprachigen Raum“ (132) und dem internationalen Forschungsstand (XII). Dabei beklagt er, es gebe „kaum empirisch-wissenschaftliche Forschung, die sich der Frage nach der Wirksamkeit verschiedener Lese-Rechtschreib-Didaktiken“ widme (2). Andererseits werden wichtige Studien, vor allem aus dem angelsächsischen Bereich von ihm gar nicht diskutiert.

Die Darstellung des Forschungsstands in Kap. 1.10 beschränkt sich auf eine tabellarische Übersicht (die sich stark an Lorenz 2017 orientiert). Sie wird auf lediglich einer Seite (66) kurz kommentiert. Durch die Beschränkung auf jeweils eine Publikation für die meisten Studien und den Verzicht auf eine differenzierte Diskussion der Befunde kommt es

dabei zu Fehleinschätzungen. So werden für „Lesen durch Schreiben“ bei May und Kirschhock schlechtere Ergebnisse unterstellt, weil lediglich Gesamtwerte zur Beurteilung herangezogen werden, die Hinweise der Autor*innen selbst auf ungünstigere soziale Bedingungen in den entsprechenden Klassen nicht berücksichtigt werden. Vom Schreibvergleich Bundesrepublik - DDR werden nur zwei zusammenfassende Berichte in Lehrerzeitschriften zitiert, an denen Kuhl kritisiert, dass „keine Angaben zu Effektstärken, Mittelwerten, Standardabweichungen“ (66) gemacht würden¹¹. Genau diese Daten sind aber in Brügelmann u. a. (1994) verfügbar, die beispielsweise Funke (2014) in seiner Metaanalyse nutzt. Aus dem Projekt „BeLesen“ werden die nach vier Ansätzen und mehreren Leistungsdimensionen sowie Zeitpunkten differenzierten Ergebnisse in der Darstellung so verkürzt, dass die entscheidende Einsicht von Schröder-Lenzen/ Merkens (2006, 15) in ihrem letzten Bericht nicht sichtbar wird:

„Varianz- und regressionsanalytisch lassen sich nur geringe Effekte hinsichtlich unterschiedlicher fachdidaktischer und pädagogischer Orientierungen des Anfangsunterrichts ausmachen. Tendenziell günstigere Lernergebnisse lassen sich für lehrgangnahe Formen des Rechtschreibunterrichts und für eher spracherfahrungsorientierte Formen des Leseunterrichts ausweisen.“

Leider hat Kuhl verfügbare Forschungsüberblicke wie Metze (2008), Brügelmann/ Brinkmann (2006/13), Lorenz (2017) und Jambor-Fahlen (2018) gar nicht oder wie Funke (2014) nur selektiv genutzt. In ihnen werden einzelne Studien auch im Detail diskutiert, so dass sie zu entsprechend differenzierteren Einschätzungen des Forschungsstands kommen. Die Interpretation mancher Befunde bleibt kontrovers und in der Tat kann man einzelne Studien unterschiedlich gewichten und deshalb zu unterschiedlichen Resümees kommen. Aber man muss sie zumindest zur Kenntnis nehmen und ihre Stärken und Schwächen explizit diskutieren, wenn man sich auf den (internationalen) Forschungsstand beruft.

Überhaupt keinen Bezug nimmt Kuhl auf die Schriftspracherwerbsforschung seit den 1980er Jahren und die auf ihrer Basis entwickelten Modelle des Lesen- und Schreibenlernens über die Differenzierung von Strategien. Für seine verallgemeinernden Folgerungen besonders problematisch ist die Vernachlässigung des internationalen Forschungsstands zum „invented spelling“, also zum Einstieg in den Schriftspracherwerb über lautorientiertes Schreiben. Für den angelsächsischen Raum hat schon das National Early Reading Panel

¹¹ Umso erstaunlicher, dass diese Kennwerte in der Dissertation selbst für den Längsschnitt fehlen.

(2008) auf die hohe Bedeutung der Einsicht in das alphabetische Prinzip verwiesen und darauf bezogen aus seinem Forschungsüberblick eine starke Beziehung zwischen Niveau der lautorientierten Verschriftung beim „invented spelling“ zu Schulbeginn und dem orthographischen Niveau in der 2. Klasse berichtet ($d = .60$). Seitdem sind über die Jahre hinweg in verschiedenen Ländern und Sprachen eine Reihe weiterer, auch experimenteller Studien durchgeführt worden (vgl. den Überblick bei Brügelmann 2019). Die Ergebnisse dieser Studien bestätigen die Annahme der Schriftspracherwerbsmodelle, dass das lautorientierte Verschriften von Wörtern trotz der phasenweisen Rechtschreibfehler ein produktiver Einstieg in den Orthographieverwerb ist. Dass Fehlertoleranz in dieser Phase nicht Verzicht auf orthographische Modelle und auf Rückmeldungen zu den Schreibversuchen der Kinder bedeutet, ist für den Spracherfahrungsansatz immer wieder betont worden (vgl. u. a. Brügelmann 1992; Brinkmann 2015b; Brinkmann/ Brügelmann 2019). Leider hat Kuhl solche methodisch konkretisierenden Texte in seiner stark verkürzten Darstellung des Spracherfahrungsansatzes in Kap. 1.6 und bei der Bewertung der Ergebnisse in Kap. 5 ebenfalls nicht berücksichtigt.

9. Zum Schluss: Wer sich ein Bild von den Stärken und Schwächen der in dieser Studie untersuchten Ansätze machen will, sollte den Hinweis des Autors ernst nehmen, dass die Rechtschreibleistung in einem Wörter- und Satz-Diktat nur ein **Teil-Aspekt schulischer Ziele** ist; die inhaltliche, sprachliche und orthographische Qualität freier Texte, die Entwicklung der Lesefähigkeit, aber auch übergreifende Kompetenzen wie die Förderung der Selbstständigkeit der Schüler*innen wurden in der Studie (bewusst) nicht erfasst.

Dankenswerterweise weist Kuhl am Ende selbst darauf hin, dass *ein* Effekt-Kriterium allein nicht über die Qualität eines Konzepts entscheiden kann (148f.):

„Zukünftige Studien könnten sich mit den Auswirkungen der Unterrichtsmethoden auf die Lesekompetenz der Kinder befassen. Vielleicht fördern die Ansätze auch die Kreativität der Kinder in unterschiedlichem Maße. Eine weitere interessante Fragestellung ist die nach der Eloquenz und der Größe des Wortschatzes der Kinder in Abhängigkeit von den verschiedenen Didaktiken. Vielleicht variiert auch die Elaboriertheit von Aufsätzen der Kinder mit der eingesetzten Unterrichtsmethode.“

Andere Studien (vgl. etwa Schröder-Lenzen/ Merckens 2006) haben diese Erkenntnis auch bereits umgesetzt. Schon der Vergleich der Rechtschreibleistungen fiel im Schreibvergleich Bundesrepublik-DDR (Brügelmann u. a. 1994) zwischen denselben Gruppen im freien Text anders aus im Diktat.

Zusätzlich sei daran erinnert, dass etwa „Lesen durch Schreiben“ ein Konzept zur Förderung des Lesenlernens ist, also in dieser Studie gar nicht in seinem zentralen Zielbereich überprüft wird. Zudem haben in diesem Ansatz pädagogische Ziele, die über das fachliche Lernen hinausgehen und in der Studie nicht untersucht wurden, einen besonders hohen Stellenwert, wie Kuhl selbst anerkennt (27):

„Auf die Entwicklung der Selbstständigkeit der Kinder wird besonderer Wert gelegt.“

Kuhl (137ff.) hält selbständiges Arbeiten und Wahlmöglichkeiten auf dieser Altersstufe und besonders für „schwache“ Schüler*innen für eine Überforderung. Auch dies ist eine Verallgemeinerung, die den unterschiedlichen Praxisformen nicht gerecht wird. Arbeits- und Verkehrsformen müssen eingeführt, besser noch: gemeinsam entwickelt und in der Gruppe verabredet, und die konkreten Anforderungen (z. B. Tages- oder Wochenplan) auf die Möglichkeiten der einzelnen Kinder abgestimmt werden (vgl. die differenzierten Hinweise in Brügelmann/ Brinkmann 2009, 10ff.).

Verdienstvoll ist immerhin, dass Kuhl versucht hat, die Lese- und Schreibmotivation der Kinder zumindest per Fragebogen zu erheben. Auch wenn sich die Ergebnisse (126ff.) nur eingeschränkt interpretieren lassen (s. zu der mangelnden bzw. nicht nachvollziehbaren Verschränkung der Stichproben oben Kap. 4), sollte das Fehlen von Vorteilen für „Lesen durch Schreiben“ in diesen Dimensionen Anlass sein, möglichen Ursachen für dieses erwartungswidrige Ergebnis nachzugehen.

Die pauschalisierende Folgerung (148)

„Nachdem sich gezeigt hat, dass ein systematischer Unterricht mit einer Fibel dem Unterricht, der auf dem Spracherfahrungsansatz aufbaut, im Hinblick auf die Rechtschreibleistung der Kinder überlegen ist, sollte sich weitere Forschung mit einer differenzierten Betrachtung und Evaluation verschiedener Fibelwerke befassen.“

ist allerdings durch die Datenlage in dieser Studie nicht zu rechtfertigen.

Fazit

Hätten der Doktorand und seine Betreuerin der Dissertation bei ihren vor anderthalb Jahren medial gepushten Vorabveröffentlichungen geschrieben: Der Forschungsstand im Bereich des Lese- und Schreibunterrichts ist nicht einheitlich, die Ergebnisse unserer Studie haben einige diskussionsbedürftige Befunde erbracht, hätten diese für die weitere Diskussion hilfreich sein können. So lohnt z. B. der Frage nachzugehen, ob Ansätze wie „Lesen durch Schreiben“ oder die Rechtschreibwerkstatt in der Breite tatsächlich so umgesetzt werden, wie ihre Autor*innen sie auf der Konzeptebene formulieren, oder welche Bedingungen dazu geführt haben könnten, dass die Fibel-Lehrgänge in der untersuchten Stichprobe bessere Ergebnisse erzielt haben, als bundesweit im Durchschnitt üblich. Vor allem stellt sich die Frage, warum in dieser Studie zwei Methoden, die beide mit dem lautorientierten Schreiben beginnen, so unterschiedlich abgeschnitten haben. Die unterschiedlichen Ergebnisse machen noch einmal deutlich, dass es nicht um das OB geht, sondern auf das WIE ankommt, das leider in dieser Studie nicht untersucht wurde.

Stattdessen wurden die Vorabveröffentlichungen zu einer Zeit, da die Datengrundlage seriös nicht nachprüfbar war, trotz ihrer begrenzten Aussagekraft medial und gegenüber den Ministerien eingesetzt, um bildungspolitische Entscheidungen gegen „Lesen durch Schreiben“ und gegen das lautorientierte Schreiben generell einzufordern. Wie die hier vorgelegte Analyse zeigt, erfüllt auch die aktuelle Publikation der Studie die Anforderung der Nachvollziehbarkeit noch nicht.

Die in diesem Beitrag offen gelegten Einschränkungen bedeuten nur teilweise eine Kritik an der persönlichen Leistung des Doktoranden, eine solche Studie als Qualifikationsarbeit auf die Beine gestellt zu haben. Aber die Grenzen ihrer Aussagekraft müssten bei den Folgerungen für die Praxis sehr viel deutlicher bedacht werden.

Kuhl schreibt, man solle die Evaluation nicht den Pädagog*innen überlassen¹² (143). Die Schwächen dieser Studie zeigen, dass es noch gefährlicher wäre, sie allein Psycholog*innen

¹² „Darüber hinaus ist es positiv zu bewerten, dass die Studie von unvoreingenommenen externen, d. h. nicht in den Schulalltag eingebundenen Psychologen geplant und durchgeführt wurde. Da Forschungsmethodik und Evaluation Schwerpunkte ihres universitären Studiums sind, eignen sich Psychologen oder andere Sozialwissenschaftler besonders für diese Art der Forschung. Für eine Evaluation im schulischen Setting erscheint es nicht sinnvoll, diese von Pädagogen planen, durchführen und auswerten zu lassen. So lässt sich eine eventuelle Verzerrung der Studienplanung und -auswertung durch ein Priming durch universitäre Lehre oder persönliche Erfahrung im eigenen Unterricht oder Werteinstellungen ausschließen.“ (143)

zu überlassen, die wegen fehlender Kenntnis der didaktischen Feinstrukturen und der Besonderheiten des alltäglichen Schulbetriebs die Bedeutung wichtiger Bedingungen für die Anlage ihrer Untersuchung und die Interpretation der Befunde übersehen.

Eine empirische Grundlage für das teilweise in den Medien geforderte und von einigen Ministerien auch umgesetzte Verbot von „Lesen durch Schreiben“ oder generell eines lautorientierten Schreibens im Anfangsunterricht liefert diese Untersuchung jedenfalls nicht.

Quellenverzeichnis

- Brinkmann, E. (1997): Rechtschreibgeschichten - Zur Entwicklung einzelner Wörter und orthographischer Muster über die Grundschulzeit hinweg. Universi Verlag: Siegen (2. Aufl. 2002). DOI: 10.13140/RG.2.2.26764.00641. Download unter: <https://t1p.de/RS-Geschichten> [Abruf: 26.4.2020]
- Brinkmann, E. (Hrsg.) (2015a): Rechtschreiben in der Diskussion - Schriftspracherwerb und Rechtschreibunterricht. Beiträge zur Reform der Grundschule, Bd. 140. Grundschulverband: Frankfurt.
- Brinkmann, E. (2015b): Richtig schreiben lernen nach dem Spracherfahrungsansatz. In: Brinkmann (2015, 44-53).
- Brinkmann, E. (2018): Der Spracherfahrungsansatz im Anfangsunterricht. In: Grundschule, 50. Jg., H. 6, 28-32. Download unter: <https://t1p.de/bri-18-spracherfahrung> [Abruf: 26.4.2020]
- Brinkmann, E./ Brügelmann, H. (2019): Empirische Studien zum Umgang mit Rechtschreibfehlern und die Bedeutung konkreter Befunde für den Rechtschreibunterricht. Download unter: <https://t1p.de/RS-Fehler> [Abruf: 26.4.2020]
- Brinkmann, E., u. a. (2008/19.): ABC-Lernlandschaft. vpm/ Klett: Stuttgart (vollst. überarb. Neuauflage 2019).
- Brügelmann, H. (1978): Veränderungen des Curriculum auf seinem Weg vom Autor zum Kind. In: Zeitschrift für Pädagogik, 24. Jg., H. 4, 601-618. Download: <https://t1p.de/Curr-Wandlungen> [Abruf: 22.4.2020]
- Brügelmann, H. (1984): Lesen- und Schreibenlernen als Denkentwicklung - Voraussetzungen eines erfolgreichen Schrifterwerbs. In: Zeitschrift für Pädagogik, 30. Jg., H. 1, 69-91.
- Brügelmann, H. (Hrsg.) (1986): ABC und Schriftsprache - Rätsel für Kinder, Lehrer und Forscher. DGLS-Jahrbuch "Lesen und Schreiben" 1. Ekkehard Faude: Konstanz (3. Aufl. 1991).
- Brügelmann, H. (1992): 20 Thesen zum Rechtschreibunterricht. In: GRUNDSCHULZEITSCHRIFT, H. 52, März 1992, 37-39. Download unter: <https://t1p.de/Thesen-RSU> [Abruf: 26.4.2020]
- Brügelmann, H. (2019): Empirische Studien zur Auswirkung eines laut-orientierten Konstruierens von Wörtern („invented spelling“) auf den Rechtschreiberwerb. Ein systematischer Überblick. Download unter: <https://t1p.de/invented-spelling> [Abruf: 26.4.2020]
- Brügelmann, H./ Brinkmann, E. (2006): Freies Schreiben im Anfangsunterricht? Eine kritische Übersicht über Befunde der Forschung. Download unter: http://www.grundschulverband.de/fileadmin/grundschulverband/Download/Forschung/Zur_Leseforschung_Feb._07.pdf [aktualisiert November 2013] [Abruf: 22.4.2020]
- Brügelmann, H./ Brinkmann, E. (2009): Öffnung des Anfangsunterrichts. Theoretische Prinzipien, unterrichtspraktische Ideen und empirische Befunde. Arbeitsgruppe Primarstufe/ Universität. Universi Verlag: Siegen (1. Aufl. 2008). Download unter: <https://t1p.de/OU-anfangsunterricht> [Abruf: 19.4.2020]

- Brügelmann, H./ Brinkmann, E. (2020): Bremer Rechtschreibforscher/innen (II). Rechtschreibgespräche in 2., 3. und 4. Klassen der Grundschule. Zwischenbericht zur Eingangserhebung im Sommer 2019. Noch unveröff. Ms. für die Senatorin für Kinder und Bildung: Bremen.
- Brügelmann, H./ Richter, S. (Hrsg.) (1994): Wie wir recht schreiben lernen. Zehn Jahre Kinder auf dem Weg zur Schrift. Libelle Verlag: CH-Lengwil (2. Aufl. 1996). Download unter: <https://www.pedocs.de/volltexte/2019/16858> [Abruf: 26.4.2020]
- Brügelmann, H./ Söhnen, R. (1981a): Ist Lesen Lehren lernbar? Plädoyer für eine vergleichende Erprobung von Lehrgängen für den Anfangsunterricht. In: IRA/D Beiträge, 4. Jg., 1/1981, 14-18. Download unter: <https://t1p.de/eva-lehrgang-1> [Abruf: 26.4.2020]
- Brügelmann, H./ Söhnen, R. (1981b): Analyse und Erprobung von Lehrwerken zum Erstlese- und Erstschreibunterricht. Eine Projektskizze. In: IRA/D Beiträge, 4. Jg., 1/1981, 18-24. Download unter: <https://t1p.de/eva-lehrgang-2> [Abruf: 26.4.2020]
- Brügelmann, H., u. a. (1994): "Schreibvergleich BRDDR" 1990/91. In: Brügelmann/ Richter (1994, 129 134).
- Downing, J. (1979): Reading and Reasoning. Chambers: Edinburgh.
- Fend, H. (2008): Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität. VS Verlag für Sozialwissenschaften: Wiesbaden.
- Frith, U. (Hrsg.) (1980): Cognitive processes in spelling. Academic Press: London.
- Funke, R. (2014): Erstunterricht nach der Methode Lesen durch Schreiben und Ergebnisse schriftsprachlichen Lernens - Eine metaanalytische Bestandsaufnahme. In: Didaktik Deutsch, 19. Jg., H. 36, 20-41.
- Grotlüschen, A./ Riekman, W. (Hrsg.) (2012): Funktionaler Analphabetismus in Deutschland. Ergebnisse der ersten leo. - Level-One Studie. Waxmann: Münster.
- Grotlüschen, A., u. a. (2019): LEO 2018 - Leben mit geringer Literalität. Pressebroschüre, Hamburg. Download unter: <http://blogs.epb.uni-hamburg.de/leo>
- Jambor-Fahlen, S. (2018): Lesen und Schreiben Lernen in der Grundschule. Fakten-Check. Mercator Institut für Sprachförderung und Deutsch als Zweitsprache. Universität: Köln. Download unter: https://www.mercator-institut-sprachfoerderung.de/fileadmin/Redaktion/PDF/Publikationen/Faktencheck_Lesen_und_Schreiben_in_der_Grundschule.pdf [Abruf: 22.4.2020]
- Kuhl, T./ Röhr-Sendmeier, U. N. (2018a): Rechtschreiberfolg nach unterschiedlichen Didaktiken. Eine kombinierte Längsschnitt-Querschnittstudie. Poster für den Bundeskongress der Schulpsychologie am 20.-22.9.2018 in Frankfurt. Download unter: <https://www.psychologie.uni-bonn.de/de/unsere-institut/abteilungen/methodenlehre-und-diagnostik/mitarbeiter-1/m.sc.-tobias-kuhl-1/poster-buko-rechtschreiberfolg-nach-unterschiedlichen-didaktiken-21.09.2018> [Abruf: 3.5.2020]
- Kuhl, T./ Röhr-Sendmeier, U. N. (2018b): Rechtschreiberfolg nach unterschiedlichen Didaktiken. Eine kombinierte Längsschnitt-Querschnittstudie. Kommentar zum Poster für den Bundeskongress der Schulpsychologie am 20.-22.9.2018 in Frankfurt. Vervielf. Ms. Institut für Psychologie der Universität: Bonn.

- Lorenz, A. (2017): Wer schreibt recht, wer schreibt schlecht? Eine Untersuchung des Zusammenhangs zwischen der Rechtschreibleistung von Erstklässlern und dem Ansatz „Lesen durch Schreiben“ von Jürgen Reichen in der Videostudie Deutsch des Projekts PERLE. Zentrum für Lehrerbildung der Universität Kassel, Reihe Studium und Forschung, Heft 27. University Press: Kassel.
- May, P. (2013): HSP 1-10. Hamburger Schreib-Probe. Manual / Handbuch: Diagnose orthografischer Kompetenz zur Erfassung der grundlegenden Rechtschreibkompetenzen. Neustandardisierung 2012 (mit separat publizierten „Hinweisen zur Durchführung und Auswertung“ für die einzelnen Jahrgänge). Verlag für Pädagogische Medien/ Klett: Stuttgart.
- Metze, W. (2008): Lernwegorientierter Schriftspracherwerb im Spiegel der Empirie und des Schulalltags. Ms. eines Vortrags in Zürich. Download unter: http://www.wilfriedmetze.de/Vortrag_Zurich_31.5.08.pdf [Abruf: 26.4.2020]
- Metze, W. (2016): Tobi. Das kann ich schon. Cornelsen: Berlin.
- National Early Literacy Panel (Hrsg.) (2008): Developing early literacy: A scientific synthesis of early literacy development and implications for intervention. National Institute for Literacy & The Partnership for Reading. Jessup, Maryland. Download unter: <https://lincs.ed.gov/publications/pdf/NELPReport09.pdf> [Abruf: 22.4.2020]
- OECD & Statistics Canada (Hrsg.) (1995): Grundqualifikationen, Wirtschaft und Gesellschaft. Ergebnisse der ersten internationalen Untersuchung von Grundqualifikationen Erwachsener. Organization of Economic Co-operation and Development: Paris.
- Scheerer-Neumann, G. (2020): Schreiben lernen nach Gehör? Freies Schreiben contra Rechtschreiben von Anfang an. Kallmeyer'sche Verlagsbuchhandlung: Seelze.
- Scheerer-Neumann, G., u. a. (1986): Andrea, Ben und Jana. Selbstgewählte Wege zum Lesen und Schreiben. In: Brügelmann (1986, 55-96).
- Schründer-Lenzen, A. (Hrsg.) (2006): Risikofaktoren kindlicher Entwicklung. Migration, Leistungsangst und Schulübergang. VS Verlag Sozialwissenschaften: Wiesbaden..
- Schründer-Lenzen, A./ Merkens, H. (2006): Differenzen schriftsprachlicher Kompetenzentwicklung bei Kindern mit und ohne Migrationshintergrund. In: Schründer-Lenzen (2006, 15-44).
- Sommer-Stumpenhorst, N. (2015): Die Rechtschreibwerkstatt. In: Brinkmann (2015a, 108-116).
- Temple, C. A., u. a. (1982): The beginnings of writing. Allyn and Bacon: London et al. (2nd rev. ed. 1988).
- Valtin, R./ Naegele, I. (Hrsg.) (1986): „Schreiben ist wichtig!“. Grundlagen und Beispiele für kommunikatives Schreibenlernen. Beiträge zur Reform der Grundschule Bd. 67/68. Arbeitskreis Grundschule: Frankfurt (4. Aufl. 1994).
- Wiemer, I./ Hüttenberger, M. (2015): Lesen durch Schreiben - Die Reichen-Methode. In: Brinkmann (2015a, 135-144).