

Sabine Röttig, Clara Schwerkolt & Guido Nottbusch

Die Entwicklung der Leseflüssigkeit in der Grundschule

Eine Longitudinalstudie über die interagierenden Dimensionen Dekodiergenauigkeit, Automatisierung, Lesegeschwindigkeit und Prosodie bei Kindern der Jahrgangsstufen 2 und 3

1 | Einleitung

Die meisten Erwachsenen haben ihren Leseprozess durch Übung automatisiert und können mit (mehr oder weniger) sinngemäßer Betonung (vor-)lesen. Doch was uns so selbstverständlich erscheint, verhält sich bei Kindern anders, z.B. zu Beginn des Lesenlernens und später bei Leseschwierigkeiten. Welche komplexen Prozesse beim Lesen ablaufen und dazu führen, dass der Lesefluss stockend, langsam oder monoton wirken kann, ist Gegenstand der Leseforschung und -didaktik. Leseflüssigkeit wird im angloamerikanischen Raum bereits seit den 1970er Jahren empirisch untersucht und theoretisch erfasst (Rosebrock & Nix 2006). Unter dem Begriff *reading fluency* werden Leseprozesse von Schülerinnen und Schülern auf Wort-, Satz- und lokaler Textebene erforscht. Im deutschsprachigen Raum liegen bislang zwar weniger entwickelte Theoriekonzepte und Befunde vor, jedoch wird der Leseflüssigkeit auch in der deutschdidaktischen Diskussion als eigenständiger Dimension der Lesekompetenz seit dem letzten Jahrzehnt vermehrt Aufmerksamkeit gewidmet (Rosebrock et al. 2017).

Die Förderung von Leseflüssigkeit sollte sich nach dem Entwicklungsstand des Kindes richten. Doch wie entwickelt sich Leseflüssigkeit? Wie genau, schnell oder expressiv lesen Kinder in welcher Altersstufe? Derzeit gibt es kaum zuverlässige Befunde, die eine Einschätzung der zeitlichen Entwicklung von Leseflüssigkeit oder die Dokumentation von Lernverläufen möglich machen (Rosebrock et al. 2017). Beim flüssigen Lesen handelt es sich um ein mehrdimensionales Konstrukt, das sich aus verschiedenen Teilfähigkeiten zusammensetzt. Es gibt daher nicht *den flüssigen Leser* oder *die unflüssige Leserin*. Manch einer liest langsam und stockend, eine andere schnell, aber fehlerhaft. Wieder ein anderer mit so vielen Pausen, dass er nicht versteht, was er liest. Welche Teilfähigkeiten bei welchen Kindern bereits oder noch nicht ausgeprägt sind und wie sich die Entwicklung von flüssig und weniger flüssig Lesenden unterscheidet, ist ebenfalls wenig erforscht.

Weitgehend übereinstimmend gilt Leseflüssigkeit als Brücke zwischen Dekodieren und Leseverstehen (Pikulski & Chard 2005), da kognitive Ressourcen für höhere Verstehensprozesse erst durch ein gewisses Maß an Leseflüssigkeit auf Wort- und Satzebene frei werden (Rosebrock et al. 2017). Durch eine erhöhte Beanspruchung der Aufmerksamkeit beim mühevollen

Entschlüsseln von Wörtern oder Sätzen kann das Verständnis längerer Textabschnitte beeinträchtigt werden. Flüssigen Leserinnen und Lesern hingegen stehen mehr Kapazitäten des Arbeitsgedächtnisses für eigentliche Verstehensprozesse zur Verfügung (Samuels 1985).

Intensiv wurde die Frage diskutiert, welche Komponenten von Lesekompetenz auf Wort- und Satzebene im Einzelnen bestimmend für das flüssige Lesen sind (Rosebrock & Nix 2006). In der gegenwärtigen Leseforschung wird jedoch eine Arbeitsdefinition weitgehend übereinstimmend geteilt (Nix 2011; Wolf & Katzir-Cohen 2001). Leseflüssigkeit wird danach als gestuftes Konstrukt verstanden, das sich aus unterschiedlichen Dimensionen zusammensetzt, und meint ein genaues, automatisiertes, schnelles und betontes Lesen (Rosebrock et al. 2017). Als wichtigste Indikatoren gelten somit die Dekodiergenauigkeit und die Automatisierung auf Wortebene sowie die Lesegeschwindigkeit und die Prosodie auf Satz- und Textebene (Gold 2009; Rasinski 2003; Rosebrock & Nix 2006). Anhand dieser Dimensionen wird Leseflüssigkeit messbar.

Mit Dekodiergenauigkeit bezeichnet man das Erzeugen der korrekten phonologischen Repräsentation eines Wortes durch den Abruf aus dem Sichtwortschatz oder durch das Synthetisieren der Lautbestandteile eines Wortes (Walter 2017). Gemäß der *Zwei-Wege-Theorie* nach Coltheart und Rastle (1994) sind unterschiedliche Repräsentationsformen von Wörtern in separaten mentalen Lexika gespeichert. Visuelle Buchstaben- und Wortmuster sind im Sichtwortschatz (orthographisches Lexikon), lautliche Repräsentationen im phonologischen Lexikon und Wortbedeutungen im semantischen Lexikon, dem Wortschatz, gespeichert. Geübte Leserinnen und Leser können flexibel auf die Speicherinstanzen zugreifen und dadurch auf zwei unterschiedliche Arten visuell Wörter erkennen: Bei der direkten, lexikalischen Route wird die Aussprache der Wörter aus dem Gedächtnis und damit dem Sichtwortschatz abgerufen, während bei der indirekten, nicht-lexikalischen Route die Aussprache mittels Graphem-Phonem-Zuordnung erschlossen werden muss (Lenhard 2013). Die indirekte Route wird beispielsweise bei unbekanntem Wörtern genutzt. Eine hohe Dekodiergenauigkeit gilt als Basiskomponente für die Ausbildung der weiteren Konstruktebenen von Leseflüssigkeit. Flüssige Leserinnen und Leser erkennen Wörter schnell, verlesen sich wenig und bemerken und korrigieren Lesefehler selbst (Rosebrock et al. 2017).

Für das Erfassen von Dekodiergenauigkeit sind Testverfahren in Form von Lautleseprotokollen (sogenannte *running records*) verbreitet, bei denen Testpersonen Wörter oder Texte lesen, während nach einem zuvor festgelegten Notationssystem Lesefehler, Auslassungen und Ersetzungen vermerkt werden. Die Dekodiergenauigkeit wird als prozentualer Anteil fehlerfreier Wörter an der Anzahl insgesamt gelesener Wörter errechnet (Nix 2011; Rieckmann 2010). Um das Textverständnis nicht zu beeinträchtigen, sollte ein kritischer Wert nicht unterschritten werden. Nach Rasinski (2003), der für das englischsprachige System eine Taxonomie entwickelt hat, kann ein Text erst bei mehr als 95 Prozent fehlerloser Wörter auf dem *independent reading level* gut und ohne Hilfe verstanden werden. Zwischen 90 und 95 Prozent Dekodiergenauigkeit wird ein Text auf dem *instructional reading level* mit unterstützender Hilfe weitgehend verstanden, bei unter 90 Prozent bleibt er jedoch auf dem *frustrational reading level* und damit unverständlich.

Weiterer Bestandteil von Leseflüssigkeit ist ein automatisierter Leseprozess. Schwächere Leserinnen und Leser können zwar häufig überwiegend akkurat dekodieren, erfassen dabei aber die Bedeutung des Gelesenen nicht, da sie noch langsam, mühevoll und stockend, also

wenig automatisiert lesen (Rosebrock & Nix 2006). Nach Samuels und Farstrup (2006) haben flüssige Leserinnen und Leser einen ausgeprägten Sichtwortschatz und erkennen Wörter als ganze Einheiten. In der Entwicklung des automatisierten Lesens wird zunächst der direkte, lexikalische Zugangsweg zum Wortschatz im Sinne der Zwei-Wege-Theorie optimiert. Bei einem weniger automatisierten Prozess werden Wörter in stärkerem Maß über den indirekten, nicht-lexikalischen Weg über Graphem-Phonem-Zuordnungen erlesen. Dies zeigt sich durch ein Lesen in Wort- oder Zweiwortschritten, bei dem einzelne Wörter zwar richtig erlesen, aber kaum in Beziehung zueinander gesetzt werden. Durch die Ausbildung von Automatisierungsfähigkeiten wird der Fokus der erfassten Einheiten immer größer (Klicpera & Gasteiger-Klicpera 1995).

Die Automatisierung kann als weitere Variable im Rahmen von Lautleseprotokollen über die Anzahl mühsam erlesener Wörter ermittelt werden. Nix (2011) und Rosebrock et al. (2017) schlagen dazu vor, solche Wörter anhand von stockendem, zögerlichem, schrittweisem Lesen, unangemessen langen Pausen und wiederholtem Ansetzen während des Lautleseprozesses zu identifizieren. Rosebrock et al. (2017) setzen dabei im Rahmen ihrer Forschungsarbeiten die Grenze zu mangelnder Automatisierung bei 5 Prozent in der beschriebenen Weise mühevoll erlesenen Wörtern, da es bislang keinen festgesetzten kritischen Wert gibt.

Die Folge einer genauen Worterfassung und eines hohen Grades an Automatisierung ist eine höhere Lesegeschwindigkeit (Landerl & Wimmer 2008). Das Lesetempo konstatiert wahrscheinlich den am besten wahrnehmbaren Unterschied zwischen flüssig und weniger flüssig Lesenden. Zu langsames Lesen verzögert zum einen die Informationsaufnahme und behindert zum anderen die Effizienz von Selbstüberwachungsprozessen beim Lesen (Nix 2011). Um die Aufmerksamkeitsspanne über einen Satz oder Text halten zu können, muss nicht unbedingt möglichst schnell, aber doch in einer Mindestgeschwindigkeit gelesen werden (Holle 2006). Neben der Dekodiergenauigkeit und dem Automatisierungsgrad ist die Lesegeschwindigkeit von der Komplexität und dem Bekanntheitsgrad des Textes sowie darüber hinaus vom Lesezweck, dem Kontext und dem Textformat abhängig (Holle 2006; Nix 2011; Rosebrock et al. 2017).

Üblicherweise wird die Lesegeschwindigkeit als gelesene Wörter pro Minute (WpM) gemessen (Nix 2011). Bisher ermittelte Richtwerte fallen sehr heterogen aus, da sie von den genannten externen Faktoren sowie der untersuchten Stichprobe abhängig sind.

Dekodiergenauigkeit und automatisiertes, schnelles Lesen werden in der Forschung als „notwendige, aber nicht hinreichende Bedingung von Leseflüssigkeit“ (Nix 2011, S. 94) betrachtet. Die vierte Dimension der Leseflüssigkeit wird als *Prosodie, phrasiertes Lesen* oder *Segmentierungsfähigkeit und Betonung* bezeichnet und meint im Allgemeinen die Fähigkeit zum betonten und sinngestaltenden Vorlesen (Rosebrock et al. 2017). Prosodisches Lesen ist wichtig, um zusammengehörende Sinnabschnitte zu bündeln, die dem Leseverstehen dienen (Rasinski 2004), und gilt daher als „one of the hallmarks of the achievement of reading fluency“ (Schwanenflugel et al. 2006, S. 119). Kuhn, Schwanenflugel und Meisinger (2010) bezeichnen Prosodie auch als „music of language“ (S. 234), da sie Rhythmus, Betonung und Intonation vereint. Prosodie umfasst das stimmliche Modulieren zusammenhängender Abschnitte, Betonungsdehnungen und Pausen und erleichtert das Entnehmen inhaltlicher Bedeutungen (Rosebrock et al. 2017). Nach Schwanenflugel et al. (2006) sind gute Dekodierfähigkeiten auf Wortebene die Voraussetzung für prosodisches Lesen und die Fähigkeit zum prosodischen

Lesen steigt mit zunehmender Lesegeschwindigkeit. Holle (2009) unterscheidet zwischen dem Aspekt der Phrasierung und dem Aspekt der Intonation des Gelesenen. Das phrasierte Lesen betrifft die sinnvolle Gliederung nach syntaktischen und semantischen Sinnunterteilungen durch Pausen, Stimmführung und die Beachtung von Satzschlusszeichen. Die Intonation beinhaltet den sprecherischen Ausdruck und bezieht sich auf inhaltliche Aspekte des Textes.

Bislang gibt es wenig formelle Verfahren, die prosodische (Vorlese-)Fähigkeiten erfassen, da dies zeitlich und inhaltlich aufwendig ist (Scheerer-Neumann 2015). Im Gegensatz zu Dekodierfähigkeiten und Lesegeschwindigkeit unterliegt die Bewertung der Prosodie stärker dem subjektiven Höreindruck. Ein von deutschen Autorinnen und Autoren adaptiertes Verfahren (vgl. Nix 2011; Rieckmann 2010) ist die Bewertung von prosodischen Fähigkeiten nach der „Oral Fluency Scale“ (Pinnell et al. 1995), der sogenannten *Pinnell-Skala*. Danach werden prosodische Aspekte auf einer vierstufigen Skala bewertet. Diese und ihr ähnliche Skalen werden in der Literatur jedoch wegen mangelnder Präzision und Reliabilität bei hohem Aufwand kritisiert (Kuhn, Schwanenflugel & Meisinger 2010; Sappok & Fay 2018; Scheerer-Neumann 2015).

Um das Konstrukt Leseflüssigkeit aus Entwicklungsperspektive zu erforschen, ist es wünschenswert, alle vier dargestellten Dimensionen einzubeziehen. Betrachtet man den derzeitigen Forschungsstand, wird schnell deutlich, dass es sowohl international als auch im deutschsprachigen Raum zum einen an Längsschnittstudien und zum anderen an der Erfassung aller vier Dimensionen von Leseflüssigkeit mangelt.

Es lässt sich feststellen, dass die meisten Studien aus dem angloamerikanischen Raum nicht Leseflüssigkeit als solche messen, sondern hauptsächlich die Entwicklung von Dekodiergenauigkeit darstellen (Landerl & Wimmer 2008). Dieser Fokus erklärt sich dadurch, dass der Schriftspracherwerb in der englischen Sprache vor allem in den ersten Phasen aufgrund der unsystematischeren Phonem-Graphem-Korrespondenzen schwieriger ist und englischsprachige Leserinnen und Leser erst später fehlerfrei zu dekodieren lernen (Landerl 1996; Nix 2011). In transparenten Orthographien wie dem Deutschen werden bezüglich der Dekodiergenauigkeit jedoch oft schon im Laufe des ersten Schuljahrs von einem Großteil der Schülerinnen und Schüler so gute Leistungen am oberen Messrand erzielt, dass dieses Maß zur Unterscheidung von starken und schwachen Leseleistungen nicht mehr aussagekräftig ist. Die meisten Untersuchungen wählen daher die Lesegeschwindigkeit oft als einziges Unterscheidungskriterium aus (Cossu, Giuliotta & Marshall 1995; Oney & Goldman 1984). Landerl und Wimmer (2008) untersuchten in einer Längsschnittstudie die Wortleseflüssigkeit¹ deutschsprachiger Kinder in Klasse 1, 4 und 8 und stellten fest, dass Werte in der Dekodiergenauigkeit immer Deckeneffekte erzielten und Unterschiede in der Leseflüssigkeit kaum beeinflussten. Zu ähnlichen Ergebnissen kamen Untersuchungen von Klicpera und Schabmann (1993), in denen ein schnell erreichtes Plateau in der Dekodiergenauigkeit bei hoher Stabilität der Leistungen deutschsprachiger Kinder bestätigt wurde. Vor dem Hintergrund des Mangels an Längsschnittstudien im deutschen Sprachraum untersuchten auch Klicpera und Gasteiger-Klicpera (1993) in ihren *Wiener Längsschnittuntersuchungen* die Entwicklung, den Verlauf und die Ursachen von Lese- und Schreibschwierigkeiten vom Ende

¹ Die *Wortleseflüssigkeit* wird bei Landerl und Wimmer (2008) als Anzahl der Lesefehler gemessen, daher wird im weiteren Verlauf der in dieser Arbeit genutzte Begriff *Dekodiergenauigkeit* verwendet.

der ersten bis zum Ende der achten Klasse in den Variablen Lesesicherheit² und Lesegeschwindigkeit. Bezüglich der Dekodiergenauigkeit ließ sich am Ende der dritten Klasse bei nicht zu schwierigen Texten ein Plateau feststellen, da die stärkeren Leserinnen und Leser diese fast fehlerfrei lasen. In der Lesegeschwindigkeit erfolgte eine gleichmäßigere Zunahme und im Unterschied zur Dekodiergenauigkeit wurden in den höheren Klassen noch erhebliche Fortschritte erzielt. Durchschnittliche Werte für die Dekodiergenauigkeit lagen in der zweiten Klasse bei 96,2 Prozent, in der dritten Klasse bei 97 Prozent und in der vierten Klasse bei 98 Prozent. Die Durchschnittswerte in der Lesegeschwindigkeit stiegen von 80 WpM in der zweiten Klasse auf 110 WpM in der dritten und auf 130 WpM in der vierten Klasse (Rosebrock et al. 2017)³.

In einer Folgestudie untersuchten Klicpera, Schabmann und Gasteiger-Klicpera (2006) den Entwicklungsverlauf von Teilleistungsschwierigkeiten beim Lesen und Schreiben. Die Ergebnisse zeigten, dass Kinder, die in der ersten Klasse Schwierigkeiten im Dekodieren der Wörter hatten, auch am Ende der Grundschulzeit einen deutlichen Rückstand aufwiesen. In der Lesegeschwindigkeit zeigte sich auch hier eine insgesamt kontinuierlichere Zunahme als bei der Dekodiergenauigkeit. Die schwächsten Leserinnen und Leser befanden sich im Durchschnitt in der vierten Klasse etwa auf dem Leistungsstand durchschnittlicher Schülerinnen und Schüler in der zweiten Klasse.

In der angloamerikanischen Forschung haben Hasbrouk und Tindal (2006) Vergleichswerte für die durchschnittliche Lesegeschwindigkeit in der Grundschule (*elementary school*) aufgestellt. In ihren Untersuchungen wurden anhand von Lautleseprotokollen jedoch nur die richtigen Wörter pro Minute (RWpM) gemessen, sodass ein direkter Vergleich mit den Ergebnissen der Wiener Längsschnittstudien nur eingeschränkt möglich ist. Durchschnittswerte in der zweiten Klasse lagen bei 89 RWpM, in der dritten Klasse bei 107 RWpM und in der vierten Klasse bei 123 RWpM.

Die dargestellten Befunde zeichnen ein recht heterogenes Bild im Hinblick auf die Vergleichbarkeit und Interpretation der Ergebnisse. Erkenntnisse zur Entwicklungsperspektive von Leseflüssigkeit liegen ausschließlich in den Dimensionen Dekodiergenauigkeit und Lesegeschwindigkeit vor. Bezüglich der Automatisierung und Prosodie fehlt es an grundlegenden empirischen Studien. Für eine differenzierte Diagnose von Leseschwierigkeiten als Voraussetzung für den Einsatz angemessener Leseförderverfahren mangelt es neben empirisch überprüften Diagnose- und Fördermethoden besonders auch an Längsschnittuntersuchungen zur Entwicklung der Leseflüssigkeit in den vier Dimensionen. Dadurch fehlen zuverlässige Richtwerte, die eine Einschätzung der zeitlichen Entwicklung von Leseflüssigkeit und die Darstellung von Erwerbsverläufen ermöglichen (Rosebrock et al. 2017). Folglich fehlt es auch an Befunden darüber, welche Teilfähigkeiten der Leseflüssigkeit bei welchen Kindern (nicht) ausgeprägt sind und wie sich flüssig und nichtflüssig Lesende in ihrer Entwicklung unterscheiden. Ziel der zugrundeliegenden Längsschnittstudie ist es daher, zu ermitteln, in welchem Alter Kinder wie genau, in welchem Maße automatisiert, wie schnell und

² Auch die *Lesesicherheit* wird bei Gasteiger-Klicpera und Klicpera (1993) als Anzahl der Lesefehler gemessen und wird im weiteren Verlauf als *Dekodiergenauigkeit* bezeichnet. Die Variable wurde anhand von Wortlisten und Texten gemessen. Vorliegend werden für einen besseren Vergleich der Ergebnisse nur die Werte der Textlesesicherheit dargestellt.

³ Diese Werte wurden aus Rosebrock et al. (2017, S. 57) entnommen, da die Werte aus den Ergebnissen bei Gasteiger-Klicpera und Klicpera (1993, S. 51) nicht eindeutig ablesbar sind.

wie angemessen intoniert lesen, wie die einzelnen Dimensionen von Leseflüssigkeit voneinander abhängen und wie stabil die Entwicklung im Verlauf der Grundschulzeit ist.

2 | Fragestellung

Der vorliegende Bericht stellt ein Zwischenfazit für die Klassenstufen 2 und 3 dar, die Teil einer bis zur 6. Klassenstufe geplanten Längsschnittstudie sind. Dabei sollen die folgenden Fragestellungen im Fokus stehen:

- (1) Wie entwickeln sich Dekodiergenauigkeit, Automatisierungsgrad, Lesegeschwindigkeit und Prosodie von der zweiten zur dritten Klassenstufe?
- (2) Welche Zusammenhänge und Interaktionen zwischen den vier Dimensionen der Leseflüssigkeit gibt es?
- (3) Gibt es typische Muster der Entwicklung der Leseflüssigkeit?
- (4) Wie stabil ist die Entwicklung auf individueller Ebene?

3 | Methode

3.1 | Versuchspersonen

Die Datenerhebungen erfolgten seit 2016 bzw. 2017 jährlich zum Schuljahresende im Rahmen einer Längsschnittuntersuchung im Bereich Grundschulpädagogik Deutsch der Universität Potsdam. Die prospektive Studie sieht eine Auswertung von Daten aus fünf Messwiederholungen der Klassenstufen 2 bis 6 vor. Die Daten werden in 12 Klassen⁴ in sieben Brandenburger Grundschulen mit einer Gesamtstichprobe von $N = 196$ Schülerinnen und Schülern erhoben (88 männlich, 108 weiblich). Von 142 Kindern (61 männlich, 81 weiblich) lagen zu beiden Testzeitpunkten vollständige Daten vor.⁵ Sie wurden zunächst im Klassenverband mit dem *Salzburger Lese-Screening für die Klassenstufen 1-4* (SLS 1-4) getestet. Im Anschluss daran folgten das individuelle Vorlesen von Texten und die Erhebung von Tonaufnahmen, welche die Grundlage für die Analyse des Lautlesens in Form von Lautleseprotokollen bildeten.

3.2 | Material

Zu allen Messzeitpunkten wurden für die Analyse des Lautlesens je zwei Lautlesetexte eingesetzt, von denen ein Text als altersangemessen und ein Text als etwas leichter eingestuft wurde. Der altersangemessene Text wurde ein Jahr später wiederholt eingesetzt und diente dann als leichterer Text. Ein neuer, altersangemessener Text kam hinzu.

⁴ Zum Teil stammen die Daten aus Flex-Klassen (jahrgangsübergreifend). Es nahmen nur die Schülerinnen und Schüler einer Klasse teil, deren Eltern eine Einverständniserklärung abgegeben hatten. Die Kinder wurden darauf hingewiesen, dass sie ihre Teilnahme verweigern oder die Testung abbrechen können.

⁵ Zu fehlenden Daten kam es wegen Nichtanwesenheit (v.a. durch Krankheit) an den Testtagen oder Schulwechsels. Nicht berücksichtigt wurden auch Lesungen, bei denen eine oder mehrere Textzeilen ausgelassen wurden oder bei denen es zu Störungen kam.

Die Auswahl der Texte richtete sich nach der Frage, was ein altersangemessener Text ist. Das Niveau von Leseflüssigkeit wird unter anderem von Textlänge, Schwierigkeitsgrad, Organisationsweise oder Explizitheit der Informationsdarbietung beeinflusst (Nix 2011). Somit ist bei der Bestimmung von Leseflüssigkeit die Text-Leser-Passung besonders relevant (Rosebrock et al. 2017). Rosebrock et al. (2017) kritisieren, dass im Unterrichtskontext oft nur „gefühl“ (S. 54) werden kann, was ein altersgemäßer Text ist, da es kaum geeignete Kriterien zur Bewertung und Adaption von Texten gibt. In der textwissenschaftlichen Forschung werden die von Groeben (1982) explizierten vier Kategorien für Textverständlichkeit *sprachliche Einfachheit*, *semantische Kürze*, *kognitive Gliederung* und *motivationale Stimulanz* unterschieden. Textverständlichkeit verbindet nach Groeben (1982) Textmerkmale und den Rezeptionsprozess der Leserin oder des Lesers. Auch Rieckmann (2018) betont, dass Textverständlichkeit im Sinne einer Text-Leser-Passung immer leserbezogene und textbezogene Anforderungen beinhaltet. Die sprachliche Einfachheit meint die Lesbarkeit eines Textes auf Wort- und Satzebene, umgesetzt durch kurze, frequente Wörter und Sätze. Die semantische Kürze bezeichnet die Informationsdichte in Bezug auf das Herstellen lokaler Kohärenz. Kognitive Gliederung beinhaltet Aspekte der Strukturierung und Organisation und soll helfen, globale Kohärenz herzustellen (Rieckmann 2018).

Text aus:	Stufe	Schwierigkeit	Wörter	LIX	Satzkomplexität
Maar: Das kleine Känguru auf Abenteuer	2	leicht	124	19	1
Boie: King-Kong, das Geheimschwein	2 3	altersgemäß leicht	148	24	4
Michaelis: Hier bei uns in Ammerlo	3	altersgemäß	219	27	6

Tab. 1: Länge, LIX und Satzkomplexität (Anzahl der Abweichungen vom Subjekt-Verb-Objekt-Satzschema) der verwendeten Texte.

Grundlage für die Lautlesetexte sind kurze, inhaltlich abgeschlossene Episoden aus verschiedenen Kinderbüchern, deren Schwierigkeit durch Bearbeitung angepasst wurde. Um neben der eher neutralen Erzählerrede auch Möglichkeiten für stärker expressive Betonung zu geben, wurden Texte mit einem hohen Dialoganteil ausgewählt. Die Lesbarkeit der Texte und damit die nach Groeben (1982) klassifizierte *sprachliche Einfachheit* wurde mithilfe des häufig verwendeten Lesbarkeitsindex (LIX) nach Bamberger und Vanecek (1984) an die Klassenstufe angepasst. Dabei wurden Durchschnittswerte für die Satzlänge und den Prozentsatz langer Wörter berechnet. Die Texte hinsichtlich des Aspekts der Textschwierigkeit (*semantische Kürze*, Groeben 1982) vergleichbar zu machen, erwies sich mangels geeigneter Verfahren zur Analyse von Satzkomplexität und vor allem fehlender Werte für einzelne Klassenstufen als schwierig. Für die Auswahl der Texte in dieser Untersuchung wurde daher ein eigenes System genutzt, bei dem Punktwerte für die Komplexität verschiedener Satzstrukturen (Anzahl der Abweichungen vom Subjekt-Verb-Objekt-Satzschema) vergeben wurden. Der Punktwert sollte – ähnlich wie die LIX-Werte – mit fortschreitendem Alter ansteigen. Somit wurde von Klasse zu Klasse der Umfang der Texte sowie deren Textschwierigkeit durch Satzlänge, Anteil langer Wörter und Satzkomplexität gesteigert. Des Weiteren wurde versucht, der *motivationalen Stimulanz* als

Aspekt von Textverständlichkeit gerecht zu werden, indem Texte aus der Lebenswelt von Grundschulkindern gewählt wurden, die möglichst spannend oder lustig sind.

3.3 | Datenauswertung

Die Auswertung erfolgte anhand der zu allen Testzeitpunkten mittels Audioaufnahmen erhobenen Lautleseprotokolle. Hieran wurden alle vier Dimensionen der Leseflüssigkeit angelegt, wie im Folgenden näher dargestellt wird.

Die **Dekodiergenauigkeit** wurde mit der in der Leseforschung überwiegend verwendeten Methode der Ermittlung von Lesefehlern im Verhältnis zur Anzahl gelesener Wörter erfasst (vgl. Nix 2011; Rieckmann 2010; Rosebrock et al. 2017). Als Verlesungen galten Aussprachefehler, Wortumstellungen und -ersetzungen sowie Auslassungen. Nur einmal gewertet wurden Wiederholungsfehler. Nicht als Fehler galten Verlesungen, die das Kind selbst korrigierte, sowie hinzugefügte Wörter, die in den Kontext passten. Als Grenzwert zu flüssiger Lektüre wurde das bei 95 Prozent angesetzte Unabhängigkeitslevel aus den Richtwerten zur Dekodiergenauigkeit von Rasinski (2003) übernommen.

Die Messung der **Automatisierung** wurde gemäß der Vorschläge von Rosebrock et al. (2017) und Nix (2011) vorgenommen, indem der Anteil an erlesenen Wörtern im Verhältnis zur Anzahl gelesener Wörter erfasst wurde. Ein Wort galt als *erlesen*, wenn es nicht auf Anheb erkannt wurde. Als Anzeichen dafür wurden wiederholtes Ansetzen, stockendes oder schrittweises Erlesen sowie das Wiederholen vorhergehender Wörter angesehen; letzteres in der Annahme, dass dadurch Zeit für das Dekodieren des folgenden Wortes gewonnen wird. Auch als unangemessen wahrnehmbare Pausen in oder vor Wörtern wurden als stumme Dekodierversuche und damit Automatisierungsdefizite gewertet (Nix 2011). Der Grenzwert wurde in Anlehnung an Rosebrock et al. (2017) bei 5 Prozent erlesenen Wörtern gesetzt; zur besseren Vergleichbarkeit mit dem positiv ausgedrückten Wert der Dekodiergenauigkeit haben wir auch die Automatisierung positiv gewendet und geben die Anzahl automatisiert gelesener Wörter an, wodurch der Grenzwert auch hier bei 95 Prozent liegt.

Für die Messung der **Lesegeschwindigkeit** wurde die in der Forschung verwendete Methode der Berechnung der gelesenen Wörter pro Minute verwendet (Nix 2011; Rieckmann 2010; Rosebrock et al. 2017). Dafür wurde die Anzahl der Wörter ins Verhältnis zur benötigten Lesezeit gesetzt. Aufgrund des Mangels an zuverlässigen Vergleichswerten wurden eigene Grenzwerte für die Lesegeschwindigkeit aufgestellt. Die niedrigste Geschwindigkeit, bei der auf allen anderen Ebenen flüssiges Lesen beobachtet werden konnte, galt als Wert, der für die jeweilige Klassenstufe mindestens zu erreichen war. Für die Klassenstufe 2 wurden auf diese Weise 80 WpM, für die Klassenstufe 3 90 WpM als Grenzwerte festgelegt.

Das **prosodische Lesen** wurde angelehnt an die Pinnell-Skala und die von Rieckmann (2010) und Nix (2011) adaptierten Verfahren bewertet. Die drei Kategorien *Größe und Anzahl der Wortgruppierungen*, *Fähigkeit zur Satzsegmentierung* und *expressive Intonation* wurden von Rieckmann (2010) übernommen. Über die Pausensetzung der Lesenden können *Größe und Anzahl der Wortgruppierungen* festgestellt werden. Rieckmann (2010) folgend wurde nach vier Levels unterschieden: Wort für Wort (Level 1), Zwei-Wort-Gruppen (Level 2), Drei- bis Vier-Wort-Gruppen (Level 3) und größere Wortgruppierungen (Level 4). In Erweiterung zu Rieckmann (2010) und zur differenzierteren Einschätzung der Leseleistung wurde bei der

Auszählung aller Wortgruppierungen ermittelt, welche Wortgruppengröße den *überwiegenden* Anteil (also > 50 Prozent aller Wortgruppen) ausmachte. Wird dieser Wert nicht erreicht, wird der Durchschnitt der beiden Wortgruppengrößen, die zusammen mindestens > 50 Prozent ergeben, ermittelt, so dass zur besseren Differenzierung des Leistungsstandes auch .5-Werte entstehen konnten. Dieser Level-Bildung liegt die Beobachtung zugrunde, dass leichte Texte der Kinderliteratur von geübten Leserinnen und Lesern zu über 75 Prozent in *größeren Wortgruppen* gelesen werden. Ein Grenzwert von > 50 Prozent für größere Wortgruppierungen ist also ein recht konservativer Wert, der jedoch weit über die bisherige Praxis, allein die zahlenmäßig häufigste Wortgruppengröße auszuzählen, hinausgeht und so zu einer verlässlicheren Einschätzung führt.

Die Fähigkeit zur Satzsegmentierung zeigt sich bei geübten Leserinnen und Lesern daran, dass durch die Pausengestaltung semantisch und syntaktisch passende Phrasen gebildet werden. Disfluent Lesende hingegen machen häufig unpassende Pausen in Satzphrasen (Schwanenflugel et al. 2006; Nix 2011), deshalb ist die Ermittlung der Fähigkeit zur Satzsegmentierung ein weiterer Schritt zur Unterscheidung von Graden der Ausprägung von Leseflüssigkeit mit dem Ziel des Textverständnisses. Dabei wird überprüft, wie viele Wortgruppierungen die Konstituentenstruktur des Textes verletzen. Hierzu werden Pausen in Konstituenten gezählt, die durch die Verschiebeprobe nicht getrennt werden können. Je nach Anzahl verletzter Konstituenten wurde auch hier nach vier Levels unterschieden: Bei hundertprozentiger Einhaltung der Konstituentenstruktur erfolgte eine Einordnung auf Level 4. Level 3 wurde bei 90 bis 99 Prozent richtiger Konstituenten, Level 2 bei 80 bis 89 Prozent richtiger Konstituenten und Level 1 bei unter 80 Prozent richtiger Konstituenten erreicht.

Als drittes wird die *expressive Intonation* eingeschätzt. Dabei dienen die Beachtung von Satzzeichen, das Absetzen wörtlicher Rede, der Umgang mit „Regieanweisungen“, der Einsatz verschiedener Lautstärken, Stimmlagen und Lesetempi sowie der Ausdruck von Emotionen als Kriterien. Die wegen mangelnder Präzision kritisierte vierstufige Pinnell-Skala (Kuhn, Schwanenflugel & Meisinger 2010; Scheerer-Neumann 2015; Sappok & Fay 2018) und ihre Adaption durch deutsche Autorinnen und Autoren (Rieckmann 2010) wurde mit Zwischenschritten bei 2,5 und 3,5 erweitert und ausdifferenziert. Hierbei wurden von uns Items, die nicht unmittelbar die expressive Intonation betreffen, ausgeschlossen, da sie bereits einzeln beurteilt wurden (Wortgruppengröße, Segmentierungsfähigkeit). Auf Level 1 unserer Skala ist gar keine angemessene Satzmelodie zu verzeichnen, auf Level 2 treten Ansätze einer angemessenen Satzmelodie bei gleichzeitig immer noch überwiegendem Überlesen von Satzzeichen auf. Level 2,5 ist beschrieben durch eine überwiegend angemessene Satzmelodie und nur noch gelegentliche Nichtbeachtung von Satzzeichen, Level 3 durch durchgehend angemessene Satzmelodie unter Beachtung aller Satzschlusszeichen. Da disfluente Leserinnen und Leser eher über Satzgrenzen hinweglesen, statt wie flüssig Lesende das Ende eines Satzes zu betonen und mit der Tonhöhe abzusinken (Clay & Imlach 1971; Cooper & Paccia-Cooper 1980), war es uns wichtig, den Grad der Beachtung von Satzgrenzen hier differenziert zu erheben und unkorrigiert bleibendes Überlesen durch Abwertungen im Level sichtbar zu machen. Die Unterschiede zwischen Level 3, 3,5 und 4 sind nur noch durch das Ausmaß expressiver Betonung gegeben. Während bei Level 3 Anzeichen expressiver Betonung zu verzeichnen sind, ist deren Anteil auf Level 3,5 ein höherer, dennoch bleiben Reserven. Für Level 4 musste der überwiegende Teil des Textes expressiv interpretiert werden.

Für einen Gesamtwert zur Einschätzung der Prosodie wurde zuletzt der Mittelwert aus den drei Aspekten *Wortgruppierungen*, Fähigkeit zur *Satzsegmentierung* und *expressive Intonation* gebildet. Im Rahmen des Forschungsprojekts wurde der Grenzwert zu flüssiger Lektüre bei 3,6 festgelegt. Diesem Grenzwert liegen folgende Annahmen zugrunde: Wenn Kinder überwiegend in 3-4-er-Wortgruppen lesen, kann das ein Anzeichen unflüssiger Lektüre sein, weil altersangemessene Texte größere Einheiten zu fast zwei Dritteln verlangen. Unkorrigiert bleibende Verstöße gegen die Konstituentenstruktur beeinträchtigen die Kohärenzbildung und sind, besonders wenn sie mehrmals auftreten, nicht mit flüssigem Lesen zu vereinbaren. Ebenso wird das Überlesen von Satzzeichen bzw. ein völliges Fehlen jedweder expressiven Intonation als unzureichend betrachtet. Eine Kombination von Reserven auf mehreren der drei Ebenen, ausgedrückt durch Level 3 oder schlechter, führt deshalb zur Diagnose eines weiteren Förderbedarfes der Prosodie. Liegt in nur einem Bereich Level 3, ansonsten Level 4 vor, wird der Grenzwert zur flüssigen Lektüre erreicht.

3.4 | Grenzwerte für flüssige Lektüre und Lesemuster

Ebene	Kürzel	Grenzwert
Dekodiergenauigkeit	D	≥ 95% fehlerfrei gelesene Wörter
Automatisierung	A	≥ 95% automatisiert gelesene Wörter
Lesegeschwindigkeit	L	≥ 80 WpM (2. Klasse)
		≥ 90 WpM (3. Klasse)
Prosodie		
Wortgruppierungen	P	≥ 3,6 (Mittelwert aus drei Aspekten der Prosodie)
Segmentierungsfähigkeit		
expressive Intonation		

Tab. 2: Verwendete Grenzwerte für die Auswertung der vier Dimensionen der Leseflüssigkeit.

Anhand der Lautleseprotokolle wurden in der Auswertung für jede Versuchsperson vier Werte für die einzelnen Subskalen ermittelt (siehe Tabelle 1). Anschließend wurde für alle Dimensionen die Kategorie + oder – hinsichtlich des Erreichens der Grenzwerte für flüssige Lektüre vergeben. Lag der erreichte Wert über dem Grenzwert, wurde ein + vergeben, sonst ein –. Damit fand für jedes Kind eine Einteilung in die Kategorien *Dekodiergenauigkeit* +/-, *Automatisierung* +/-, *Lesegeschwindigkeit* +/-, und *Prosodie* +/- statt. Daraus ergab sich für jede Versuchsperson ein Lesemuster (D A L P), das von ++++ für flüssige Leserinnen und Leser (Grenzwert erreicht in allen Dimensionen) bis zu ---- für nichtflüssige Leserinnen und Leser (auf keiner der Ebenen ausreichende Fähigkeiten) reichte. Durch die unterschiedlichen Werte der Schülerinnen und Schüler entstanden verschiedene mehr oder weniger typische Lesemuster (siehe Ergebnisse).

3.5 | Reliabilitätsprüfung durch Zweitratings

Die Diagnose der Leseflüssigkeit beinhaltet zum einen quantitative Bestimmungen wie z.B. der Lesegeschwindigkeit (WpM) und der Dekodiergenauigkeit (Anzahl der Lesefehler) und zum anderen die qualitative Auswertung des Automatisierungsgrades und der Prosodie. Schon die Einschätzung des Automatisierungsgrades unterliegt aber sowohl quantitativen als auch qualitativen Bestimmungen: Die Frage, ob die Pausenlänge vor einem Wort als unangemessen lang und damit als Zeichen mangelnder Automatisierung gewertet wird, ist abhängig vom Höreindruck des Raters. Da Ergebnisse aus solchen qualitativen Verfahren grundsätzlich auch subjektiv geprägt sind, erfordern sie, um die Kriterien der Auswertungsobjektivität und vor allem der Reliabilität zu erfüllen, die Berechnung einer Inter-Rater-Reliabilität, ermöglicht durch die Durchführung von Zweitratings (Rieckmann 2010; Scheerer-Neumann 2015). Vorliegend wurden die Bewertungen der Dimensionen Automatisierung und Prosodie auf Übereinstimmung durch Zweitratings überprüft. Die Dekodiergenauigkeit und – in noch stärkerem Maße – die Lesegeschwindigkeit sind weitgehend objektiv zu beurteilen und werden daher hier nicht überprüft.

Für die Reliabilitätsprüfung wird konventionell ein Ausschnitt des Materials der Haupterhebung in Höhe von 10 bis 20 Prozent verwendet (Inter-Kodierer-Übereinstimmung, Bortz & Döring 2016). Für die vorliegende Analyse wurden von 68 der 142 Kinder die jeweils altersgemäßen Texte in den Klassenstufen 2 und 3 durch eine zweite Person ausgewertet. Als Reliabilitätsmaß wurde der Kendall'sche Rangkorrelationskoeffizient berechnet (Kendalls Tau), der für die Analyse von Daten auf mindestens Ordinalskalen-Niveau geeignet ist und in der verbreiteten Variante b auch mit gleichen Rängen (ties) zurechtkommt (vgl. McLeod 2011, S. 4f.). Theoretisch hätte für den Automatisierungsgrad auch ein Verfahren gewählt werden können, das Intervallskalen-Niveau voraussetzt, zur besseren Vergleichbarkeit mit den Prosodie-Bewertungen, für die dieses Niveau nicht vorliegt, wurde aber Kendalls Tau gewählt. Für den Automatisierungsgrad wurde $\tau = .74$ berechnet ($p < .0001$), was einer guten Übereinstimmung entspricht. Auch der Mittelwert für die Beurteilungen der drei Aspekte der Prosodie bei beiden Ratern korreliert hoch mit $\tau = .77$, wobei die Übereinstimmung für die Wortgruppierungen ($\tau = .6$) und die Expressivität ($\tau = .68$) allein geringer ausfallen als die für die Segmentierungsfähigkeiten ($\tau = .8$).

4 | Ergebnisse

Zur Betrachtung der Entwicklung der Leseflüssigkeit jeweils isoliert in den Dimensionen Dekodiergenauigkeit, Automatisierungsgrad, Lesegeschwindigkeit und Prosodie wurden die Daten von 142 Kindern ausgewertet, für die Daten zu allen Texten und Messzeitpunkten vorlagen. Für die statistische Analyse wurde das Paket lme4 (Bates et al. 2015) für R verwendet. Hiermit haben wir gemischte lineare Modelle zum Einfluss der Klassenstufe und der Textschwierigkeit auf die vier Dimensionen der Leseflüssigkeit berechnet. In den finalen Modellen wurden jeweils Klassenstufe und Textschwierigkeit als fixe Faktoren und die Versuchspersonen als Zufallsfaktor eingesetzt. Modelle mit in Klassen und/oder Schulen geschachtelten Versuchspersonen als Zufallsfaktor führten nicht zu signifikant besseren Ergebnissen (*likelihood ratio tests*). Der Einfluss des jeweiligen Textes ist über die Faktoren Stufe und Schwierigkeit bereits in den Modellen enthalten. Als Signifikanzkriterium wird ein t-Wert (absolut) von > 2 verwendet.

4.1 | Entwicklung der Dekodiergenauigkeit

Betrachtet man die Dekodiergenauigkeit isoliert, zeigt sich ein signifikanter Unterschied sowohl zwischen den Klassenstufen ($t = -4,9$) und den Schwierigkeitsgraden ($t = -5,2$) als auch eine signifikante Interaktion zwischen Klassenstufe und Schwierigkeit ($t = 3,4$). Aus der Darstellung der vom Modell geschätzten Werte (siehe Abb. 1 oben links) wird ersichtlich, dass bei der Dekodiergenauigkeit bereits ein Deckeneffekt vorliegt: Bis auf wenige Ausnahmen haben alle Kinder den Grenzwert von 95 Prozent korrekt dekodierter Wörter überschritten. Der im Kontrast gefundene Unterschied zwischen dem leichten und dem altersgemäßen Text in Klassenstufe 2 ist vermutlich der besonderen Einfachheit des leichten Lesetextes geschuldet. Für den zweiten Text, der in der zweiten Klassenstufe als altersgemäß und in der dritten Klassenstufe als leichter Text gelesen wurde, gibt es keinen signifikanten Unterschied mehr, aber auch innerhalb der dritten Schulstufe gibt es fast gar keine Unterschiede zwischen dem leichten und dem altersgemäßen Text.

4.2 | Entwicklung des Automatisierungsgrades

Nach dem gerechneten Modell ergibt sich von Klassenstufe 2 zu Klassenstufe 3 insgesamt ein Zuwachs von 1,2 Wörtern pro 100 Wörter, die automatisiert gelesen werden (siehe Abb. 1 oben rechts). Dieser Zuwachs ist signifikant ($t = 4,0$). Auch der Unterschied zwischen den jeweils leichten und altersgemäßen Texten ist signifikant ($t = -6,8$). In den schwereren Texten werden knapp zwei Wörter weniger automatisiert gelesen (pro 100 Wörter). Die Interaktion ist nicht signifikant ($t < 0,5$), der Einfluss der Textschwierigkeit ist in beiden Klassenstufen ähnlich.

4.3 | Entwicklung der Lesegeschwindigkeit

Die Lesegeschwindigkeit entwickelt sich von der zweiten zur dritten Klassenstufe ausgesprochen stark (siehe Abb. 1 unten links). So nimmt die Anzahl der gelesenen Wörter pro Minute nach dem Modell um 19,4 Wörter zu ($t = 17,4$). Zwischen den Texten verschiedener Schwierigkeitsgrade ergibt sich über die Klassenstufen hinweg ein Unterschied von 7,6 Wörtern pro Minute ($t = -6,7$), wobei der Geschwindigkeits-Unterschied zwischen dem leichten und dem altersgemäßen Text in der dritten Klassenstufe deutlich größer ausfällt (16,3 WpM) als in der zweiten Klassenstufe (6,3 WpM), was sich in der signifikanten Interaktion ausdrückt ($t = -5,7$).

4.4 | Entwicklung der Prosodie

Die Bewertung der Prosodie (siehe Abb. 1 unten rechts) für die Kinder der dritten Klassenstufe fällt signifikant besser aus als diejenige für die Kinder der zweiten Klassenstufe ($t = 6,0$). Der Unterschied liegt bei 0,19 Punkten. Die Lesungen der jeweils leichten Texte werden ebenfalls signifikant besser bewertet (um 0,2 Punkte) als die jeweils altersgemäßen Texte ($t = 6,1$). Der Unterschied zwischen dem leichten und dem altersgemäßen Text fällt in der dritten Klassenstufe zwar etwas größer aus, ist aber nicht signifikant ($t = 1,7$). Beim leichten Text in der Klassenstufe 3 stellen sich schon bei einigen Kindern Deckeneffekte ein, wenn auch immer noch über die Hälfte der Drittklässler den festgelegten Grenzwert von 3,6 nicht erreichen kann. In der zweiten Klassenstufe sowie in der dritten beim altersgemäßen Text erreichen weniger als $\frac{1}{4}$ der Kinder den Grenzwert.

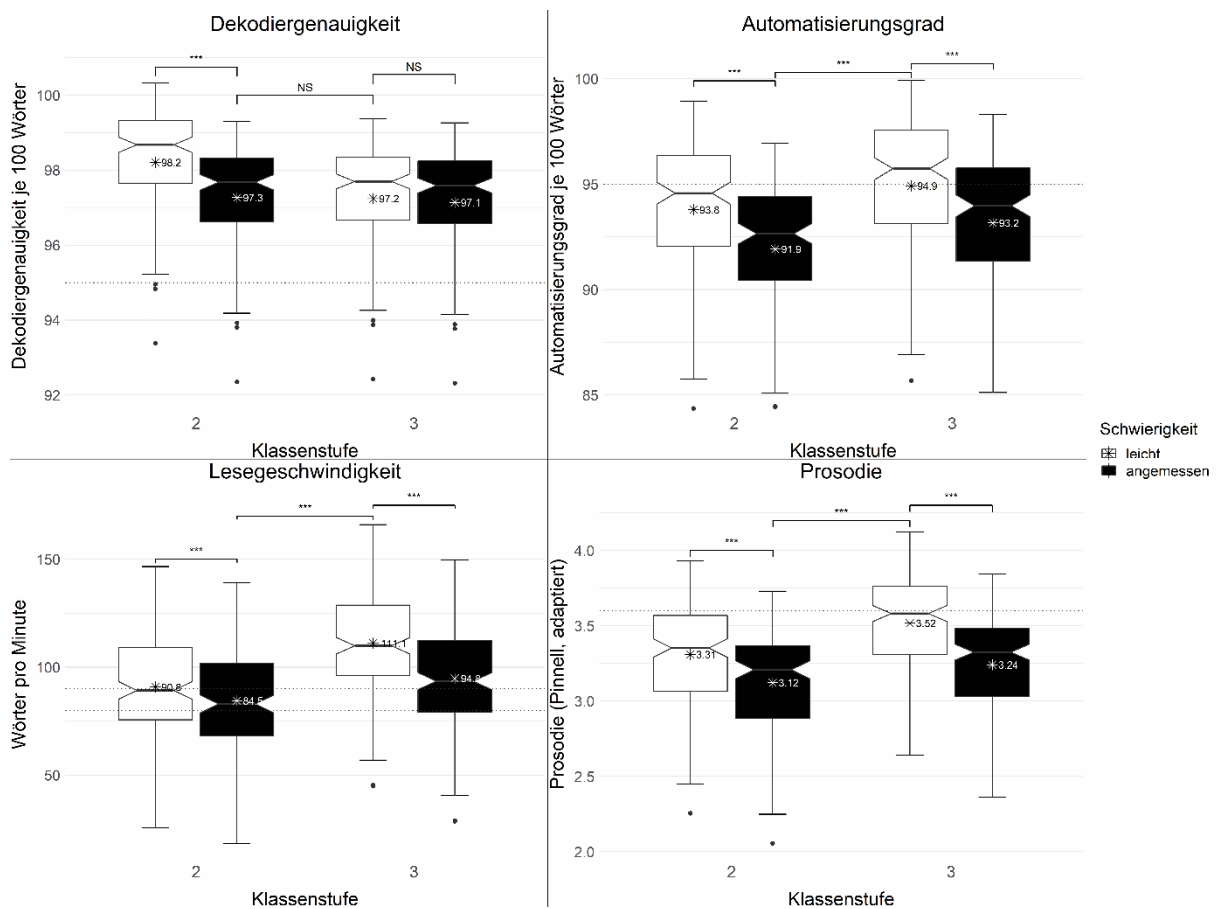


Abb. 1: Vom jeweiligen Modell vorhergesagte Werte für die Entwicklung der Dekodiergenauigkeit, des Automatisierungsgrades, der Lesegeschwindigkeit und der Prosodie von der zweiten zur dritten Klassenstufe, jeweils für einen leichten und einen altersgemäßen Text (der altersgemäße Text aus Klassenstufe 2 entspricht dem ein Jahr später gelesenen leichten Text). Die Boxen zeigen jeweils den Median und die mittleren 50 Prozent der Daten (zwischen dem unteren und dem oberen Quartil); der Mittelwert ist zusätzlich als * eingezeichnet. Die Antennen zeigen Werte mit einem Abstand von maximal 1,5 IQR, Ausreißer werden als Punkte dargestellt.

4.5 | Interaktionen zwischen den Dimensionen

Aufgrund der gefundenen Deckeneffekte bei der Dekodiergenauigkeit erscheint eine Betrachtung der Korrelation mit den anderen drei Dimensionen wenig zielführend. Von den verbleibenden drei Dimensions-Paaren wiederum sind vor allem diejenigen mit der Lesegeschwindigkeit interessant, nicht zuletzt da diese als einzige normalverteilt ist. Über alle mit dem gemischten linearen Modell vorhergesagten Daten (alle Texte in beiden Klassenstufen) beträgt der Korrelationskoeffizient für Automatisierungsgrad * Lesegeschwindigkeit $R = .85$ ($df = 693$, $p < .0001$, d.h. etwa 72 Prozent der Varianz können erklärt werden). In Abb. 2 ist getrennt nach Textschwierigkeit und Klassenstufe anhand der jeweiligen Streuung zu erkennen, dass in beiden Klassenstufen bei den leichten Texten der Grenzwert für die Automatisierung erst ab dem Grenzwert der Lesegeschwindigkeit erreicht wird. Mit anderen Worten: Fast alle Zweitklässler, die über 95 Prozent der Wörter automatisiert lesen, zeigen eine Lesegeschwindigkeit über dem Grenzwert von 80 Wörtern pro Minute. Mit wenigen Ausnahmen lesen Drittklässler, die über 95 Prozent der Wörter automatisiert lesen, mit über 90 Wörtern pro Minute. Beim altersgemäßen bzw. relativ schweren Text hingegen erreichen

überhaupt nur wenige Kinder der zweiten Klassenstufe den Grenzwert von 95 Prozent automatisiert gelesener Wörter – und dies sind im Wesentlichen die Kinder, die schon in der zweiten Klasse einen relativ schweren Text mit über 100 Wörtern pro Minute lesen. Oberhalb der Grenzwerte wird (grob geschätzt) pro Steigerung der WpM um 10 Wörter der Automatisierungsgrad um 0,75 Prozentpunkte erhöht; unterhalb der Grenzwerte ist diese Erhöhung etwa doppelt so groß.

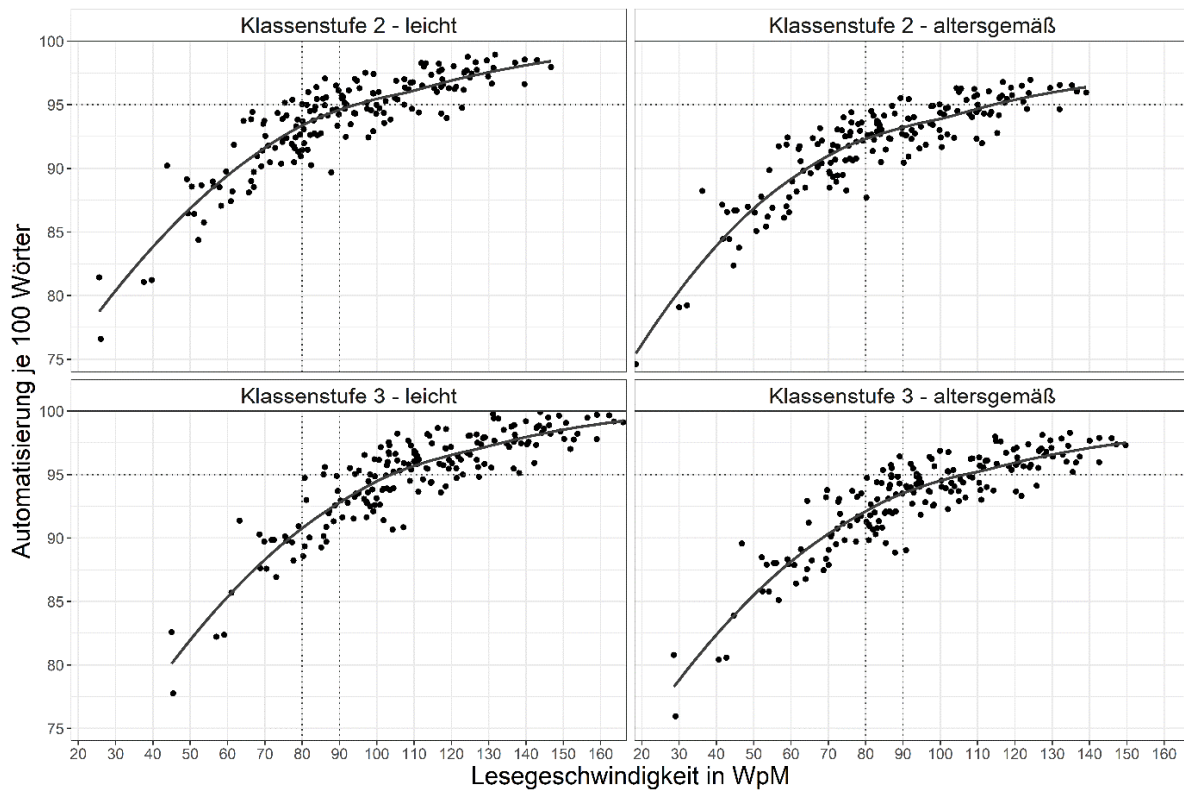


Abb. 2: Streudiagramme für Automatisierungsgrad * Lesegeschwindigkeit getrennt nach Textschwierigkeit und Klassenstufe (aus dem GLM vorhergesagte Werte). Zusätzlich eingezeichnet ist die Kurve der lokal gewichteten Streudiagrammglättung (LOWESS). Die gepunkteten Linien zeigen horizontal den Grenzwert für den Automatisierungsgrad (95 Prozent) sowie vertikal für die Lesegeschwindigkeit nach Klassenstufen (2: 80 WpM; 3: 90 WpM).

Auch für den Zusammenhang zwischen dem Automatisierungsgrad und der Bewertung der Prosodie findet sich ein signifikanter Korrelationskoeffizient ($R = .69$, $df = 693$, $p < .0001$)⁶.

Die Streudiagramme für Prosodie * Lesegeschwindigkeit in Abb. 3 zeigen einen jeweils sehr ähnlichen, aber versetzten Verlauf, wobei die Verteilungen für den leichten Text in Klassenstufe 2 und den altersgemäßen Text in Klassenstufe 3 sehr ähnlich ausfallen. Insgesamt zeigt sich ein Korrelationskoeffizient von $R = .79$ ($df = 693$, $p < .0001$), d.h. ca. 62 Prozent der Varianz können

⁶ Auf eine Darstellung als Streudiagramm wird an dieser Stelle verzichtet, da die Streuung der Prosodie-Bewertungen unterhalb von 95 Prozent Automatisierung bei den leichten und unter ca. 92 Prozent bei den altersgemäßen Texten recht groß wird.

erklärt werden. Ähnlich wie bei der Automatisierung lesen Zweitklässler, die bezüglich der Prosodie eine Bewertung über dem Grenzwert von 3,6 erhielten, einen leichten Text mit einer Geschwindigkeit von über 80 Wörtern pro Minute. Bei den schwereren altersgemäßen Texten erreichten nur sehr wenige Kinder beider Klassenstufen den Prosodie-Grenzwert, von denen fast alle mit über 100 WpM lasen. Beim für die Drittklässler leichten Text lagen fast alle Kinder, die eine Prosodie-Bewertung über 3,6 erreichten, auch über dem Grenzwert der Lesegeschwindigkeit von 90 WpM. Andererseits ist das Erreichen des Grenzwertes für die Lesegeschwindigkeit selbst bei den leichten Texten keine Garantie, dass auch prosodisch flüssig vorgelesen wird.

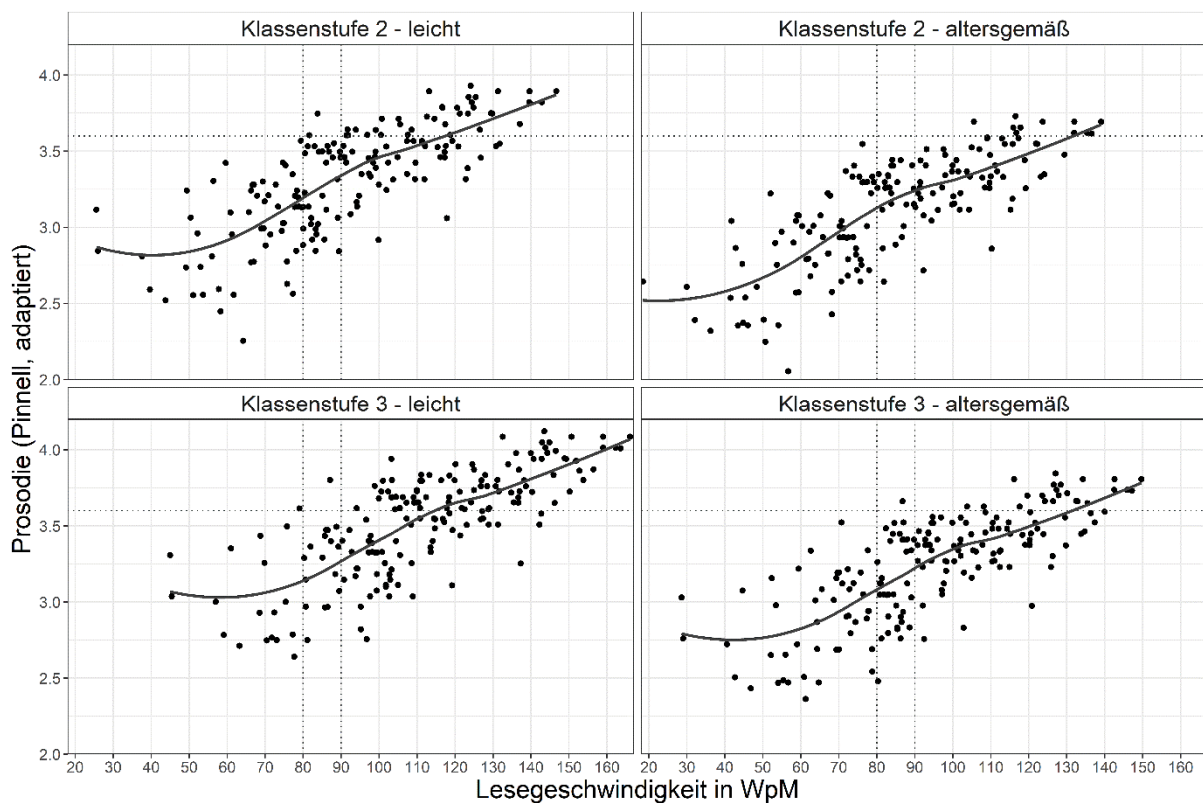


Abb. 3: Streudiagramme für Prosodie * Lesegeschwindigkeit getrennt nach Textschwierigkeit und Klassenstufe (aus dem GLM vorhergesagte Werte). Zusätzlich eingezeichnet ist die Kurve der lokal gewichteten Streudiagrammglättung (LOWESS). Die gepunkteten Linien zeigen horizontal den Grenzwert für die Prosodie (3,6) sowie vertikal für die Lesegeschwindigkeit nach Klassenstufen (2: 80 WpM; 3: 90 WpM).

4.6 | Muster der Entwicklung der Leseflüssigkeit

Um typische Entwicklungsmuster herauszufinden, wurden die o.g. Grenzwerte für die einzelnen Dimensionen der Leseflüssigkeit an die Daten der Kinder angelegt und damit für jeden gelesenen Text zu jedem Testzeitpunkt Muster von ++++ für flüssige Leserinnen und Leser (Grenzwert erreicht für D, A, L und P) bis zu ---- (kein Grenzwert erreicht) gefunden.⁷

⁷ Dieses Verfahren ist angelehnt an Nix 2011: 200 f., der an einer kleinen Stichprobe fünf Muster exemplarisch nachweist. In Vorstudien zu unseren noch nicht abgeschlossen ausgewerteten empirischen Untersuchungen zu Fördermethoden konnten wir alle 16 möglichen Muster, z.T. in Einzelfällen, nachweisen.

Um die Stabilität der vorgefundenen Muster im Verlauf eines Schuljahres zu überprüfen, wurden die Muster für den altersgemäßen Text zum ersten Messzeitpunkt in der zweiten Klassenstufe (Text 2) einerseits mit den Mustern für denselben (dann leichten) Text zum zweiten Messzeitpunkt ein Jahr später verglichen (Text 2) und andererseits mit dem ebenfalls in der dritten Klassenstufe gelesenen altergemäßen Text (Text 3). In die Vergleiche einfließen konnten nur die Muster von Kindern, die die Texte jeweils zu beiden Testzeitpunkten gelesen hatten. Von 153 Kindern lag Text 2 sowohl aus Stufe 2 als auch Stufe 3 vor (siehe Abb. 4, links). Für eine bessere Übersichtlichkeit wurden die sechs häufigsten Muster ausgewählt, die bei 134 Kindern (87,6 Prozent) für beide Bewertungen vorlagen. Keines der ausgeschlossenen Muster kam bei mehr als drei Kindern zu einem Testzeitpunkt vor. Für den zweiten Vergleich des altersgemäßen Textes in Klassenstufe 2 mit dem ebenfalls altersgemäßen Text in Klassenstufe 3 lagen Leseproben von 155 Kindern vor (siehe Abb. 4, rechts). Bei 133 Kindern (86,9 Prozent) lagen beide Bewertungen innerhalb der häufigsten Kategorien, 22 Kinder zeigten zu mindestens einem Testzeitpunkt eines der weniger häufigen Muster (die Muster und die Kinder überschneiden sich zum Großteil mit denen aus Abb. 4 links, sind aber nicht identisch). Erneut kam keines der ausgeschlossenen Muster bei mehr als drei Kindern zu einem Testzeitpunkt vor.

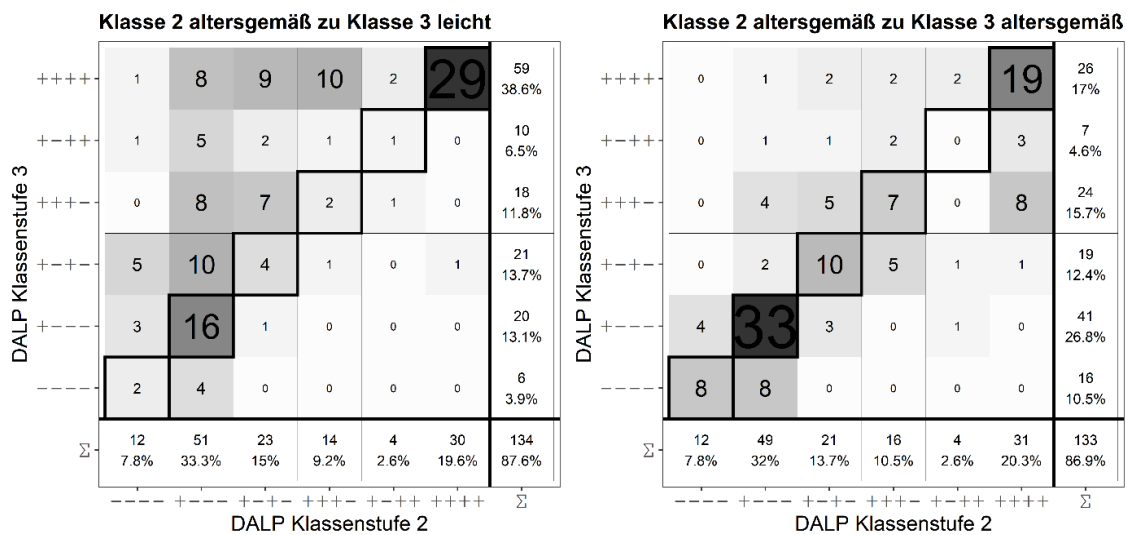


Abb. 4: Kontingenztabelle für die absoluten Häufigkeiten von Entwicklungsmustern von der Klassenstufe 2 zur Klassenstufe 3 für denselben Text (links) und jeweils altersgemäße Texte (rechts). Die Tabellen sind jeweils von unten nach oben zu lesen. In der zweiten Spalte der linken Tabelle stehen z.B. alle Kinder, die den Text in der zweiten Klasse mit dem Muster "+ - - -" gelesen haben. Vier von ihnen haben denselben Text ein Jahr später schlechter (mit "- - - -") gelesen, 16 Kinder erhielten dieselbe Bewertung wie ein Jahr zuvor und die Kinder in den Zellen darüber haben sich ein wenig bis sehr stark verbessert. In der hervorgehobenen Diagonalen sind die Häufigkeiten aufgeführt, bei denen es keine Änderung gegeben hat; Werte oberhalb der Diagonalen zeigen eine Verbesserung, unterhalb eine Verschlechterung des Musters. Die Summen und prozentualen Anteile der Spalten (Klasse 2) und Zeilen (Klasse 3) sind unten und rechts angegeben; die Prozentwerte beziehen sich auf die Gesamtzahl von 153 (links), bzw. 155 (rechts) Kindern. Die zur Marke von 100 Prozent fehlenden Kinder zeigten zu mindestens einem Zeitpunkt ein anderes Muster.

Von den 134 Kindern, die im Vergleich "Klasse 2 altersgemäß zu Klasse 3 leicht" eines der sechs häufigsten Muster gezeigt haben, konnte sich über die Hälfte (53,7 Prozent) verbessern, 40,3 Prozent zeigten eine konstante Leistung und nur 6 Prozent der Kinder verschlechterten sich. Im Vergleich der Muster für die jeweils altersgemäßen Texte in Klasse 2 und 3 ändert sich

das Bild: Je ca. 1/5 der Kinder verschlechtert oder verbessert sich und 3/5 erhielten dieselbe Beurteilung wie ein Jahr zuvor (Prozentwerte, schwächer: 22,6; gleich: 57,9; stärker: 19,5). Hierbei haben wir das Muster + – ++ stärker eingeschätzt als das Muster +++ –, da die Ansprüche an das prosodische Lesen höher sind als der Automatisierungsgrad.

Das Muster + – – – ist das häufigste Muster bei altersgemäßen Texten sowohl in Klasse 2 (hier findet es sich bei etwa einem Drittel der Kinder) als auch in Klasse 3 (noch bei gut einem Viertel). Das Muster ist gekennzeichnet durch hinreichend genaues Lesen, das aber nicht automatisiert abläuft sowie mit einer langsamen Lesegeschwindigkeit und mangelnden prosodischen Fähigkeiten einhergeht. Bei den 49 Kindern mit diesem Muster (Abb. 4, rechts) folgt auf die (frühe) Ausbildung von Dekodierfähigkeiten bis zum Ende der 2. Klasse kein (musterrelevanter) Zuwachs auf anderen Ebenen. Fünf von sechs dieser Kinder bleiben auch in Klasse 3 in diesem Muster oder verschlechtern sich (16,7 Prozent). Nur jedes sechste Kind entwickelt in Klasse 3 auf mindestens einer weiteren Ebene ausreichende Fähigkeiten. In Klasse 3 bilden zwar vier Kinder erstmals ausreichende Dekodierfähigkeiten aus (steigen also aus Muster – – – – auf), andere vier verschlechtern sich aber unter den gestiegenen Anforderungen, so dass nun immer noch mehr als ein Viertel der Kinder lediglich mit ausreichender Dekodiergenauigkeit liest.

An zweiter Stelle folgt das Muster flüssigen Lesens + + + +. Etwa 20 Prozent der Kinder gelingt es bereits in Klasse 2 auch bei einem schwierigeren Text, alle Normwerte zu erfüllen; unter den gestiegenen Anforderungen in Klasse 3 sind es jedoch nur noch 17 Prozent. Statt einer zu erwartenden Zunahme flüssiger Leserinnen und Leser mit steigender Leseerfahrung und -übung sinkt deren Zahl. Deutlich größer ist der Anteil derer, denen flüssiges Lesen bei gleichbleibender Textschwierigkeit in Klasse 3 gelingt (38,6 Prozent). Die meisten Kinder verbessern sich aus dem Muster + + + –, können also ihre prosodischen Fähigkeiten ausbauen. Aber auch eine große Anzahl von Kindern mit „multiplen“ Problemen auf mehreren Ebenen können ihre Fähigkeiten innerhalb eines Jahres bei gleichbleibenden Anforderungen bis zu flüssigem Lesen ausbauen. Das macht deutlich, dass sich auch bei schwachen Leserinnen und Lesern innerhalb eines Jahres starke Entwicklungen vollziehen können, die jedoch nicht mit den wachsenden Anforderungen an Textschwierigkeit Schritt halten können. Eventuell macht sich hier aber auch eine in Vergleich zu Klasse 2 zu starke Anhebung der Schwierigkeit des Lautlesetextes bemerkbar. Dafür spräche auch der Rückgang an flüssigen Leserinnen und Lesern in Klasse 3. Besonders die Zunahme an langen Wörtern könnte dazu geführt haben, dass der Text relativ schwerer zu lesen war.

Etwa jedes zehnte Kind liest in Klasse 2 hinreichend schnell, automatisiert und genau, jedoch mit niedrigen Prosodiewerten (+ + + –). Ein Jahr später steigt der Wert auf 15,5 Prozent, was zu fast gleichen Teilen an Kindern liegt, die ihre Automatisierungsfähigkeiten verbessert bzw. sich in ihren prosodischen Fähigkeiten verschlechtert haben. Die Entwicklungsperspektive zeigt, dass dieses Muster weniger eine schon gute Vorstufe flüssigen Lesens ist, sondern oft Ausdruck einer unterschwelligen Überforderung mit Automatisierungsprozessen. Im Detail zeigen Leserinnen und Leser dieses Musters oft ein Lesen in 3-4-er Wortgruppen, Fehler in der syntaktischen Strukturierung sowie überlesene Satzzeichen. Hier ist also nicht in erster Linie bzw. nicht nur eine Förderung der Prosodie angezeigt, sondern der weitere Ausbau der Automatisierung, damit die Lesegeschwindigkeit weiter steigen kann und überhaupt erst Raum für prosodische Produktion geschaffen wird.

5 | Diskussion

Die Zusammenschau der Entwicklung auf den einzelnen Ebenen von Leseflüssigkeit zeigt, dass die Fähigkeit zum prosodischen Lesen mit zunehmender Lesegeschwindigkeit steigt (vgl. hierzu bereits Schwanenflugel et al. 2006). Die Lesegeschwindigkeit wiederum resultiert aus einem hohen Automatisierungsgrad des Worterkennens (vgl. bereits Landerl & Wimmer 1998). Hierbei spielen die Grenzwerte eine wichtige Rolle, bei ihrer Festlegung ist aber nicht nur die Klassenstufe der Kinder zu berücksichtigen, sondern auch die Textschwierigkeit. Oberhalb der Grenzwerte erhöht die Lesegeschwindigkeit sich pro Prozentpunkt Automatisierungsgrad nur noch etwa halb so viel wie unterhalb der Grenzwerte.

Bei der Zugehörigkeit zu einem Muster der Leseflüssigkeit gibt es eine hohe Stabilität. Wer Ende Klasse 2 nicht flüssig liest, hat demnach eine nur geringe Chance, dies Ende der 3. Klasse zu tun. Nur einem Fünftel der Kinder in Klasse 2 mit geringfügigen Problemen bezüglich der Prosodie und/oder der Automatisierung gelingt es, diese innerhalb eines Jahres zu beheben. Kein Kind mit einer (zu geringen) Lesegeschwindigkeit unter dem Grenzwert von 80 WpM in Klasse 2 liest ein Jahr später flüssig. Unter dem Grenzwert in Klasse 2 liegende Lesegeschwindigkeiten haben demnach prognostische Qualität – Geschwindigkeitsdefizite in Klasse 2 sind persistent und verhindern die Ausbildung flüssigen Lesens in Klasse 3. Zwei Dritteln der Kinder, die Ende Klasse 2 keine andere Ebene als die der Automatisierung ausgebildet haben (– + – –), gelingt das auch ein Jahr später nicht. Aus dem übrigen Drittel rutscht bei gestiegenen Anforderungen die Hälfte sogar in unzureichende Dekodierfähigkeiten ab; nur ca. 16 Prozent gelingt innerhalb eines Jahres die Ausbildung einer oder mehrerer weiterer Ebenen.

Die Prosodie als anspruchsvollste Ebene, die Textverstehen sowohl begünstigt als auch voraussetzt („Henne-Ei-/Ei-Henne-Prinzip“), ist offenbar abhängig von der Ausbildung hierarchieniedrigerer Prozesse. Besonders der Automatisierung kommt hier ein Schlüsselmoment zu. Sowohl ein vergrößerter Sichtwortschatz bei zunehmender Leseerfahrung, steigendem Wortschatz und Vorwissen, als auch die Optimierung der indirekten Route beim Erlesen von Wörtern spielen dabei eine Rolle. Das zunächst deutlich mühsame und langsame, schrittweise Erlesen unbekannter Wörter geht zunehmend über in kürzere Stockungen in oder vor unbekanntem Wörtern, so dass beides – eine geringere Anzahl erlesener Wörter sowie ein schnelleres (wenn auch nicht flüssiges) Erlesen – zu Zuwächsen in der Lesegeschwindigkeit mit steigender Klassenstufe führt. Angesichts dessen, dass auch in Klasse 3 noch weit über die Hälfte der Kinder über keinen ausreichenden Automatisierungsgrad verfügt und die Varianz sich noch vergrößert, ist nicht damit zu rechnen, dass sich bei steigender Textschwierigkeit in der Entwicklung der Leseflüssigkeit alsbald Deckeneffekte bei der Automatisierung einstellen – vielmehr werden nicht-automatisierte Wörter zunehmend effektiver erlesen.

Automatisiertes und schnelles Lesen bedingen einander und sind Voraussetzungen für prosodisches Lesen. Eine möglichst hohe Geschwindigkeit ist jedoch keineswegs Garant noch notwendig für Leseflüssigkeit. In den Daten lassen sich vielmehr Mindestgeschwindigkeiten identifizieren, ab denen flüssiges Lesen möglich wird. Obgleich die durchschnittliche Geschwindigkeit in Klasse 2 bereits 84,5 WpM beträgt, ist in der Stichprobe flüssiges Lesen bereits ab 80 WpM beobachtbar. Schwache Leserinnen und Leser weisen von Klasse 2 zu Klasse 3 zwar starke Geschwindigkeitszuwächse auf, holen ihre in Klasse 2 besser lesenden Mitschüler jedoch nicht ein, vielmehr sind die Zuwächse an Lesegeschwindigkeit über die gesamte Gruppe

relativ linear. Geschwindigkeitsdefizite scheinen persistent zu sein (Klicpera 2006), was sich auch in der kaum veränderten großen Varianz in beiden Klassen zeigt. In Klasse 3 ist flüssiges Lesen ab 90 WpM zu beobachten, obwohl die durchschnittliche Geschwindigkeit auch hier um ca. 5 WpM höher liegt. Besonders stark sichtbar ist der Anstieg der Lesegeschwindigkeit bei gleichbleibender Textschwierigkeit. Nur hier werden Werte (111 WpM) erreicht, die mit denen aus der Wiener Längsschnittstudie (Klicpera & Gasteiger-Klicpera 1995) (110 WpM) zu vergleichen sind.⁸ Der altersangemessene Text wird in unserer Stichprobe mit 94,9 WpM deutlich langsamer gelesen. Erst ab Klasse 4 wird zu erwarten sein, dass der Richtwert von 100 WpM als Mindestgeschwindigkeit für flüssiges Lesen, der mangels empirischer Daten bisher in der deutschen Lesedidaktik lediglich vorgeschlagen wurde (Rosebrock 2017), erreicht wird.

Prosodisches Lesen nach den (relativ) strengen Regeln unserer Skala ist für den Großteil der Kinder in Klasse 2 und 3 noch nicht vollständig umsetzbar. Hier verhindert wohl vor allem die Konzentration auf Automatisierungsprozesse bessere prosodische Leistungen. Bei Ausschaltung von textseitigen Schwierigkeiten durch einen leichteren Text wird sichtbar, dass ein größerer Teil der Zweit- und Drittklässler Fähigkeiten zur expressiven Betonung und syntaktischen Gliederung zeigt. Immerhin befinden sich in Klasse 3 drei Viertel der Kinder auf Prosodie-Level 3 oder höher – das heißt, die Satzmelodie wird immer beachtet und es liegen zumindest Ansätze expressiver Betonung vor.

Kaum eine Rolle bei der weiteren Entwicklung von Leseflüssigkeit spielt hingegen die Dekodiergenauigkeit. Sie erreicht früh ein gewisses Plateau und beeinflusst die Leseflüssigkeit bei den meisten Kindern kaum (Landerl & Wimmer 1998; 2008). Je nach Textschwierigkeit lesen deutlich mehr als die Hälfte (leichter Text) oder knapp darunter (altersgemäßer Text) der Kinder bereits Ende der 2. Klasse nur zwei Prozent oder weniger Wörter falsch. Allerdings ist der Wert von ca. 11 Prozent von Kindern mit mangelnden Dekodierfähigkeiten in Klasse 3 Hinweis darauf, dass etwa jedes zehnte Kind von einer Progression der Leseflüchtigkeitsentwicklung abgeschnitten ist, weil grundlegende Dekodierfähigkeiten nachhaltig nicht ausgebildet werden. Kinder mit mangelnder oder nur knapp ausgebildeter Dekodierfähigkeit bedürfen einer frühen Intervention bereits in Klasse 2.

Die Einschätzung der Leseflüchtigkeit mit Hilfe überarbeiteter Lautleseprotokolle hat zu einer differenzierten Einschätzung individueller Lesemuster und Entwicklungsverläufe geführt. Trotzdem bleibt zu konstatieren, dass die zuverlässige Auswertung größere Erfahrung und Übung voraussetzt und zeitaufwändig bleibt. Selbst bei einer scheinbar objektiven Auflistung der zu zählenden Dekodierfehler ist zu beobachten, dass Höraufmerksamkeit auch Grenzen hat (bzw. trotz hochwertiger Tonaufnahmen manche Wortrealisierungen schwer zu hören sind) und es zuweilen zu minimalen Abweichungen bei der Einschätzung durch zwei Rater kommt. Subjektive Höreindrücke spielen aber vor allem auf den Ebenen Automatisierung und Prosodie eine Rolle (siehe 3.5).

Sobald weitere Messzeitpunkte komplett ausgewertet sind, planen wir komplexere statistische Analysen, die einerseits Interaktionen zwischen den vier Dimensionen der Leseflüchtigkeit

⁸ Hasbrouk & Tindal (2006) geben mit 107 RWpM – richtig gelesenen Wörtern – einen etwas niedrigeren Wert an. Da aber nur die richtig gelesenen Wörter gezählt werden, ist zu vermuten, dass ein etwas höherer WpM-Wert vorgelegen hat.

gezielter betrachten sollen und andererseits die vorhandenen Deckeneffekte ausgleichen. Tobit growth curve Modelle (Wang et al. 2008) erscheinen hier vielversprechend.

6 | Literatur

- Bamberger, R. & Vanecek, E. (1984). *Lesen – Verstehen – Lernen – Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache*. Jugend und Volk.
- Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48.
- Bortz, B. & Döring, N. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer.
- Clay, M. & Imlach, R. (1971). Juncture, stress and pitch as reading behavior variables. *Journal of verbal learning behavior and verbal behavior*, 10, 133–139.
- Coltheart, M. & Rastle, K. (1994). Serial Processing in Reading Aloud: Evidence for Dual-Route Models of Reading. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1197–1211.
- Cooper, W. E. & Paccia-Cooper, J. (1980). *Syntax and speech*. Harvard University Press.
- Gold, A. (2009). Leseflüssigkeit. Dimensionen und Bedingungen bei leseschwachen Hauptschülern. In: A. Bertschi-Kaufmann & C. Rosebrock (Hrsg.), *Literalität. Bildungsaufgabe und Forschungsfeld* (S. 151–164). Juventa.
- Groebe, N. (1982). *Leserpsychologie: Textverständnis – Textverständlichkeit*. Aschendorff.
- Hasbrouck, J. & Tindal, G. A. (2006). Oral Reading Fluency Norms: A Valuable Assessment Tool for Reading Teachers. *The Reading Teacher*, 59(7), 636–644.
- Holle, K. (2006). Flüssiges und phrasiertes Lesen (fluency). In: S. Weinhold (Hrsg.), *Schriftspracherwerb empirisch. Konzepte – Diagnostik – Entwicklung* (S. 87–119). Schneider Verlag Hohengehren.
- Holle, K. (2009). Psychologische Lesemodelle und ihre lesedidaktischen Implikationen. In: C. Garbe, K. Holle & T. Jesch (Hrsg.), *Texte lesen. Textverstehen, Lesedidaktik, Lesesozialisation*, 2 (S. 103–165). Schöningh.
- Klicpera, C. & Gasteiger-Klicpera, B. (1993). *Lesen und Schreiben. Entwicklung und Schwierigkeiten. Die Wiener Längsschnittuntersuchungen über die Entwicklung, den Verlauf und die Ursachen von Lese- und Schreibschwierigkeiten in der Pflichtschulzeit*. Verlag Hans Huber.
- Klicpera, C. & Gasteiger-Klicpera, B. (1995). *Psychologie der Lese- und Schreibschwierigkeiten. Entwicklung, Ursachen, Förderung*. Beltz.
- Klicpera, C. & Schabmann, A. (1993). Do German-speaking children have a chance to overcome reading and spelling difficulties? A longitudinal survey from the second until the eighth grade. *European Journal of Psychology of Education*, 8, 307–323.
- Klicpera, C., Schabmann, A. & Gasteiger-Klicpera, B. (2006). Die mittelfristige Entwicklung von Schülern mit Teilleistungsschwierigkeiten im Bereich der Lese- und Rechtschreibschwierigkeiten. *Kindheit und Entwicklung*, 15(4), 216–227.
- Kuhn, M. R., Schwanenflugel, P. J. & Meisinger, E. B. (2010). Alligning Theory and Assessment of Reading Fluency: Automaticity, Prosody, and Definitions of Fluency. *Reading Research Quarterly*, 45(2), 230–251.
- Landerl, K. (1996). *Legasthenie in Deutsch und Englisch*. Peter Lang.

- Landerl, K. & Wimmer, H. (2008). Development of Word Reading Fluency and Spelling in a Consistent Orthography: An 8-Year Follow-Up. *Journal of Educational Psychology*, 100(1), 150–161.
- Landis, J. R. & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174.
- Lenhard, W. (2013). *Leseverständnis und Lesekompetenz. Grundlagen – Diagnostik – Förderung*. Kohlhammer.
- McLeod, A. I. (2011). *Kendall: Kendall rank correlation and Mann-Kendall trend test*. <https://cran.r-project.org/web/packages/Kendall/Kendall.pdf> [16.09.2021].
- Nix, D. (2011). *Förderung der Leseflüssigkeit. Theoretische Fundierung und empirische Überprüfung eines kooperativen Lautlese-Verfahrens im Deutschunterricht*. Juventa.
- Pikulski, J. J. & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, 58, 510–518.
- Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campell, J. R., Gough, P. B. & Beatty, A. S. (1995). Listening to Children read aloud: Data from NAEP's Integrated Reading Performance (IRPR) at Grade 4. *Office of Educational Research and Improvement*. U.S. Department of Education.
- Rasinski, T. V. (2003). *The Fluent Reader. Oral Strategies for Building Word Recognition, Fluency, and Comprehension*. Scholastic Professional Books.
- Rasinski, T. V. (2004). Creating Fluent Readers. A growing body of evidence points to reading fluency as an important factor in student reading success, *What Research Says About Reading*, 61, 46–51.
- Rieckmann, C. (2010). *Leseförderung in sechsten Hauptschulklassen. Zur Wirksamkeit eines Vielleseverfahrens*. Schneider Verlag Hohengehren.
- Rieckmann, C. (2018). *Grundlagen der Lesedidaktik*. Band 2. Eigenständiges Lesen (2. korrigierte Aufl.). Schneider Verlag Hohengehren.
- Rosebrock, C. & Nix, D. (2006). Forschungsüberblick: Leseflüssigkeit (Fluency) in der amerikanischen Leseforschung und –didaktik. *Didaktik Deutsch*, 20, 90–113.
- Rosebrock, C. & Nix, D. (2017). *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung* (8. korrigierte Aufl.). Schneider Verlag Hohengehren.
- Rosebrock, C., Nix, D., Rieckmann, C. & Gold, A. (2017). *Leseflüssigkeit fördern. Lautleseverfahren für die Primar- und Sekundarstufe*. Klett Kallmeyer.
- Samuels, S. (1985). Word Recognition. In: H. Singer & R.B. Ruddell (Hrsg.), *Theoretical Models and Processes of Reading* (S. 256-275). International Reading Association.
- Samuels, S. & Farstrup, A. (2006). *What research has to say about fluency instruction*. International Reading Association.
- Sappok, C. & Fay, J. (2018). Prosodische Aspekte von Leseflüssigkeit messen. Evaluation einer Ratingprozedur mit Audioaufnahmen von DrittklässlerInnen. *Didaktik Deutsch*, 44, 61–83.
- Scheerer-Neumann, G. (2015). *Lese-Rechtschreib-Schwäche und Legasthenie. Grundlagen, Diagnostik und Förderung*. Kohlhammer.
- Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M. & Stahl, S. A. (2004). Becoming a Fluent Reader: Reading Skill and Prosodic Features in the Oral Reading of Young Readers. *Journal of Educational Psychology*, 96(1), 119–129.
- Walter, J. (2017). Effektivität der Förderung der Leseflüssigkeit mit Hilfe von Hörbüchern bei Grundschulern: Zwei Pilotstudien. *Zeitschrift für Heilpädagogik*, 68, 104–123.