

SLLD (B) SPRACHLICH-
LITERARISCHES
LERNEN UND
DEUTSCHDIDAKTIK
BÄNDE

Weiterführende Grundlagenforschung in Lesedidaktik und Leseförderung

Theorie - Empirie - Anwendung

Steffen Gailberger &
Christopher Sappok
(Hrsg.)

BAND 1

Weiterführende Grundlagenforschung in Lesedidaktik und Leseförderung

Theorie – Empirie – Anwendung

Herausgegeben von

Steffen Gailberger & Christopher Sappok

Open Access verfügbar unter
<https://www.doi.org/10.46586/SLLD.189>

www.slld.eu

Das Werk und seine Teile sind unter der Creative Commons Lizenz CC BY-SA „Namensnennung - Weitergabe unter gleichen Bedingungen 4.0 International“ veröffentlicht.

<https://creativecommons.org/licenses/by-sa/4.0/deed.de>



Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

© Steffen Gailberger, Christopher Sappok

Dieser Titel steht zum Download bereit unter: <https://omp.ub.rub.de/index.php/SLLD/catalog/series/SLLD-B>

ISBN: 978-3-96955-004-5 (digital)

ISSN: 2701-0600 (digital)

DOI: <https://doi.org/10.46586/SLLD.189>

Veröffentlichung der elektronischen Version durch:



Universitätsbibliothek der Ruhr-Universität Bochum

Universitätsstr. 150

D-44801 Bochum

<https://www.ub.ruhr-uni-bochum.de/>

SLLD wurde bei der Deutschen Forschungsgemeinschaft beantragt von Michael Beißwenger (Universität Duisburg-Essen), Steffen Gailberger (Bergische Universität Wuppertal), Miriam Morek (Universität Duisburg-Essen) und Björn Rothstein (Ruhr-Universität Bochum).

Gefördert durch

DFG Deutsche
Forschungsgemeinschaft

Redaktionen

SLLD (Z) ZEITSCHRIFT FÜR
SPRACHLICH-
LITERARISCHES
LERNEN UND
DEUTSCHDIDAKTIK

Jörn Brüggemann
Jochen Heins
Miriam Morek
Juliane Stude

SLLD (B) SPRACHLICH-
LITERARISCHES
LERNEN UND
DEUTSCHDIDAKTIK
BÄNDE

Steffen Gailberger
Michael Krelle
Swantje Weinhold
Thomas Zabka

SLLD (E) SPRACHLICH-
LITERARISCHES
LERNEN UND
DEUTSCHDIDAKTIK
EINFÜHRUNGEN

Birgit Mesch
Florian Radvan
Björn Rothstein

SLLD (U) SPRACHLICH-
LITERARISCHES
LERNEN UND
DEUTSCHDIDAKTIK
UNTERRICHTSVORHABEN

Michael Beißwenger
Juliane Dube
Steffen Gailberger
Boris Körkel
Kirsten Schindler
Benjamin Uhl
Johannes Wild
Heike Wirthwein

Beirat

Cordula Artelt (Bamberg)
Michael Becker-Mrotzek (Köln)
Albert Bremerich-Vos (Essen)
Christian Dawidowski (Osnabrück)
Ricarda Freudenberg (Weingarten)
Christine Garbe (Köln)
Ingrid Gogolin (Hamburg)
Cornelia Gräsel (Wuppertal)
Elke Grundler (Ludwigsburg)
Matthias Hölzner (Essen)
Michael Kämper-van den Boogaart (Berlin)
Beate Leßmann (Kiel)
Christine Pauli (Fribourg)
Susanne Prediger (Dortmund)
Susanne Riegler (Leipzig)
Knut Schwippert (Hamburg)
Torsten Steinhoff (Siegen)
Maja Wiprächtiger-Geppert (FHNW Windisch)
Arne Ziegler (Graz)
Evelyn Ziegler (Essen)

Inhalt

| | |
|--|-----|
| <i>Steffen Gailberger & Christopher Sappok</i> Weiterführende Grundlagenforschung in Lesedidaktik und Leseförderung. Einführung in den Band | 2 |
| <i>Britta Juska-Bacher, Martina Röthlisberger, Ladina Brugger & Christoph Zangger</i> Die Bedeutung von phonologischer Bewusstheit, Benennungsgeschwindigkeit und Wortschatz | 11 |
| <i>Sabine Röttig, Clara Schwerkolt & Guido Nottbusch</i> Die Entwicklung der Leseflüssigkeit in der Grundschule | 27 |
| <i>Sabine Stephany, Markus Linnemann, Evghenia Goltsev & Michael Becker-Mrotzek</i> Prosodische Aspekte der Leseflüssigkeit als Indikator für Lesekompetenz | 48 |
| <i>Christopher Sappok</i> Exploring Advanced Prosody – eine Best-Practice-Untersuchung zum lauten Lesen in der weiterführenden Schule | 68 |
| <i>Gerrit Helm</i> Das Ausmaß der Unterschiede in den Lese-Items des Salzburger Lese-Screenings 2–9 | 98 |
| <i>Anke Schmitz, Fabiana Karstens & Jörg Jost</i> Vermittlung von Lesestrategien im Umgang mit Sachtexten im Deutschunterricht | 126 |
| <i>Ruven Stahns, Svenja Rieser & Anke Hußmann</i> Zur Passung von Leseförderung und Schülerkompetenzen im Leseunterricht in Deutschland | 146 |
| <i>Steffen Gailberger, Britta Pohlmann, Laura Reichenbach, Franziska Thonke & Jonna Wolters</i> Zum nachhaltigen Einfluss von Lautleseverfahren auf Leseflüssigkeit, Leseverstehen, Rechtschreibung sowie Kompetenzen jenseits des Deutschunterrichts | 166 |
| <i>Daria Ferencik-Lehmkuhl & Ilka Fladung</i> Lesediagnostik und -förderung im heterogenen Deutschunterricht | 192 |
| <i>Maik Philipp</i> Liest du noch oder sourcst du schon? Sourcing als neuer Bestandteil kompetenten Lesens (multipler, teils digitaler Dokumente) | 214 |
| <i>Carolin Meier</i> Leseförderung mit Antolin? | 231 |
| <i>Miriam Dittmar</i> Biologielehrmitteltexte sinnvoll lesen und verstehen | 249 |

Steffen Gailberger & Christopher Sappok

Weiterführende Grundlagenforschung in Lesedidaktik und Leseförderung

Einführung in den Band

Was bedeutet Grundlagenforschung? Mit dieser ‚großen‘ Frage in einen lesedidaktischen Sammelband zu starten, scheint (neben des expliziten Rekurses auf seinen Titel) insofern legitim und naheliegend zu sein, da die Deutsch- bzw. Lesedidaktik u.E. bislang zu wenig für eine eingehende terminologische Klärung in diese Richtung unternommen hat. So wird auf Grundlagenforschung als wissenschaftstheoretischer Begriff innerhalb der deutschdidaktischen Methodendiskussion im Grunde gänzlich verzichtet – und zwar sowohl hinsichtlich der theoretischen wie empirischen Methodendiskussion (vgl. etwa Neumann & Mahler 2014; Kämper-van den Boogaart & Spinner 2019; Feilke & Pohl 2014; Pohl & Ulrich 2011; Bredel & Reißig 2011 oder Kammler & Knapp 2002). In jenen Publikationen wiederum, in denen der Begriff Grundlagenforschung durchaus Verwendung findet und die in nuce der Deutschdidaktik (bzw. an den Randbezirken ihrer Nachbarschaftsdisziplinen wie bspw. der empirischen Lehr-Lern-Forschung) zuzurechnen sind, geschieht dies aber weitgehend nur beispielhaft oder implizit, sehr selten aber systematisch und explizit (vgl. etwa Pflugmacher 2016; Nickel-Bacon 2006; Fricke, Kammler & Rupp 2012; Bayrhuber et al. 2012; Koch-Priewe 2004; Walter 2001; erkennbare Ausnahmen hinsichtlich dieser Beobachtung sind bei Köster 2016 und Frederking 2016 auszumachen).

Dabei ist es gar nicht so aufwendig, den Begriff Grundlagenforschung auf die Deutsch- bzw. Lesedidaktik zu beziehen (vgl. hierzu etwa grundlegend Gailberger 2018), um erkennen zu können, dass in ihnen seit über zwei Jahrzehnten bereits Grundlagenforschung betrieben wird. Zur Verdeutlichung dessen zeichnen wir zunächst den wissenschaftshistorischen Prozess nach, der von einer Fachdisziplin (wie bspw. der Lesedidaktik) durchlaufen werden muss, um (wissenschaftstheoretisch gesprochen) jene Position zu erlangen, die es ihr überhaupt erst erlaubt, grundlagenintendiert Forschungsfragen und Ergebnisse zu generieren.

Der wissenschaftshistorische Prozess einer (neuen) Fachdisziplin

Eine nachhaltig einflussreiche, historisch perspektivierte wissenschaftstheoretische Sicht auf diesen Klärungsversuch wurde von Wolfgang van den Daele (1975) in seinem mittlerweile als kanonisch geltenden Beitrag „Autonomie contra Planung: Scheingefecht um die Grundlagenforschung?“ vorgenommen. In ihm zeichnet der mittlerweile emeritierte Professor für Wissenschaftsforschung und Soziologie den Entwicklungsprozess einer sich im Konstituierungsvorgang befindlichen Fachdisziplin mithilfe eines Drei-Phasen-Modells nach, an dessen (vorläufigem)

Ende überhaupt erst die Voraussetzungen für grundlagenintendiertes Forschen erfüllt seien. Dieses (freilich nur vorläufige) Ende wird von ihm definiert als die „Reife“ einer Wissenschaft (ebd.: 32), die sich dann zeige, wenn von externer Seite (d.h. politisch, ökonomisch oder sozial) versucht werde, planerisch auf ihre Fragen, Methoden und Ergebnisse Einfluss zu nehmen, sie zu lenken, ggf. auch für eigene Zwecke zu beeinflussen. Diese Reife steht also für die Güte einer betreffenden Fachdisziplin und kann in der Modellierung von den Daeles durch einen sich in drei Phasen beschreibbaren Entwicklungsprozess erreicht werden, wie wir es im Folgenden einmal am Beispiel der germanistischen Fach- bzw. Lesedidaktik nachvollziehen wollen.

Die (explorative) Phase 1

Die erste Phase einer sich gerade neu konstituierenden Fachdisziplin ist dadurch gekennzeichnet, dass ihre ersten Forschungsvorhaben ein „exploratives, empiristisches [oder auch] vorparadigmatisches“ Vorgehen offenbaren (ebd.: 27). Im Sinne einer Klärung ihrer Grundlagen wird zwar auch hier bereits so etwas wie ‚Grundlagenforschung‘ betrieben, strenggenommen kann dies aber freilich nur in einem sehr weiten, mehr oder weniger vorwissenschaftlichen Sinne als ‚Art of the Soluble‘ gelten. Was einer betreffenden Fachdisziplin in diesem frühen Stadium v.a. noch fehlt, ist eine sich progressiv auswirkende Theoriedynamik, da die Ausformulierung der Theorie(n) des Gegenstandsbereichs/der Gegenstandsbereiche zu diesem Zeitpunkt ja überhaupt erst noch das Ziel der betroffenen Wissenschaft(ler*innen) darstellt, so dass hier Entdeckungen noch einen Vorrang vor Erklärungen haben (ebd.). Überträgt man diese allgemeinwissenschaftlichen Gedanken auf den Deutschunterricht und seine damals noch nicht existente Fach- bzw. Lesedidaktik, wie bspw. Kämper-van den Boogaart (2019) es tut, ließe sich diese erste Phase auf die Geschichte von Deutschunterricht und Deutschdidaktik bis zum Jahr 1970 datieren, in der Vertreter wie etwa Robert Ulshöfer (1952 ff.) und Hermann Helmers (1966), bzw. ein Jahrhundert zuvor Robert Heinrich Hieckes (1841) oder Philipp Wackernagel (1842) erstmals den Versuch unternahmen, eine fachdidaktische Fundierung des Deutschunterrichts vorzunehmen.

Die (theoriedynamische) Phase 2

Die anschließende zweite Phase des fachspezifischen Entwicklungsprozesses nach van den Daele zeichnet sich durch eine deutliche Zunahme an kognitiver Eigendynamik, d.h. also durch Theoriedynamik und Paradigmatisierung innerhalb der Disziplin, aus (van den Daele 1975: 30). Sie hat die erste Phase abzulösen bzw. zu überwinden, wenn das Forschungsprogramm nicht mehr durch explorativ-empiristische Ansätze abgedeckt werden kann, sondern wenn es durch die Implikationen und Probleme von Erklärungsansätzen (d.h. durch Modelle, Hypothesen und deren Stützung oder Verwerfung etc.) bestimmt wird. In dieser Phase sind begriffliche Ungeheimheiten der Erklärungsansätze aufzubereiten, auf verwandte oder benachbarte Phänomene auszudehnen und modellhaft in andere theoretische Konzepte zu transformieren und zu integrieren. Was die Phase zwei aber vor allem von der Phase eins unterscheidet, ist eine erste experimentelle Realisierung ihrer Methoden und Ansätze durch konkrete Anwendung im Feld. Van den Daele nennt diese Phase auch die Phase der Autonomie, was bedeuten soll, dass sie in diesem Stadium noch vor externen Planungsintentionen gefeit ist, so lange nämlich, wie der theoriedynamische Prozess anhält und noch zu keinen eindeutigen (d.h. hier: empirischen) Ergebnissen geführt hat. Für die Deutsch- bzw. Lesedidaktik ist diese Phase auf die drei Jahrzehnte nach 1970 zu datieren, Jahrzehnte der (fast ausschließlich) theoretischen Auseinandersetzungen, die vor dem Hintergrund von Ideologiekritik, Strukturalismus, Rezeptionsästhetik,

Poststrukturalismus, Dekonstruktion etc. geführt wurden (vgl. Kammler 2019) und die der hier besonders fokussierten Lesedidaktik noch nicht den empirischen und sozialisationswissenschaftlichen Spin geben konnten, von dem sie nach 2000 so sehr profitierte.

Die (finalisierende) Phase 3

Mit der lesedidaktischen Zeitenwende nach 2000, die in der letztgemachten Anspielung auf die erste PISA-Studie durchschimmert, verweisen wir auf den dialektischen Umstand, dass eine betreffende Fachdisziplin die soeben als konstitutiv angenommene Autonomie in Phase zwei bei zunehmendem ‚Erfolg‘ wieder verliert, womit schließlich die „Finalisierungsphase“ (van den Daele 1975: 30), spricht: die Phase ihrer ‚Reife‘ (s. oben), erreicht wäre. Final ist diese Phase drei deswegen zu nennen, da sie auf der als gelungen geltenden theoretischen Grundierung des Faches fußend den relativen Abschluss der Theoriedynamik und ihre gleichzeitige Überwindung markiert. Dieser Abschluss zeigt sich schließlich in einer weitergehenden Spezifikation der zuvor entwickelten theoretischen Grundlagen durch **weiterführende** Forschung, die an dieser Stelle erstmals als theoriegestützte und/oder empirische **Grundlagenforschung** im eigentlichen Sinne bezeichnet werden kann (vgl. hierzu grundlegend auch Mittelstraß 1992). Der dabei zu beobachtende (potenzielle oder reale) Verlust von Autonomie einer Wissenschaft z.B. durch politische oder soziale Interessen und Einflussnahmen (Stichwort: KMK, PISA, IGLU, VERA, IQB-Bildungstrend etc.) ist in der Perspektive van den Daeles (1975) aber, anders als etwa bei Wintersteiner (2007), nicht von vornherein als negativ zu bewerten. Er erkennt darin vielmehr jenen Ausweis von „Reife“ der betreffenden Wissenschaft (ebd.: 32), die sich (im und durch den Prozess ihrer Reifung) in der Überwindung der explorativen wie der theoriedynamischen Phasen Bahn bricht und damit nunmehr nicht allein die Darstellung der theoretischen Grundlagen eines Faches ermöglicht, sondern die tieferliegende und weiterführende Durchdringung dieser als Grundlagenforschung vorantreibt.

Zur aktuellen Situation der deutsch- bzw. lesedidaktischen Grundlagenforschung

Wie wir meinen, ist damit die aktuelle Situation weiter Teile der Deutsch- bzw. Lesedidaktik umrissen, verstanden als die Wissenschaft nicht nur zur konzeptionellen Entwicklung und theoretischen Fundierung (im Sinne der Phase zwei), sondern ebenso auch zur weiteren Durchdringung und schließlich zur empirischen Erforschung des Lesens und seiner Förderung innerhalb und außerhalb des Deutschunterrichts. Viel ist dazu bereits geleistet worden, nicht umsonst gilt das Lesen von Schüler*innen als die mittlerweile am besten erforschte sprachliche Teilkompetenz im Deutschen. Mit primarstufenspezifischen Studien aus bildungswissenschaftlicher Perspektive wie IGLU 2001 bis 2016 (vgl. zuletzt Hußmann et al. 2017), dem IQB-Ländervergleich 2011 (vgl. Stanat et al. 2012), dem IQB-Bildungstrend 2016 für die Primarstufe (vgl. Stanat et al. 2017) oder den bisherigen VERA-3-Untersuchungen (vgl. etwa Groß Ophoff et al. 2008 oder Krelle 2015) auf der einen Seite, sowie einzelnen Projekten wie jenes von Gold et al. (2015) zur Förderung der Leseflüssigkeit, von Marx, Weber und Schneider (2005a; 2005b) zu Aspekten phonologischer Bewusstheit, von Klicpera, Gasteiger-Klicpera und Schabmann (1993) zu Längsschnittentwicklungen bei Lese- und Rechtschreibschwäche oder von Helmke und Schrader (1998) zur entwicklungs- und motivationspsychologischen Entwicklung von Grundschüler*innen auf der anderen Seite werden seit über zwei Jahrzehnten immer wieder neue Aspekte des Lesens und Lernens in Schule und Unterricht in den Fokus gerückt, sodass wir mittlerweile auf ein sehr breites Wissen über die Modellierung von Lesekompetenz (vgl. im Überblick Gailberger 2019) und ihrer Förderung (vgl. im Überblick Nix 2019) zurückgreifen können.

An dieser Stelle möchte der vorliegende Sammelband zur „Weiterführenden Grundlagenforschung in Lesedidaktik und Leseförderung“ anknüpfen, den theoretischen und empirischen Forschungsstand ergänzen und die Erkenntnisse auf Schule und Unterricht übertragen.

Britta Juska-Bacher, Martina Röthlisberger, Ladina Brugger & Christoph Zangger

Den Beginn machen Britta Juska-Bacher, Martina Röthlisberger, Ladina Brugger und Christoph Zangger. In ihrer Studie „*Lesen im 1. Schuljahr: Die Bedeutung von phonologischer Bewusstheit, Benennungsgeschwindigkeit und Wortschatz*“ steht die Frage im Zentrum, welchen Einfluss Wortschatzkompetenzen auf den frühen Leseerwerb haben. Untersucht werden, neben zahlreichen Kontextvariablen, Wortschatzumfang, -tiefe und -organisation in einem auf weiterführenden Längsschnitt angelegten Szenario. Mit dem Wortschatz nehmen die Autor*innen einen im deutschsprachigen Raum z.B. gegenüber der phonologischen Bewusstheit vernachlässigten Faktor genauer in den Blick, der bislang eher als Voraussetzung für hierarchiehöhere Lesekompetenzen erforscht wurde, und plädieren dafür, diesem Faktor auch in frühen Förderkontexten einen festeren Platz zuzuweisen.

Mit der **Leseflüssigkeit** steht ein *weiterer* Faktor im Zentrum gleich vierer Beiträge, der in seiner Bedeutung für Leseerwerbsprozesse bis vor ca. 15 Jahren hierzulande kaum Beachtung fand. In den Bildungsstandards Deutsch für den Primarbereich (KMK 2005) taucht der Begriff noch nicht explizit auf und es kann als das Verdienst von Karl Holle angesehen werden, seine Relevanz bereits sehr früh erkannt und die vorher hauptsächlich im angelsächsischen Raum erfolgte Forschung für die Deutschdidaktik fruchtbar gemacht zu haben (Holle 2006).

Sabine Röttig, Clara Schwerkolt & Guido Nottbusch

Sabine Röttig, Clara Schwerkolt und Guido Nottbusch untersuchen in „*Die Entwicklung der Leseflüssigkeit in der Grundschule [...]*“ Audioaufnahmen von Zweit- und Drittklässler*innen, ebenfalls in einem auf weiterführenden Längsschnitt angelegten Szenario. Hier steht die Frage im Zentrum, wie einzelne Aspekte von Leseflüssigkeit im Zuge des Erwerbsprozesses aufeinander aufbauen bzw. welche Blockaden entstehen können, wenn die Entwicklung bzgl. hierarchieniedriger Aspekte wie Automatisierung stockt. Bestätigung finden dabei Tendenzen in der einschlägigen internationalen Forschung, dass die Prosodie eine besondere Rolle in Hinblick auf das Textverstehen spielt und somit auch ein besonderes diagnostisches Potenzial birgt. Dieses weiter auszuschöpfen ist noch an die Klärung vor allem methodischer Desiderate geknüpft.

Sabine Stephany, Markus Linnemann, Evghenia Goltsev & Michael Becker-Mrotzek

Mit „*Prosodischen Aspekte der Leseflüssigkeit als Indikator für Lesekompetenz*“ beschäftigen sich Sabine Stephany, Markus Linnemann, Evghenia Goltsev und Michael Becker-Mrotzek. Die Autor*innen stellen ein diagnostisches Instrument vor, das auf einen Einsatz durch Lehrkräfte im Klassenzimmer angelegt ist. Bei Audioaufnahmen von Viertklässler*innen geht es um die Frage, ob unterschiedlichen lokalen prosodischen Herausforderungen stimmlich Rechnung getragen wird. Die Evaluation des Instruments bestätigt den Ansatz, dass – anders als bei holistischeren Verfahren zur Diagnostik von Leseflüssigkeit – Subkomponenten von Prosodie zu unterscheiden sind. Diese weisen offenbar unterschiedliche Beziehungen zum Textverstehen auf und müssen aus methodisch-diagnostischer, aber auch aus Förderperspektive unterschiedlich behandelt werden.

Christopher Sappok

Diagnostik bei Subkomponenten von Prosodie stellt auch der Beitrag von Christopher Sappok, *„Exploring Advanced Prosody – eine Best-Practice-Untersuchung zum lauten Lesen in der weiterführenden Schule“*, in den Vordergrund. Untersucht wurden Audioaufnahmen von bereits flüssig lesenden Fünft- bis Siebtklässler*innen, um Ziele des Erwerbsprozesses und konkrete Lernerwartungen zu definieren. Dabei konnte gezeigt werden, dass in der weiterführenden Schule Prosodie als Ursache für Leistungsunterschiede ein großes Thema ist, die Lesegeschwindigkeit aber nicht mehr. Der Autor plädiert dafür, zwischen „Lese Flüssigkeit“ und einer konzeptionell anders gewichteten „Vorlesekompetenz“ bei Teenagern zu unterscheiden.

Gerrit Helm

Lesegeschwindigkeit ist es nun gerade, worauf es bei einem der verbreitetsten Diagnose-Instrumente seiner Art ankommt. Vor diesem Hintergrund vertieft Gerrit Helm in dem Beitrag *„Das Ausmaß der Unterschiede in den Lese-Items des Salzburger Lese-Screenings 2–9“* die Frage, welche Rolle die Bearbeitungszeit der lesenden Schüler*innen spielt. Sie wurde bei Jungen und Mädchen der Primarstufe, die an einem Tablet arbeiteten, itemspezifisch gestoppt. Daraus ergeben sich wertvolle Hinweise auf problematische Anwendungs- bzw. Interpretationskontexte bei einem Instrument, das sich in dieser Perspektive vielleicht nicht *ganz* umsonst durch maximale Anwendungsökonomie auszeichnet.

Bei den soweit zur Leseflüssigkeit vorgestellten Beiträgen steht Diagnostik im Vorder-, und die Didaktik im Hintergrund. Methoden und Konstrukte werden häufig noch hinterfragt und optimiert, wie das Beispiel Prosodie zeigt. Die vier nachfolgenden Beiträge beziehen sich demgegenüber unmittelbar auf die Ausgestaltung von Gelingensbedingungen mit dem Schwerpunkt **Textverstehen**, wobei nicht allein die Schüler*innen, sondern auch die Lehrkräfte fokussiert werden.

Anke Schmitz, Fabiana Karstens & Jörg Jost

Anke Schmitz, Fabiana Karstens und Jörg Jost untersuchen in ihrem Beitrag *„Vermittlung von Lese Strategien im Umgang mit Sachtexten im Deutschunterricht“*, welche Rolle das intensiv beforschte selbstregulierte Lernen im Kontext Lesedidaktik in der Praxis spielt. Verglichen werden dazu zwei Perspektiven: Lehrkräfte werden befragt und ihr Unterricht zur Arbeit an Sachtexten in Jahrgangsstufe 5 wird genau beobachtet. Klarer wird so nicht nur, inwiefern Lehrkräfte Lese Strategien initiieren und unterstützen, sondern auch, was sie darüber wissen. Dies hilft zu begreifen, auch mit Blick auf die universitäre deutschdidaktische Ausbildung, welche Engpässe auf dem Weg von der Forschung zur Praxis noch bestehen.

Ruven Stahns, Svenja Rieser & Anke Hußmann

Sehr ähnlich gelagert ist das Interesse von Ruven Stahns, Svenja Rieser und Anke Hußmann in dem Beitrag *„Zur Passung von Leseförderung und Schülerkompetenzen im Leseunterricht in Deutschland“*. In ihm werden Ergebnisse aus einer Lehrkräftebefragung (IGLU 2016) mit Daten zum Unterrichtskontext verglichen. Dabei analysieren die Autor*innen den Stellenwert von Lese Strategien und Stillesephasen und ob dieser dem Leistungsprofil des Klassenverbands entspricht. Auch dieser Beitrag stellt einen erhöhten Bedarf an Durchlässigkeit zwischen Forschung und Praxis fest, kann aber zeigen, dass es um die Diagnostik im Klassenzimmer besser bestellt ist als um die gezielte Berücksichtigung diagnostischer Information.

Steffen Gailberger, Britta Pohlmann, Laura Reichenbach, Franziska Thonke & Jonna Wolters

Einen über den Lese- und Deutschunterricht hinaus wirksamen Ansatz verfolgen Steffen Gailberger, Britta Pohlmann, Laura Reichenbach, Franziska Thonke und Jonna Wolters mit der Langzeitstudie „*Zum nachhaltigen Einfluss von Lautleseverfahren auf Leseflüssigkeit, Leseverstehen, Rechtschreibung sowie Kompetenzen jenseits des Deutschunterrichts*“. Ausgewählte Grundschulen der Freien und Hansestadt Hamburg wurden über drei Jahre mit einem breitgefächerten Interventionsangebot methodisch ausgestattet und diagnostisch begleitet. Leseschwache Schüler*innen und Transfereffekte auf Domänen auch jenseits des Lesens (hier: Rechtschreibung) und jenseits des Deutschunterrichts (hier: Mathematik) wurden dabei besonders fokussiert. Verdeutlicht und bestätigt werden konnte u.a. die Bedeutung einer kontinuierlichen, hoch frequenten und langanhaltenden Dauer von Fördermaßnahmen für deren nachhaltigen Erfolg.

Daria Ferencik-Lehmkuhl & Ilka Fladung

Auch Daria Ferencik-Lehmkuhl und Ilka Fladung berichten in ihrem Beitrag „*Lesediagnostik und -förderung im heterogenen Deutschunterricht*“ von einem Angebot für Schulen, das in einem differenzierten und auf mehrere Jahre ausgelegten Komplex aus Diagnostik und Förderung bestand. Damit sollten gymnasiale Ganztagschulen vor allem darin unterstützt werden, Schüler*innen mit sozioökonomisch bedingten Schwächen noch stärker in ihrer Perspektive auf den Erwerb der Hochschulreife zu stärken. Überraschend ist bei den Ergebnissen z.B. das Ausmaß an Heterogenität in der Leseleistung noch in Jahrgangsstufe 7, aber auch eine ausgeprägte Heterogenität seitens der einbezogenen Lehrkräfte und Schüler*innen, wenn es darum geht, Unterstützungsangebote überhaupt konsequent wahrzunehmen.

Die letzten drei Beiträge in diesem Band erweitern unsere Sicht auf das Lesen und seine Bedeutung in verschiedener Hinsicht, v.a. tritt hier der Aspekt des Lesens *als mit Texten und Medien umgehen* in den Vordergrund.

Maik Philipp

Maik Philipp bietet mit seinem Beitrag „*Liest du noch oder sourcst du schon?*“ einen Überblick über den Stand der Forschung zum Thema „Sourcing“, womit grob die Wahrnehmung und Verwertung von impliziten oder expliziten Informationen über (multiple) Texte und deren Ursprung gemeint ist. Dies ist besonders relevant im Angesicht des Internets als übermächtigem, aber zweifelhaftem Textlieferant. Zweifel zu kultivieren bzw. ihnen systematisch (*educated*) Rechnung zu tragen, erfordert eine ganz eigene Form von Kompetenz. Deren Erforschung steht – vor allem im Hinblick auf ihre Didaktik – erst am Anfang, wobei davon auszugehen ist, dass ihre Bedeutung weiterhin rasant zunimmt.

Carolin Meier

Einen ganz spezifischen Zweifel formuliert Carolin Meier schon im Titel ihres Beitrags „*Leseförderung mit Antolin?*“. Antolin ist ein kommerzielles, auch in schulischen Kontexten sehr weit verbreitetes Angebot, bei dem online Fragen zu Büchern zu beantworten und so Punkte zu sammeln sind. Was sagen die Schüler*innen dazu? Die Auswertung von leitfadengestützten Gruppendiskussionen und Befragungen in den Jahrgangsstufen 4 und 7 bestätigen, dass die

vom Anbieter gemachten Versprechen, die z.B. mit der Steigerung der intrinsischen Lesemotivation über eine Steigerung des Lesevolumens hinausgehen, kaum eingelöst werden.

Miriam Dittmar

Kommerzielle Angebote stehen auch im Zentrum des Beitrags „*Biologielehrmitteltexte sinnvoll lesen und verstehen*“ von Miriam Dittmar. Besonderes Interesse gilt dabei den in die Lehrwerke eingebundenen Abbildungen bzw. der Frage, wie diese dann in den Leseprozess integriert werden. Entsprechende Aktivitäten werden bei Siebtklässler*innen nicht nur genau beobachtet, sondern es werden auch Überarbeitungsprinzipien abgeleitet und in ihrer Effektivität evaluiert. Auch wenn hier ein herkömmlicheres Format untersucht wird, liegt der Fokus auf Text-Bild-Interaktionen für einen Bereich, dessen Bedeutung im digitalen Zeitalter zu-, wenn nicht gar überhandnimmt.

Wir freuen uns, mit den hier freilich nur unvollständig skizzierten Beiträgen einen umfassenden Querschnitt durch die, wie wir meinen, im besten Sinne aufregende (Grundlagen-)Forschungslandschaft des schulischen Lesens und seiner Förderung präsentieren zu können. Zugleich sind wir gespannt, inwiefern die uns gegenwärtig begleitenden Corona-bedingten Wandlungen der schulischen und außerschulischen Arbeitsbedingungen, die auch diesem Band in seiner Entstehung und Realisierung zu schaffen gemacht haben, zu *neuer* weiterführender Forschung, z.B. zur digitalen Lesedidaktik und Leseförderung, führen werden.

Für das gründliche Lektorat des Manuskripts danken wir Laura Reichenbach, Raphaela Rödeler und Kim Stroeks sehr herzlich.

Literatur

- Bayrhuber, H., Harms, U., Muszynski, B., Ralle, B., Rothgangel, M., Schön, L., Vollmer, H. & Weigang, H.-G. (Hrsg.) (2012): *Empirische Fundierung in den Fachdidaktiken*. Waxmann.
- Bredel, U. & Reißig, T. (Hrsg.) (2011): *Weiterführender Orthographieunterricht*. Schneider Hohengehren.
- Feilke, H. & Pohl, T. (Hrsg.) (2014): *Schriftlicher Sprachgebrauch – Texte verfassen*. Schneider Hohengehren.
- Frederking, V. (2016): Allgemeine Fachdidaktik – Metatheorie und Metawissenschaft der Fachdidaktiken. Begründungen und Konsequenzen. In Vorstand der Gesellschaft für Fachdidaktik (Hrsg.), *Auf dem Weg zu einer Allgemeinen Fachdidaktik*. (179–204). Waxmann.
- Frickel, D., Kammler, C. & Rupp, G. (Hrsg.) (2012): *Literaturdidaktik im Zeichen von Kompetenzorientierung und Empirie. Perspektiven und Probleme*. Fillibach Verlag.
- Gailberger, S. (2018): Grundlagenforschung. Wissenschaftstheoretische Reflexionen aus deutschdidaktischer Perspektive. In J. Boelmann (Hrsg.), *Empirische Forschung in der Deutschdidaktik*. (S. 17–48). Schneider Verlag.
- Gailberger, S. (2019): Modellierung von Lesekompetenz. In M. Kaemper-van den Boogaart & K. Spinner (Hrsg.), *Deutschunterricht in Theorie und Praxis. Band 11/1: Lese- und Literaturunterricht. Konzeptionelle und empirische Grundlagen*. (S. 273–346). Schneider Verlag.

- Gold, A., Behrendt, S., Lauer-Schmaltz, M. & Rosebrock, C. (2015): Förderung der Leseflüssigkeit in dritten Grundschulklassen. In C. Rosebrock & A. Bertschi-Kaufmann (Hrsg.): *Literalität erfassen: bildungspolitisch, kulturell, individuell*. (S. 203–218). Beltz-Juventa.
- Groß Ophoff, J., Isaac, K., Hosenfeld, I. & Eichler, W. (2008): Erfassung von Leseverständnis im Projekt VERA. In B. Hofmann & R. Valtin (Hrsg.), *Checkpoint Literacy*. (S. 36-51). Deutsche Gesellschaft für Lesen und Schreiben (DGLS).
- Helmers, H. (1966): *Didaktik der deutschen Sprache. Einführung in die Theorie der muttersprachlichen und literarischen Bildung*. Klett.
- Helmke, A. & Schrader, F.-W. (1998): Entwicklung im Grundschulalter. Die Münchner Studie „SCHOLASTIK“. *Pädagogik*, 34(6), 24–28.
- Hiecke, R. H. (1841): *Der deutsche Unterricht auf deutschen Gymnasien*. Eduard Eisenach.
- Holle, K. (2006): Flüssiges und phrasiertes Lesen (fluency). Lesetheoretische Grundlagen und unterrichtspraktische Hinweise. In: S. Weinhold (Hrsg.): *Schriftspracherwerb empirisch*. (S. 87–119). Schneider.
- Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, Da., Lankes, E.-M., Mcelvany, N., Stubbe, T. C. & Valtin, R. [Hrsg.] (2017): *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Waxmann.
- Kämper-van den Boogaart, M. & Spinner, K. (Hrsg.) (2019): *Deutschunterricht in Theorie und Praxis (DTP). Band 11/1-3: Lese- und Literaturunterricht*. Schneider Verlag.
- Kämper-van den Boogaart, M. (2019): Geschichte des Lese- und Literaturunterrichts. In M. Kämer-van den Boogaart & K. Spinner (Hrsg.), *Deutschunterricht in Theorie und Praxis (DTP). Band 11/1: Lese- und Literaturunterricht. Konzeptionelle und empirische Grundlagen*. (S. 3–88). Schneider Verlag.
- Kammler, C. (2019): Literaturtheorie und Literaturdidaktik. In M. Kämer-van den Boogaart & K. Spinner (Hrsg.), *Deutschunterricht in Theorie und Praxis (DTP). Band 11/1: Lese- und Literaturunterricht. Konzeptionelle und empirische Grundlagen*. (S. 202–236). Schneider Verlag.
- Kammler, C. & Knapp, W. (2002): *Empirische Unterrichtsforschung und Deutschdidaktik*. Schneider Hohengehren.
- Klicpera, C., Gasteiger-Klicpera, B. & Schabmann, A. (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*. Huber Verlag.
- KMK (Kultusministerkonferenz, Beschlüsse der) (2005): *Bildungsstandards im Fach Deutsch für den Primarbereich (Jahrgangsstufe 4) – Beschluss vom 15.10.2004*. Luchterhand.
- Koch-Priewe, B. (2004): Professionsforschung und Didaktik der LehrerInnenbildung. In B. Koch-Priewe, F.-U. Kolbe & J. Wildt (Hrsg.), *Grundlagenforschung und mikrodidaktische Reformansätze zur Lehrerbildung*. (S. 7–21). Julius Klinkhardt.
- Köster, J. (2016): Die dilemmatische Disziplin – Deutschdidaktik zwischen Eklektizismus und Partialisierung. In: C. Bräuer (Hrsg.): *Denkrahmen der Deutschdidaktik: Die Identität der Disziplin in der Diskussion*. (S. 59–78). Peter Lang.
- Krelle, M. (2015): Leseverstehen im Kontext der Vergleichsarbeiten für die dritte Klasse – Leistungen und Grenzen eines diagnostischen Instruments zur Sprachförderung. *Leseforum Schweiz. Literalität in Forschung und Praxis*, 1, 1–27. http://www.leseforum.ch/myUploadData/files/2015_1_Krelle.pdf [16.09.2021].

- Marx, P., Weber, J. & Schneider, W. (2005a): Langfristige Auswirkungen einer Förderung der phonologischen Bewusstheit bei Kindern mit Defiziten in der Sprachentwicklung. *Die Sprachheilarbeit*, 50(6), 280–285.
- Marx, P., Weber, J. & Schneider, W. (2005b): Phonologische Bewusstheit und ihre Förderung bei Kindern mit Störungen der Sprachentwicklung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37(2), 80–90.
- Mayringer, H. & Wimmer, H. (2014): *Salzburger Lese-Screening für die Klassenstufen 2–9*. Hogrefe.
- Mittelstraß, J. (1992): *Leonardo-Welt. Über Wissenschaft, Forschung und Verantwortung*. Suhrkamp.
- Neumann, A. & Mahler, I. (Hrsg.) (2014): *Empirische Methoden der Deutschdidaktik: Audio- und videografierende Unterrichtsforschung*. Schneider Hohengehren.
- Nickel-Bacon, I. (2006): Positionen der Literaturdidaktik – Methoden des Literaturunterrichts. Ein heuristischer Explikationsversuch für die empirische Grundlagenforschung. In: N. Groeben & B. Hurrelmann (Hrsg.), *Empirische Unterrichtsforschung: Literatur- und Lesedidaktik*. (S. 95–114). Juventa.
- Nix, D. (2019): Lesekompetenz systematisch und systemisch fördern. In M. Kämer-van den Boogaart & K. Spinner (Hrsg.), *Deutschunterricht in Theorie und Praxis (DTP). Band 11/1: Lese- und Literaturunterricht. Konzeptionelle und empirische Grundlagen*. (S. 165–216). Schneider Verlag.
- Pflugmacher, T. (2016): Nähe durch Abstand – Abstand durch Nähe. Deutschdidaktik als reflexive Wissenschaft. In: C. Bräuer (Hrsg.), *Denkrahmen der Deutschdidaktik: Die Identität der Disziplin in der Diskussion*. (S. 79–94). Peter Lang.
- Pohl, I. & Ulrich, W. (Hrsg.) (2011): *Wortschatzarbeit*. Schneider Hohengehren.
- Stanat, P., Pant, H. A., Böhme, K. & Richter, D. (Hrsg.) (2012): *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011*. Waxmann.
- Stanat, P., Schipolowski, S., Rjosk, C., Weirich, S. & Haag, N. (Hrsg.) (2017): *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der vierten Jahrgangsstufe*. Waxmann.
- Ulshöfer, R. (1952–57): *Methodik des Deutschunterrichts*. 3 Bände. Unterstufe. Klett.
- van den Daele, W. (1975): Autonomie contra Planung: Scheingefecht um die Grundlagenforschung? In *Wirtschaft und Wissenschaft*, 23(2), 29–32.
- Wackernagel, P. (1842): *Der Unterricht in der Muttersprache, 4. Teil des Deutschen Lesebuchs*. Liesching.
- Walter, J. (2001): *Förderung bei Lese- und Rechtschreibschwäche: Grundlagenforschung, methodische Konsequenzen, Praxisbeispiele und mediendidaktische Anregungen auf der Basis empirischer Forschungsmethoden*. Hogrefe.
- Wintersteiner, W. (2007): Die Innenwelt der Außenwelt der Innenwelt. Deutschdidaktik im Sog gesellschaftlicher Interessen. In: *Didaktik Deutsch* 22/2007, 51–70.

Britta Juska-Bacher, Martina Röthlisberger, Ladina Brugger & Christoph Zangger

Lesen im 1. Schuljahr: Die Bedeutung von phonologischer Bewusstheit, Benennungsgeschwindigkeit und Wortschatz

Die Vermittlung der Lesefähigkeit ist eine zentrale Aufgabe der Volksschule. Sie hat über das Fach Deutsch hinaus in allen Schulfächern wie auch außerschulisch große Bedeutung. Wird zudem berücksichtigt, dass die in den unteren Klassen erworbenen basalen Lesekompetenzen die wesentliche Grundlage und zugleich den limitierenden Faktor für den Aufbau hierarchiehöherer Lesekompetenzen in den oberen Primarklassen darstellen (z.B. Perfetti 1985, Rosebrock et al. 2010, Verhoeven & Van Leeuwe 2008), können die Untersuchung und Förderung der Lesekompetenz am Schulbeginn nicht genug Gewicht erhalten.

Das Forschungsprojekt „Die Entwicklung von Wortschatz und Lesen. Eine Untersuchung auf der Unterstufe“ (EnWoLe)¹ ist eben diesem frühen Bereich gewidmet, mit dem Ziel, die gegenseitige Beeinflussung von basalen Lesekompetenzen, den Kontextvariablen phonologische Bewusstheit und schnelles Benennen (von Farben, Buchstaben und Zahlen) sowie Wortschatzkompetenzen im ersten bis dritten Schuljahr zu untersuchen. Sowohl das Lesen als auch der Wortschatz werden in mehrere Teilkompetenzen untergliedert. Im Folgenden werden, nach einer kurzen Einführung ins Thema und der damit verbundenen Klärung der Forschungsfrage, die Ergebnisse der Datenerhebung im ersten Schuljahr vorgestellt.

1 | Forschungsstand hierarchieniedrige Leseprozesse

1.1 | Kontextvariablen (phonologische Bewusstheit und schnelles Benennen)

In der psychologisch geprägten Leseforschung gut untersucht ist insbesondere der Einfluss der *phonologischen Bewusstheit*. Darunter wird die Fähigkeit verstanden, die Aufmerksamkeit auf die formale Seite der Sprache zu lenken und diese zu analysieren und zu manipulieren (zu den Teilfähigkeiten siehe Muter et al. 1998, Schnitzler 2008). Die phonologische Bewusstheit gilt als zentrale Vorläuferfertigkeit des Lesens (z.B. Torgesen et al. 1994). Ihre wichtige Bedeutung für das Rekodieren als Teilbereich des Wortlesens konnten zahlreiche Studien besonders für die Anfangsphase des Schriftspracherwerbs nachweisen (für das Deutsche u.a. Ennemoser et al.

¹ Wir danken dem Schweizerischen Nationalfonds für die Förderung dieses Projekts sowie den beteiligten Lehrpersonen und Kindern ganz herzlich für ihre engagierte Mitarbeit bzw. die begeisterte und geduldige Teilnahme.

2012, Fricke et al. 2016, Juska-Bacher et al. 2016, Schneider & Näslund 1993, Stock 2009). Weniger häufig berücksichtigt wurde im deutschsprachigen Diskurs die Zugriffsgeschwindigkeit auf die Wortbedeutung, gemessen durch das schnelle Benennen oder Rapid Automated Naming (= RAN)². Das *schnelle Benennen* misst, wie viel Zeit eine Versuchsperson benötigt, eine Reihe bekannter und gleichzeitig präsentierter Stimuli (Abbildungen, Symbole, Buchstaben oder Zahlen) zu benennen. Dafür muss sie die Stimuli identifizieren, Zugriff auf die entsprechenden Wörter im mentalen Lexikon haben und diese artikulieren. Den wichtigen Einfluss des schnellen Benennens auf die Lesegeschwindigkeit ergaben u.a. Studien von Ennemoser et al. (2012) und Fricke et al. (2016). Verschiedentlich wurde in der internationalen Forschung die verstärkte Einbeziehung dieses Faktors postuliert (u.a. von Cain et al. 2004, Schmitt 2010). Es ist davon auszugehen, dass eine schnelle Zugriffsgeschwindigkeit beim Lesen eine kognitive Entlastung darstellt, die anderen Prozessen zugutekommt. Ennemoser et al. (2012) konnten zeigen, dass phonologische Bewusstheit und Zugriffsgeschwindigkeit insbesondere in den ersten zwei Schuljahren eine zentrale Bedeutung haben und danach – wenn die Texte länger und schwieriger werden – ihre Bedeutung zugunsten von Wortschatz und Grammatik immer mehr zurückgeht.

1.2 | Mündlich erworbene Sprachkompetenzen: Wortschatz

In Verbindung mit dem anfänglichen Lesen wurden im Deutschen bisher im Mündlichen erworbene Sprachkompetenzen wie der Wortschatz kaum systematisch berücksichtigt. In den USA hingegen wird Wortschatz seit der Jahrtausendwende in die großen nationalen Studien immer häufiger als wichtiger Einflussfaktor der Lesekompetenz einbezogen und seine Bedeutung bestätigt (z.B. National Center for Education Statistics 2012). Verschiedene Studien konnten einen engen Zusammenhang von Wortschatz- und Lesekompetenzen in den ersten Schuljahren belegen, es konnte aber nicht abschließend geklärt werden, wie genau die verschiedenen Teilkompetenzen von Lesen und Wortschatz zusammenhängen (z.B. Ricketts et al. 2007). Ouellette (2006) forderte daher explizit eine Aufgliederung von Wortschatz und Lesen in ihre Teilkompetenzen (siehe auch Fricke et al. 2016).

Im englischsprachigen Raum wird Wortschatz seit einigen Jahrzehnten nach Umfang und Qualität unterschieden (Überblick siehe Read 2004). Wortschatzumfang umfasst die Zahl der lexikalischen Einträge im mentalen Lexikon, die einen minimalen Bedeutungsgrad beinhalten (Anderson & Freebody 1981). Zu Wortschatzqualität wird semantisches Wissen, formales Wissen und die Vernetzung der Wörter im mentalen Lexikon gerechnet (Juska-Bacher & Jakob 2014).

Wortschatzumfang und frühes Lesen: Der Einfluss des Wortschatzumfangs auf das Lesen am Schulanfang wurde im Englischen u.a. von Ouellette (2006), Scarborough (2005) sowie Sénéchal et al. (2006), im Deutschen u.a. von Juska-Bacher et al. (2016) aufgezeigt. Die Autor*innen wiesen meist keinen direkten, sondern einen indirekten Einfluss des Wortschatzumfangs über die phonologische Bewusstheit auf das Wortlesen nach und erklärten diesen Einfluss durch die Reorganisation des mentalen Lexikons bei Erreichen eines kritischen Wortschatzumfangs (lexical restructuring model). Verhoeven et al. (2011) konnten bei niederländischen Kindern in einer Längsschnittstudie vom 1. bis 6. Schuljahr nachweisen, dass der Wortschatzumfang bei Schulbeginn das spätere Wortlesen voraussagt. Ebenfalls aus den Niederlanden

² Der Begriff „Rapid Automated Naming“ wurde von Denckla und Rudel (1974) geprägt, die den ursprünglich für Aphasie-Patienten entwickelten Test erstmals an Vorschulkindern als Lese-Prädiktor anwandten.

stammt eine Studie von De Jong & Van der Leij (1999), in der Wortschatzumfang (und -qualität) im Rahmen des Einflusses der phonologischen Bewusstheit auf das Lesen erhoben wurden und ein Zusammenhang zwischen Wortschatzumfang im Kindergarten und dem frühen Lesen nachgewiesen werden konnte. Andere Autor*innen betonten, dass Wortschatzumfang erst ab dem zweiten Schuljahr an Bedeutung gewinnt, wenn die Anforderungen vom basalen Entschlüsseln zum verstehenden Lesen übergehen (z.B. Fricke et al. 2016, Nation & Snowling 2004, Whitehurst & Lonigan 1998). Dabei ist bisher nicht geklärt, ob Wortschatzkompetenzen direkt auf das Leseverstehen wirken oder ob eher von einem beide Faktoren beeinflussenden Mittler wie der allgemeinen verbalen Kompetenz auszugehen ist.

Wortschatzqualität und frühes Lesen: Im Vergleich zum Wortschatzumfang ist der Zusammenhang von Wortschatzqualität und Lesekompetenzen noch einmal deutlich seltener untersucht. Vermutet wird ein solcher Zusammenhang von Sénéchal et al. (2006: 180), wenn diese schreiben: "the depth and organization of a child's semantic knowledge may also be predictive of reading". Der *lexical quality hypothesis* von Perfetti (2007) zufolge hat die Wortschatzqualität, i.S. präziser und flexibler Repräsentationen, einen nachweisbaren Einfluss auf das Lesen. Mit Hilfe von Regressionsanalysen konnten bspw. Roth et al. (2002) semantisches Wortwissen auf der Kindergartenstufe als Prädiktor für das Lesen in der 2. Klasse belegen (auch Snow et al. 1995). Für den deutschen Sprachraum ist die Frage des Zusammenhangs zwischen Wortschatzqualität (semantisches Wissen) und den Lesekompetenzen bisher unbeantwortet. Im Niederländischen – einer Orthografie mit einer dem Deutschen vergleichbaren Transparenz (Borgwaldt et al. 2005, Seymour et al. 2003) – wiesen De Jong & Van der Leij (1999) einen Zusammenhang zwischen Wortschatzqualität und Lesefähigkeit nach, der allerdings vom Kindergarten an eine abnehmende Tendenz zeigte.

Neben dem Einfluss sprachlicher bzw. sprachnaher Fähigkeiten sind auch einige nicht-sprachliche Fähigkeiten gut untersucht. Insbesondere die Wirksamkeit der fluiden Intelligenz, d.h. der Fähigkeit des Problemlösens mit wenig Bezug zum Wissen (vgl. Rost 2013), auf das Lesen ist gut belegt (z.B. Ramseier & Brühwiler 2003). Das Projekt *EnWoLe* setzt mit einer Untersuchung der ersten drei Schuljahre unter Berücksichtigung verschiedener Teilkompetenzen des Lesens und des Wortschatzes sowie unter Einbeziehung von phonologischer Bewusstheit und RAN genau hier an.

2 | Design und Durchführung

Im Projekt *EnWoLe* wurden im Zeitraum von Mitte März bis Ende Mai 2018 348 Kinder getestet. Nachdem vier unvollständige Datensätze von der Datenanalyse ausgeschlossen wurden, blieben Daten von 344 Kindern (197 Mädchen und 147 Jungen, Durchschnittsalter: 7;5 Jahre, Spannweite: 6;0–8;6 Jahre) aus 39 ersten Klassen übrig. Die Kinder besuchten eine Primarschule in den Deutschschweizer Kantonen Bern, Basel-Stadt und -land oder Luzern, sprachen Deutsch als Erstsprache und wiesen keine diagnostizierten Sprachentwicklungsbeeinträchtigungen auf. In den Klassen wurde mit neun verschiedenen Erstlese-Lehrmitteln gearbeitet, deren Einfluss statistisch kontrolliert wurde.

In Einzeltestungen kamen die Papierversionen folgender Tests zum Einsatz:

Lesen

Das Wort-, Satz- und Textlesen wurde mit dem ELFE II (Lenhard & Schneider 2017) gemessen. In diesem Test geht es um verstehendes Lesen auf der basalen Ebene des Wortlesens (Rekodieren, Lautsynthese und Dekodieren) sowie auf der Satz- und Textebene (Sinnentnahme, Textverständnis). Die Schüler*innen lesen in den drei Teiltests innerhalb einer vorgegebenen Zeit von drei bzw. sieben Minuten möglichst viele Einheiten: Beim Wortlesen sehen die Kinder ein Bild, zu dem vier Wortalternativen gelesen und das richtige Wort ausgewählt werden soll. Beim Satzlesen liest das Kind einen Satz, bei dem für einen Teil (meist ein Wort) fünf Alternativen vorgegeben sind, aus denen es die richtige bestimmen soll. Im Textlese-Untertest liest das Kind schließlich einen kurzen Text und eine/mehrere Fragen dazu und soll aus jeweils vier vorgegebenen Antworten die richtige ankreuzen.

Für die Erhebung der Leseflüssigkeit kam eine Kurzfassung des SLRT-II (Moll & Landerl 2014) zum Einsatz, in der die Kinder die Aufgabe haben, innerhalb einer Minute so viele der aufgelisteten wortähnlichen Pseudowörter wie möglich zu lesen.

Wortschatz

Wortschatzumfang wurde mit dem PPVT-4 (Lenhard et al. 2015) erhoben. Dieser adaptive Test misst den rezeptiven Wortschatz. Die Kinder hören dafür jeweils ein Wort und müssen ihm eine von vier im Testheft abgebildeten Bedeutungen zuordnen. Aus Zeitgründen haben wir mit einer Testhalbierung gearbeitet und nur die ungeraden³ Items verwendet.

Das Testen der **Wortschatztiefe** erfolgte mit dem WISC-IV (Petermann & Petermann 2007), Untertest 6 Wortschatztest, Wortaufgaben 7-35 (Testhalbierung, so dass nur mit 15 ungeraden Items gearbeitet wurde⁴). Der IQ-Test beinhaltet einen Untertest, in dem die Proband*innen Wörter wie „Regenschirm“ definieren sollen. Unter vorgegebenen Bedingungen darf einmal nachgefragt werden. Die Antworten werden dem Handbuch zufolge mit 0-2 Punkten bewertet. Ziel dieses Tests ist es, anhand von 15 Untertests einen IQ-Wert zu ermitteln, das Testergebnis zeigt keinen detaillierten Wert für die Wortschatztiefe. Da in *EnWoLe* aber eben diese detailliert und nach objektiven, reliablen und validen Kriterien bewertet werden sollte, wurden im Projekt für jedes Item individuelle Bewertungskriterien auf semantischer Grundlage erarbeitet und in einem Manual festgehalten. Dies war u.a. deshalb nötig, weil im Test verschiedene Wortarten (Nomen: konkret vs. abstrakt, Adjektive und Verben) vorkamen. Die Antworten der Schüler*innen wurden gemäß diesem Manual jeweils von zwei unabhängigen Raterinnen mit 0-3 Punkten bewertet. Bei Abweichungen wurden die Ratings von einer dritten Person auf die Umsetzung des Manuals überprüft, das Manual ggf. angepasst und die Abweichungen entsprechend neu bewertet. Die Übereinstimmung der Ratings war sehr gut bis ausgezeichnet, in zwei von 15 Fällen ausreichend.⁵

Wortschatzorganisation wurde mit Hilfe des WWT 6-10 (Glück 2011), Kurzfassung 1 expressiv für das Alter 7;0 bis 8;11 (die 40 Items T16-55) erhoben. Der WWT ist ein Diagnoseinstrument zur gezielten Überprüfung von semantisch-lexikalischen Fähigkeiten. Das Kind wird bildunterstützt mit vier Aufgabentypen konfrontiert: Es soll Objekte (Nomen) und Handlungen (Verben)

³ Da der PPVT-4 adaptiv ist, richtet sich der Einstieg nach dem Alter des Kindes und das letzte Set hängt von dem Wissen resp. den Antworten der Kinder ab. Aus diesem Grund ist die Anzahl N an diesem Punkt schwierig zu benennen.

⁴ Item 9 „Fahrrad“ wurde durch den Helvetismus „Velo“ ersetzt.

⁵ Alle Cohen's Kappa waren $> .69$, mit Ausnahme von „Velo“, $\kappa = .65$ und „fortziehen“, $\kappa = .59$.

benennen, Antonyme (Adjektive) und Hyperonyme (Nomen) bilden. Da der WWT Wortschatzdefizite aufdecken soll, die *EnWoLe*-Stichprobe aber aus Kindern ohne Sprachverzögerungen besteht, wurde nur die allgemeine Abrufhilfe zur Frage nach dem Zielwort verwendet, darüber hinaus auf weitere (semantische und phonologische) Abrufhilfen verzichtet. Für die hier vorgestellte Auswertung werden die Gesamtleistung sowie die Antonyme und Hyperonyme (als Bereiche der Wortschatzorganisation) verwendet.

| Variable | Testinstrument |
|-------------------------------|---|
| 1 Wortschatz | |
| 1.1 Wortschatzumfang | Lenhard et al. (2015): PPVT (Testhalbierung) |
| 1.2 Wortschatztiefe | Wechsler (2011): WISC-IV (Testhalbierung) |
| 1.3 Wortschatzorganisation | Glück (2011): WWT (Kurzfassung) |
| 2 Lesen | |
| 2.1 Wortlesen | Lenhard & Schneider (2017): ELFE II |
| 2.2 Satzlesen | Lenhard & Schneider (2017): ELFE II |
| 2.3 Textlesen | Lenhard & Schneider (2017): ELFE II |
| 2.4 Pseudowörterlesen | Moll & Landerl (2014): SLRT-II |
| 3 Kontextvariablen | |
| 3.1 phonologische Bewusstheit | Stock et al. (2017): BAKO (Teiltests Vokalersetzung, Restwortbestimmung und Phonemvertauschung) |
| 3.2 RAN | Mayer (2016): Tephobe (Teiltests Farben, Buchstaben und Zahlen) |
| 3.3 Intelligenz | Weiß, R. & Osterland, J. (2013): CFT 1-R (Teiltest Matrizen) |

Tab. 1: Untersuchte Variablen und Testinstrumente

Kontextvariablen

Phonologische Bewusstheit im engeren Sinne (nach Skowronek & Marx 1989) wurde mit dem BAKO 1-4, Stock et al. 2017), Teiltests Vokalersetzung, Restwortbestimmung und Phonemvertauschung erhoben, die Auswahl dieser Teiltests mit insgesamt 30 Items ist auf die schwierigste Stufe der phonologischen Bewusstheit, d.h. die Phonemmanipulation beschränkt, die Abfolge ist mit steigenden Anforderungen verbunden (zuerst Substitution, dann Elision und schließlich Permutation; vgl. Schnitzler 2008: 28). Wir entsprechen mit der Auswahl zugleich der Forderung von Stanovich et al. (1984), mindestens drei verschiedene Teiltests zur Bestimmung der phonologischen Bewusstheit durchzuführen.

Für den **RAN** wurde der Tephobe (Mayer 2016), Teiltests Farben, Buchstaben und Zahlen eingesetzt. Darin benennen die Kinder nach zweimaligem Üben möglichst schnell Items in zehn Zeilen mit je fünf Itemeinträgen. Die Versuchsleiter*innen stoppen die Zeit und notieren eventuelle Fehler.

Die **Intelligenz** schließlich wurde mit dem Untertest 6, Matrizen aus dem CFT 1-R (Weiß & Osterland 2013) erhoben. Aufgabe der Schüler*innen ist es hier, innerhalb von drei Minuten zu jeweils fünf figuralen Vorgaben die korrekte Fortsetzung zu finden (hoher Anteil an fluider Intelligenz). Dieser Testteil wurde ausgewählt, weil er am höchsten mit dem Gesamtergebnis korreliert (Weiß & Osterland 2013).

Eine Zusammenfassung des Studiendesigns findet sich in Tabelle 1.

3 | Ergebnisse

Zur Datenanalyse wurden R (R Core Team 2014) und SPSS 25 verwendet. Als zu erklärende Variablen wurden die Lesevariablen definiert, Wortschatz- und Kontextvariablen als Einflussgrößen auf das Lesen. Vorgängig wurden allfällige Einflüsse von Geschlecht, Alter und Lehrmittel auf die Leistungen in den Bereichen Lesen, Wortschatz und Kontextvariablen kontrolliert. In keiner der drei Variablengruppen fand sich ein signifikanter Unterschied in Bezug auf das Geschlecht oder das relative Alter. Das Lehrmittel zeigte einen schwachen Zusammenhang mit dem Wortschatzumfang, alle anderen Variablen zeigten keinen Zusammenhang mit dem Lehrmittel.⁶ Die Richtwerte der Leistungen im Lesen, Wortschatz und den Kontextvariablen sind in Tabelle 2 zusammengefasst.

| Variable | Richtwerte | | | | <i>maximal möglich</i> |
|---------------------------|------------|------------|-------------------|-----------|------------------------|
| | <i>Min</i> | <i>Max</i> | <i>Mittelwert</i> | <i>SD</i> | |
| Wortlesen | 2 | 56 | 20 | 7.9 | 75 |
| Satzlesen | 0 | 27 | 5.2 | 4 | 36 |
| Textlesen | 0 | 16 | 2.8 | 2.8 | 26 |
| Pseudowortlesen | 2 | 52 | 19.7 | 7.7 | 156 |
| Wortschatzumfang | 90 | 197 | 142.5 | 19 | n.a. ¹ |
| Wortschatztiefe | 2 | 29 | 14.5 | 4.7 | 45 |
| Wortschatzorganisation | 1 | 36 | 17.9 | 6.6 | 40 |
| Phonologische Bewusstheit | 0 | 29 | 13.1 | 7.3 | 30 |
| RAN | 1.4 | 6.1 | 3.3 | 0.7 | n.a. ² |

Anmerkungen: N = 344

¹ Die maximale Anzahl ist abhängig vom Anfangsset, das wiederum altersabhängig ist.

² Richtwert: Anzahl korrekt benannte Items pro Sekunde

Tab. 2: Deskriptive Statistik

⁶ Der Einfluss des Geschlechts wurde mit Chi-Quadrat Tests überprüft, alle $p < .16$, der Einfluss des Alters (mittelwertzentriert) mit Spearman Korrelationen, alle $p < .13$. Die Rolle des Lehrmittels wurde mit Somers d berechnet, $d = .115$, $p = .012$ für den Wortschatzumfang, für alle anderen Variablen $p < .13$.

Um die Einflüsse der einzelnen Variablen aufeinander zu prüfen, wurden in einer ersten Analyse die Korrelationen der Variablen innerhalb und über die Variablengruppen hinaus analysiert. Da die Daten mit Ausnahme der RAN-Gesamtwerte nicht normalverteilt waren, wurde ein nicht-parametrisches Verfahren gewählt. Innerhalb der Variablengruppen waren die Korrelationen hochsignifikant, alle $p < .001$. Die interne Konsistenz der Werte war dementsprechend hoch, alle Cronbachs Alpha $> .68$. Der Intelligenztest korrelierte signifikant mit allen Variablengruppen, alle $r_s > .13$, alle $p < .02$. In den folgenden Analysen wurde der CFT nicht berücksichtigt.

Die 15 Items der Variable Wortschatztiefe wurden ebenfalls auf ihre interne Konsistenz hin überprüft. Ein Item („Anomalie“) wurde ausgeschlossen, da es dafür keine Datenpunkte gab. Cronbachs Alpha blieb niedrig, $\alpha = .60$, wobei das Item „fortziehen“ die schlechteste Trennschärfe aufwies ($r_{it} = .07$). Ein Ausschluss dieses Items erhöhte Cronbachs Alpha auf $\alpha = .62$.

Im Vergleich zwischen den Variablengruppen zeigte sich, dass die Korrelation zwischen der Summe der Lesevariablen und derjenigen der Kontextvariablen hochsignifikant waren, $N = 344$, $r_s = .54$, $p < .001$ für den RAN und $r_s = .41$, $p < .001$ für die Maße der phonologischen Bewusstheit (zu den einzelnen Korrelationen siehe Tab. 3).

| Kontextvariablen | Lesevariablen | | | |
|---------------------------|------------------|------------------|------------------|------------------------|
| | <i>Wortlesen</i> | <i>Satzlesen</i> | <i>Textlesen</i> | <i>Pseudowortlesen</i> |
| RAN | .46** | .44** | .43** | .50** |
| Phonologische Bewusstheit | .36** | .46** | .38** | .34** |

Anmerkungen: N = 344; ** $p < .001$

Tab. 3: Korrelationen zwischen Lese- und Kontextvariablen

Die Summe der Lesevariablen und der Wortschatzvariablen korrelierten ebenfalls, $r_s = .16$, $p = .002$. Zu den Korrelationen der einzelnen Variablen siehe Tabelle 4.

| Wortschatzvariablen | Lesevariablen | | | |
|-------------------------------------|------------------|------------------|------------------|------------------------|
| | <i>Wortlesen</i> | <i>Satzlesen</i> | <i>Textlesen</i> | <i>Pseudowortlesen</i> |
| Wortschatzumfang | .04 | .16** | .12* | .15** |
| Wortschatzorganisation (insgesamt) | .10 | .20*** | .21*** | .14* |
| Wortschatzorganisation (Antonyme) | .07 | .19*** | .18*** | .11* |
| Wortschatzorganisation (Hyperonyme) | .12* | .18** | .18*** | .09 |
| Wortschatztiefe | .09 | .20*** | .20*** | .170** |

Anmerkungen: N = 344; * $p < .05$; ** $p < .01$; *** $p < .001$

Tab. 4: Korrelationen zwischen Lese- und Wortschatzvariablen

Die Kontextvariablen korrelierten nur teilweise mit den Wortschatzvariablen, $r_s = .193$, $p < .001$ für die phonologische Bewusstheit, aber $r_s = -.04$, $p = .465$ für die RAN-Werte (siehe Abb. 1).

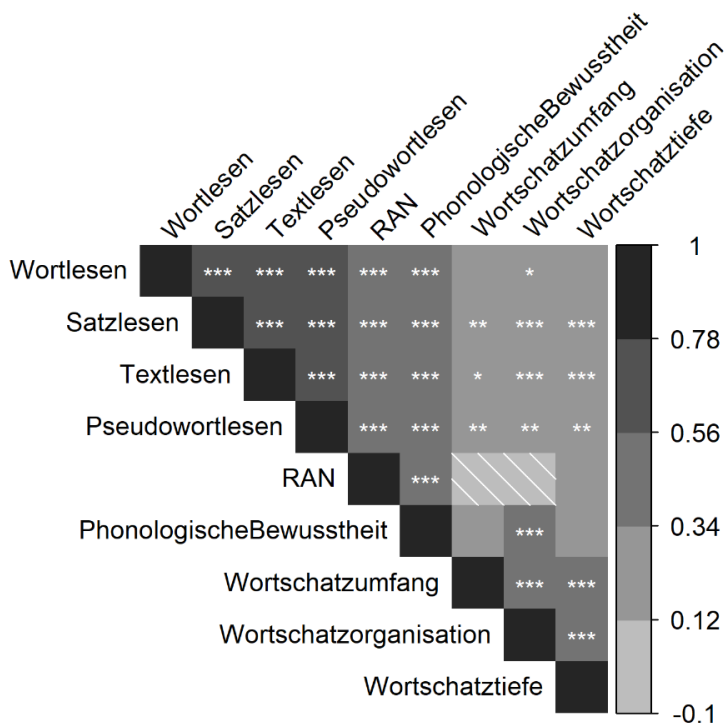


Abb. 1: Korrelationen unter den Variablen-Gruppen. * $p < .05$ ** $p < .001$; *** $p < .0001$

Zur Überprüfung, ob die Testinstrumente tatsächlich die theoretischen Konstrukte Lesen, Wortschatz und die Kontextvariablen RAN und phonologische Bewusstheit abbildeten, wurde eine explorative Faktorenanalyse verwendet (Backhaus et al. 2008). Der CFT als autonomes Konstrukt wurde nicht in der Faktorenanalyse berücksichtigt. Als Gütekriterien werden dabei einerseits Testgrößen wie das *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* (KMO) herangezogen, andererseits erfolgt diese auch visuell mittels Scree-Plots, welche die Anzahl zu extrahierender Faktoren im Hinblick auf die erklärten Varianzanteile in den Items wiedergeben. Das Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO = .811) und der Bartlett Sphärität-Test, $\chi^2(105) = 2475,61$, $p < .001$ wiesen darauf hin, dass die Struktur der Variablen mit einer Hauptkomponentenanalyse vereinfacht werden durfte (zum KMO s. Kaiser 1960, aber auch Fabrigar et al. 1999; zum Bartlett Sphärität-Test Bartlett 1950). Die Lesevariablen wurden zu einem Faktor zusammengefasst, der 78 Prozent der Varianz erklärt (zu den Ladungen der Variablen siehe Abb. 2.A.).⁷ Aufgrund des Scree-Plots und der theoretischen Struktur wurde eine Drei-Faktor-Lösung gewählt mit den Faktoren Wortschatz, RAN und phonologische Bewusstheit (siehe Abb. 2.B.). Dabei wurden 41 Prozent der Varianz durch den Wortschatz, je 29 Prozent durch den RAN und die phonologische Bewusstheit erklärt.

⁷ Die Komponentenanalyse für den Faktor Lesen wurde nicht rotiert. Für die beiden anderen Variablen Gruppen, den Wortschatz und die Kontextvariablen (RAN und phonologische Bewusstheit), wurde eine oblique Rotation (promax) gewählt, da die Komponenten-Matrix Werte über .32 enthielt.

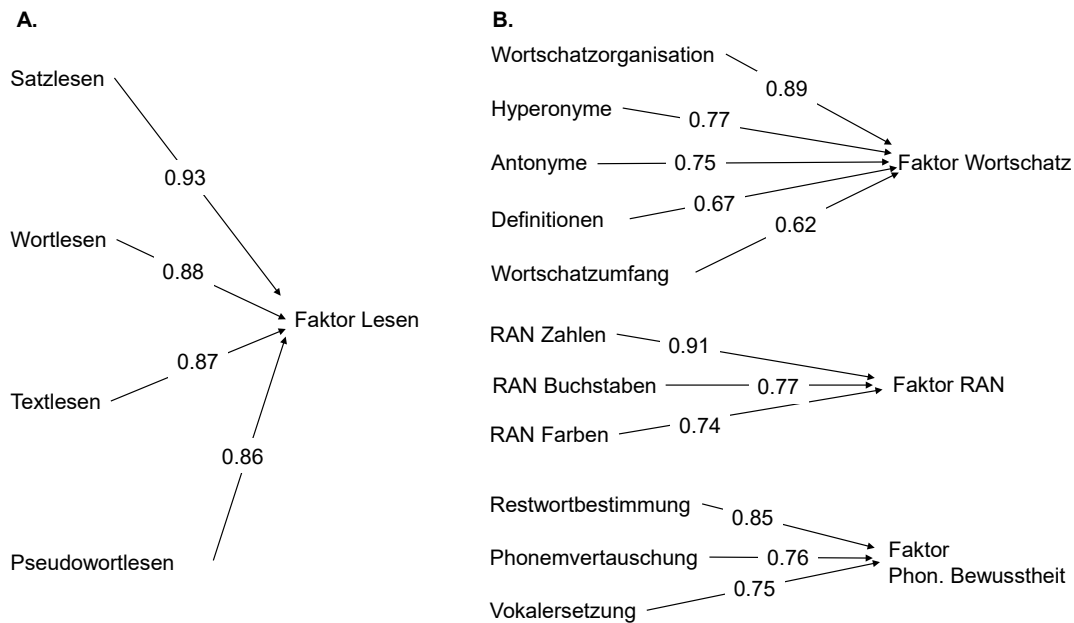


Abb. 2: Die Ladungen der Lesevariablen auf den Faktor Lesen im nicht rotierten Modell (A.). Die Ladungen der Variablen auf die drei Faktoren Wortschatz, RAN und phonologische Bewusstheit im rotierten Modell (B.).

Um die Determinanten der Lesefähigkeiten der Erstklässler*innen multivariat zu testen, wurden multiple lineare Regressionsmodelle verwendet. Um allfälligen Lehrpersoneneffekten und der hierarchischen Datenstruktur Rechnung zu tragen, wurde in einem ersten Schritt geprüft, ob ein Modell mit einem zusätzlichen zufälligen Effekt auf der Ebene der Lehrpersonen (Mixed effect models, siehe Pinheiro & Bates 2006) besser zu den Daten passt. Hierzu wurden neben den entsprechenden AIC- und BIC-Werten der Modelle – ohne und mit einem solchen Effekt – zusätzlich ein Likelihood-Ratio Test verwendet. Die nicht-signifikanten Ergebnisse des Tests wie auch die konsequent tieferen AIC- und BIC-Werte wiesen darauf hin, dass das Modell ohne Effekt letzteres im vorliegenden Fall die Daten und Zusammenhänge adäquater abbildet.

| Modell | R ² | Faktor | B | Standardfehler | β | Sig. |
|--------|----------------|-------------------|-------|----------------|-------|--------|
| 1 | 0.20 | RAN | 0.452 | 0.048 | 0.452 | < .001 |
| 2 | 0.300 | RAN | 0.370 | 0.047 | 0.369 | < .001 |
| | | phon. Bewusstheit | 0.321 | 0.047 | 0.320 | < .001 |
| 3 | 0.325 | RAN | 0.387 | 0.047 | 0.387 | < .001 |
| | | phon. Bewusstheit | 0.267 | 0.049 | 0.267 | < .001 |
| | | Wortschatz | 0.167 | 0.047 | 0.167 | < .001 |

Anmerkungen: abhängige Variable: Faktor Lesen. Der CFT leistete keinen signifikanten Beitrag zu R² und wurde nicht zum Modell hinzugefügt.

Tab. 4: Regressionskoeffizienten (B) und standardisierte Regressionskoeffizienten der Prädiktoren RAN, phonologische Bewusstheit und Wortschatz.

In der Folge wurden die Faktoren Wortschatz, RAN und phonologische Bewusstheit mit einer stufenweisen Regression auf ihre Güte als Prädiktoren für den Faktor Lesen analysiert. Der CFT wurde als manifeste Größe direkt in das Regressionsmodell aufgenommen. Das Modell mit den drei Prädiktoren Wortschatz, RAN und phonologische Bewusstheit erwies sich als signifikant, $F(3, 339) = 54.36, p < .001$. Die Resultate zeigten, dass alle drei Prädiktoren statistisch signifikant das Kriterium Lesen voraussagten.

4 | Diskussion

Die Ergebnisse der statistischen Analyse der Projektdaten zum ersten Zeitpunkt (1. Schuljahr) legen nahe, dass nicht nur die Kontextvariablen phonologische Bewusstheit und RAN einen Zusammenhang mit der Lesefähigkeit haben, sondern – wenn auch in geringerem Maße – ebenso der Wortschatz.

Zu den Kontextvariablen

Erwartungsgemäß zeigten die Kontextvariablen hochsignifikante Korrelationen mit den Variablen der Gruppe Lesen $r_s = .55$ für den RAN, $r_s = .41$ für die phonologische Bewusstheit, beide $p < .001$. Was die phonologische Bewusstheit angeht, lässt der Vergleich mit den Resultaten aus den Studien von Ennemoser et al. (2012) und Juska-Bacher et al. (2015, 2016) nur geringe Abweichungen erkennen, obwohl teilweise unterschiedliche Testinstrumente genutzt wurden und in Ennemoser et al. (2012) die phonologische Bewusstheit im Vorschulalter, in Juska-Bacher et al. (2015) am Anfang des ersten Schuljahrs, erhoben wurde. Auch in der vorliegenden Studie kann damit der hohe Einfluss der phonologischen Bewusstheit auf die frühe Lesefähigkeit nachgewiesen werden. Für die etwas schwächere Korrelation mit dem Pseudowörterlesen ist die „Natur“ der Lesemaße verantwortlich. Während der ELFE prominenter das Leseverständnis abbildet, misst der SLRT die Leseflüssigkeit. Bereits in früheren Studien korrelierte die phonologische Bewusstheit stärker mit dem Leseverständnis als mit der Lesegeschwindigkeit (z.B. Ennemoser et al. 2012, Fricke et al. 2016, für eine Übersicht siehe Pfof 2015). Zu beachten ist, dass die Daten unterschiedliche Verteilungen aufwiesen. Die rechtslastige Verteilung der Restwortbestimmung im Vergleich zu der eher linkslastigen der Phonemvertauschung und der stark linkslastigen der Vokalersetzung lässt den Schluss zu, dass der Schwierigkeitsgrad der Aufgaben unterschiedlich war. Die Vokalersetzung hatte bereits in Juska-Bacher et al. (2015) ähnliche Resultate gezeitigt. Denkbar ist, dass die Kinder trotz Übungsbeispielen Probleme mit der Aufgabenstellung hatten. Nichtsdestotrotz liegen die Mittelwerte dieser Studie im mittleren Bereich der Normwerte der jeweiligen Aufgabe (Stock, Marx & Schneider 2017).

Ähnlich wie in der Studie von Fricke et al. (2016) korrelierte der RAN hochsignifikant mit allen Lesemaßen und zeigte zudem eine hohe Prädiktion für das Lesen in der Regressionsanalyse. Im Vergleich zu den Normwerten (Mayer 2016) lagen die Kinder der vorliegenden Studie beim Benennen der Buchstaben und Zahlen leicht über dem Durchschnitt, während sie bei den Farben durchschnittlich abschnitten. Da die Stichprobe nur deutschsprachige Kinder ohne Spracherwerbsprobleme umfasste, ist dies durchaus folgerichtig. Der Vergleich zwischen den Wortschatzvariablen und den Kontextvariablen zeigte hingegen ein heterogenes Bild. So wiesen der Wortschatz und die phonologische Bewusstheit analog zu der Studie von Juska-Bacher et al.

(2016)⁸ einen Zusammenhang auf, während Wortschatz und RAN nicht miteinander korrelieren.

Zum Wortschatz

Die Korrelationen zwischen Lese- und Wortschatzmaßen sind im Hinblick auf die Wortschatzorganisation am stärksten. Ganz offensichtlich unterstützten robuste lexikalische und semantische Repräsentationen nicht nur die Dekodierung von Wörtern, sondern auch die Rekodierung, was sich in der Korrelation mit dem Pseudowortlesen manifestiert. Der Zusammenhang, der für die englische Sprache bereits von Roth et al. (2002) nachgewiesen wurde, gilt also auch für das Deutsche.

Ebenso vergleichbar ist das Verhältnis der Korrelationsstärke zwischen Wortschatzumfang und -organisation mit den Resultaten von Roth et al. (2002). Der Wortschatzumfang korreliert weniger stark mit dem Wortlesen als in der Vorgängerstudie (Juska-Bacher et al. 2016), was in den unterschiedlichen Testinstrumenten begründet sein mag. Es muss allerdings berücksichtigt werden, dass in der ersten Klasse vor allem das Wort- und das Pseudowortlesen konsistente Resultate geliefert haben, weil viele Kinder – für die 1. Klasse durchaus nicht ungewöhnlich – an komplexeren Konstrukten noch gescheitert sind. So werden die Resultate des Satz- und des Textlesens durch einige sehr gute Leser*innen sehr stark beeinflusst. Es kann davon ausgegangen werden, dass es sich dabei um Kinder handelt, die bereits dekodierend lesen. Dem heterogenen Bild wurde mit der Faktorenanalyse Rechnung getragen.

Die Wortschatztiefe zeigt ähnliche Werte wie der Wortschatzumfang, auch hier kann geschlossen werden, dass produktive semantische Kenntnisse kaum Einfluss auf die Rekodierfähigkeit haben, mit zunehmendem dekodierendem Lesen aber wichtiger werden dürften.

Bei der Betrachtung der Faktorladungen zeigt sich wie oben dargelegt, dass die Wortschatzmaße nicht homogen sind. Während die Wortschatzorganisation insgesamt auf den Faktor Wortschatz lädt, ist dies bei Wortschatzumfang und -tiefe weniger stark der Fall. Nichtsdestotrotz sind die Ladungen hoch und die Faktorbildung eindeutig. Dies gilt auch für die Kontextvariablen RAN und phonologische Bewusstheit, die eindeutig auf je einen Faktor laden. Insgesamt ist das Modell konsistent.

Die Regression zeigt, dass der Wortschatz ebenfalls ein Prädiktor für das frühe Lesen ist, wenn auch in schwächerem Ausmaß als der RAN und die phonologische Bewusstheit. Dieses Resultat weicht von denjenigen früherer Studien ab, die den Einfluss des Wortschatzes nur indirekt über die phonologische Bewusstheit nachweisen konnten (z.B. Ouellette 2006, Scarborough 2005 fürs Englische, Sénéchal 2006 fürs Französische). Dabei gilt es zu bedenken, dass es sich in diesen beiden Fällen um Sprachen mit einer intransparenten Orthografie handelt. Der Leseerwerb in intransparenten Sprachen erfordert mehr Zeit (Mann & Wimmer 2002), weswegen sich die Zusammenhänge zwischen Wortschatz und Leseverständnis erst zu einem späteren Zeitpunkt manifestieren dürften. Diese These wird gestützt durch die Resultate von Studien aus Sprachen mit transparenter Orthografie. So fanden De Jong & Van der Leij (1999) Korrelationen zwischen sowohl rezeptivem als auch produktivem Wortschatz im Kindergarten und der Lesefähigkeit in der 1. Klasse. RAN und phonologische Bewusstheit waren hochsignifikant als Prädiktoren, wie

⁸ In dieser Studie konnte anhand der Längsschnittdaten aus der 1. und 2. Klasse ein indirekter Einfluss des Wortschatzumfangs über die phonologische Bewusstheit auf das Lesen nachgewiesen werden (Juska-Bacher et al. 2016).

es aufgrund der Korrelationen und früherer Studien nicht anders zu erwarten war. Auf der anderen Seite wurde auch in der Vorgängerstudie der hier präsentierten Forschung für deutschsprachige Kinder nur ein indirekter Zusammenhang des Wortschatzes über die phonologische Bewusstheit gezeigt (Juska-Bacher et al. 2016). Die Daten des zweiten Erhebungszeitpunktes von *EnWoLe* werden zeigen, ob sich im Längsschnitt der direkte Zusammenhang von Wortschatz und Lesen bestätigen lässt.

5 | Fazit: Die Bedeutung von Kontextvariablen und Wortschatz beim Lesen im 1. Schuljahr

In der vorliegenden Studie konnte die zentrale Bedeutung der Kontextvariablen *RAN* und *phonologische Bewusstheit* für das Lesen von Erstklässler*innen für das Deutsche bestätigt werden. Dabei zeigte sich gegen Ende des Schuljahres ein stärkerer Zusammenhang zwischen *RAN* und Lesen als zwischen der phonologischen Bewusstheit und dem Lesen.

Neben diesen auch in anderen Sprachen vielfach untersuchten Kontextvariablen wurden in der beschriebenen Untersuchung auch verschiedene Wortschatzmaße, nämlich *Wortschatzumfang*, *-tiefe* und *-organisation* systematisch erfasst. Während Einigkeit darüber herrscht, dass Wortschatzkompetenzen eine zentrale Voraussetzung für hierarchiehöhere Lesekompetenzen darstellen, wurde ihre Bedeutung für das beginnende Lesen im deutschsprachigen Raum bisher kaum untersucht. Die systematische Einbeziehung der verschiedenen Wortschatzmaße zeigte aber bereits auf dieser frühen Stufe einen Zusammenhang mit dem Lesen, wenn dieser auch schwächer ausgeprägt war als bei den Kontextvariablen. Der Längsschnittverlauf dieser Studie soll den vernachlässigten Faktor Wortschatz genauer beleuchten und einen Beitrag zum Verständnis des Leseerwerbs im deutschsprachigen Raum leisten. Dieses Verständnis ist Voraussetzung für eine noch gezieltere Förderung beginnender Leser*innen, die für die gesamte schulische Laufbahn und darüber hinaus von Bedeutung ist.

Hinsichtlich der Didaktik des frühen Lesens weisen die bisherigen Projektergebnisse aus dem ersten Schuljahr in folgende Richtung:

In Bezug auf die Kontextvariable *phonologische Bewusstheit* belegen die Ergebnisse dieser Studie einmal mehr, dass ihre Förderung im Kindergarten und am Beginn der Schulzeit für den Schriftspracherwerb von zentraler Bedeutung ist. Dies wurde auch von der klinischen Linguistik erkannt, so dass bereits zahlreiche Förderinstrumente für den Kindergarten und den Schulanfang zur Verfügung stehen (z.B. Küspert & Schneider 2018, Rickli 2018).

Mit der Kontextvariable *RAN* gibt es eine weitere Kontextvariable, die im ersten Schuljahr einen ähnlich hohen Zusammenhang mit dem Lesen aufweist wie die phonologische Bewusstheit. Allerdings konnte bisher weder in der Forschung überzeugend aufgezeigt werden, dass sich *RAN*-Training erfolgreich auf die Leseleistung auswirkt (u.a. Berglez 2002, De Jong & Vrielink 2004, Mayer 2018), noch gibt es entsprechende Trainingsmaterialien für die Schule (in dieser Hinsicht noch nicht überzeugende Ausnahme: Holzer et al. 2018).

Der im Vergleich zu den Kontextvariablen zwar weniger starke, aber dennoch deutliche Zusammenhang zwischen den *Wortschatzvariablen* und dem Lesen schließlich zeigt, dass Wortschatzförderung in der Unterstufe auch unter dem Aspekt der Leseförderung und nicht nur – wie es häufig geschieht – im Hinblick auf die Förderung des Rechtschreibwortschatzes sinnvoll ist. Die

auffällige Korrelation zwischen Wortschatzorganisation und Lesen weist darauf hin, dass es keineswegs nur darum geht, dass Kinder in ihrem mentalen Lexikon über möglichst viele Wörter verfügen, sondern dass es ebenso wichtig ist, dass diese Einträge gut miteinander vernetzt sind. Die bekannten Übungen zu Wortfeldern, Hyperonymen, Synonymen und Antonymen leisten also auch einen Beitrag zur Förderung des frühen Lesens. Gleichzeitig hat – das konnte bereits in einer Vorgängerstudie gezeigt werden – der Wortschatzumfang einen positiven Einfluss auf die Entwicklung der phonologischen Bewusstheit und trägt somit indirekt zur Entwicklung basaler Lesefertigkeiten bei (Juska-Bacher et al. 2016). Systematische Wortschatzarbeit gehört also ab Schulbeginn auch im Hinblick auf die Leseförderung in den Deutschunterricht.

6 | Literatur

Verwendete Testinstrumente & genannte Trainingsprogramme

- Glück, C. W. (2011). *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige (WWT 6–10)*. Urban & Fischer.
- Holzer, J., Marischler, E., Schöfl, M. & Steinmair, G. (2018). *Vorläuferheft Lesen & Schreiben: Üben mit System*. PL & L-Pädagogische Lehr- und Lehrmittel Verlags GmbH.
- Küspert, P. & Schneider, W. (2018). *Hören, lauschen, lernen, Sprachspiele für Kinder im Vorschulalter Würzburger Trainingsprogramm zur Vorbereitung auf den Erwerb der Schriftsprache*. Vandenhoeck & Ruprecht.
- Lenhard, A., Lenhard, W., Segerer, R. & Suggate, S. (2015). *Peabody Picture Vocabulary Test – Revision IV (Deutsche Adaption)*. Pearson Assessment.
- Lenhard, W. & Schneider, W. (2006). *ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler*. Hogrefe.
- Mayer, A. (2016). *TEPHOBE. Test zur Erfassung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit*. Reinhardt.
- Moll, K. & Landerl K. (2010). *SLRT-II Lese- und Rechtschreibtest. Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT)*. Huber.
- Petermann, F. & Petermann, U. (2007). *Hamburg-Wechsler-Intelligenztest für Kinder – IV. Übersetzung und Adaptation des WISC-IV® von David Wechsler*. Huber.
- Rickli, U. (2018). *Hörschlau*. Lehrmittelverlag.
- Stock, C., Marx, P. & Schneider, W. (2003). *BAKO 1-4. Basiskompetenzen für Lese-Rechtschreibleistungen. Ein Test zur Erfassung der phonologischen Bewusstheit vom ersten bis vierten Grundschuljahr*. Beltz Test.
- Weiß, R. & Osterland, J. (2013). *CFT 1-R. Grundintelligenztest Skala 1. Revision*. Hogrefe.

Verwendete Forschungsliteratur

- Anderson, R. C. & Freebody, P. (1981). Vocabulary knowledge. In: J. T. Guthrie (Hg.), *Comprehension and Teaching: Research Reviews*, 77–117. International Reading Association.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2008). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Springer.

- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of statistical psychology*, 3(2), 77–85.
- Berglez, A. (2002). *Prävention von Lese-Rechtschreibschwierigkeiten. Ein Training der Benennungsgeschwindigkeit. Unveröff. Dissertation. Universität Bielefeld.* <https://pub.uni-bielefeld.de/download/2302377/2302380/0030.pdf> [16.09.2021].
- Borgwaldt, S. R., Hellwig, F. M. & De Groot, A. M. B. (2005). Onset entropy matters - Letter-to-phoneme mappings in seven languages. *Reading and Writing: An Interdisciplinary Journal*, 18(3), 211–229.
- Cain, K., Oakhill, J. & Bryant, P. (2004). Children's Reading Comprehension Ability: Concurrent Prediction by Working Memory, Verbal Ability, and Component Skills. *Journal of Educational Psychology*, 96(1), 31–42.
- De Jong, P. F. & Van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology*, 91(3), 450–476.
- De Jong, P.F. & Vrielink, L.O. (2004). Rapid automatic naming: Easy to measure, hard to improve. *Annals of Dyslexia*, 54(1), 65–88.
- Denckla, M. B. & Rudel, R. (1974). Rapid "automatized" naming of pictured objects, colors, letters and numbers by normal children. *Cortex*, 10(2), 186–202.
- Ennemoser, M., Marx, P., Weber, J. & Schneider, W. (2012). Spezifische Vorläuferfertigkeiten der Lesegeschwindigkeit, des Leseverständnisses und des Rechtschreibens: Evidenz aus zwei Längsschnittstudien vom Kindergarten bis zur 4. Klasse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44(2), 53–67.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods*, 4(3), 272.
- Fricke, S., Szczerbinski, M., Fox-Boyer, A. & Stackhouse, J. (2016). Preschool predictors of early literacy acquisition in German-speaking children. *Reading Research Quarterly*, 51(1), 29–53.
- Juska-Bacher, B., Beckert, C., Gyger, M., Jakob, S., & Schneider, H. (2015). Leserelevante Kompetenzen an der Schwelle vom Kindergarten zur Schule. *Zeitschrift für Literaturwissenschaft und Linguistik*, 45(2), 28–42.
- Juska-Bacher, B., Beckert, C., Stalder, U. & Schneider, H. (2016). Die Bedeutung des Wortschatzes für basale Lesekompetenzen. *Didaktik Deutsch*, 40, 20–39.
- Juska-Bacher, B. & Jakob, S. (2014). Wortschatzumfang und Wortschatzqualität und ihre Bedeutung im fortgesetzten Spracherwerb. *Zeitschrift für Angewandte Linguistik*, 61(1), 49–75.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1), 141–151.
- Mann, V. & Wimmer, H. (2002). Phoneme awareness and pathways into literacy: A comparison of German and American children. *Reading and Writing*, 15(7), 653–682.
- Mayer, A. (2018). Benennungsgeschwindigkeit und Lesen* Rapid automatized Naming (RAN) and Reading. *Forschung Sprache*, 1/2018, 20–42.
- Muter, V., Hulme, C., Snowling, M. & Taylor, S. (1998). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of Experimental Child Psychology*, 71(1), 3–27.
- Nation, K. & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*, 27(4), 342–356.

- National Center for Education Statistics (2012). *The Nations Report Card: Vocabulary Results from the 2009 and 2011 NAEP Reading Assessments: Executive Summary*. <https://nces.ed.gov/nationsreport-card/pubs/main2011/2013452.asp> [16.09.2021].
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554–566.
- Perfetti, C. (1985). *Reading ability*. Oxford University Press.
- Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, 11(4), 357–383.
- Pfost, M. (2015). Children's Phonological Awareness as a Predictor of Reading and Spelling. *Zeitschrift Für Entwicklungspsychologie und Pädagogische Psychologie*, 47(3), 123–138.
- Pinheiro, J. & Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.
- Ramseier, E. & Brühwiler, C. (2003). Herkunft, Leistung und Bildungschancen im gegliederten Bildungssystem. Vertiefte PISA-Analyse unter Einbezug der kognitiven Grundfähigkeiten. *Schweizerische Zeitschrift für Bildungswissenschaften*, 25(1), 37.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In: P. Bogaards & B. Laufer (Hrsg.), *Vocabulary in a Second Language*. (S. 209–227). John Benjamins.
- Ricketts, J., Nation, K. & Bishop, D. V. M. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, 11(3), 235–257.
- Rosebrock, C., Rieckmann, C., Nix, D. & Gold, A. (2010). Förderung der Leseflüssigkeit bei leseschwachen Zwölfjährigen. *Didaktik Deutsch*, 16(28), 33–58.
- Rost, D. H. (2013). *Interpretation und Bewertung pädagogisch-psychologischer Studien: Eine Einführung*. UTB GmbH.
- Roth, F. P., Speece, D. L. & Cooper, D. H. (2002). A Longitudinal Analysis of the Connection between Oral Language and Early Reading. *The Journal of Educational Research*, 95(5), 259–272.
- Scarborough, H. S. (2005). Developmental relationships between language and reading: Reconciling a beautiful hypothesis with some ugly facts. In: H. W. Catts & A. G. Kamhi (eds.), *The connections between language and reading disabilities*. (S. 3–24). Lawrence Erlbaum Associates Publishers.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
- Schneider, W. & Näslund, J. C. (1993). The impact of early metalinguistic competencies and memory capacity on reading and spelling in elementary school: Results of the Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC). *European Journal of Psychology of Education*, 8(3), 273.
- Schnitzler, C. D. (2008). *Phonologische Bewusstheit und Schriftspracherwerb*. Thieme.
- Sénéchal, M. (2006). Testing the Home Literacy Model: Parent Involvement in Kindergarten is Differentially Related to Grade 4 Reading Comprehension, Fluency, Spelling, and Reading for Pleasure. *Scientific Studies of Reading*, 10(1), 59–87.
- Sénéchal, M., Ouellette, G. & Rodney, D. (2006). The Misunderstood Giant: On the Predictive Role of Early Vocabulary to Future Reading. In: D. K. Dickinson & S. B. Neuman (Hrsg.): *Handbook of Early Reading Research*, 173–182.
- Seymour, P. H. K., Aro, M. & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(Pt 2), 143–174.

- Skowronek, H. & Marx, H. (1989). Die Bielefelder Längsschnittstudie zur Früherkennung von Risiken der Lese-Rechtschreibschwäche: Theoretischer Hintergrund und erste Befunde. *Heilpädagogische Forschung* 15, 38–49.
- Snow, C. E., Tabors, P. O., Nicholson, P. A. & Kurland, B. F. (1995). SHELL: Oral Language and Early Literacy Skills in Kindergarten and First-Grade Children. *Journal of Research in Childhood Education*, 10(1), 37–48.
- Stanovich, K. E., Cunningham, A. E. & Cramer, B. B. (1984). Assessing Phonological Awareness in Kindergarten Children: Issues of Task Comparability. *Journal of Experimental Child Psychology*, 38, 75–190.
- Stock, C. (2009). Der Zusammenhang zwischen phonologischer Bewusstheit und Leseleistung bei Grundschulkindern. In: W. Lenhard & W. Schneider (Hrsg.). *Diagnostik und Förderung des Leseverständnisses* (S. 81-96). Hogrefe.
- Torgesen, J. K., Wagner, R. K. & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of learning disabilities*, 27(5), 276–286.
- Verhoeven, L. & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, 22(3), 407–423.
- Verhoeven, L., Van Leeuwe, J. & Vermeer, A. (2011). Vocabulary Growth and Reading Development across the Elementary School Years. *Scientific Studies of Reading*, 15(1), 8–25.
- Whitehurst, G. J. & Lonigan, C. J. (1998). Child Development and Emergent Literacy. *Child Development*, 69(3), 848–872.

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, Giuliotto & 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 Längsschnittunter-*

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.

suchungen die Entwicklung, den Verlauf und die Ursachen von Lese- und Schreibschwierigkeiten vom Ende 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

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.

welchem Maße automatisiert, wie schnell und 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 „gefühlte“ (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 Anhieb 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.

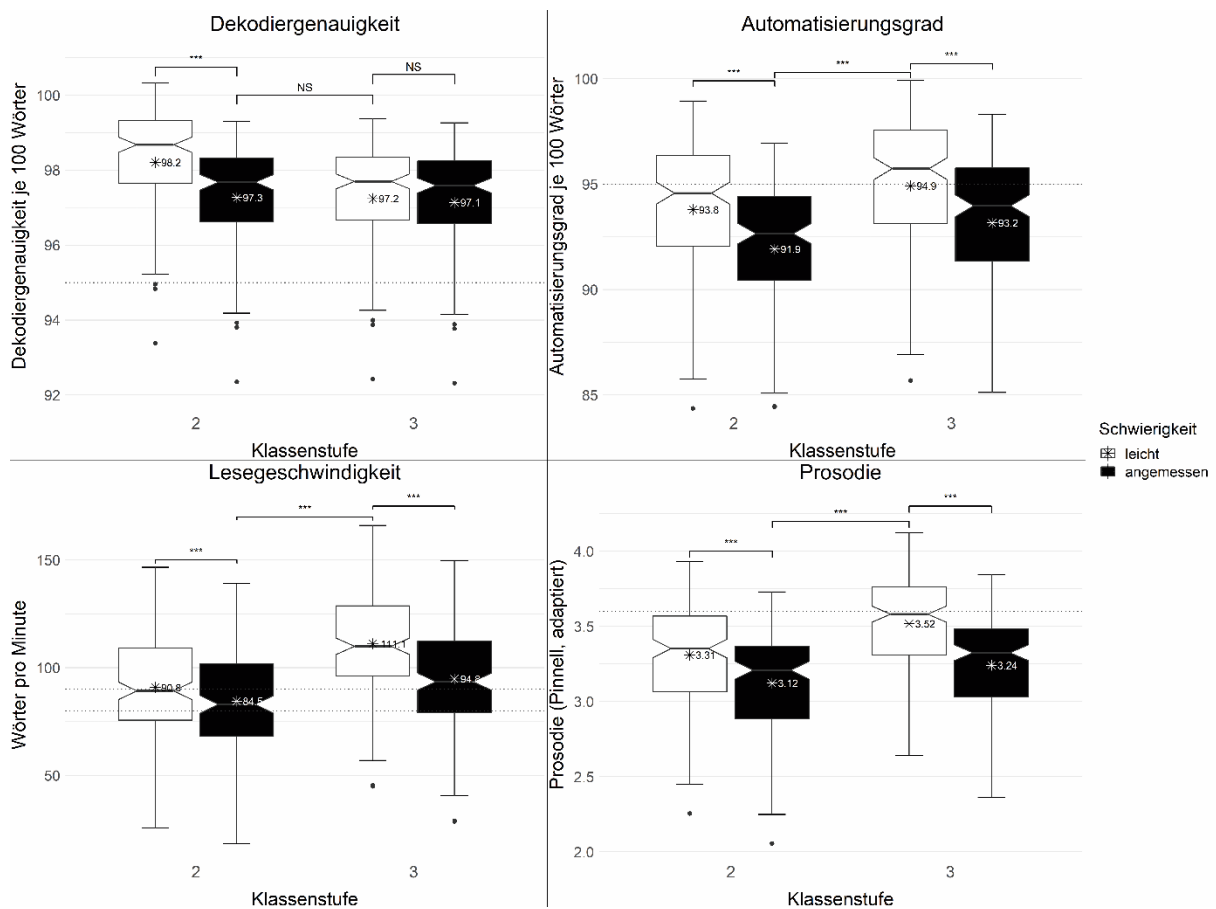


Abb. 1: Vom jeweiligen Modell vorhergesagte Werte für die Entwicklung der Dekodiergenauigkeit, des Automatisierungsgrades, der Lesegeschwindigkeit und der Prosodie von der zweiten zur dritten Klassenstufe, jeweils für einen leichten und einen altersgemäßen Text (der altersgemäße Text aus Klassenstufe 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.

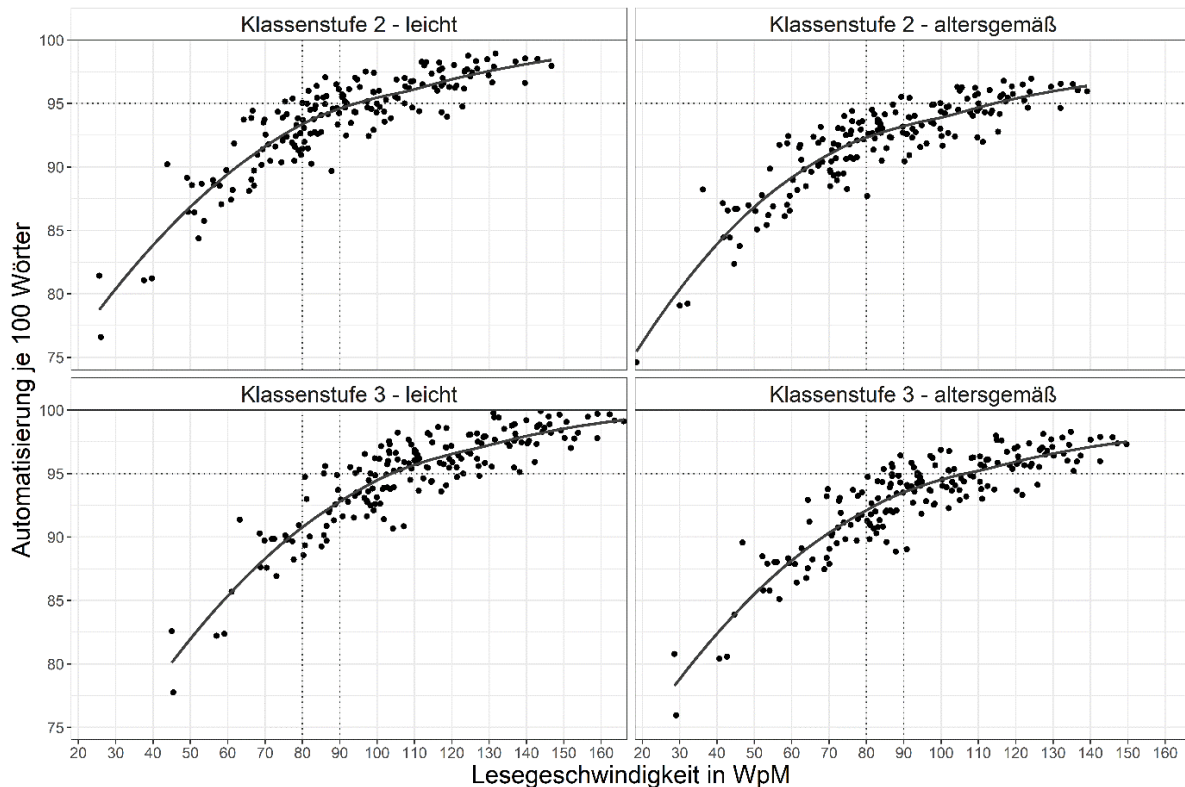


Abb. 2: Streudiagramme für Automatisierungsgrad * Lesegeschwindigkeit getrennt nach Textschwierigkeit und Klassenstufe (aus dem GLM vorhergesagte Werte). Zusätzlich eingezeichnet ist die Kurve der lokal gewichteten Streudiagrammglättung (LOWESS). Die gepunkteten Linien zeigen horizontal den Grenzwert 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.

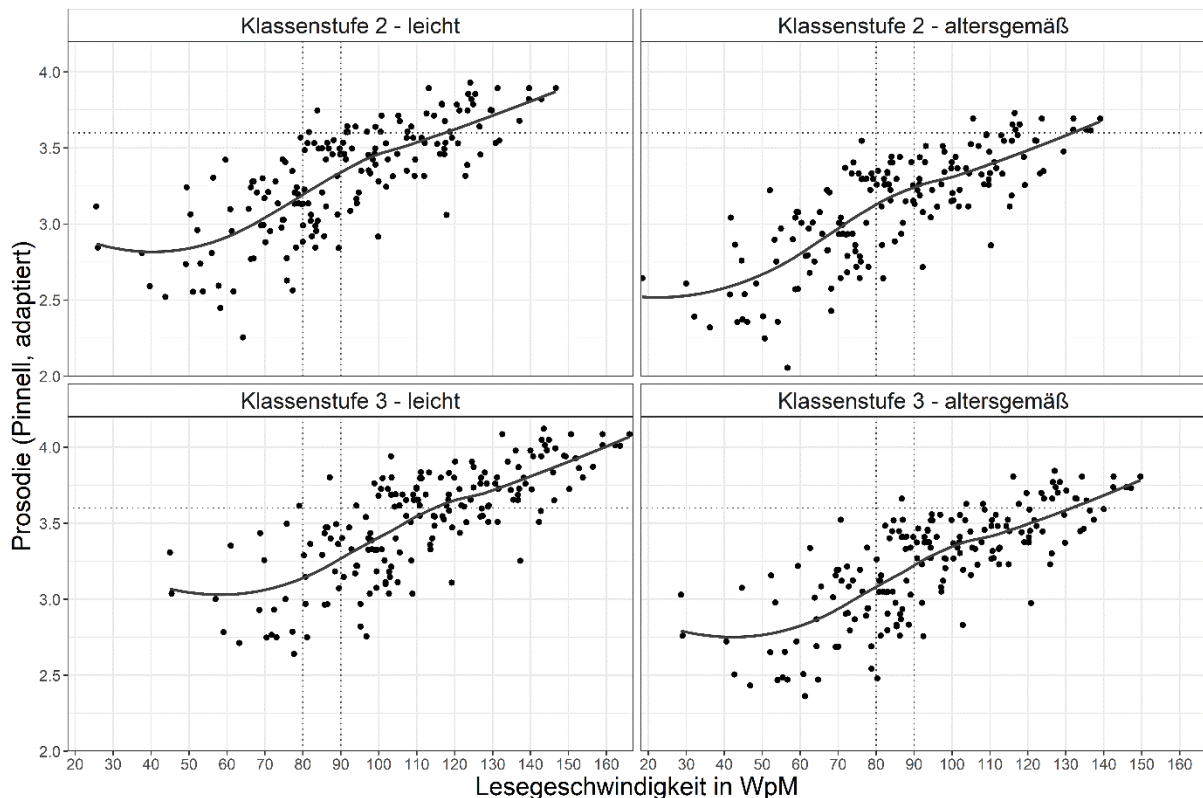


Abb. 3: Streudiagramme für Prosodie * Lesegeschwindigkeit getrennt nach Textschwierigkeit und Klassenstufe (aus dem GLM vorhergesagte Werte). Zusätzlich eingezeichnet ist die Kurve der lokal gewichteten Streudiagrammglättung (LOWESS). Die gepunkteten Linien zeigen horizontal den Grenzwert für die Prosodie (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 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)

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.

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.

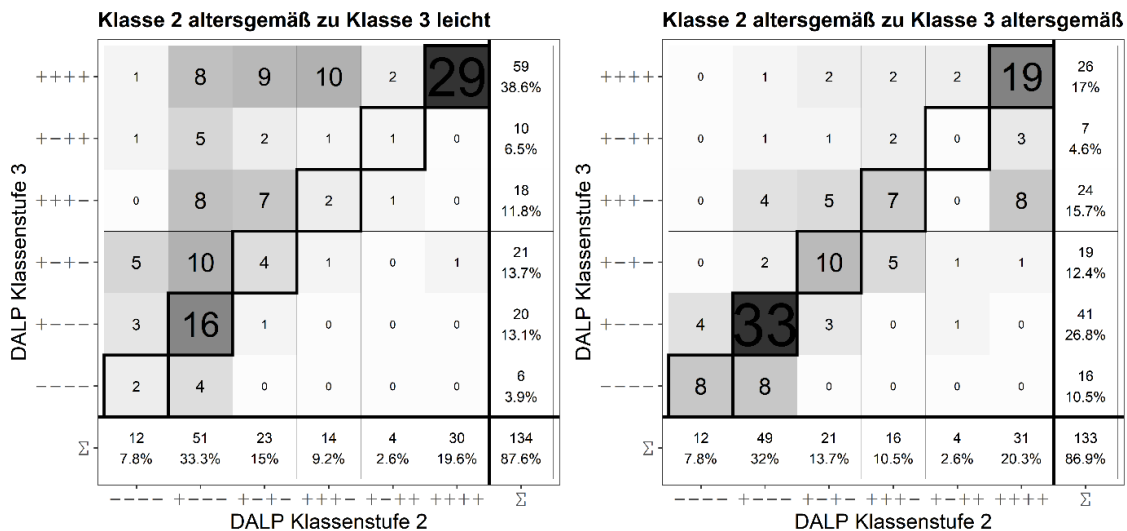


Abb. 4: Kontingenztabelle für die absoluten Häufigkeiten von Entwicklungsmustern von der Klassenstufe 2 zur Klassenstufe 3 für denselben Text (links) und jeweils altersgemäße Texte (rechts). Die Tabellen sind jeweils von unten nach oben zu lesen. In der zweiten Spalte der linken Tabelle stehen z.B. alle 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 unterschweligen 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 gezielter betrachten sollen und andererseits die vorhandenen Deckeneffekte ausgleichen. Tobit growth curve Modelle (Wang et al. 2008) erscheinen hier vielversprechend.

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.

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). Aligning 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 Rating-prozedur 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.
- Wang, L., Zhang, Z., McArdle, J. J. & Salthouse, T. A. (2008). Investigating Ceiling Effects in Longitudinal Data Analysis. *Multivariate Behavioral Research* 43(3), 476–496.
- Wolf, M. & Katzir-Cohen, T. (2001). Reading Fluency and Its Intervention. *Scientific Studies of Reading*, 5(3), 211–239.

Sabine Stephany, Markus Linnemann, Evghenia Goltsev & Michael Becker-Mrotzek

Prosodische Aspekte der Leseflüssigkeit als Indikator für Lesekompetenz

Analysen mithilfe prosodischer Lupenstellen

1 | Einleitung

In der deutschen Lesedidaktik hat die Leseflüssigkeit in der letzten Dekade mehr und mehr an Prominenz gewonnen. Während Verfahren ihrer Förderung, wie z.B. die Lautleseverfahren, allen voran die Lautlesetandems, mittlerweile gut untersucht sind (u.a. Chard, Vaughn & Tyler 2002; Lauer-Schmaltz, Rosebrock & Gold, 2014; Nix 2011; Rosebrock et al. 2010) sind diagnostische Verfahren, die durch das Bereitstellen von Informationen über die individuelle Ausprägung der Leseflüssigkeit eine adäquate Förderung erst ermöglichen, noch wenig zugänglich (vgl. hierzu auch den Beitrag von Helm in diesem Band). Das Erheben der Leseflüssigkeit liegt zumeist in der Hand der Lehrkraft und geschieht i.d.R. als informelle, ganzheitliche Beobachtung während des Lesens im laufenden Unterricht. Aus einer diagnostischen Perspektive wäre eine präzisere und differenzierte Messung aller Aspekte der Leseflüssigkeit notwendig. Vor allem die reine Lesegeschwindigkeit lässt sich vergleichsweise gut erheben, auch wenn kein standardisiertes Instrument vorliegt. Leseflüssigkeit beinhaltet neben dem schnellen Lesen jedoch auch die Komponente des sinngestaltenden oder prosodischen Lesens. Dieses zu erheben fällt selbst in Forschungsprojekten aufgrund einer fehlenden einheitlichen Konzeption und fehlender Instrumente schwer, noch problematischer ist es für die Diagnostik in der Schule, ein adäquates Instrument zu finden. Ziel der vorgestellten empirischen Studie ist es daher, einen Beitrag zu zweierlei Aspekten zu leisten. Zum einen geht es darum, eine auf ‚Lupenstellen‘ abzielende Verfahrensweise zur Erstellung eines diagnostischen Instrumentes vorzustellen und zu eruieren, ob eine Weiterentwicklung lohnenswert ist, um ein für den Schulgebrauch handhabbares Instrument zu entwickeln. Zum anderen soll untersucht werden, inwiefern prosodische Aspekte der Leseflüssigkeit ein Indikator für Leseverstehen sein können. Denn das prosodische Lesen als eine Komponente der Leseflüssigkeit sollte nicht isoliert betrachtet werden. Für schulischen Erfolg und gesellschaftliche Teilhabe ist letztlich nicht das flüssige Lesen entscheidend, sondern das Verstehen von Texten. Leseflüssigkeit und damit auch das prosodische Lesen sollte daher vor dem Hintergrund seiner Relevanz für das Textverstehen betrachtet werden. Ein besseres Verständnis der Zusammenhänge zwischen dem Verstehen von Texten und der Leseflüssigkeit, insbesondere des prosodischen Lesens, ist jedoch ein weiteres Desiderat der Leseforschung.

Im Folgenden wird zunächst das Konstrukt Leseflüssigkeit dargelegt, bevor in den darauffolgenden Kapiteln insbesondere auf die Dimension des prosodischen Lesens eingegangen wird. Die

sich anschließende empirische Studie basiert auf Daten, die im BMBF-Forschungsprojekt „Lese- und Schreibflüssigkeit – Konzeption, Diagnostik, Förderung“¹ entstanden sind, und zeigt auf, wie die oben genannte Verfahrensweise aussieht und welche Eigenschaften ein so entwickeltes Instrument hat, bevor der Text mit einer Diskussion der Ergebnisse abschließt.

2 | Leseflüssigkeit

Textverstehen ist ein komplexer Vorgang, bei dem Leserinnen und Leser auf Grundlage ihres eigenen Wissens neue Informationen aus einem Text verarbeiten, sodass im besten Falle ein kohärentes mentales Modell des gelesenen Textes entsteht (Kintsch 1998). Die komplexen Prozesse des Textverstehens erfordern ausreichend kognitive Ressourcen. Nur wenn basale Lese- prozesse wie das Dekodieren oder der Aufbau lokaler Bezüge innerhalb und zwischen Sätzen mühelos und ohne zu stocken erfolgen, sind genügend Arbeitsgedächtniskapazitäten vorhanden für hierarchiehöhere Prozesse zum Aufbau eines mentalen Modells (Just & Carpenter 1992; LaBerge & Samuels 1974; Samuels 1994). Flüssiges Lesen ist somit unabdingbar für verstehen- des Lesen. Es bildet eine Brücke zwischen den basalen Prozessen des Dekodierens und den komplexeren Prozessen des Textverstehens (Kuhn & Stahl 2003; Rosebrock & Nix 2006) wie das Ziehen von Inferenzen und das Herstellen lokaler Bezüge.

Leseflüssigkeit lässt sich als dreidimensionales Konstrukt darstellen, das Kuhn, Schwanenflugel und Meisinger (2010: 240) folgendermaßen definieren:

Fluency combines accuracy, automaticity, and oral reading prosody, which, taken together, facilitate the reader's construction of meaning. It is demonstrated during oral reading through ease of word recognition, appropriate pacing, phrasing, and intonation. It is a factor in both oral and silent reading that can limit or support comprehension.

Flüssiges Lesen meint somit sowohl eine akkurate Wortdekodierung, d.h. die Leserin oder der Leser ist in der Lage, Wörter zu erkennen und auf Graphem- bzw. Phonemebene größtenteils fehlerfrei wiederzugeben, als auch eine weitgehende Automatisierung der basalen Prozesse. Letzteres bedeutet, dass kognitive Prozesse wie das Dekodieren von Wörtern oder Wortgruppen schnell und mühelos, d.h. ohne größeren kognitiven Aufwand ausgeführt werden (Schneider & Shiffrin 1977; Shiffrin & Schneider 1977). Akkuratheit und Automatisierung lassen sich auch als Lesegeschwindigkeit zusammenfassen. Eine geringe Lesegeschwindigkeit zeigt sich am langsamen und mühevollen Dekodieren einzelner Wörter und einer vermehrten Fehleranfälligkeit auf der Silben- und Wortebene. Bei Leseanfängern lässt sich gut beobachten, dass sie am Ende eines Satzes wieder vergessen haben, was sie zu Beginn des Satzes gelesen haben, weil das Dekodieren der einzelnen Wörter im Satz noch viele kognitive Ressourcen benötigt und somit das Verstehen des Textes erschwert wird.

Lesegeschwindigkeit ist zwar eine notwendige, aber keine hinreichende Bedingung für flüssiges Lesen. Neben der Akkuratheit und der Automatisierung spielt für das spätere Leseverstehen auch das prosodische bzw. sinngestaltende Lesen eine wichtige Rolle als Teilaspekt der Leseflüssigkeit. Während des Leseprozesses muss nicht nur das Dekodieren von Wörtern, sondern auch das Erlesen größerer Einheiten in den Blick genommen werden. Es müssen sowohl

¹ Das durch das BMBF-geförderte Forschungs- und Entwicklungsprojekt „Lese- und Schreibflüssigkeit – Konzeption, Diagnostik, Förderung“ (FluLeS) wurde von 2015-2019 an der Universität zu Köln am Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache durchgeführt (Projektleitung Jun.-Prof. Dr. Markus Linnemann, Prof. Dr. Michael Becker-Mrotzek, Prof. Dr. Hans-Joachim Roth, Prof. in Dr. Hilde Haider).

Wörter zu sinnvollen Phrasen zusammengefasst und morpho-syntaktische Hinweise im Text zur Segmentierung genutzt als auch angemessen betont, rhythmisch und mit sinnvoller Intonationskurve gelesen werden (Chafe 1988). Erst durch das sinngestaltende Lesen verschiebt sich somit der Fokus vom Wortlesen hin zum Satz- oder Textlesen. Kuhn, Schwanenflugel und Meisinger (2010) sprechen bei der Prosodie entsprechend von einer „critical component of reading fluency“ (ebd.: 233). Groen, Veenendaal und Verhoeven (2019) weisen dem prosodischen Lesen sogar eine wichtigere Bedeutung für das Textverstehen zu als dem Dekodieren: „The construction of meaning seems more closely tied to text reading prosody than to decoding efficiency, at least, when children have mastered automaticity in reading“ (ebd.: 16).

Trotz seiner Relevanz steht das prosodische Lesen bisher kaum im Fokus von Leseforschung, Unterrichtspraxis und Diagnostik. Im Folgenden wird mit dem Ziel, ein praxistaugliches Diagnoseinstrument zum prosodischen Lesen zu entwickeln, genauer auf die prosodischen Aspekte der Leseflüssigkeit eingegangen, die operationalisierbar sind und entwicklungs sensitiv einen Beitrag zum Textverstehen liefern.

3 | Prosodische Aspekte der Leseflüssigkeit

3.1 | Prosodie

Grundsätzlich findet sich in der Forschung eine Vielfalt an Definitionen und Bezeichnungen des Phänomens ‚Prosodie‘. Zu den am häufigsten verwendeten Begriffen gehören ‚Prosodie‘ und der Sammelbegriff ‚Suprasegmentalia‘. Beide werden häufig synonym oder aber mit unterschiedlichen Denotationen verwendet (Bose et al. 2016; Noak 2016). An dieser Stelle wird nicht vertiefend auf die entsprechenden Argumentationen und Diskurse eingegangen, sondern eine dieser Studie zugrunde gelegte Arbeitsdefinition vorgestellt. Dabei wird, analog zu der in der Leseforschung üblichen Bezeichnungsweise, der Begriff ‚Prosodie‘ gewählt.

In Anlehnung an Spillmann (2000) wird mit Prosodie ein Komplex von Merkmalen bezeichnet, der einzelnen Lauten übergeordnet ist. Diese Merkmale sind an sprachliche Einheiten – Silbe, Wort, Phrase und Satz – gebunden und erfüllen alleinstehend oder miteinander verknüpft bestimmte Funktionen in mündlichen Äußerungen (vgl. hierzu auch Bose et al. 2016; Bußmann 2002; Dahmen & Weth 2017; Noak 2016). Der Umfang dieses Komplexes wird unterschiedlich bestimmt. So werden u.a. Lautheit, Dauer, Sprechgeschwindigkeit, Sprechspannung sowie Stimmqualität und Stimmausdruck dazu gezählt. Nach Ott (2012) und Bose et al. (2016) kann zwischen der linguistischen, der emotionalen und der pragmatischen Funktion der Prosodie unterschieden werden. Zur linguistischen Funktion gehört z.B. die Fokussierung durch Betonung und Strukturierung durch Pausen, die emotionale Funktion umfasst das Ausdrücken von Gefühlen und Haltungen und die pragmatische Funktion z.B. indirekt geäußerte Illokutionen.

Im Kontext der Leseflüssigkeit und ihrer Diagnostik steht die linguistische Funktion der Prosodie mit ihren Merkmalen Satzintonation (Interpunktion), Satzbetonung (Fokus im Satz) und Pausen im Vordergrund. Auf diese wird im Folgenden genauer eingegangen.

- Mit der Intonation ist der Tonhöhenverlauf bzw. der Verlauf der Sprechmelodie innerhalb eines Satzes bzw. einer Äußerung gemeint. Sie erfüllt im Deutschen eine syntaktische Funktion, z.B. die Kennzeichnung des Satzmodus. Es wird dabei zwischen unterschiedlichen Kon-

turen wie fallend oder steigend und ihrem jeweiligen Verlauf in einer Äußerung unterscheiden. So markiert die satzschließende, fallende Intonation eine Aussage – *sie spielt* ↓ – und eine steigende Intonation eine nicht abgeschlossene Einheit, wie z.B. eine Frage – *sie spielt* ↑? (Noak 2016; Spillmann 2000).

- Unter Betonung oder Akzent wird eine Hervorhebung oder Prominenz eines Abschnitts des Sprachsignals aufgefasst. Dies kann u.a. durch Intensität (Lautstärke) und durch Dauer erreicht werden (Ramers 2007). Mit Hilfe des Satzaketes, der stärker ist als der Wortakzent, wird dem Hörenden im Deutschen auf syntaktischer Ebene mitgeteilt, welche Komponente im Satz eine besondere Relevanz trägt bzw. im Fokus steht (Noak 2016), z.B. ‚SIE kauft das Buch‘ (wichtig ist hier die handelnde Person), ‚sie KAUFF das Buch‘ (wichtig ist hier die Handlung) oder ‚sie kauft DAS Buch/das BUCH‘ (wichtig ist hier die Art des Objektes oder das Objekt).
- Die Pause oder auch Unterbrechung beim Sprechen oder Lesen besitzt im Deutschen ebenfalls eine syntaktische Funktion. Sie wirkt segmentierend und informationssteuernd (Spillmann 2000). Mit ihr werden u.a. Phrasengrenzen gekennzeichnet und es kann ebenfalls Bedeutung unterschieden werden, z.B. ‚Er /+/ dachte sie /+/ wird kommen‘ vs. ‚Er dachte /+/ sie wird kommen‘ (Er, dachte sie, wird kommen. vs. Er dachte, sie wird kommen.) (ebd.: 71). Dabei ist jedoch wichtig zu berücksichtigen, dass Pausen ebenfalls auf anderen Ebenen wie der Wort- oder Textebene vorkommen, und dabei u.a. durch Länge unterschieden werden, sowie ebenfalls als nicht geplante Atempausen – als physiologische Gegebenheit – auftreten können (Peters 2012). Zugleich können Pausen auch rhetorisch-stilistische Funktionen haben und z.B. eine besondere Bedeutung des gerade Geäußerten vermitteln (Institut für Deutsche Sprache 2013).

3.2 | Prosodie und Lesen

Prosodie und Lesen sind eng miteinander verbunden (Koriat, Greenberg & Kreiner 2002; Stolterfoht et al. 2007). Ergert (2012) spricht bei ausgebauten prosodischen Kenntnissen sogar von einer „essentiellen Voraussetzung“ (ebd.: 43) für das Lesenlernen. Denn beim Lesen geht es um wesentlich komplexere Vorgänge als nur das Erlesen bzw. Dekodieren von Graphem- oder Phonemfolgen. Hier sind verstärkt Kenntnisse von prosodischen Regularitäten vonnöten. Zu solchen gehören u.a. das Wissen um die schriftliche Markierung der Wortgruppen: Nominalgruppen werden im Deutschen mit einem groß geschriebenen Wort abgeschlossen. Geübte Lesende können mit diesen visuellen Hinweisen schnell die Realisierungsregularitäten – Fokussierung des letzten Wortes sowie fallende Intonation und Pausierung – verbinden (ebd.). Dieser Prozess wird als phonologische Kodierung (*phonological coding*, Pollatsek, Rayner & Lee 2000) bezeichnet. Ergert (2012) weist ferner darauf hin, dass dieses Wissen zum einen das fehlerfreie Erlesen erleichtert, zum anderen kann sich die erfolgreiche Produktion und Rezeption positiv auf die Verarbeitung von sprachlichen Einheiten und den weiteren Spracherwerb auswirken. Dies wird für die Verarbeitung auf syntaktischer Ebene u.a. in den Studien von Fodor (2002) und Steinhauer (2003) gezeigt. Basierend auf orthografischen, syntaktischen sowie auch semantischen und pragmatischen Indikatoren im Geschriebenen manifestiert sich prosodisches Lesen u.a. in Form von angemessener Betonung, Intonation, Pausengestaltung und Rhythmus. Rasinski (2004: 46) nimmt an, dass erst durch eine solche Strukturierung und Zerlegung in kleinere kohärente Sinneinheiten ein Text und seine Bedeutung erschlossen wird:

The reader must parse the text into syntactically and semantically appropriate units. If readers read quickly and accurately but with no expression in their voices, if they place equal emphasis on every word and have no sense of phrasing, and if they ignore most punctuation, blowing through periods and other markers that indicate pauses, then it is unlikely that they will fully understand the text.

Auch wenn das genaue Zusammenwirken von Prosodie und Textverstehen noch ein Desiderat ist, zeigen sich in den wenigen vorhandenen Studien positive Effekte des prosodischen Lesens auf das Textverstehen (Groen, Veenendaal & Verhoeven 2019; Klauda & Guthrie 2008; Paige et al. 2017; Veenendaal, Groen und Verhoeven 2015). Groen, Veenendaal und Verhoeven (2019) können belegen, dass diese Zusammenhänge auch bestehen bleiben, wenn die Dekodierfähigkeit konstant gehalten wird. Linnemann et al. (in Review) zeigen einen seriellen Effekt von Lesegeschwindigkeit, Prosodie und Textverstehen: Eine angemessene Lesegeschwindigkeit ist Voraussetzung für prosodisches Lesen, das wiederum einen starken Einfluss auf das Textverstehen hat. Weitere Studien zeigen, dass prosodisches Lesen ein wichtiger Prädiktor für zukünftige Leseleistungen ist. So sagt das prosodische Lesen im ersten Schuljahr beispielsweise das Verstehen von Texten im zweiten Schuljahr voraus (Kuhn, Schwanenflugel & Meisinger 2010).

Damit kann prosodisches, sinngestaltendes Lesen einerseits als Indikator, zugleich aber auch als Voraussetzung des flüssigen, verstehenden Lesens angesehen werden. In diesen beiden Funktionen sollte es sowohl im Rahmen der Diagnostik als auch im Rahmen von Sprachbildung und -förderung berücksichtigt werden (Ergert 2012; Richter 2007).

3.3 | Entwicklung prosodischer Fertigkeiten

Wie sich das prosodische Lesen im Zusammenspiel mit anderen Aspekten des Leseprozesses entwickelt, ist bisher ein Desiderat. Die Forschung zur Entwicklung der Prosodie bei Kindern befasst sich kaum mit dem Lesen, sondern meist mit dem mündlichen kommunikativen Sprachgebrauch. Hier zeigt sich, dass Kinder schon sehr früh sensibel für prosodische Elemente sind und prosodische Muster nutzen. Im Rahmen des erstsprachlichen Erwerbs wird die Verarbeitung von prosodischen Merkmalen als eine der frühesten Kompetenzen betrachtet, die bereits bei Neugeborenen bzw. sogar vor der Geburt vorhanden ist (Schröder & Höhle 2012). Erste Vorkommen der Rhythmus- und Intonationsproduktion werden ab ca. einem Jahr beobachtet und entwickeln sich graduell weiter (Prieto & Esteve-Gibert 2018). Diese frühe prosodische Sensibilität scheint wichtig für die Entwicklung von phonologischer Bewusstheit und letztlich auch für akkurates Lesen zu sein (Paige et al. 2017).

Beim Lesenlernen können Kinder nicht mehr auf prosodische Hinweise zurückgreifen, die im mündlichen Sprachgebrauch vorliegen. Sie müssen lernen, die syntaktische Struktur von geschriebenen Sätzen zu erkennen und dafür Hinweise wie grafische Signale (Interpunktion) oder morphologische Marker zu nutzen. Schreiber (1991) nimmt an, dass Kinder, bevor sie lernen, diese Hinweise zu nutzen, langsam, unbetont und eher Wort-für-Wort lesen. Im Regelunterricht ist dies vor allem in der ersten Klasse der Fall, bevor sich mit fortschreitender Automatisierung prosodisches Lesen einstellt. Paige et al. (2017) konnten in einer Studie mit Erst-, Zweit- und Drittklässlern zeigen, dass prosodisches Lesen im ersten Schuljahr den größten Anstieg verzeichnet und in der dritten Klasse stagniert. Trotz dieses Anstiegs gibt es Hinweise, dass der Einfluss des prosodischen Lesens auf das Textverstehen zu Beginn der Grundschulzeit geringer

ist als der des Wortdekodierens (Miller & Schwanenflugel 2008). Denn Schülerinnen und Schüler müssen zunächst noch viele kognitive Ressourcen für das Dekodieren aufbringen und richten ihre Aufmerksamkeit noch nicht auf syntaktische Strukturen. Über die dritte Klasse hinaus konnten Sappok, Linnemann und Stephany (2020) in einer echten Longitudinalstudie mit Schülerinnen und Schülern der dritten bis siebten Jahrgangsstufe zeigen, dass das prosodische Lesen im Mittel *bis* zur siebten Klasse ansteigt und in der dritten und vierten Klasse auch ein substantieller Zusammenhang mit dem Textverstehen vorliegt. Ab der fünften Klasse nimmt dieser Zusammenhang wieder leicht ab, möglicherweise weil prosodische Lesefertigkeiten im Mittel soweit erworben sind, dass der unterscheidende Einfluss auf das Textverstehen nachlässt. Darüber hinaus zeigen die Autoren und die Autorin (ebd.) anhand prototypischer Lernverläufe, dass es eine substantielle Gruppe von Schülerinnen und Schülern gibt, deren prosodische Lesefertigkeiten bis in die Sekundarstufe auf niedrigem Niveau stagnieren.

Insgesamt zeigt die Forschungslage, dass prosodische Fertigkeiten durchgängig wichtig sind, zunächst um Muster zu erkennen und so die Sprache zu lernen (Ergert 2012), im weiteren Verlauf der Kindheit dann auch für das Lesenlernen. Aber es zeigt sich auch, dass die Prosodieentwicklung im Kontext des Lesenlernens heterogen verläuft. Dies macht deutlich, wie wichtig auch im Bereich des prosodischen Lesens eine differenzierte Diagnostik ist, um potentielle Risikoschülerinnen und -schüler frühzeitig erkennen und fördern zu können.

4 | Messung und Diagnostik von Leseflüssigkeit

Eine valide, objektive und zuverlässige Lesediagnostik ist eine notwendige Vorbedingung für eine gezielte Lesedidaktik und Leseförderung. Dabei geht es in der Regel um mehr als das Testen eines engen Konstruktes wie z.B. des Dekodierens. Für eine adäquate Lesedidaktik und -förderung sind nicht alleine ein Lesetest und ein Förderprogramm notwendig, denn im Unterricht sind viele Gelingensbedingungen und personale Faktoren der Schülerinnen und Schüler mit zu berücksichtigen: z.B. literare und literale Erfahrungen, der jeweilige Stand der Lesekompetenz als Anknüpfungspunkt für folgendes Lernen sowie Vorwissen über bestimmte Leseinhalte. Eine umfassende, unterrichtsintegrierte Lesediagnostik beinhaltet alltägliche Lernbeobachtung genauso wie punktuell eingesetzte umfangreiche, standardisierte Tests. Neben diesen Verfahren können im Unterricht eingesetzte (halb-)standardisierte Verfahren nutzbringend sein, um z.B. blinde Flecken hinsichtlich bestimmter, engumrissener Fertigkeiten bei Schülerinnen und Schülern schnell sichtbar zu machen – vorausgesetzt, ein solches diagnostisches Instrument lässt sich ökonomisch in den Unterricht einbeziehen. Zu diesen Verfahren gehören mit ihren je spezifischen Herausforderungen die Ansätze zur Messung der verschiedenen Dimensionen der Leseflüssigkeit. Hierzu werden zumeist verschiedene Skalen genutzt.

- *Akkuratheit* wird mit Hilfe eines von der Probandin oder dem Probanden vorgelesenen Texts als „richtig gelesene Wörter“ bezogen auf die Gesamtzahl der gelesenen Wörter in einem Text gemessen, wobei Selbstkorrekturen erlaubt sind. Es ist gleichermaßen möglich, die Messung bereits beim ersten Lesen (*prima vista*) oder nach mehrfachem, meist leisem Lesen, durchzuführen.
- Die *Automatisierung* wird zumeist als „richtig gelesene Wörter pro Zeiteinheit“ gemessen.
- Für die Messung des *prosodischen Lesens* liegen im angloamerikanischen Raum mehrere Skalen vor, die jeweils verschiedene Schwerpunkte hinsichtlich der linguistischen Dimension der

Prosodie setzen. Die Leseflüssigkeitsskala des US-amerikanischen *National Assessment of Educational Progress (NAEP)* (Daane et al. 2005) ist vierfach abgestuft und fokussiert vor allem das phrasierte Lesen: Während des Leseprozesses wird die richtige Gruppierung von Wörtern in größere, aussagekräftige Phrasen beurteilt (s. Abbildung 1). Auch wenn diese Skala in den Studien, in denen sie eingesetzt wurde, gute Kennwerte hinsichtlich ihrer Testgüte besitzt, ist sie inhaltlich nicht sehr stringent konstruiert und vermischt in ihren Abstufungen verschiedene prosodische Merkmale (z.B. Phrasierung, Betonung). Als flüssige Leserin bzw. flüssiger Leser gilt, wer drei oder vier Punkte erlangt hat.

| | | |
|------------------|----------------|--|
| Fluent | Level 4 | Reads primarily in larger, meaningful phrase groups. Although some regressions, repetitions, and deviations from text may be present, these do not appear to detract from the overall structure of the story. Preservation of the author's syntax is consistent. Some or most of the story is read with expressive interpretation. |
| | Level 3 | Reads primarily in three- or four-word phrase groups. Some small groupings may be present. However, the majority of phrasing seems appropriate and preserves the syntax of the author. Little or no expressive interpretation is present. |
| Nonfluent | Level 2 | Reads primarily in two-word phrases with some three- or four-word groupings. Some word-by-word reading may be present. Word groupings may seem awkward and unrelated to larger context of sentence or passage. |
| | Level 1 | Reads primarily word-by-word. Occasional two-word or three-word phrases may occur – but these are infrequent and/or they do not preserve meaningful syntax. |

Abb. 1: NAEP-Skala der Leseflüssigkeit (Daane et al. 2005: 28).

Etwas ausdifferenzierter und somit inhaltlich enger an die Merkmale der Prosodie angelehnt ist die *Multidimensional Fluency Scale (MDFS)* nach Rasinski (2004). Die Skala (s. Abbildung 2) besteht aus vier vierfach abgestuften Subskalen zu den Dimensionen Betonung, Phrasierung, Rhythmus und Geschwindigkeit, bei denen jeweils vier Punkte, also insgesamt 16 Punkte, zu erreichen sind. Nach Rasinski (2004) liest jemand dann einen Text flüssig bzw. kann als flüssige Leserin oder flüssiger Leser bezeichnet werden, wenn sie oder er mindestens acht Punkte erreicht hat.

Bei der Messung der Leseflüssigkeit handelt es sich nicht um ein standardisiertes und normiertes Verfahren, für das sich prinzipiell testtheoretische Kennwerte angeben ließen. Die in Studien angegebenen Werte zur Reliabilität sind vielfach ausreichend hoch, die Messfehler somit ausreichend kontrolliert. Bezogen auf die Objektivität scheint die Durchführung und Auswertung der Messung insgesamt weniger problematisch. Hinsichtlich der Interpretationsobjektivität, also der Interpretation der inhaltlichen Tragkraft der gemessenen Werte, ist jedoch zweifelhaft, ob die in der Literatur angegebenen Grenzen einer strengen Überprüfung standhalten können. Dies ist z.T. der Problematik geschuldet, dass für die Testung keine Texte vorgegeben werden. Daraus ergibt sich für die Auswahl der Texte die Problematik der Angemessenheit für eine bestimmte, zu testende Zielgruppe und das Problem der Vergleichbarkeit der Skalenergebnisse bei verschiedenen Texten. Dies wiederum führt u.a. zu einem Fehlen von kriterialen

und sozialen Normwerten für die Akkuratheit und Automatisierung als Lesegeschwindigkeit auf der einen Seite und für die prosodischen Aspekte auf der anderen Seite. Wie viele korrekt gelesene Wörter pro Minute jeweils altersangemessen sind und wie schnell ein Text gelesen werden muss, um ihn auch zu verstehen, ist umstritten und hängt stark von der Textschwierigkeit ab (Rosebrock et al. 2016). Gleiches dürfte für die NAEP-Skala und die Mehrdimensionale Flüssigkeitsskala gelten. Die kriterialen Grenzwerte, die jemand überschritten haben muss, um als flüssige Leserin oder als flüssiger Leser zu gelten, sind eher aus der Praxis oder aus Forschungskontexten heraus entstanden, deren Grundlage eine für die jeweilige Zielgruppe sinnvolle Textauswahl dargestellt haben dürfte.

- | | |
|--------------------------|--|
| A. Expression and Volume | <p>(1) Reads words as if simply to get them out. Little sense of trying to make text sound like natural language. Tends to read in a quiet voice.</p> <p>(2) Begins to use voice to make text sound like natural language in some areas of the text but not in others. Focus remains largely on pronouncing the words. Still reads in a quiet voice.</p> <p>(3) Makes text sound like natural language throughout the better part of the passage. Occasionally slips into expressionless reading. Voice volume is generally appropriate throughout the text.</p> <p>(4) Reads with good expression and enthusiasm throughout the text. Varies expression and volume to match his or her interpretation of the passage.</p> |
| B. Phrasing | <p>(1) Reads in monotone with little sense of phrase boundaries; frequently reads word-by-word.</p> <p>(2) Frequently reads in two- and three-word phrases, giving the impression of choppy reading; improper stress and intonation fail to mark ends of sentences and clauses.</p> <p>(3) Reads with a mixture of run-ons, mid-sentence pauses for breath, and some choppiness; reasonable stress and intonation</p> <p>(4) Generally reads with good phrasing, mostly in clause and sentence units, with adequate attention to expression.</p> |
| C. Smoothness | <p>(1) Makes frequent extended pauses, hesitations, false starts, sound-outs, repetitions, and/or multiple attempts.</p> <p>(2) Experiences several "rough spots" in text where extended pauses or hesitations are more frequent and disruptive.</p> <p>(3) Occasionally breaks smooth rhythm because of difficulties with specific words and/or structures.</p> <p>(4) Generally reads smoothly with some breaks, but resolves word and structure difficulties quickly, usually through self-correction.</p> |
| D. Pace | <p>(1) Reads slowly and laboriously.</p> <p>(2) Reads moderately slowly.</p> <p>(3) Reads with an uneven mixture of fast and slow pace.</p> <p>(4) Consistently reads at conversational pace: appropriate rate throughout reading.</p> |

Abb. 2: Mehrdimensionale Flüssigkeitsskala (Rasinski 2004).

Zur Erfassung der relativen Unterschiede zwischen den Schülerinnen und Schülern hinsichtlich der Leseflüssigkeit ist keine Normierung notwendig. Daher können die Skalen prinzipiell zur Ermittlung individueller Unterschiede eingesetzt werden, vorausgesetzt es stellen sich weder Boden- noch Deckeneffekt bei der zu untersuchenden Gruppe ein².

Die Tatsache, dass es sich bei der Erhebung der Leseflüssigkeit nicht um einen standardisierten Test handelt, führt zu einer weiteren Schwierigkeit in der Anwendung im Unterricht. Während in wissenschaftlichen Kontexten das Einschätzen der verschiedenen Dimensionen der Leseflüssigkeit meist anhand von (wiederholt abspielbaren) Lautleseprotokollen geschehen kann, lässt sich die Erhebung der Leseflüssigkeit im Unterricht oft nur durchführen, während die Schülerinnen und Schüler "live" einen Text vorlesen und gleichzeitig bestimmte Parameter erhoben werden. Hier stellt sich jedoch das Problem des Umgangs mit einer zu hohen Dichte an diagnostischen Informationen. Akkuratheit und Automatisierung können dabei als Lesegeschwindigkeit gemeinsam erhoben werden. Sie müssen jedoch getrennt von den prosodischen Merkmalen diagnostiziert werden, um den Diagnostiker nicht zu überfordern und die Gütekriterien der Messung sicherzustellen. Hinzu kommt ein weiteres Problem: Während die Erhebung der Lesegeschwindigkeit durch eine Lehrkraft im Unterricht mit einem adäquaten Text im Wesentlichen ohne große (Mess-)Fehler durchgeführt werden kann, lässt sich das prosodische Lesen mit ihren Teilkomponenten nicht ohne Weiteres mit jedem Text durchführen – mindestens müsste der Text auf seine prosodischen Merkmale hin untersucht werden, was nicht aufwandslos machbar ist. Relevant wäre es also, für den Unterricht eine handhabbare Möglichkeit anzubieten, das prosodische Lesen in den wichtigsten Facetten mit Hilfe eines (halb-)standardisierten Verfahrens "live", also während des Vorlesens einer Schülerin oder eines Schülers, zu erheben. Ein solches Verfahren wird im Folgenden vorgestellt und diskutiert.

5 | Fragestellung

Die vorliegende Untersuchung soll einen Beitrag dazu leisten, wie prosodisches Lesen valide und reliabel und zudem möglichst ökonomisch gemessen werden kann. Zum einen konnte theoretisch gezeigt werden, dass sich bestimmte Merkmale der Prosodie während des Lesens funktional auf den zu lesenden Inhalt beziehen. Zum anderen ist die bisherige Leseflüchtigkeitsdiagnostik, die sehr auf eine heuristische Beobachtung mehrerer Dimensionen der Flüssigkeitsleistung gleichzeitig ausgerichtet ist, gerade im Einsatz im Unterricht zu träge und stellt den Beobachter eines Leseprozesses vor die Herausforderung, zu viel gleichzeitig beobachten zu müssen. Es stellt sich somit die Frage, ob eine Reduktion der beobachtbaren Informationen und eine Konzentration auf bestimmte Stellen des Vorgelesenen eine ökonomische Variante der Leseflüchtigkeitsdiagnostik sein könnte, oder genauer:

- Lässt sich mit Hilfe von ‚Lupenstellen‘, also anhand ausgewählter prosodischer Stellen im Text, valide, reliabel und ökonomisch prosodisches Lesen messen?
- Ließe sich ein solches Instrument im Unterricht einsetzen?

Die in Absatz 3.2 genannten Studien konnten zeigen, dass prosodisches Lesen und Textverstehen zusammenhängen. Als weitere Frage kann daher formuliert werden:

² Nach strengen testtheoretischen Anforderungen müssten die Skalen allerdings raschskaliert sein, um relative Unterschiede zwischen Personen deuten zu können.

- Sind prosodische Lupenstellen-Skalen in der Lage, hinsichtlich des Textverstehens gute von weniger guten Leserinnen und Lesern zu differenzieren?

Im Folgenden wird eine Pilotstudie vorgestellt, die diesen Fragen nachgeht.

6 | Studie

6.1 | Design und Methode

Im Folgenden werden erste Entwicklungsschritte hin zu einem diagnostischen Tool vorgestellt, das das Potenzial besitzt, prosodisches Lesen im Unterricht durch Lehrkräfte ökonomisch, objektiv, reliabel und valide zu erheben. Hierzu wurden auf der Grundlage theoretischer Erwägungen zur Leseflüssigkeit generell und zur Prosodie im Besonderen Stellen in einem zielgruppenadäquaten Text aufgedeckt, die distinkt bestimmte prosodische Merkmale markieren, sogenannte ‚Lupenstellen‘. Diese Lupenstellen wurden mit Hilfe deskriptiver Kennwerte der klassischen Testtheorie an einer Stichprobe von laut gelesenen Texten analysiert und zu Skalen zusammengefasst, die wiederum hinsichtlich ihrer Reliabilität und Validität überprüft wurden. Hierzu wurden Reliabilitätskoeffizienten und korrelative Zusammenhänge mit etablierten Skalen des prosodischen Lesens innerhalb des Konstrukts der Leseflüssigkeit betrachtet (konvergente Validität) sowie Zusammenhänge mit dem Leseverstehen hergestellt.

6.2 | Material

Das in dieser Studie verwendete Material zur Erfassung der prosodischen Aspekte der Leseflüssigkeit umfasste zum einen einen expositorischen Text, der von Schülerinnen und Schülern gelesen wurde. Zur Bewertung des prosodischen Lesens wurde ein Instrument entwickelt, das einer Studierendengruppe zur Verfügung gestellt wurde. Da dieses Instrument zentral für diese Studie ist, wird seine Entwicklung ausführlich vorgestellt. Zur Erfassung des Leseverstehens wurde ein standardisierter Lesetest eingesetzt.

Audioaufnahmen gelesener Texte

Zur Erhebung der für die Leseflüssigkeit relevanten prosodischen Merkmale, Betonung, Intonation und Pausierung, wurde ein expositorischer Text mit 150 Wörtern zum Thema „Krötenwanderung“ eingesetzt, den die Schülerinnen und Schüler laut vorlasen. Der Text ist inhaltlich etwa für die dritte bis sechste Klasse geeignet.

Instrument DisiLe

Das Instrument „Diagnose sinngestaltenden Lesens“ (*DisiLe*) beruht zum einen auf den theoretischen Erwägungen zu den Merkmalen der Prosodie beim Textlesen, zum anderen wurden durch Lesungen mehrerer kompetenter Leserinnen und Leser Lupenstellen ausfindig gemacht, d.h. Stellen, an denen sich die genannten prosodischen Merkmale beim lauten Lesen manifestieren. Grundlegende Bedingungen für die Erstellung der Lupenstellen waren wie folgt:

- Die Lupenstelle bezieht sich dezidiert auf eins der oben beschriebenen prosodischen Merkmale (Intonation, Betonung und Pausierung). Sie erfüllt zudem eine relevante linguistische Funktion³.
- Die Lupenstelle ist eindeutig, d.h. z.B., dass in einem Satz nicht verschiedene Möglichkeiten bestehen, ein Wort prominent zu betonen.
- Kompetente Leserinnen und Leser setzen die Merkmale an den Lupenstellen durchgängig und gleichartig um, d.h. dass z.B. eine Pause im Text auch von allen kompetenten Leserinnen und Lesern während des lauten Lesens an der gleichen Stelle gemacht wird.
- Kompetente Hörerinnen und Hörer gelesener Texte können die Merkmale in ihrer Ausprägung auch wahrnehmen, so sollten z.B. Pausen im gelesenen Text auch von allen als solche wahrgenommen und identifiziert werden können.
- Von den Lupenstellen wird angenommen, dass sich die Phänomene auf der Ebene der Phrase und des Satzes zeitökonomisch und ohne professionelle Hilfsmittel auditiv erfassen lassen.

Tabelle 1 beschreibt Beispiele für Lupenstellen der Merkmale Betonung, Intonation und Pausierung inklusive ihrer Begründungen. Insgesamt enthält der Text 34 Items. 21 Lupenstellen entfallen auf das Merkmal Betonung (davon zwölf zu betonende und neun nicht zu betonende Stellen), neun Lupenstellen auf die Intonation (vier davon aufsteigende oder gleichbleibende Intonation, fünf abfallende) und vier Stellen auf relevante Pausen.

| Merkmals | Beispielsatz | Beispielitem | Begründung |
|------------|---|-------------------------|--|
| Betonung | Die Erdkröten müssen oft Straßen überqueren, um an ihr Ziel zu kommen. | Straßen (betont) | Die Akzentuierung deutet hier auf eine neue und relevante Information. Diese neue Information ist der Aufhänger für den folgenden inhaltlichen Absatz. |
| | Die Erdkröten müssen oft Straßen überqueren, um an ihr Ziel zu kommen. | oft (unbetont) | Die semantische Einheit liefert keine neue und für den folgenden Gedankengang ausschlaggebende Information. |
| Intonation | An vielen Straßen gibt es deshalb kleine Zäune und Tunnel. | Zäune (steigend) | Hierbei wird signalisiert, dass die Aufzählung noch nicht abgeschlossen ist. |
| | An vielen Straßen gibt es deshalb kleine Zäune und Tunnel . | Tunnel (fallend) | Es handelt sich um eine satzschließende Intonation. |
| Pause | Das ist sehr gefährlich. An vielen Straßen... | Pause nach „gefährlich“ | Die Pause signalisiert eine Phrasengrenze und hat zugleich auch eine rhetorisch-stilistische Funktion der Hervorhebung. |

Tab. 1: Beispiele für Items zur Betonung, Intonation und Pausierung.

³ An der Stelle möchten wir darauf hinweisen, dass hinsichtlich des Status dieser Phänomene – z.B. ihrer Betrachtung als semantisch differenzierend - durchaus verschiedene Sichtweisen existieren (Spielmann 2000).

Aus den entwickelten Lupenstellen wurde das Instrument ‚DisiLe‘ erstellt. Die Lupenstellen wurden zur praktischen Umsetzung während des Hörens wie folgt auf einem DIN A4-Bogen gekennzeichnet (s. Abbildung 3): Stellen, die betont werden müssen, wurden fett und unterstrichen, nicht zu betonende Lupenstellen gestrichelt markiert. Zum Ankreuzen dieses Merkmals stand ein quadratisches Kästchen zur Verfügung. Intonatorische Lupenstellen wurden durch Pfeile über der jeweiligen Stelle angegeben. Die Richtung des Pfeils entsprach dabei der Richtung der Intonationskontur, so zeigten aufwärtszeigende Pfeile eine aufsteigende Intonationskontur an. Kreise zum Ankreuzen dienten der Markierung, ob die geforderte Intonationskontur eingehalten wurde. Eine senkrechte Linie mit zugehörigem Rechteck zum Ankreuzen diente der Markierung von einzuhaltenden Pausen.

Krötenwanderung

Im Frühjahr verlassen die Erdkröten ihr Winterquartier und machen sich auf den Weg zu einem Gewässer. Sie suchen den Teich oder Tümpel , in dem sie selbst aufgewachsen sind. Dort wollen sie ihre Eier ablegen . Das nennt man laichen. Meist sind sie abends und nachts unterwegs.

Die Erdkröten müssen oft Straßen überqueren, um an ihr Ziel zu kommen. Das ist sehr gefährlich. An vielen Straßen gibt es deshalb kleine Zäune und Tunnel .

Abb. 3: Ausschnitt aus dem Instrument DisiLe (ausgefüllt).

Weitere Maße der Leseflüssigkeit

Zur Erfassung des prosodischen Lesens wurde auf eigene Übersetzungen englischsprachiger Instrumente zurückgegriffen. Eingesetzt wurden Übersetzungen der *NAEP-Skala* zur Leseflüssigkeit (Daane et al. 2005) und der *Multi-dimensional fluency scale MDFS* (Rasinski 2004). Um möglichst valide Ergebnisse zu erhalten, wurden beide Prosodieskalen eingesetzt, die jeweils eine ausreichend hohe Interraterübereinstimmung zwischen zwei geschulten Gutachterinnen erreichten ($ICC_{NAEP} = .84$; $ICC_{MDFS} = .97$). Für alle Schülerinnen und Schüler lagen somit Werte für ihre Fertigkeit im prosodischen Lesen auf verschiedenen Subskalen vor.

Textverstehen

Zur Erhebung des Textverstehens wurde der standardisierte Lesetest ELFE – Ein Leseverständnistest für Erst- bis Sechstklässler (Lenhard & Schneider 2006) durchgeführt. Der Test besteht aus den Untertests Wort-, Satz- und Textverstehen.

6.3 | Stichprobe und Durchführung

Durch das beschriebene Design ergaben sich zwei Stichproben: Die Stichprobe der Schülerinnen und Schüler, die den Text laut lasen, und eine Stichprobe von Studierenden, die die prosodischen Aspekte anhand der Lesungen bewertet haben.

Schülerinnen und Schüler

Insgesamt lasen 40 Schülerinnen und Schüler der 4. Klasse aus zwei Grundschulen den expository Text „Krötenwanderung“ in Einzelsitzungen laut vor, 55 Prozent davon waren familiär monolingual Deutsch, das Geschlechterverhältnis war ausgeglichen. Die prima vista-Lesungen wurden in einer Audioaufzeichnung festgehalten, sodass für die spätere Auswertung die Möglichkeit bestand, den gelesenen Text mehrmals erneut zu hören. Der Lesetest ELFE wurde im Klassenverband durchgeführt.

Studierende

Insgesamt wurden 17 Lehramtsstudierende aus einem Seminar zur Lesediagnostik und -förderung zur Einschätzung der prosodischen Stellen innerhalb der gelesenen Texte rekrutiert. Studierende mit geringer Übereinstimmung pro Skala als Indiz für entweder zufällig ausgefüllte Kästchen oder zu geringer Kompetenz, die entsprechenden Merkmale wahrzunehmen, wurden aus der Stichprobe entfernt (s. u.). Es verblieben 16 Studierende, die jeweils 15 gelesene Texte in Form von Audiodateien zur Einschätzung erhielten, so dass jeder Text von drei Studierenden bewertet wurde. Jeder Text wurde von einer Person drei Mal angehört: Zunächst wurden alle Lupenstellen zur Intonation durchgehend bewertet, danach diejenigen zur Betonung und schließlich beim dritten Durchgang die Stellen, an denen das Vorhandensein oder die Abwesenheit von Pausen bewertet werden musste.

Vor dieser Prozedur wurden die Studierenden in ihre Aufgabe eingewiesen, indem sie sich mit dem Testbogen DisiLe, der Legende und den Anweisungen vertraut machten (z.B. „Kreuzen Sie in den Quadraten an, wenn die Schülerin oder der Schüler die entsprechende Stelle betont liest. Machen Sie kein Kreuz, wenn die Schülerin oder der Schüler an der entsprechenden Stelle nicht betont liest.“). Bevor die Studierenden mit der eigentlichen Analyse starteten, übten Sie das Verfahren an zwei Beispielaudiodateien.

6.4 | Ergebnisse

Bevor Kennwerte für die einzelnen Texte bzw. Schülerinnen und Schüler über die Studierendenbewertungen hinweg gemittelt wurden, wurde die Stichprobe der Studierenden genauer untersucht, um ggf. diejenigen zu ermitteln, die Auffälligkeiten aufweisen, die auf zufälliges Ankreuzen hindeuteten. Hierzu wurden die individuellen Daten der Studierenden über die bewerteten 15 Texte hinweg auf kohärentes Bewerten geprüft. Beruhend auf der Annahme, dass innerhalb der gehörten Textauswahl prosodische Strukturen vorhanden sind, deutet eine fehlende interne Konsistenz auf „zufällig“ angekreuzte Lupenstellen hin. Möglicherweise konnten die betreffenden prosodischen Merkmale in ihrer Variabilität von den betreffenden Studierenden tatsächlich nicht ausreichend gut wahrgenommen oder bewertet werden. Bei dieser Analyse zeigte sich, dass eine Studentin statistische Kennwerte aufwies, die sich von den anderen Studierenden unterschieden. Diese Bewerterin wurde aus der Stichprobe entfernt.

Itemanalysen und Reliabilität

Es stellte sich die Frage, ob prosodisches Lesen reliabel durch Lupenstellen erhoben werden kann. Tabelle 2 zeigt die Schwierigkeiten, Standardabweichungen und Trennschärfen für die Items getrennt nach den Skalen Betonung, Intonation und Pausen, wie sie sich nach Mittelung über die einzelnen Bewerterinnen und Bewerter zeigen. Einige Items (in der Tabelle mit * gekennzeichnet) mussten aufgrund zu geringer Trennschärfen aus der Analyse ausgeschlossen werden. Die Reliabilitäten lagen bei Cronbachs $\alpha = .71$ (Betonung), $\alpha = .40$ (Intonation) und $\alpha = .56$ (Pausen). Während die Betonung sich also vergleichsweise reliabel messen ließ, lagen die Merkmale Intonation und Pausen hinsichtlich der Reliabilität im mittleren Bereich. Die Skala Pausen ist zudem mit drei verbleibenden Items sehr klein, der Wert daher nicht sehr aussagekräftig. Die Gesamtskala aller Lupenstellen abzüglich der in der Tabelle vermerkten Items mit schlechten Kennwerten betrug Cronbachs $\alpha = .73$.

Die Schwierigkeiten der Lupenstellen wiesen einige Unterschiede auf. Beim Merkmal Betonung schwankten die Schwierigkeiten zwischen .44 und .94, d.h. die Lupenstelle 19 („springen“) im Satz „Die Kröten können nicht über die Zäune springen.“ wurde nur von 44% der Schülerinnen und Schülern betont gelesen. Das Wort „springen“ hat hier eine besondere Relevanz und muss hervorgehoben werden, da es einen Kontrast mit der Information „kriechen“ im nächsten Satz darstellt: „Sie kriechen so lange daran entlang [...]“. Die Lupenstelle 34 („zurück“) in der Phrase „in den Wald zurück“ wurde richtigerweise von 94% der Schülerinnen und Schüler nicht betont, das Wort „zurück“ trägt keine substantielle Information. Ähnliches gilt für die Intonation: Bei Lupenstelle 2 („Tümpel“) in der Phrase „Sie suchen den Teich oder Tümpel, in dem sie selbst[...]“ stieg bei einer substantiellen Anzahl der Schülerinnen und Schüler die Intonationskontur nicht, während sie bei Lupenstelle 7 („laichen“) in „Das nennt man laichen“ zumeist korrekt eine abfallende Kontur realisierten.

| L-Stelle | <i>P</i> | <i>SD</i> | <i>r_{it}</i> | L-Stelle | <i>P</i> | <i>SD</i> | <i>r_{it}</i> | L-Stelle | <i>P</i> | <i>SD</i> | <i>r_{it}</i> |
|-----------------|----------|-----------|-----------------------|-------------------|----------|-----------|-----------------------|---------------|----------|-----------|-----------------------|
| Betonung | | | | 22 | .89 | .13 | .36 | 12 | .80 | .20 | .21 |
| 3 | .63 | .31 | .29 | 23 | .58 | .33 | .55 | 14 | .65 | .29 | .10 |
| 4 | .76 | .22 | .13 | 25 | .67 | .25 | .26 | 15* | .93 | .40 | -.27 |
| 5* | .79 | .19 | .07 | 28 | .81 | .23 | .27 | 18* | .62 | .31 | .02 |
| 6 | .81 | .21 | .23 | 30 | .62 | .29 | .32 | 24* | .56 | .24 | .02 |
| 8* | .62 | .25 | -.05 | 31 | .86 | .18 | .23 | 27* | .91 | .20 | -.12 |
| 9 | .59 | .25 | .17 | 32 | .78 | .21 | .28 | Pausen | | | |
| 10 | .64 | .25 | .35 | 33* | .74 | .22 | .02 | 13 | .49 | .42 | .49 |
| 11* | .70 | .24 | .05 | 34 | .94 | .11 | .19 | 20 | .46 | .34 | .45 |
| 16 | .56 | .25 | .37 | Intonation | | | | 26 | .32 | .28 | .46 |
| 17 | .65 | .24 | .33 | 1 | .82 | .21 | .29 | 29* | .76 | .30 | .05 |
| 19 | .44 | .29 | .20 | 2* | .57 | .34 | .01 | | | | |
| 21 | .72 | .23 | .55 | 7 | .88 | .21 | .21 | | | | |

Anmerkungen: L-Stelle = Lupenstelle, *P* = Schwierigkeit, *SD* = Standardabweichung, *r_{it}* = Trennschärfe, * = Item wurde für Cronbachs α aufgrund zu geringer Trennschärfe entfernt.

Tab. 2: Schwierigkeiten, Streuung und Trennschärfen für die einzelnen Lücken

Zusammenhänge mit weiteren Maßen der Leseflüssigkeit

Die Frage nach der konvergenten Validität lässt sich durch Zusammenhänge mit verschiedenen Maßen des gleichen Konstruktes beantworten. Zum gleichen Konstrukt gehört im weiteren Sinne die Leseflüssigkeit insgesamt (d.h. einschließlich der Akkuratheit und Automatisierung), im engeren Sinne die prosodischen Skalen innerhalb der Leseflüssigkeit. Tabelle 3 zeigt die Zusammenhänge der Lupenstellen-Skalen mit den Skalen der Lesegeschwindigkeit, der mehrdimensionalen Flüssigkeitsskala und der NAEP-Skala. Die Zusammenhänge dieser Skalen mit den Lupenstellen-Skalen Betonung und Intonation sind insgesamt nur von mittlerer Stärke, was z.T. daran liegt, dass in die Berechnung der Leseflüssigkeitsskalen NAEP und MDFS verschiedene prosodische Merkmale eingehen. Der Zusammenhang zwischen den Pausen und den jeweiligen Skalen ist allerdings negativ und somit schwer zu interpretieren, inhaltlich bedeutet dieser Zusammenhang, dass diejenigen, die an den eigentlich richtigen Stellen pausieren, geringe Werte in der prosodischen Leseflüssigkeit besitzen.

| prosodisches Merkmal | Akkuratheit | Automatisierung | Mehrdimensionale Flüssigkeitsskala | | | | | NAEP |
|----------------------|-------------|-----------------|------------------------------------|----------|-------------|----------|-----------------|-------|
| | | | Gesamt | Betonung | Phrasierung | Rhythmus | Geschwindigkeit | |
| Betonung | .41** | .58** | .49** | .56** | .45** | .41** | .41** | .51** |
| Intonation | .38* | .35* | .38* | .38* | .38* | .31* | .35* | .36* |
| Pausierung | -.13 | -.26 | -.17 | -.11 | -.13 | -.13 | -.24 | -.38* |
| Gesamt | .29 | .21 | .23 | .56 | .45 | .41 | .41 | .10 |

Anmerkung: *p < .05; ** p < .01

Tab. 3: Korrelationen der prosodischen Merkmale mit Skalen der Leseflüssigkeit

Wie schon bei den Schwierigkeiten wiesen die einzelnen Lupenstellen z.T. erhebliche Unterschiede im Zusammenhang mit den globalen Skalen der Leseflüssigkeit auf. So schwanken die Zusammenhänge zwischen $r = .56$ bei Lupenstelle 16 (für den Fortgang des Textes relevante Betonung auf „Zäune und Tunnel“ im Satz „An vielen Straßen gibt es deshalb kleine Zäune und Tunnel“) und $r = .00$.

Zusammenhänge mit dem Textverstehen

Ausgehend von den Studien, die zeigen, dass prosodisches Lesen und Textverstehen zusammenhängen, stellte sich die Frage, ob die Lupenstellen-Skalen in der Lage sind, hinsichtlich des Verstehens gute von weniger guten Leserinnen und Lesern zu differenzieren. Die Korrelationen zeigen, dass sich dies am besten durch die Skala Betonung realisieren lässt, d.h. Schülerinnen und Schüler mit guter Betonung haben eine höhere Kompetenz, Texte zu verstehen. Geringer ist der Zusammenhang mit der Intonation, hier zeigt sich, dass der Zusammenhang mit dem Textverstehen gegenüber dem Wortverstehen geringer ausfällt. Wie bereits bei der Korrelation

mit den anderen Flüssigkeitsskalen ist auch hier die Lupenstellen-Skala des Pausierens negativ korreliert mit dem Leseverstehen. Tabelle 4 fasst die Korrelationen der Skalen mit dem Leseverstehen zusammen.

| prosodisches Merkmal | ELFE | | | Gesamtwert |
|----------------------|----------------------|----------------------|----------------------|------------|
| | Wortverständ- nis | Satzverständ- nis | Textverständ- nis | |
| Betonung | .52** | .43** | .45** | .51** |
| Intonation | .45** | .24 | .19 | .32* |
| Pausierung | -.08 | -.35* | -.27 | -.26 |
| Gesamt | .29 | .11 | .12 | .19 |

Anmerkung: * $p < .05$; ** $p < .01$

Tab. 4: Korrelationen der prosodischen Merkmale mit prosodischen Skalen der Leseflüssigkeit.

Betrachtet man einzelne Lupenstellen, so ist ihr Beitrag zur Erklärung des Textverstehens unterschiedlich groß. Lupenstelle 16 zur Betonung (s. vorangegangener Absatz) korreliert z.B. mit $r = .61$ mit dem Textverstehen, die Lupenstelle 30 zur Betonung mit $r = .41$, andere Stellen korrelieren weitaus geringer.

7 | Diskussion und Ausblick

Innerhalb der Leseflüssigkeit ist die Dimension des prosodischen Lesens in der Forschung, der Lesedidaktik, der Diagnostik und in der Förderung weitgehend ein Desiderat. Die wenigen Studien zur Entwicklung des prosodischen Lesens und zu Zusammenhängen mit dem Textverstehen zeigen jedoch die Relevanz, die prosodischen Aspekte stärker in den Vordergrund zu rücken. In der vorliegenden Studie ging es darum, auf der Grundlage theoretischer Erwägungen zum prosodischen Lesen ein Instrument zu entwickeln, mit dessen Hilfe prosodisches Lesen während des lauten Lesens im Klassenraum erhoben werden kann. Hierzu wurde ein Instrument mit ‚Lupenstellen‘ entwickelt, mit dessen Hilfe anhand ausgewählter Stellen in einem Text die Fertigkeit des prosodischen Lesens in den Merkmalen *Betonung*, *Intonation* und *Pausierung* gemessen werden soll. In dieser Studie ging es darum, die Methode der ‚Lupenstellen‘ zu testen, die dann die Grundlage eines anwendungsorientierten, ökonomisches Instruments bilden können.

Die Ergebnisse zeigten, dass die Lupenstellen-Skalen zur *Betonung* und *Intonation* substantielle Reliabilitäten aufwiesen. Sie lagen im mittleren Bereich. Die Skalen korrelierten zudem mit den einschlägigen Skalen zur Leseflüssigkeit im mittleren Bereich. Eine Betrachtung der einzelnen Lupenstellen brachte eine hohe Bandbreite sowohl in ihrer Schwierigkeit als auch in ihren Zusammenhängen mit den Leseflüssigkeitsskalen zu Tage. Dies hängt möglicherweise damit zusammen, dass bei einer Bewertung der Leseflüssigkeit auf globaler Ebene die für einen Text *wichtigen* Betonungen prominenter sind und stärker in die Bewertung des prosodischen Lesens eingehen. Für eine Weiterentwicklung des Instruments bedeutet dies, relevante Lupenstellen zu finden bzw. zu erstellen, die eng mit einem globalen Eindruck des prosodischen Lesens zusammenhängen.

Ähnliche Ergebnisse zeigten sich hinsichtlich der Zusammenhänge der Lupenstellen-Skalen *Betonung* und *Intonation* mit dem Textverstehen. Insgesamt sind die Zusammenhänge von mittlerer bis geringer Größe. Dabei zeigt sich, dass *Betonungen* stärker mit dem Textverstehen zusammenhängen als die *Intonation*. Möglicherweise benötigt man für eine ausreichend adäquate Intonationskontur weniger das Verstehen des Textes, denn sie lässt sich z.T. auch an der Interpunktion ablesen. Auf Lupenstellen-Ebene zeigt sich auch hier wieder eine große Bandbreite der Zusammenhänge mit z.T. hohen Werten. Dies kann zwar ein Artefakt der Untersuchung oder statistische Fluktuation auf Einzelebene sein, es könnte aber auch sein, dass Betonungen im Text verschiedene Funktionen erfüllen und daher für das Textverstehen mehr oder weniger relevant sind. Verschiedene hochkorrelierende Lupenstellen mit inhaltlich relevanter Information für den Fortgang des Textes weisen darauf hin. Textverstehen lässt sich also auch durch Lupenstellen an wenigen neuralgischen Punkten ablesen. Da hier aber das Textverstehen nicht innerhalb des Textes erhoben wurde, der laut vorgelesen und dessen Lupenstellen bewertet wurden, lässt sich nicht zeigen, dass das richtige Betonen und Intonieren zu einem besseren Verständnis *dieses* Textes geführt hat oder, auch dies ist möglich, dass das Textverstehen *dieses* Textes zu einer besseren Leistung im prosodischen Lesen geführt hat.

Die Skala *Pausen* hat durchweg entweder schlechte Kennwerte oder sie verhält sich gegenteilig zu den anderen Skalen. Dies liegt möglicherweise daran, dass Pausen deutlich weniger festgelegt sind als die beiden anderen Merkmale. Sie sind nicht immer obligatorisch und semantisch relevant. Sie können zudem durch Atempausen entstehen, als rhetorisches Mittel eingesetzt werden und generell die Sprechweise einer Person auszeichnen. Da erfahrene Hörende implizit oder auch explizit mit diesen Prinzipien vertraut sein dürften, könnte dies die Erklärung für die wenig nutzbringenden Ergebnisse hinsichtlich dieser Skala sein. Für eine Weiterentwicklung des Instruments müsste man diese Skala möglicherweise ausschließen.

Insgesamt zeigt sich, dass es lohnenswert ist, die Herangehensweise der ‚Lupenstellen‘ weiterzuerfolgen, weitere Erkenntnisse müssen die Weiterentwicklung jedoch flankieren. Hierzu gehört zu untersuchen, ob alle Hörerinnen und Hörer in der Lage sind (auch nach einer Trainingsphase) die Lupenstellen wahrzunehmen, denn aus der Stichprobe der Studierenden musste eine Studentin ausgeschlossen werden, da sie zu inkonsistent bewertet hatte. Schließt man Motivationsmangel aus, stellt sich die Frage, ob Bewertende die Merkmale nicht richtig beurteilen, weil sie sie nicht wahrnehmen. Es kann davon ausgegangen werden, dass nicht nur das allgemeine Hörvermögen, sondern auch Kompetenzen in der Wahrnehmung und Diskrimination einzelner Phänomene im Bereich der Prosodie variieren können.

Genauer untersucht werden muss auch die differenzierende Funktion verschiedener Lupenstellen im Text, denn einige Lupenstellen haben eine höhere prognostische Kraft als andere. Sind beispielsweise bestimmte Intonationskonturen und Betonungen im Text ‚wichtiger‘ als andere? Gibt es verschiedene linguistische Funktionen von Betonungen, die verschiedene andere Merkmale im Leseprozess vorhersagen? Unterscheiden sich bestimmte Funktionen und die sprachlichen Einheiten, die diese jeweils tragen, hinsichtlich der Salienz? Ist die fehlende Betonung bei einer absolut neuen Information auffälliger als bei einer zwar relevanten, aber bereits eingeführten?

Eine Beantwortung dieser und ähnlicher Fragen könnte helfen, Texte zu erstellen, die ganz gezielt bestimmte prosodische Phänomene in den Vordergrund stellen, und weiter aufzuklären, inwieweit prosodisches Lesen an dezidierten Stellen das Textverstehen prognostizieren kann.

Dies würde einen genaueren Einblick in die Funktion einzelner prosodischer Merkmale hinsichtlich des Textverstehens geben. Gelänge dies, könnte auch die Anzahl der Lupenstellen weiter reduziert und auf die relevanten beschränkt werden, damit in einem praxisnahen Tool, anders als in dieser Pilotstudie, schon beim ersten Lesedurchgang der Schülerin bzw. des Schülers ohne Audioaufnahme alle Stellen möglichst messfehlerreduziert erhoben werden können.

Auch wenn es gelingt, ein solches Instrument weiterzuentwickeln, bleiben zwei Herausforderungen bestehen: Die erste, und diese ist die kleinere von beiden, betrifft das Problem der Normierung: Welche Skalenwerte hinsichtlich des prosodischen Lesens müssen die Schülerinnen und Schüler erreichen, um einen Text zu verstehen? Eine zweite Herausforderung umfasst die Richtung der gezeigten Zusammenhänge: Führt Leseflüssigkeit zum Textverstehen oder muss ein Text (wenigstens ansatzweise) verstanden sein, um ihn flüssig, also akkurat, zügig und prosodisch weitgehend korrekt, zu lesen. Denn wie viele andere Studien zum Themenkomplex *Leseflüssigkeit – Prosodie – Textverstehen* ist auch diese Studie eine korrelative Studie, d.h. kausale Zusammenhänge zwischen der Prosodie und dem Textverstehen können auch hier nicht abgeleitet werden.

8 | Literatur

- Bose, I., Hirschfeld, U., Neuber, B. & Stock, E. (2016). *Einführung in die Sprechwissenschaft. Phonetik, Rhetorik, Sprechkunst*. Narr Francke Attempto.
- Bußmann, H. (2002). *Lexikon der Sprachwissenschaft*. Alfred Kröner.
- Chafe, W. (1988). Punctuation and the prosody of written language. *Written Communication*, 5(4), 396–426.
- Chard, D. J., Vaughn, S. & Tyler, B. J. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities*, 35(5), 386–406.
- Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading*. U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. Government Printing Office.
- Dahmen, S. & Weth, C. (2017). *Phonetik, Phonologie und Schrift*. UTB GmbH.
- Ergert, M. (2012). *Prosodie & Didaktik. Neue Ansätze für erfolgreichen Sprach- und Schrifterwerb*. Schneider Hohengehren.
- Esteve-Gibert, N. & Prieto, P. (2018). Early development of the prosody-meaning interface. In: P. Prieto, & N. Esteve-Gibert (Eds.). *The Development of Prosody in First Language Acquisition* (S. 227-246). John Benjamins Publishing Company.
- Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. *Proceedings of the 1st international conference on speech prosody*, 83–88.
- Groen, M. A., Veenendaal, N. J. & Verhoeven, L. (2019). The role of prosody in reading comprehension: evidence from poor comprehenders. *Journal of Research in Reading*, 42(1), 37–57.

- Institut für Deutsche Sprache (2013). "Kontrastive Sicht". *Grammatisches Informationssystem grammis*. DOI: 10.14618/eurogramm, Permalink: <https://grammis.ids-mannheim.de/kontrastive-grammatik/4866> [16.09.2021]
- Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. University Press.
- Klauda, S. L. & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology*, 100(2), 310–321.
- Konopczynski, G. (1995). A developmental model of acquisition of rhythmic patterns: Results from a cross-linguistic study. *Proceedings of the Intonational Congress of Phonetic Sciences – Stockholm, 4*, 22–25.
- Koriat, A., Greenberg, S., & Kreiner, H. (2002). The extraction of structure during reading: Evidence from reading prosody. *Memory and Cognition*, 30(2), 270–280.
- Kuhn, M. R. & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, 95(1), 3-21.
- Kuhn, M. R., Schwanenflugel, P. J. & Meisinger, E. B. (2010). Aligning Theory and Assessment of Reading Fluency: Automaticity, Prosody, and Definitions of Fluency. *Reading Research Quarterly*, 45, 230–251.
- LaBerge, D. & Samuels, S. J. (1974). Toward a theory of automatic information process in reading. *Cognitive Psychology*, 6(2), 293–323.
- Lauer-Schmaltz, M., Rosebrock, C. & Gold, A. (2014). Lautlesetandems in der Grundschule - Bedingungen und Grenzen ihrer Wirksamkeit. *Didaktik Deutsch*, 37, 44–61.
- Lenhard, W. & Schneider, W. (2006). *ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler*. Hogrefe.
- Linnemann, M., Stephany, S., Lemke, V., Bulut, N., Haider, H., Roth, H.-J. & Becker-Mrotzek, M. (in Review). The Dimensionality of Writing and Reading Fluency and its Impact on Composition and Comprehension. *Journal of Writing Research*.
- Miller, J. M. & Schwanenflugel, P. J. (2008). A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading Research Quarterly*, 43(4), 336–354.
- Nix, D. (2011). *Förderung der Leseflüssigkeit. Theoretische Fundierung und empirische Überprüfung eines kooperativen Lautleseverfahrens im Deutschunterricht*. Juventa.
- Noak, C. (2016). *Phonologie*. Universitätsverlag Winter.
- Ott, S. (2012). Prosodie - eine Einführung. In: C. Kauschke, S. Ott & V. Maihack (Hrsg.), *Prosodie und Kindersprache. Die Bedeutung der Prosodie und ihrer Schnittstellen für Spracherwerb und Sprachtherapie* (S. 26–29). ProLog.
- Paige, D. D., Rupley, W. H., Smith, G. S., Rasinski, T. V., Nichols, W. & Magpuri-Lavell, T. (2017). Is prosodic reading a strategy for comprehension? *Journal for educational research online*, 9(2), 245–275.
- Peters, B. (2012). *Form und Funktion prosodischer Grenzen im Gespräch*. Südwestdeutscher Verlag für Hochschulschriften.
- Pollatsek, A., Rayner, K., & Lee, H.-W. (2000). Phonological coding in word perception and reading. In: A. Kennedy, R. Radach, D. Heller, & J. Pynte (Hrsg.), *Reading as a perceptual process* (S. 399–425). Elsevier.

- Post, B. & Payne, E. (2018). Speech rhythm in development. What is the child acquiring? In: P. Prieto & N. Esteve-Gibert (Hrsg.), *The Development of Prosody in First Language Acquisition* (S. 125–144). John Benjamins Publishing Company.
- Ramers, K. H. (2007). Phonologie. In: J. Meibauer, U. Demske, J. Geilfuß-Wolfgang, J. Pafel, K.-H. Ramers, M. Rothweiler & M. Steinbach (Hrsg.), *Einführung in die germanistische Linguistik* (S. 70–120). Metzler Verlag.
- Rasinski, T. (2004). Creating fluent readers. *Educational Leadership*, 61(6), 46–51.
- Richter, N. (2007). Prosodie neutraler und nicht-neutraler Äußerungen der russischen Standardsprache im Kontext des Fremdspracherwerbs. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*. <https://tujournals.ulb.tu-darmstadt.de/index.php/zif/article/view/293/285> [16.09.2021].
- Rosebrock, C. & Nix, D. (2006). Forschungsüberblick: Leseflüssigkeit (Fluency) in der amerikanischen Leseforschung und -didaktik. *Didaktik Deutsch*, 20, 90–112.
- Rosebrock, C., Nix, D., Rieckmann, C. & Gold, A. (2016). *Leseflüssigkeit fördern. Lautleseverfahren für die Primar- und Sekundarstufe*. Kallmeyer.
- Rosebrock, C., Rieckmann, C., Nix, D. & Gold, A. (2010). Förderung der Leseflüssigkeit bei leseschwachen Zwölfjährigen. *Didaktik Deutsch*, 28, 33–58.
- Samuels, S. J. (1994). Toward a theory of automatic information processing in reading, revisited. In: R. B. Ruddell, M. R. Ruddell & H. Singer (Hrsg.): *Theoretical models and processes of reading* (S. 816–837). International Reading Association.
- Sappok, C., Linnemann, M. & Stephany, S. (2020). Leseflüssigkeit – Prosodie – Leseverstehen. Eine Longitudinalstudie zur Entwicklung der Leseflüssigkeit von Jahrgangstufe 3 bis 7. In: I. Rautenberg (Hrsg.), *Evidenzbasierte Forschung zum Schriftspracherwerb* (S. 175–209). Schneider.
- Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84(1), 1–66.
- Schreiber, P. A. (1991). Understanding prosody's role in reading acquisition. *Theory into Practice*, 30(3), 158–164.
- Schröder, C. & Höhle, B. (2011). Prosodische Wahrnehmung im frühen Spracherwerb. *Sprache Stimme Gehör*, 35, 91–98.
- Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84(2), 127–190.
- Spillmann, H. O. (2000). *Einführung in germanistische Linguistik*. Langenscheidt.
- Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, 86(1), 142–164.
- Steinhauer, K., & Friederici, A. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, 30(3), 267–295.
- Stolterfoht, D., Friederici, A. D., Alter, K. & Steube, A. (2007). Processing focus structure and implicit prosody during reading: Differential ERP effects. *Cognition*, 104(3), 565–590.
- Veenendaal, N. J., Groen, M. A., & Verhoeven, L. (2015). What oral text reading fluency can reveal about reading comprehension. *Journal of Research in Reading*, 38(3), 213–225.

Christopher Sappok

Exploring Advanced Prosody – eine Best-Practice-Untersuchung zum lauten Lesen in der weiterführenden Schule

1 | Einführung

Der vorliegende Beitrag stellt eine Ratingprozedur mit drei Skalen zur Erhebung von fortgeschrittener Vorlesekompetenz und ihren prosodischen Komponenten vor. Anhand von 156 kurzen Aufnahmeausschnitten wurden „Vorlesewettbewerb“ (Skala C) und als hypothetische Komponenten von C die Faktoren „Stimmführung“ (Skala B) und „Hervorhebung“ (Skala A) erhoben. Dies geschah zum einen, um die Skalen zu evaluieren und zum anderen, um unseren Begriff von voll ausgebildeter Leseflüssigkeit zu konkretisieren¹. Generell gilt nach aktuellem Forschungsstand nach wie vor das *Henne-Ei-Dilemma* (Nix 2011; Kuhn, Schwanenflugel & Meisinger 2010): Hilft *reading prosody* beim Textverstehen oder setzt *reading prosody* ein Textverstehen voraus? Damit stellt sich die Frage, welchen über sich selbst hinausweisenden, also auf Textverstehen bezogenen diagnostischen Wert die Ermittlung und welchen didaktischen Wert die Vermittlung von Vorlese-Skills hat (Funke 2018).

Spätestens mit Holle (2006) wird die Relevanz von *reading aloud* oder *oral reading* (Daane et al. 2005) auch im deutschsprachigen Raum immer höher gewertet (Rosebrock & Nix 2006; Rosebrock et al. 2016). Untersucht werden in diesen Kontexten hauptsächlich Audiodaten lesender Grundschüler*innen (Sappok & Fay 2018; Röttig et al. und Stephany et al. in diesem Band). Längsschnittstudien (Röttig et al. in diesem Band) sind die Ausnahme und für die Zeit nach der Grundschule liegen kaum empirische Untersuchungen vor, auch international (Godde, Bosse & Bailly 2020). Im vorliegenden Beitrag werden Schüler*innen untersucht, die in einer früheren Untersuchung (Sappok, Linnemann & Stephany 2020) als besonders flüssig identifiziert wurden (Level 4 auf der Pinnell-Skala, s. Abbildung 1). Die Untersuchung einer Best-practice-Stichprobe soll Aufschluss darüber geben, was flüssige Leser*innen bereits können und worin sie sich dabei immer noch unterscheiden. Dies geschieht auch in Hinblick auf die Operationalisierung in Form von signalphonetischen und damit in Zukunft auch automatisch analysierbaren Parametern. Vor diesem Hintergrund sollen zunächst einige Vorüberlegungen angestellt werden.

¹ Die Audiodaten können für wissenschaftliche Zwecke beim Autor zum Download angefordert werden.

| | |
|--|----------------|
| <p>4 = Level 4: Die Schülerin / der Schüler liest weitgehend in größeren, semantisch sinnvollen Worteinheiten. Trotz gelegentlicher Rücksprünge im Text, Wiederholungen und Abweichungen wird dem übergeordneten Verlauf des Textes Rechnung getragen. Ein Großteil des Textes wird expressiv interpretiert (verschiedene Lesegeschwindigkeiten, Lautstärken, Stimmlagen, emotionale Beteiligung etc.).</p> <p>3 = Level 3: Die Schülerin / der Schüler liest überwiegend in Dreier- oder Vierer-Wortgruppen; gelegentlich treten auch kleinere Wortgruppen auf. Die Mehrheit der Wortgruppierung ist (trotzdem) angemessen und entspricht der Syntax des Textes. Wenig oder keine expressive Interpretation (Verschiedene Lesegeschwindigkeiten, Lautstärken, Stimmlagen, emotionale Beteiligung etc.).</p> <p>2 = Level 2: Die Schülerin / der Schüler liest überwiegend in Zweier-Wortgruppen. Dreier- und Vierer-Wortgruppen treten gelegentlich auf. Ab und zu kommt auch ein Wort-für-Wort Lesen vor. Die Wortgruppierungen erscheinen ungeschickt und stehen in keinem Zusammenhang zur Syntax des Textes.</p> <p>1 = Level 1: Die Schülerin / der Schüler liest den Text hauptsächlich Wort für Wort. Nur gelegentlich treten Zweier- oder Dreier-Wortgruppierungen auf. Die wenigen Wortgruppierungen sind unregelmäßig und unterstützen nicht die Syntax des Textes.</p> | <p>ICC = 1</p> |
|--|----------------|

Abb. 1: Die NAEP-Leseflüssigkeitsskala in adaptierter Form (Pinnell et al. 1995; Rosebrock & Nix 2015). Drei von drei Rater*innen haben die unten untersuchten Schüler*innen mit NAEP-Fluency = 4 bewertet (Sappok, Linnemann & Stephany 2020), was einem Reliabilitätswert von (post hoc) ICC = 1 entspricht.

Der Titelbegriff *advanced prosody* setzt einen *early prosody*-Begriff voraus. Für den deutschsprachigen Raum kann die prosodische Dimension des Konstrukts Leseflüssigkeit mit *early prosody* insoweit gleichgesetzt werden, als dass Leseflüssigkeit überwiegend im Primarkontext bzw. im Kontext frühen Textverstehens fokussiert wird. Neuere Erkenntnisse sprechen jedoch dafür, hiervon *advanced prosody* dezidiert zu unterscheiden.

Einen weitgehend anerkannten Ausgangspunkt stellt die Fluency-Definition mit drei Dimensionen von Kuhn, Schwanenflugel und Meisinger (2010, Herv.: CS) dar:

„Fluency combines *accuracy, automaticity, and oral reading prosody*, which, taken together, facilitate the reader’s construction of meaning. It is demonstrated during oral reading through ease of word recognition, appropriate pacing, phrasing, and intonation. It is a factor in both oral and silent reading that can limit or support comprehension.“

Groen, Veenendaal & Verhoeven (2018) kritisieren an dieser Definition, dass *automaticity* und *reading prosody* ein gleichrangiger Stellenwert in Bezug auf *reading comprehension* zugewiesen wird. Sie verglichen eine Gruppe niederländischer, älterer *poor comprehenders* (Jahrgangsstufe 5), die dabei altersgemäß automatisiert lasen, mit einer gleichaltrigen Gruppe, bei der auch *comprehension* altersgemäß war. Die Autor*innen stellten fest, dass bei den *poor comprehenders* auch der Faktor *text reading prosody*, der sehr differenziert erhoben wurde, „poor“ ausgeprägt ist, und fassen zusammen:

„It has been proposed that reading fluency – as a combination of accuracy, automaticity and text reading prosody – facilitates the reader’s construction of meaning (Kuhn et al. 2010). The results from the current study [...] suggest that the ‘automaticity aspect’ of reading is a distinct process from the construction of meaning. The construction of meaning seems more closely tied to text reading prosody than to decoding efficiency, at least, when children have mastered automaticity in reading” (Groen, Veenendaal & Verhoeven 2018: 16).

Die zuletzt genannte Bedingung kann für die Stimmen, die auf den Aufnahmen der vorliegenden Untersuchung zu hören sind, als voll erfüllt gelten. So können prosodische Merkmale in den Blick genommen werden, ohne dass ihre Wahrnehmung durch andere Defizite gestört wird. Zur Vertiefung soll zunächst eine differenzierte Klärung zum allgemeinen Prosodiebegriff vorgenommen werden.

Um Prosodie und deren Verhältnis zu linguistischen Einheiten zu verdeutlichen, ist die Unterscheidung von vier *perceptual domains* von Laver (1994) die beste Grundlage: „There [are] only four perceptual domains available to the human auditory system for differentiating the elements of speech. These [are] the domains of perceptual quality, duration, pitch and loudness” (Laver 1994: 431).

Quality steht hier für das, *was* gesagt wird und damit für den linguistischen Gehalt, wie ihn der Wortlaut des Textstimulus liefert. Prosodie hingegen steht als Sammelbegriff für die prosodischen Basismerkmale *duration* (Dauerphänomene), *pitch* (Tonhöhenbewegung) und *loudness* (Lautheitsphänomene), kurz: dafür, *wie* etwas gesagt wird.

Abbildung 2 stellt die drei prosodischen Basismerkmale als nominal skalierte Z-Achse in erweiterte Zusammenhänge.

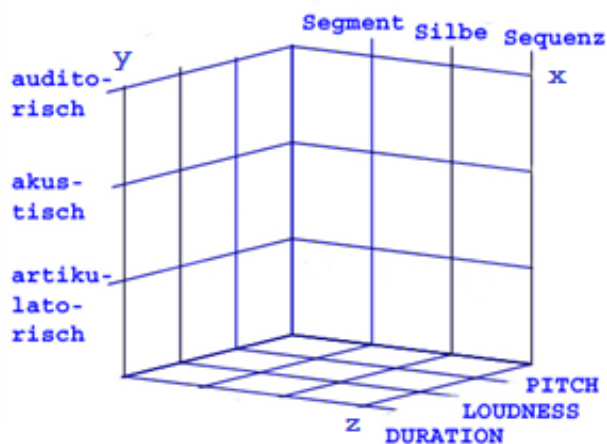


Abb. 2: Prosodie als dreidimensionales Konstrukt. Die X-Achse bezeichnet Größenordnungen (Tillmann & Mansell 1980), die Y-Achse bezeichnet Korrelate (Pompino-Marschall 2009) und die Z-Achse bezeichnet Basismerkmale (Laver 1994) von Prosodie. Es ergibt sich ein Raum zur Verortung von Bestimmungstücken prosodischer Forschung.

Die X-Achse Größenordnung unterscheidet nicht nur die Bezugseinheiten Segment, Silbe und Sequenz, sondern auch unterschiedliche Zeitfenster der kognitiven Sprachverarbeitung (Miller 1956; Kegel 1990; Pöppel 1997; Sappok 2010). Geht man hier hinsichtlich der Fensterbreite von

Segment = 30 ms, Silbe = 300 ms und Sequenz = 3000 ms aus, kann die X-Achse als logarithmisch skaliert gelten. Sie geht auf die Prosodiekonzeption von Tillmann zurück (Tillmann & Mansell 1980).

Hinter der Y-Achse stehen drei Typen von phonetischen Korrelaten. Diese Unterscheidung leitet sich aus der nachrichtentechnischen Unterscheidung von Sender, Kanal und Empfänger her (Weaver & Shannon 1949). Auch wenn das Kommunikationskettenmodell eine zeitliche Abfolge suggeriert, ist die Y-Achse als nominal skaliert zu bezeichnen, weil die Einteilung vor allem Methodenunterschiede bei der artikulatorischen, akustischen und auditiven Phonetik (Pompino-Marschall 2009) in den Vordergrund rückt, die in keinem quantifizierbaren Verhältnis zueinander stehen.

Damit ergeben sich 27 hypothetische prosodische Einzelmerkmale als Punkte im XYZ-Raum. Sie sind nicht alle gleichermaßen relevant, bieten aber einen differenzierten Rahmen zur Verortung von Bestimmungstücken prosodischer Forschung allgemein. Gravierend zu nennen ist die Einschränkung, dass pitch, loudness und duration hier in allen Kontexten gleichermaßen isoliert dastehen. Mit aufsteigender Größenordnung gilt jedoch mehr und mehr, dass die drei Basismerkmale zusammenwirken. Dennoch wird pitch in der internationalen Prosodieforschung eine besonders große Rolle zugeschrieben und auch in Hinblick auf advanced prosody kann pitch als besonders aufschlussreich gelten.

In der Grundlagenforschung zur Schnittstelle von Prosodie und Sprachverstehen steht die Untersuchung explizit vorgelesener Sprache nicht im Vordergrund (eine Ausnahme ist Blaauw 1995), sondern Sprache im Diskurs. Einen aktuellen Forschungsüberblick hierzu bietet Dahan (2015). Dort heißt es:

“Prosody is characterized as an abstract structure composed of discrete tonal elements aligned with the segmental composition of the sentence organized in constituents of increasing size, and this structure is influenced by the phonological, syntactic, and informational structures of the sentence” (Dahan 2015: 441).

Hervorzuheben ist hier, dass *informational structure* (oder Fokus, vgl. Fisseni 2011) besonders nah mit Textverstehen und einer „singgestaltenden“ oder gar „sinnstiftenden“ Leseprosodie in Verbindung gebracht werden kann.

Ein Modell zu pitch, das im Kontext der Sprachsyntheseforschung besonders einflussreich ist, ist das Fujisaki-Modell (Fujisaki 1988; für das Deutsche: Möbius 1993, s. Abbildung 3), an dem sich die vorliegende Untersuchung orientiert. Im Fujisaki-Modell wird pitch als das additive Zusammenwirken von zwei Steuermechanismen, *phrase component* (Sequenz) und *accent component* (Silbe), modelliert. Addiert ergeben die Outputs dieser Mechanismen Grundfrequenz, also einen akustischen Parameter (F0). Die erste Komponente liefert Eckparameter für eine globale Tendenz (T01 und T02 in Abbildung 3), die meist abfallend ist („Deklination“, Pompino-Marschall 2009). Die zweite liefert Eckparameter für diskrete Auslenkungsintervalle (T11 bis T32 in Abbildung 3). Das additive Zusammenspiel der Komponenten wird Superimposition genannt (Fujisaki 1988).

Im vorliegenden Kontext ist für die Accent-Komponente die Skala A zuständig, beschränkt sich aber im Gegensatz zum Fujisaki-Modell auf einen Hauptakzent. Skala B nimmt dem Modell entsprechend die Phrasenkomponente in den Blick.

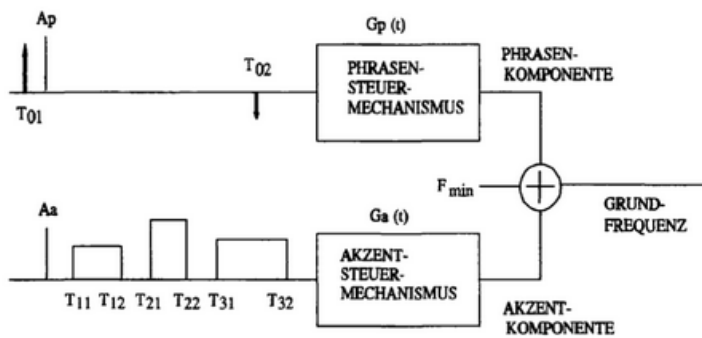


Abb. 3: Fujisaki-Modell mit zwei additiven Komponenten von Grundfrequenz (Fujisaki 1988, Reproduktion aus Möbius 1993: 68).

Geht es nun speziell um das laute Lesen in didaktischen Kontexten, spielen der Text und seine spezifischen prosodischen Herausforderungen eine große Rolle. In Sappok & Fay (2018) wurde mit dem einfacheren narrativen Text „Der Hase und der Schneemann“ mit 12 Drittklässler*innen gearbeitet. Erhoben und über signalphonetische Messungen rekonstruiert wurde, wie diese Schüler*innen Sprecherwechsel in reinem Dialog gestalten (Diskursgliederung). Hierfür wurde jede Gesamtaufnahme in ca. 2 bis 4 Sekunden lange Doppelsequenzen geschnitten. Zu den Doppelsequenzen wurden Ratings durchgeführt und signalphonetisch mit Bezug zu pitch, loudness und duration akustische Korrelate von Diskursgliederung gemessen (Stimme-Verstellen und Pausen). Ein wesentlicher Befund bestand darin, dass auch innerhalb von Pinnell-Level 4 ein breites Leistungsspektrum aufgezeigt werden konnte. Unter ähnlichen methodischen Bedingungen wurde die hier vorgestellte Ratingprozedur organisiert. Hierauf wird nun kurz eingegangen:

Der Text „Dolmetscher“ (s. drittes Unterkapitel), der im Zentrum der vorliegenden Untersuchung steht, kann im Gegensatz zu „Der Hase und der Schneemann“ (Sappok & Fay 2018) als besonders schwierig gelten. Statt prosodischer Diskursgliederung steht mit Skala C eine maximal holistische Vorlesekompetenz als nicht weiter in Frage gestellte abhängige Variable im Zentrum. Vorlesekompetenz ist hier also genau das, was die Vorlesekompetenz-Skala misst.

Vor diesem Hintergrund lässt sich für die vorliegende Untersuchung zusammenfassen:

- zentrale Hypothese: Die erhobenen Ratingdaten können zu einem explorativen Modell nach dem Muster $C \sim B + A + \text{weitere Faktoren}$ integriert und methodisch und didaktisch weiterführend interpretiert werden.

Die Differenzierung dieser Komponenten stellt hohe methodische Anforderungen. Dies gilt besonders für Faktor A (Sappok 2002; Sappok 2010; Sappok & Arnold 2012a, 2012b).

Aus dieser Grundlagen-Sicht heraus bzw. im Hinblick auf eine Evaluation der Skalen werden neben den Lautlese-Performanzen von Schüler*innen auch die Hör-Performanzen von Rater*innen (N = 39 Lehramtstudierende) genauer in den Blick genommen. Im Hintergrund stehen dabei zwei Annahmen, die als Heurismen wie folgt formuliert werden können:

- Heurismus I: Das besondere Problem bei der diagnostischen Analyse von Audiodaten besteht eher in einem Überangebot an potenziell relevanter Information als in einem Defizit.

- Heurismus II: Vorlesekompetenz und ihre Komponenten lassen sich effektiver auf Phrasen-Ebene anstatt auf Ebene des gesamten Textes erfassen.

Die Motivation des Schneidens *in* bzw. Beurteilen-Lassens *von* 2-4 Sekunden Sequenzen besteht darin, dass damit Entscheidungen stimuliert werden, die direkt aus dem Arbeitsgedächtnis abgerufen werden können (entspricht einem Fenster von etwa 3 Sekunden „unmittelbarer Gegenwart“, Pöppel 1997). Das Rating beschränkt sich also auf die Sequenzebene und erspart es dem Hörenden, über den ganzen Text zu integrieren (eine Textebene gibt es wohlweislich nicht in Abbildung 2). Erwartet wird so eine höhere Differenziertheit der Ratings.

Aus weiterführend-didaktischer Sicht stehen hier wie bei Sappok & Fay (2018) im Hintergrund zwei Annahmen, die als spekulative Axiome wie folgt formuliert werden können:

- Axiom I: Kompetenz beim lauten Lesen ist ein Indikator für Aspekte von Textverstehen.
- Axiom II: Umgekehrt können solche Aspekte über die Förderung von lautem Lesen mitgefördert werden („bootstrapping“).

In vorliegender Untersuchung steht die Diagnostik im Vordergrund, über die Axiome kann nur weiterführend spekuliert werden. Dies wird unter Einbeziehung der didaktischen Perspektive im fünften und letzten Unterkapitel aufgegriffen. Zunächst (Unterkapitel 2) werden Stichprobe und Aufnahmebedingungen näher beschrieben. Dazu werden die Anknüpfungspunkte der Einführung zu konkreteren Hypothesen weiterentwickelt. Das dritte Unterkapitel behandelt den Textstimulus und erläutert die drei Ratingskalen sowie deren computergestützten Einsatz bei der Evaluation. Im vierten Unterkapitel werden die dabei erhobenen Ratingdaten analysiert und interpretiert. Im Zentrum des Abschlusskapitels steht, wie gesagt, die Deutschdidaktik.

2 | Stichprobe und Hypothesen

Die 26 Aufnahmen von 16 Schüler*innen, die hier untersucht werden, stammen aus einem Korpus mit der Bezeichnung LAUDIO (= „Longitudinal Audio“), das insgesamt knapp 1000 Aufnahmen von 61 Schüler*innen enthält, die vom Autor dieses Beitrags über einen Zeitraum von mehreren Jahren beim lauten Lesen aufgezeichnet wurden. Dabei kamen verschiedene Textstimuli zum Einsatz, die eine große Bandbreite von relevanten Textmerkmalen repräsentieren (Schwierigkeitsgrad, Länge, Genre usw.). Bei dem Textstimulus „Dolmetscher“ (im Folgenden: DOL), mit dem die hier untersuchten Aufnahmen entstanden sind, handelt es sich um einen kurzen, aber vergleichsweise anspruchsvollen Sachtext, der im Gegensatz zu den anderen verwendeten Texten als Ankertext bei allen Aufnahmesitzungen zum Einsatz kam (außer wenn ad hoc eindeutig entschieden werden konnte, dass dieser Text eine massive Überforderung darstellen würde, v. a. in Jahrgangsstufe 3). Auf diesen Text wird im nächsten Unterkapitel näher eingegangen. Im Folgenden wird zunächst die Stichprobe der gesamten LAUDIO-Datenerhebung und die Auswahl der hier untersuchten Aufnahmen näher beschrieben.

Die LAUDIO-Datenerhebung begann gegen Ende des ersten Halbjahres 2014/2015. Es wurden zwei dritte und zwei vierte Klassenverbände einer inklusiven Grundschule im ländlichen Raum von Rheinland-Pfalz aufgenommen, und zwar immer in Einzelsitzungen mit dem Versuchsleiter (CS) und immer mit der mündlichen Instruktion: „Hier hast du einen Text. Lies mir den Text jetzt bitte mal vor.“ bzw. beim zweiten Lesen – stets unabhängig von der Qualität der eben erfolgten

Performance: „Super gemacht! Lies mir den Text jetzt bitte noch mal vor, da klappt’s dann bestimmt *noch* besser.“

Insgesamt hatten die vier Klassenverbände 73 Schüler*innen, wovon zum ersten Erhebungszeitpunkt 61 Schüler*innen an der Untersuchung teilnahmen. Die Differenz ist auf fehlendes Einverständnis der Eltern oder auf krankheitsbedingtes Fehlen zurückzuführen und wird im Folgenden als zufallsbedingt erachtet. Begleitend zur ersten Audio-Datenerhebung wurden die jeweiligen Deutschlehrkräfte gebeten, zu jedem Kind Einschätzungen in Bezug auf acht Teilkompetenzen des Deutschunterrichts auf der klassischen Notenskala von *sehr gut* bis *ungenügend* abzugeben. Erfragt wurden hier die Aspekte *Leseflüssigkeit*, *Leseverstehen*, *Schreibkompetenz Texte*, *Rechtschreibung*, *Gesprächskompetenz*, *Zuhörkompetenz*, *Sprachreflexionskompetenz* und *Wortschatz* (vgl. KMK 2004, Mittelwerte hieraus erscheinen als Deutschleistung in Tabelle 1). Außerdem wurden Einschätzungen zum Sprachniveau Deutsch, zur Sozialkompetenz (hier nicht einbezogen) und zum sozioökonomischen Hintergrund (hier nicht einbezogen) der Kinder eingeholt. Zusätzlich wurden mit allen jeweils aufgenommenen Kindern zum ersten und allen folgenden Erhebungszeitpunkten schriftliche Tests zur Erfassung verschiedener Aspekte von Lesekompetenz durchgeführt, nämlich ELFE 1 - 6² (Lenhard & Schneider 2006) und SLS 2 – 9 (Wimmer & Mayringer 2014).

In den anschließenden drei Jahren wurden möglichst viele dieser Schüler*innen in durchschnittlich etwa 15-minütigen Einzelsitzungen einmal jährlich erneut aufgenommen, so dass mit dem Ende des Erhebungszeitraums Anfang 2018 Längsschnittdaten zu einer „Kohorte 3 bis 6“ und einer „Kohorte 4 bis 7“ vorliegen. Für eine erste Überblicksuntersuchung (Sappok, Linnemann & Stephany 2020) wurden aus der ursprünglichen Stichprobe insgesamt 31 Schüler*innen ausgewählt ($n_{3-6} = 13$, $n_{4-7} = 18$) und deren DOL-1-Performanzen analysiert, d. h. zum ersten Erhebungszeitpunkt „prima vista“ bzw. „prima vista seit letztem Jahr“ für alle weiteren Erhebungszeitpunkte. Einbezogen wurden in Sappok, Linnemann & Stephany (2020) genau diejenigen Schüler*innen, von denen mindestens zum ersten und zum letzten der vier Erhebungszeitpunkte mindestens eine DOL-Aufnahme vorliegt. Die Ausfälle kommen dadurch zustande, dass zum ersten Erhebungszeitpunkt keine DOL-Aufnahme gemacht wurde (extrem schwache Schüler*innen). Abgesehen davon werden alle anderen Ausfallgründe für die Analyse als zufallsbedingt erachtet (dass einzelne Kinder auf weit entfernte weiterführenden Schulen übergegangen sind, dass das Elterneinverständnis entzogen wurde usw.).

Um diese Zufallsbedingtheit zu überprüfen und damit auch, inwiefern diese Stichprobe trotz der Ausfälle als repräsentativ angesehen werden kann, wurden die ELFE-Testdaten des ersten Erhebungszeitpunkts mit den betreffenden Angaben zur Normierungsstichprobe von ELFE verglichen. Der Vergleich zeigte, dass die Stichprobe keine nennenswerten Abweichungen aufweist, außer dass die Streuung bei den Drittklässler*innen beim Wort-, Satz- und Textverstehen überall leicht höher, bei den Viertklässler*innen hingegen leicht niedriger ausfällt als bei den Normierungen. Insgesamt spricht also wenig gegen die Annahme, dass die Ausprägungen der verschiedenen Aspekte von Lesekompetenz innerhalb der DOL-Stichprobe aus vier Klassenverbänden ähnlich verteilt sind wie in den beiden Grundgesamtheiten (Drittklässler*innen und Viertklässler*innen).

² Nur bei den Schüler*innen der Kohorte 4 bis 7 zum letzten Erhebungszeitpunkt wurde der ELFE 1 – 6 nicht durchgeführt, da er nicht mehr indiziert ist.

| Ko- horte | Schüler* innen-ID | Ge- schl. | Muttersp. Deutsch | Deutsch- leistg. | weit. Schule | Jgst 3 | Jgst 4 | Jgst 5 | Jgst 6 | Jgst 7 |
|--------------|----------------------|--------------|----------------------|---------------------|-----------------|-----------|-----------|-----------|-----------|-----------|
| 3 - 6 | AIG-03 | w | 1 | 1, 0 | GYM | x | x | x | x | |
| | AIG-09 | m | 1 | 2, 3 | RS+ | x | x | x | x | |
| | AIG-17 | m | 1 | 1, 1 | GYM | x | x | x | x | |
| | AIG-19 | m | 1 | 1, 5 | GYM | x | x | x | x | |
| | AIG-23 | w | 1 | 2, 0 | RS+ | x | | x | x | |
| 4 - 7 | AIG-42 | m | 1 | 2, 0 | GYM | | x | | x | x |
| | AIG-46 | m | 1 | 2, 4 | RS+ | | x | | x | x |
| | AIG-48 | w | 1 | 2, 4 | RS+ | | x | | x | x |
| | AIG-51 | w | 3 | 2, 8 | RS+ | | x | | x | x |
| | AIG-55 | w | 1 | 1, 6 | GYM | | x | | x | x |
| | AIG-57 | w | 1 | 1, 6 | RS+ | | x | | x | x |
| | AIG-61 | m | 1 | 1, 9 | GYM | | x | | x | x |
| | AIG-63 | m | 1 | 2, 0 | RS+ | | x | | x | x |
| | AIG-68 | w | 1 | 2, 9 | RS+ | | x | | x | x |
| | AIG-72 | w | 1 | 2, 6 | GYM | | x | | | x |
| AIG-73 | w | 2 | 2, 5 | GYM | | x | | x | x | |

Tab. 1: Merkmale und Kontexte der 26 untersuchten Aufnahmen. In den Zeilen sind solche Schüler*innen aufgeführt, von denen mind. eine Aufnahme bei maximaler Übereinstimmung auf Pinnell-Level 4 eingeordnet wurde (grau unterlegt). Ohne Unterlegung angekreuzt sind vorliegende Aufnahmen, die keine derart optimale Bewertung erhalten haben. Weiterhin aufgeführt sind die wichtigsten Rahmendaten zu den beteiligten Schüler*innen: Deutschkenntnisse und Deutschleistung wurden von der jew. Lehrkraft zum ersten Erhebungszeitpunkt eingeschätzt, und zwar einmal auf einer Skala von 1 (= Muttersprache) bis 4 (= kaum vorhanden) und einmal auf Schulnotenskala als Mittelwert aus acht Einzeleinschätzungen (s. o.). Bei der weiterführenden Schule steht GYM für Gymnasium und RS+ für Realschule Plus.

Von jeder/jedem der 31 Schüler*innen liegen aufgrund von vereinzelt Ausfällen innerhalb des Erhebungszeitraums im Durchschnitt drei (statt idealerweise vier) Aufnahmen vor, bei denen der DOL-Text gelesen wurde (insgesamt 99 Aufnahmen). Diese Aufnahmen wurden in Sappok, Linnemann & Stephany (2020) hinsichtlich Automatisiertheit und Akkuratheit analysiert sowie im Rahmen einer größer angelegten Ratingprozedur (insgesamt kamen 256 Aufnahmen und N = 51 Rater*innen zum Einsatz) mit Lehramtsstudierenden des Faches Deutsch einer umfassenden Bewertung unterzogen. Auf die verschiedenen Kriterien und näheren Umstände soll hier nicht erschöpfend eingegangen werden (Einzelheiten s. ebd.). Wichtig ist für den vorliegenden Kontext nur, dass jede Aufnahme von drei Rater*innen bewertet wurde. Dabei wurde u. a. die vierstufige NAEP-Skala zur globalen Erfassung von Leseflüssigkeit (Pinnell et al. 1995, Daane et al. 2005) angewendet, und zwar in der von Rosebrock & Nix (2015) für das Deutsche adaptierten Form³. Die Inter-Rater-Reliabilität erwies sich als hoch (ICC = .80).

Für die vorliegende Untersuchung wurden nun aus den 99 Aufnahmen genau diejenigen ausgewählt, die von allen dreien der jeweiligen Rater*innen auf Leseflüchtigkeitslevel 4 eingestuft

³ Einziger Unterschied: Die Stufenbeschreibungen von Rosebrock & Nix (2015) sind auf narrative Texte ausgerichtet und es wird auf das Gelesene als „die Geschichte“ referiert. Dies wurde im vorliegenden Kontext durch „der Text“ ersetzt, da ein Sachtext beurteilt werden sollte.

wurden. Dies war bei 26 Aufnahmen der Fall (Tabelle 1). Diese Aufnahmen zeichnen sich also dadurch aus, dass sie im Rahmen der verwendeten Skala mit maximaler Übereinstimmung die maximal positive Bewertung erhalten haben (post hoc ICC = 1, s. Abbildung 1).

Der so definierte Auswahlprozess bringt also ein Höchstmaß an „positiver Homogenität“ bzgl. Leseflüssigkeit auf der NAEP-Skala mit sich. Damit geht ein hohes Maß an Heterogenität im Hinblick auf andere Aspekte einher. So beinhaltet die Teilmenge Aufnahmen von sämtlichen untersuchten Zeitpunkten (Jahrgangsstufe 3 bis 7) und einzelne Schüler*innen sind mit einer bis hin zu vier Aufnahmen vertreten. Tabelle 1 fasst wichtige Merkmale der untersuchten Aufnahmen in einem Schüler*innen-zentrierten Arrangement zusammen

Auch wenn bei der „Ziehung“ dieser Stichprobe der Zufall nicht unter voll kontrollierten Bedingungen gewaltet hat, kann davon ausgegangen werden, dass damit ungefähr die langfristig lesestärkere Hälfte aus den ursprünglichen vier Grundschulklassen repräsentiert ist (unter Ausblendung von Schüler*innen mit Inklusionshintergrund). Als Ergebnis der Vorarbeiten kann diese Feststellung ins Negative gewendet und als weiterführende Hypothese folgendermaßen zusammengefasst werden:

- Hypothese 1: Aus ca. 50% aller Grundschülerinnen und Grundschüler werden bis Mitte Sekundarstufe I keine flüssigen Leser*innen.

Dieser Hypothese lässt sich entgegenhalten, dass die Stufenbeschreibung von Level 4 maximal hohe Anforderungen aufführt, indem mit expressiver Interpretation auf der Ebene des prosodischen Lesens nicht nur ein sinn- und syntaxgemäßes, sondern darüber hinaus auch ein sinn-gestaltendes Lesen eingefordert wird (s. Abbildung 1). Eigene Vorarbeiten haben allerdings mehrfach gezeigt, dass, zumindest von studentischen Rater*innen, Level 4 relativ wenig streng zugeteilt wird, wenn das Referenzfeld wie hier breit ist. So muss bei der Pinnell-Skala von einem ausgeprägten Deckeneffekt gesprochen werden (Sappok & Fay 2018, Sappok, Linnemann & Stephany 2020). Diese Feststellung kann wie folgt als weiterführende Hypothese ausformuliert werden:

- Hypothese 2 a) Das Überschreiten der Schwelle zu NAEP-Level 4 bedeutet, lediglich early prosody erworben zu haben.
- Hypothese 2 b) Innerhalb von NAEP-Level 4 (advanced prosody) ist ein breites Spektrum aus-zudifferenzieren.

Wichtig für das weitere Vorgehen ist die Unterscheidung, dass weniger eine Stichprobe von Schüler*innen, wie im Arrangement von Tabelle 1 aus Übersichtlichkeitsgründen nahegelegt, als eine Stichprobe von Aufnahmen behandelt wird. Aus dieser Perspektive stellt das lesende Individuum ein eher nachrangiges von mehreren Merkmalen von Aufnahmen dar. Die Aufnahmenstichprobe repräsentiert somit die Grundgesamtheit von solchen „Lautleseereignissen“ (bzw. Lernständen zu einem bestimmten Zeitpunkt) aus dem „Altersraum“ bis Mitte Sekundarstufe I, die mit maximaler Eindeutigkeit auf Pinnell-Level 4 (s. Abbildung 1) eingeordnet werden können. Diese Perspektivsetzung hat diverse Konsequenzen für das methodische Vorgehen. Ein Vorteil besteht z. B. darin, dass die Lückenhaftigkeit der Datenerhebung statistisch umso weniger ins Gewicht fällt (leere Felder in Tabelle 1), je stärker das lesende Individuum und auch dessen Jahrgangsstufe in den Hintergrund gerückt werden.

3 | Eigenschaften des Textes und der Skalen

Der als Lesestimulus eingesetzte Text stammt aus einem Schülerlexikon mit der Altersangabe „ab 9 Jahren“ (Beuschel-Menze et al. 2009, Klappentext). Es handelt sich um den Eintrag zum Begriff „Dolmetscher“:

„**Dolmetscher** sind Übersetzer von einer Sprache in die andere. Während aber ein Übersetzer Texte nur schriftlich übersetzt, muss der Dolmetscher auch in mündlichen Verhandlungen übersetzen können. Ein Simultan-Dolmetscher übersetzt sogar laufend, während weitergesprochen wird. Das seltsame Wort ist aus der osmanisch-türkischen Sprache abgeleitet. Der *tilmac* war dort ein Vermittler.“ (aus „Schülerlexikon von A - Z“, Beuschel-Menze et al. 2009: 64, Herv. i. Orig.)

Der Text weist eine hohe Informationsdichte auf. Die durchschnittliche Satzlänge beträgt 9,8 Wörter, der Anteil an langen Wörtern (Buchstabenanzahl > 6) beträgt 42,8%. Es ergibt sich ein LIX-Wert (Lesbarkeitsindex nach Björnsson, zit. nach Lenhard & Lenhard 2014) von 52,6, was auf einen erhöhten Schwierigkeitsgrad hinweist. Die Ausführungen und Unterscheidungen des Textes sind vielschichtig, indem sie Merkmale einer anspruchsvollen Definition aufweisen: Im ersten Satz wird der Übersetzer-Begriff als *genus proximum* vorgestellt; im zweiten Satz wird der Gegensatz Schriftlichkeit-Mündlichkeit als *differentia specifica* zur Abgrenzung von einem damit implizit revidierten Übersetzerbegriff eingeführt, was als ziemlich problematisch gelten kann. Im dritten Satz wird über das seltene Wort „simultan“ noch einmal eine Ebene tiefer ausdifferenziert. Im vierten und fünften Satz wird eine etymologische Einordnung vorgenommen, wobei die Lautung des Ursprungsbegriffs „tilmac“ nicht eindeutig geklärt wird. Vor diesem Hintergrund muss der Text als *kein* guter Beitrag zu einem Schülerlexikon „ab 9 Jahren“ (Beuschel-Menze et al. 2009, Klappentext) angesehen werden. Gerade vor diesem Hintergrund aber, nämlich weil er so auch für Siebtklässler*innen noch schwierig genug ist, kann der Text als ein gut geeigneter Stimulus für die Diagnostik von advanced prosody gelten.

Dazu lässt sich festhalten, dass der Text einige bemerkenswerte prosodische Herausforderungen enthält. Die Verhältnisse auf Ebene der Informationsstruktur sind bei aller Komplexität relativ eindeutig gelagert, sodass sich ein Mitvollziehen dieser Verhältnisse seitens des Lesenden auf der Ebene der fokussierenden stimmlichen Hervorhebung bestimmter Wörter niederschlagen kann (z. B. „Dolmetscher sind Übersetzer von EINER Sprache in die andere.“ vs. „Dolmetscher sind Übersetzer von einer SPRACHE in die andere.“). Dies ist als besonders anspruchsvoller Aspekt von prosodischem Lesen bzw. als potenzieller Indikator für komplexe Aspekte von Textverstehen anzusehen: „By highlighting key information, good prosody in the reading of linguistic focus features can signal the ongoing construction of a discourse structure during fluent reading“ (Schwanenflugel, Westmoreland & Benjamin 2013, weitere Ausführungen hierzu in Sappok, Linnemann & Stephany 2020).

Die untersuchten Audioaufnahmen sind mit einem Headsetmikrofon vom Typ Sennheiser ME 3-ew entstanden, das mit einem digitalen Aufnahmegerät vom Typ Olympus LS-11 verbunden war. Das Format der Audiodateien ist überall Mono, 44.1-kHz, WAV 16-bit. Die Dauern der Aufnahmen liegen zwischen 20 und 34 Sekunden. Hierfür sind Unterschiede im Eigentempo der Schüler*innen sowie Stockungen und Wiederholungen (zwischen 0 und 6 Lesefehler pro Aufnahme) verantwortlich zu machen. Für die Audioanalyse im Rahmen der nun vorgestellten Ratingprozedur wurden sämtliche Aufnahmen an den Satzgrenzen in Sequenzen geschnitten, die im Folgenden als Audiophrasen bezeichnet werden sollen. Der längste Satz (Satz 2) wurde an

der Kommasetzung auch noch einmal geteilt, um die Ähnlichkeit in der Länge der somit resultierenden 6 Audiophrasen je Aufnahme zu erhöhen (Länge zwischen 17 und 22 Silben für die ersten fünf Phrasen; einzig die sechste Phrase (der letzte Satz) weicht mit einer Länge von 9 Silben nennenswert nach unten ab).

Alle Rater*innen hatten damit dieselben $26 * 6 = 156$ Audiophrasen zu beurteilen, allerdings nach unterschiedlichen Kriterien und in individuell randomisierter Abfolge. Es kamen drei Skalen zum Einsatz, mit denen drei gleichgroße Gruppen à 13 Rater*innen gearbeitet haben. Sämtliche Ratingsitzungen fanden gleichzeitig in einem Zeitfenster von ca. 60 min statt, und zwar an Rechnerarbeitsplätzen in einer mittels der Softwareumgebung Praat/ExperimentMfC (Boersma & Weenink 2020) automatisierten Ablauforganisation auf einer graphischen Benutzeroberfläche. Die Rater*innen erhielten ein Daten- und Instruktionpaket zum Download und arbeiteten zuhause. Sie waren angewiesen, möglichst hochwertige Kopfhörer einzusetzen. Über Modalität und Qualität der von den Rater*innen verwendeten Audiotechnik konnte darüber hinaus keine Kontrolle ausgeübt werden.

Der mausgesteuerte Bildschirm- und Audio-Ablauf bestand darin, dass alle Rater*innen nach einer teamspezifischen Start-Instruktionsfolie (s. Abbildung 4) zuerst den ersten Satz des DOL-Textes in 26 Audiophrasen in individuell randomisierter Reihenfolge präsentiert bekamen, dann den ersten Teil des zweiten Satzes in 26 Audiophrasen, dann den zweiten Teil des zweiten Satzes usw (s. Tabelle 2).

Der Start einer einzelnen Audiophrasenpräsentation bestand in einem automatischen Abspielen der Phrase. Dazu wurde die per Mausklick zu wählenden Option gegeben, die Phrase noch einmal zu hören (bis zu drei Wiederholungen waren möglich). Daneben war die entsprechende Fragestellung aufgeführt, zusammen mit Buttons zur Beantwortung. Wenn eine Beantwortung erfolgt war, konnte sie korrigiert oder aber es konnte per Mausklick zur nächsten Audiophrasen-Präsentation übergegangen werden, die dann wieder mit dem automatischen Abspielen startete. Die Rater*innen hatten in mehreren Online-Seminarsitzungen eine basale Einführung in den Umgang mit der Software Praat und den Kontext „Hörexperiment“ bekommen, wobei zum Kennenlernen des Textes auch schon andere DOL-Aufnahmen aus dem LAUDIO-Korpus zum Einsatz gekommen waren.

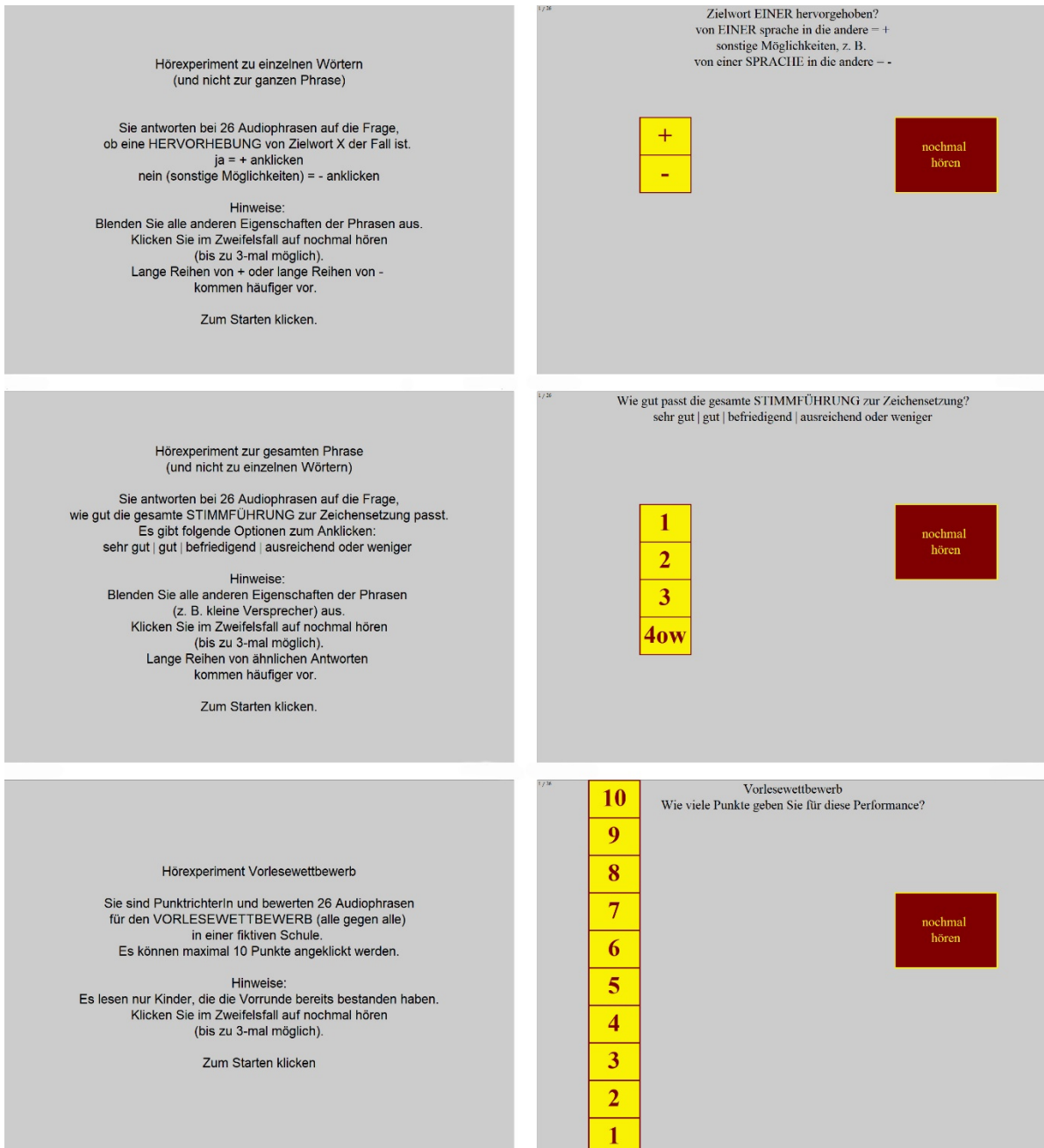


Abb. 4: Die unter ansonsten identischen Bedingungen zum Einsatz gekommenen Ratingskalen in Form der Bildschirmanzeige, mit der die jew. Sitzungen starteten und der jew. ersten Bildschirmanzeige zur ersten Audiophrasen-Präsentation.

Die Instruktionen A und B sind mit dem Ziel formuliert, bei den Rater*innen die Konzentration auf den erfragten Aspekt zu maximieren. Die Instruktion C ist im Gegensatz dazu mit dem Ziel formuliert worden, ein maximal holistisches Urteil zu erhalten. Die Fokussierung auf HERVORHEBUNG in Skala A zielt auf das Konzept der phonetischen Prominenz auf Wortebene ab. Bewusst verzichtet wurde hier auf die zusätzliche Fokussierung auf die semantische bzw. informationsstrukturelle Motiviertheit der anvisierten Prominenzverhältnisse, weil dies zu umständlich zu vermitteln und außerdem zu befürchten ist, dass ansonsten der Faktor des Rater*innen-

eigenen Textverständnisses störenden Einfluss gewinnen würde (im Sinne einer Rater*in-eigenen „Informationsstruktur-Kompetenz“). Im Ablauf wurde dann je Phrase die Hervorhebung eines bestimmten Wortes im Gegensatz zu dessen Nicht-Hervorhebung erfragt. Zusätzlich wurde jew. eine Gegenbeispielvariante angegeben. Die zugrundeliegende Bestimmung der Zielwörter war davon geprägt, dass deren Hervorhebung im Zentrum eines Prominenzmusters steht, das maximal im Sinne der Informationsstruktur ist (Muster „+ Informationsstruktur“). Im Einzelnen handelt es sich um folgende Muster und damit Angaben in den betreffenden Positionen der aktuellen Anzeige auf der Benutzeroberfläche (vgl. Abb. 4 A rechts):

| Phrase | Muster „+ Informationsstruktur“ | Muster „- Informationsstruktur“ |
|--------|----------------------------------|----------------------------------|
| 1 | von EINER sprache in die andere | von einer SPRACHE in die andere |
| 2-1 | nur SCHRIFTLICH übersetzt | nur schriftlich ÜBERSETZT |
| 2-2 | auch in MÜNDLICHEN verhandlungen | auch in mündlichen VERHANDLUNGEN |
| 3 | während WEITER gesprochen wird | während weiter GESPROCHEN wird |
| 4 | das seltsame WORT ist | das seltsame wort ist |
| 5 | der TILMAC war | der tilmac war |

Tab. 2: „Sinngestaltende Prosodie“. Die auf der binären Skala A erfragten Angaben zu den in der jeweiligen Audiophrase aktualisierten Prominenzverhältnissen.

Bei den Phrasen 4 und 5 wurde wegen mangelnder klarer Alternativen auf Ebene der Informationsstruktur für die Gegenüberstellung eine hervorhebungsneutrale prosodische Gestaltung und damit das anvisiert, was alltagssprachlich „monotones“ Lesen genannt wird. Die Buchstabenfolge <tilmac> wird als eine Art Pseudowort angesehen, von dem auch ohne eindeutige Lautung klar sein sollte, welche Rolle es im Satz spielt. Zusammenfassend lässt sich festhalten, dass es sich bei Skala A um eine exkludierende („sonstige Möglichkeiten“) binäre Skala handelt, die zunächst als zweistufige Nominalskala (Reliabilitätsprüfung über Light’s Kappa) behandelt wird (s. Tabelle 3).

Die Fokussierung STIMMFÜHRUNG bei Skala B bezieht sich auf die Gesamtkontur der Intonationsphrase; mit der zusätzlichen Fokussierung „satzzeichengemäß“ werden die von syntaktischen Zusammenhängen bestimmten Intonationsverhältnisse, v. a. Deklination (Pompino-Marschall 2009) fokussiert. Da die Satzzeichenverhältnisse bei allen Phrasen recht klar sind, wurde erwartet, dass eine Rater*innen-eigene „Satzzeichenkompetenz“ hier weniger störend sein würde als eine Rater*innen-eigene „Informationsstruktur-Kompetenz“ bei Skala A. Um einer durch den Begriff „Satzzeichen“ nahegelegten Beschränkung der Fokussierung auf das Ende oder verstärkt die zweite Hälfte der Audiophrase entgegenzuwirken, hieß die Instruktion, die bei jeder Präsentation erschien: „Wie gut passt die *gesamte* STIMMFÜHRUNG zur Zeichensetzung?“ (vgl. Abbildung 4 B rechts). Zusammenfassend lässt sich festhalten, dass es sich bei Skala B um eine trunkierte („ausreichend oder weniger“) Likert-Skala auf Ordinalniveau handelt, die im Folgenden aus methodischen Gründen als vierstufige Intervallskala behandelt wird (vgl. hierzu die Diskussion in Rietveld & Chen (2006: 303f.)).

Die Fokussierung VORLESEWETTBEWERB bei Skala C bezieht sich auf den oben formulierten Heurismus 1. Vorausgesetzt wird, dass schon eine isolierte Audiosequenz von wenigen Sekunden Dauer ausreichend diagnostische Information enthält, wenn es um die Kodierung einer globalen Vorlesekompetenz geht. Damit ist auch mit einem erhöhten Einfluss der Rater*innen-eigenen Zuhörkompetenz zu rechnen. Über die Hinweise „(alle gegen alle) in einer fiktiven Schule“ (vgl. Abb. 4 C links) sollte ein Casting-Szenario wie aus den Medien bekannt suggeriert werden. Dabei sollte dem Einfluss entgegengewirkt werden, dass die teilweise gut hörbaren Altersunterschiede bei den Audiophrasen in die Beurteilung als Gewichtung einfließen. Die Benutzeroberfläche bei den einzelnen Präsentationen war im Gegensatz zu A immer dieselbe und lautete „Wie viele Punkte vergeben Sie für diese Performance?“ (vgl. Abbildung 4 C rechts).

Drei der 39 Rater*innen lieferten technisch fehlerhafte Daten. Bei der Analyse von Inter-Rater-Reliabilität wurden dann zwei weitere Rater*innen auffällig, da sie in einer ihrer Skalen-Phrasenkombination immer denselben Wert angeklickt hatten, was zu einer Fehlermeldung wegen Nullvarianz führte. Die betroffenen Datensätze wurden manuell entfernt. Weitere Maßnahmen zu Identifikation und Ausschluss unkooperativer oder überforderter Rater*innen fanden nicht statt.

| | A | A | B | C | A | B | C |
|---------------|------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Phrase | Light's | ICC _{tcs} | ICC _{tcs} | ICC _{tcs} | ICC _{tca} | ICC _{tca} | ICC _{tca} |
| 1 | .28 | .26 | .33 | .59 | .82 | .86 | .95 |
| 2-1 | .29 | .31 | .39 | .55 | .86 | .89 | .93 |
| 2-2 | .27 | .28 | .25 | .47 | .84 | .81 | .92 |
| 3 | .21 | .23 | .30 | .57 | .79 | .85 | .94 |
| 4 | .11 | .10 | .36 | .70 | .59 | .88 | .96 |
| 5 | .16 | .19 | .33 | .60 | .75 | .87 | .94 |
| <i>Median</i> | .24 | .25 | .33 | .58 | .81 | .87 | .94 |

Tab. 3: Inter-Rater-Reliabilität der drei Skalen im Spiegel unterschiedlicher Indizes. Werte bei ICC_{tcs} < .5 lassen relativ eindeutig auf Verbesserungsbedarf bei der Skalenkonstruktion schließen.

Aufgrund der grundsätzlich symmetrischen und stark kontrollierten Bedingungen der Gesamt-Ratingprozedur konnten besonders differenzierte Analysen der Inter-Rater-Reliabilität durchgeführt werden. Für die binäre Skala A wurde Light's Kappa (nach Hallgren 2012) eingesetzt, ein Parameter für nominal skalierte Faktoren, dessen Algorithmus für jede mögliche Rater*innen-Paarung innerhalb eines Phrasensets einen einfachen Kappa-Wert und aus all diesen dann den Mittelwert liefert. Analysen wurden in der Softwareumgebung R (R Core Team 2014) mithilfe des Paketes „irr“ (Gamer, Lemon & Singh 2019) durchgeführt. Dabei liegen, wie bei den anderen im folgenden verwendeten Indizes, die Ergebnisse i. d. R. zwischen 0 (keine Übereinstimmung) bis 1 (maximale Übereinstimmung).

Die Resultate für Skala A müssen teilweise als nicht ausreichend für eine vollwertige weiterführende Analyse dieser Daten angesehen werden, zumal hier der naheliegendste nächste Schritt in einer Reduktion auf eine binäre Ausprägung pro Audiophrase besteht (Item gelöst vs. nicht gelöst). Die Phrasen 1 bis 3 erreichen Light's Kappa Werte zwischen .2 und .3 und damit das

Level „fair agreement“ (nach Hallgren 2012). Die Phrasen 4 und 5 liegen noch einmal merklich niedriger. Dies könnte darauf zurückzuführen sein, dass bei 4 und 5 im Ablauf keine eindeutige Hervorhebungsalternative aufgezeigt war (siehe Abbildung 4).

Die Inter-Rater-Reliabilität der Daten aller Skalen wurde außerdem mit ICC-Analysen geprüft. Es zeigte sich für Skala A eine starke Übereinstimmung mit Light's Kappa, sodass der unmittelbare Vergleich zu B und C möglich ist. Hierbei spielen zwei Perspektiven eine Rolle. Um Aufschlüsse zur Skalenkonstruktion und zu den Hörkompetenzen der Rater*innen zu erhalten, wurden die Parameter „twoway, consistency, single unit“ (= ICC_{tcs}) gesetzt, was zu konservativen Ergebnissen im Hinblick auf die hypothetische Reliabilität der Daten führt, aber feiner differenziert. Bei B zeigt sich bis auf Phrase 2-2 durchgängig eine etwas höhere Tendenz als bei A. Bei Skala C hingegen zeigt sich klar ein insgesamt höheres Niveau. Diese Verhältnisse lassen sich wie folgt zusammenfassen: $IRR.A < IRR.B \ll IRR.C$.

Mit Perspektive auf die weiterführende Analyse der Daten bzgl. der Lesekompetenzen der Schüler*innen ist zu berücksichtigen, dass das Setting mit sehr vielen Rater*innen darauf ausgelegt ist, die rohen Ratingdaten zu Audiophrasen-spezifischen Mittelwerten zusammenzufassen (s. nächstes Unterkapitel). Ausschlaggebend ist deshalb nach Hallgren (2012; vgl. Sappok & Fay 2018) die Analyse, bei der die Parameter „twoway, consistency, average“ (ICC_{tca} in Tabelle 3) gesetzt sind, was zu progressiven Ergebnissen im Hinblick auf die hypothetische Reliabilität der Daten führt. Alle ermittelten ICC_{tca} -Werte für die Skalen B und C sind größer als .75 und genügen damit hohen Ansprüchen („excellent“ nach Hallgren 2012).

Festgehalten werden kann somit, dass die Skala C anhand kurzer Audiosequenzen maximal reliable Daten für die weiterführende Analyse liefert. Die Inter-Rater-Reliabilität dürfte auch einer merklichen Verringerung der Rater*innen-Anzahl unter ansonsten günstigeren Bedingungen (z. B. Präsenz, Schulung, hochwertige Audiotechnik) standhalten. Geringer ist die Inter-Rater-Reliabilität für die Skala B einzustufen, was in der vorliegenden Untersuchung durch eine hohe Rateranzahl tendenziell aufgefangen wird. Die Skala A muss als wenig reliabel eingestuft werden, wobei Phrase 4 und Phrase 5 als besonders wenig reliabel identifiziert wurden. Die Ratingdaten A für diese Phrasen wurden für den zweiten Schritt der weiterführenden Analyse (Zusammenfassung zu Aufnahmenkennwerten, s. u.) ausgeschlossen.

4 | Ergebnisse

Für die Analyse werden die Ratingdaten in zwei Schritten durch Mittelwertbildung kondensiert und weiteren Merkmalen gegenübergestellt. Der erste Schritt fokussiert die 156 Audiophrasen (Mittelwert aus max. 13 Ratings pro Skala und Audiophrase), der zweite Schritt fokussiert die 26 Gesamtaufnahmen (Mittelwert aus max. 78 Ratings pro Skala). Im Zentrum stehen dabei die Ratings der Skala C und die Frage nach Hinweisen dazu, worauf sich eine mehr oder weniger hohe C-Punktzahl einer Aufnahme im Einzelnen zurückführen lässt.

4.1 | Ergebnisse auf Ebene der Audiophrasen

Pro Audiophrase liegen bzgl. der C-Skala 13 Einzelratings vor, die zu einem Mittelwert zusammengefasst wurden. Es ergibt sich eine Verteilung von 156 Werten mit global (s. Tabelle 4, unterste Zeile): $M = 6,74$, $SD = 1,48$, Range = 3,67 bis 9,42. Abbildung 5 zeigt diese Werte im Ranking. So erweist sich, dass die Skala breit und relativ gleichmäßig ausgenutzt wurde.

| Phrasennummer | Wörter Anzahl | Silben Anzahl | Einzel- ratings Anzahl | C M (SD) | B M (SD) | A Σ | E M (SD) | S M (SD) |
|---------------|------------------|------------------|------------------------------|----------------|----------------|---------------|------------------|----------------|
| 1 | 9 | 18 | 13 | 7,10 (1,54) | 7,64 (1,25) | 3 | - 0,04 (0,20) | 4,57 (0,69) |
| 2-1 | 8 | 17 | 13 | 7,08 (1,41) | 7,46 (1,37) | 16 | - 0,46 (0,65) | 4,40 (0,70) |
| 2-2 | 9 | 20 | 13 | 7,07 (1,21) | 7,68 (1,08) | 18 | - 0,38 (0,64) | 4,59 (0,80) |
| 3 | 9 | 22 | 13 | 5,53 (1,35) | 7,11 (1,20) | 16 | - 0,77 (0,65) | 3,86 (0,83) |
| 4 | 10 | 20 | 13 | 6,76 (1,42) | 7,80 (1,20) | - | - 0,62 (0,80) | 3,96 (0,65) |
| 5 | 6 | 9 | 13 | 6,90 (1,43) | 7,22 (1,07) | - | - 0,12 (0,33) | 3,46 (0,73) |
| Gesamt | 51 | 106 | 78 | 6,74 (1,48) | 7,65 (1,23) | 53 | - 0,40 (0,63) | 4,14 (0,83) |

Anmerkung: Die A-Daten zu den Phrasen 4 und 5 wurden wegen mangelnder Reliabilität ausgeschlossen.

Tab. 4: Kennwerte der Audiophrasen (N = 156) auf Phrasenebene und insgesamt.

Vergleichbar wurde mit den Ratings der Skala B verfahren. Erst wurden die als Schulnoten erfragten Daten in Punktzahlen umgerechnet („sehr gut“ = 4 Punkte bis „ausreichend oder weniger“ = 1 Punkt; $M = 3,06$ (i. e. „gut“), $SD = 0,49$, Range = 1,62 bis 3,92). Multipliziert man die Punktzahlen mit dem Faktor 2,5, ergeben sich die besser auf C beziehbaren B-Kennwerte auf einer Skala von 1 bis 10 (B global: $M = 7,65$ und $SD = 1,23$ und Range = 4,04 bis 9,81 (siehe Tabelle 4, unterste Zeile)). Für die rechnerischen Analysen ist diese Transformation irrelevant. Der höhere Mittelwert bei B im Gegensatz zu C bei geringerer Streuung zeigt, dass die B-Skala weniger gleichmäßig, sondern eher im höheren Bereich ausgenutzt wurde. Die nach rechts spitz zulaufende Punktwolke in Abbildung 5 zeigt eine Tendenz abnehmender Abweichung zwischen B und C, d. h. je höher C, desto enger geht C mit einer „insgesamt satzzeichengemäßen Stimmführung“ einher. Hier spielt augenscheinlich auch ein Deckeneffekt eine Rolle.

Auch die Ratings der Skala A wurden zu Mittelwerten zusammengefasst. Die Fragestellung nach der Hervorhebung eines Wortes in einer Phrase im Gegensatz zu allen anderen möglichen Hervorhebungsverhältnissen lässt allerdings keine sinnvolle Interpretation der Anteilhaftigkeit zu,

die mit diesen Mittelwerten suggeriert wird. Deshalb wurde gerundet, d. h. war der Mittelwert aus den 13 Ratings $\geq 0,5$, erhielt die jew. Audiophrase die Ausprägung $A = 1$ (= Hervorhebung informationsstrukturgemäß), sonst 0. Da die Werte zu Phrase 4 und 5 aus der Analyse ausgeschlossen wurden, liegen 104 Ausprägungen vor, global hiervon 53-mal 1 (s. Tabelle 4, unterste Zeile). So erscheinen in Abbildung 5 zwei Kettenstrukturen auf der 0- und auf der 1-Linie, die allerdings von fehlenden Werten (Phrase 4 und Phrase 5) durchsetzt sind. Ein Zusammenhang mit den C-Daten zeichnet sich dahingehend ab, dass etwa ab $C > 7$ eine höhere „Kettendichte“ auf der 1-Linie als auf der 0-Linie ersichtlich ist, d. h. mit hoher C-Punktzahl geht meist auch eine informationsstrukturgemäße Worthervorhebung einher.

Als weitere Merkmale werden auf der Ebene der Audiophrasen die Fehleranzahl E („Errors“) und die Sprechgeschwindigkeit S [Silben/sek] einbezogen (Korrelate von Akkuratheit und fortgeschrittener Automatisierung). Als Fehler gezählt wurden Auslassungen, Ersetzungen und Hinzufügungen auf Wortebene (Daane et al. 2005). Da es sich bei den Audiophrasen um kurze Sequenzen handelt, sind die meisten fehlerfrei (106-mal 0 Fehler). Von den verbleibenden 50 Aufnahmen weisen 38 einen und 12 zwei Fehler auf. Für Abbildung 5 (rote Kettenstrukturen) wurden diese Werte durch $(E * (-1))$ ersetzt, um die positive Konnotation der Y-Achse aufrecht zu erhalten, und 0 Fehlerpunkte wurden herausgelöscht, damit sie nicht auch als rote Punkte und damit optisch als Minuspunkte in Erscheinung treten. Die unausgewogene Verteilung schlägt sich in der auffälligen Mittelwert-Streuungskombination von $M = 0,4$ und $SD = 0,63$ nieder (s. Tabelle 4, unterste Zeile). Abbildung 5 lässt klar erkennen, dass die Fehler schlüssigerweise fast ausschließlich im unteren Bereich von C auftreten.

Die Sprechgeschwindigkeit S wurde in Sprechsilben/Sekunde gemessen. Für die fehler- und pausenfreien Audiophrasen wurde die absolute Dauer von Sprechbeginn bis Sprechende gemessen und mit der Silbenanzahl der Textphrase verrechnet. Jede von Fehlern betroffene Phrase wurde auf zusätzliche oder fehlende Sprechsilben sowie merkliche Pausen geprüft und das Ergebnis mit der Silbenanzahl der jew. Textphrase und dann nach dem beschriebenen Muster mit der absoluten Dauer, ggf. abzüglich Pausen, verrechnet [tats. Sprechsilben/sek]. Es ergaben sich die globalen Kennwerte von $M = 4,14$ und $SD = 0,83$ und Range = 2,02 bis 6,73 (s. Tabelle 4, unterste Zeile). Gestört wird S durch Stockungen im Redefluss, wie sie oft mit Fehlern einhergehen und nicht als Pausen in Erscheinung treten, und zwar im Hinblick auf das auf Konstruktebene avisierte latente Personenmerkmal Eigentempo.

Zum Begriff Eigentempo ist ein kurzer Exkurs angebracht: Ein intraindividuell konstantes Eigentempo kann als prosodisches Gestaltungsmittel bei fortgeschrittenem Vorlesen gelten und ist auch phonetisch klar bestimmbar, allerdings nur bei fehler- und stockungsfreien Passagen. Eine transparentere Maßeinheit wäre hier $(S * 60) / 2$, also Doppelsilben pro Minute. Unter den verallgemeinernden Annahmen, dass im Durchschnitt jede zweite Silbe betont ist und dass ein Wort aus durchschnittlich zwei Silben besteht, lässt sich $(S * 60) / 2$ ungefähr auf die Einheit beats per minute (BPM) bzw. words per minute (WPM) beziehen. So ergibt sich für BPM oder WPM global: $M = 124$, $SD = 24$, Range = 61 bis 202. Über die entsprechenden Schwellwerte zu den Tempo-Begriffen aus der Musik ist im Hinblick auf ein latentes Personenmerkmal Eigentempo festzuhalten, dass die Messungen zu den 156 Audiophrasen den Intervallen andante (76-108 BPM), moderato (108-120 BPM), allegro (120-168 BPM) und presto (168-200 BPM) zugeordnet werden können. Die Transformation ist rechnerisch irrelevant und wird in den Darstellungen nicht berücksichtigt, weil Silben/sek auf einer 10er-Skala graphisch gut eingeordnet werden

können. Abbildung 5 zeigt einen Zusammenhang zwischen S und C dahingehend, dass zwischen den auf Skala C ganz schwachen und ganz starken Phrasen ein Anstieg von ca. $S = 4$ auf $S = 5$ einhergeht.

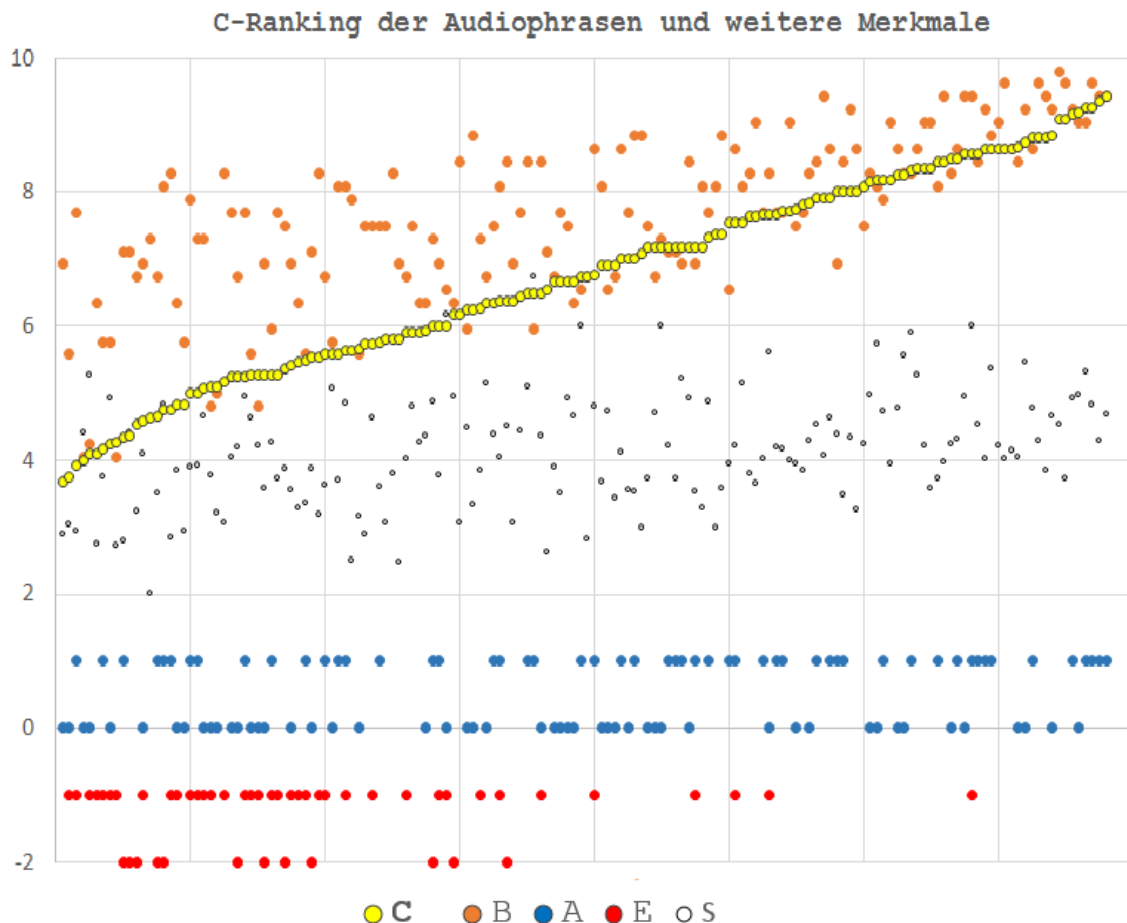
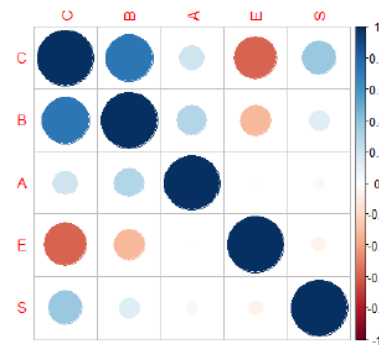


Abb. 5: Datenübersicht auf Ebene der 156 Audiophrasen im Ranking nach C. Erfasste Merkmale zu $n = 156$ Audiophrasen: C = Rating-Mittelwerte („Vorlesewettbewerb“), B = Rating-Mittelwerte („Stimmführung“), A = Rating-Mittelwerte („Hervorhebung“), E = - Fehlersumme („Errors“), S = Silben/sek („Speed“).

Die nach Textphrasen aufgeschlüsselten Analysen (s. Tabelle 2) lassen Aussagen zur oben angesprochenen Rater*innen- bzw. Konstruktions-Perspektive zu. Auf Textphrasenebene sind spaltenweise Abweichungen unerwünscht, weil sie auf Einflüsse schließen lassen, die Inhalt oder Stellung der Phrasen auf die Bewertung durch die Rater*innen haben. Der Faktor Phrasenlänge ist bis auf Phrase 5 wohlkontrolliert und Skala C und B weisen kaum Abweichungen auf. Bei A fällt nur Phrase 1 auf, bei der den Ratings zufolge in nur 3 von 26 Fällen informationsstrukturgemäß hervorgehoben wurde. Allerdings zeigen sich die Phrasen stark unterschiedlich anfällig für Fehler, was auf unterschiedliche Leseschwierigkeit hinweist und nicht überraschend ist. Hinsichtlich der Sprechgeschwindigkeit ist einzig Phrase 5 auffällig. Somit spricht genügend Evidenz dafür, die in Tabelle 4 vorgestellten Daten für die nachfolgenden Analysen ohne weiteren Datenausschluss zusammenzufassen.

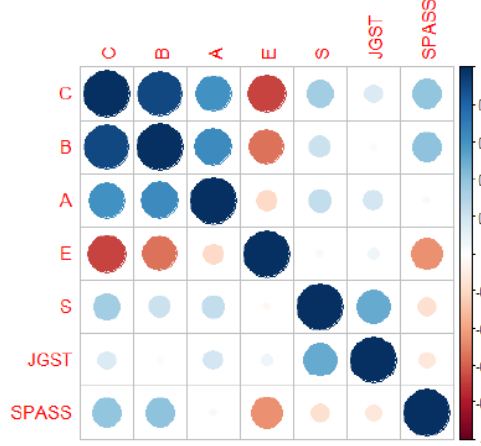
a) Phrasen

| | C | B | A | E | S |
|---|------|------|-----|------|---|
| C | 1 | | | | |
| B | .73 | 1 | | | |
| A | .21 | .30 | 1 | | |
| E | -.58 | -.32 | .01 | 1 | |
| S | .37 | .14 | .04 | -.07 | 1 |



b) Aufnahmen

| | C | B | A | E | S | JGST | SPASS |
|-------|------|------|------|------|------|------|-------|
| C | 1 | | | | | | |
| B | .90 | 1 | | | | | |
| A | .60 | .63 | 1 | | | | |
| E | -.67 | -.54 | -.19 | 1 | | | |
| S | .34 | .21 | .24 | -.03 | 1 | | |
| JGST | .15 | .02 | .18 | .07 | .50 | 1 | |
| SPASS | .40 | .40 | .03 | -.45 | -.16 | -.12 | 1 |



Anmerkungen: Rating-Mittelwerte C = „Vorlesewettbewerb“, B = „Stimmführung“, A = „Hervorhebung“, E = Fehlersumme („Errors“), S = Silben/sek („Speed“), Jahrgangsstufe (JGST) und SPASS („Engagement“, s. Sappok, Linnemann & Stephany 2020).

Abb. 6: Korrelationskoeffizienten und Visualisierung:

- a) für n = 156 Audiophrasen
- b) für n = 26 Aufnahmen

Die in Abbildung 5 visualisierten Verhältnisse sollen nun über Korrelationen genauer in den Blick genommen werden. Hierfür wurden die fehlenden A-Werte für Phrase 4 und 5 durch den Aufnahme-internen Mittelwert aus den vorliegenden Phrasen 1 bis 3 ersetzt, um sie rechnerisch zu neutralisieren. Abbildung 6 a) zeigt die Korrelationsmatrix für die auf Ebene der Audiophrasen einbezogenen Faktoren und die Visualisierung der Koeffizienten. Die Ergebnisse fügen sich zu einem klaren Bild. Hinsichtlich der Frage nach der Einflussgröße auf C zeigt sich eine Rangfolge von $B > E > S > A$, wobei alle untersuchten Faktoren, auch der niedrigste von $r = .21$ für A, auf $p < .01$ -Niveau signifikant sind (p-Werte sind deshalb in Abb. 6 nicht mit dargestellt).

Damit ist die Analyse auf der Ebene der Audiophrasen abgeschlossen. Sie mündet vor allem in die Frage nach der Trennschärfe zwischen C und B (Korrelation zu hoch) und die Frage nach der Rolle von Faktor A (Korrelation zu niedrig). Ein differenzierteres Bild zeigt sich nach der Zusammenfassung der Phrasenmittelwerte zu Aufnahme-Mittelwerten (bis auf S, das auf Ebene der

Aufnahmen neu berechnet wurde, da so der Faktor unterschiedliche Phrasenlängen neutralisiert werden konnte).

4.2 | Ergebnisse auf Ebene der Aufnahmen

Abbildung 6 b) zeigt die Korrelationsmatrix für die auf Ebene der Gesamtaufnahmen einbezogenen Faktoren und die Visualisierung der Koeffizienten. Im Vergleich mit der Phrasenebene zeigen sich insgesamt klar höhere r -Werte. Grund ist, dass die Zusammenführung der 6 Audiophrasen zu einer Aufnahme Störvariablen zunehmend unter Kontrolle bringt.

Auf der Ebene der Aufnahmen können nun auch Merkmale einbezogen werden, die auf Phrasenebene nicht berücksichtigt wurden, weil sie bei allen Phrasen einer Aufnahme gleich ausgeprägt sind (z. B. Engagement („SPASS“ in Abbildung 6, nähere Erläuterung unten; diese Daten wurden bereits in der Vorgängeruntersuchung (Sappok, Linnemann & Stephany 2020) erhoben).

Zur Schulform kann ausgesagt werden, dass acht der 16 beteiligten Kinder auf ein Gymnasium (GYM) und acht in eine Realschule Plus (RS+) übergegangen sind (s. Tabelle 1), die Ausprägungen sind also gleich verteilt. Weitere Aufschlüsse werden im vorliegenden Kontext von diesem Faktor aber nicht erwartet, zumal auch Aufnahmen aus der Grundschule einbezogen sind. Zum Geschlecht kann festgestellt werden, dass 11 Aufnahmen von Schülern und 15 von Schülerinnen stammen. Der C-Mittelwert für die Schüler ist $M = 6,96$, $SD = 0,72$, Range = 5,60 bis 8,12, für die Schülerinnen ergibt sich: $M = 6,59$, $SD = 1,14$, Range = 4,97 bis 8,43. Der Leistungsunterschied ist nicht erheblich, die Streuung ist bei den Schülerinnen bedeutend höher, was auf unkontrollierte Einflüsse beim Zustandekommen der Aufnahmenstichprobe zurückgeführt wird. Eine sinnvolle Interpretation kann nicht gefunden werden. Das Geschlecht ist im vorliegenden Kontext als wenig C-relevant anzusehen und wird im Folgenden nicht berücksichtigt.

Im Hinblick auf Alter bzw. Jahrgangsstufe der lesenden Schüler*innen zum Zeitpunkt der Aufnahme (vgl. Tabelle 1) wurde Jahrgangsstufe (JGST) der Variable „Alter in Tagen zum Aufnahmezeitpunkt“ (ebenfalls erhoben) vorgezogen. JGST und Alter korrelieren mit $r = .97$, doch JGST ist an sich relevanter für didaktische Interpretationskontexte. Zu den entsprechenden ganzzahligen Werten (vgl. Tabelle 1) wurde 0,5 hinzuaddiert, da die Aufnahmen stets zum Halbjahreswechsel des betreffenden Schuljahrs gemacht wurden (Transformation für die Analyse irrelevant). Bei JGST ist $M = 6,38$, $SD = 1,14$, Range = 3,5 bis 7,5.

Bei „SPASS“ handelt es sich um diejenige Skala, die in der Vorgängeruntersuchung (Sappok, Linnemann & Stephany 2020) nach „NAEP-Lese flüssigkeit“ die zweitreliabelste war. In der seinerzeit eingesetzten Ratingprozedur waren die Aufnahmen analog zu NAEP-Lese flüssigkeit (hier immer 4, vgl. Abbildung 1) nach dem Kriterium zu beurteilen gewesen: „Spaß: Wie gut hat dem Kind die ‚Vorlesearbeit‘ gefallen?“, und zwar auf einer Skala von „gar nicht“ = 1 bis „sehr gut“ = 4. In der Längsschnittperspektive hatte sich gezeigt, dass SPASS eher ein über die Zeit intraindividuell konstantes Personenmerkmal ist, als dass in den untersuchten Biographien die SPASS-Ausprägung mit der individuellen Entwicklung bei NAEP-Lese flüssigkeit ansteigt (ebd.). Bei SPASS ist $M = 3,02$, $SD = 0,56$, Range = 2 bis 4.

Im Hinblick auf die Korrelationsverhältnisse zeigt der Vergleich von Abbildung 6 a) und 6 b) eine etwas andere Rangfolge bei den einzelnen Einflussfaktoren auf C: Abb. 6 a) $B > E > S > A$; Abb. 6 b) $B > E > A > S$.

Außerdem muss die Höhe der Korrelation $C \sim B$ in 6 b) als Anzeichen für mangelnde Trennschärfe in den Blick genommen werden.

S korreliert in beiden Perspektiven auffällig schwach mit C. Weiterhin korreliert S, ähnlich gering, mit JGST (siehe Abbildung 6 b). SPASS korreliert mit C in merklichem, aber ähnlich geringem Maße wie S. Vergleichbar korreliert SPASS mit B und E, aber nicht mit A und S.

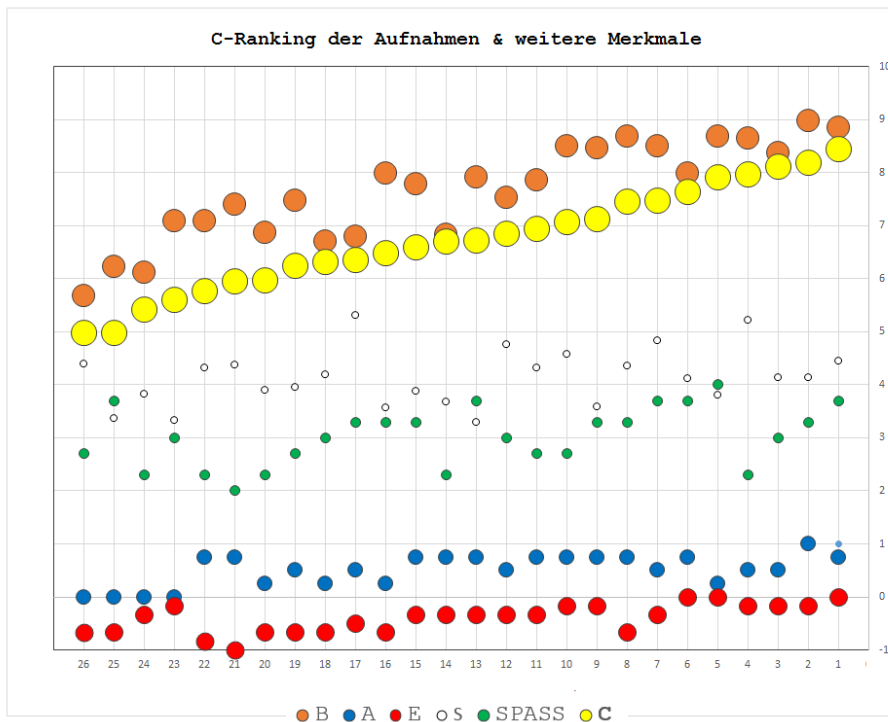
JGST fällt vor allem dadurch auf, dass kaum eine Korrelation mit C vorliegt. Dies kann darauf zurückgeführt werden, dass so viele Aufnahmen aus 6 und 7 vorliegen. Weiter fällt JGST einzig durch seine Korrelation mit S auf. Um diese Tendenz zu quantifizieren, sollen hier die JGST-spezifischen S-Mittelwerte präsentiert werden, als Referenzpunkte für Anschlussuntersuchungen mit höheren und gleichmäßigeren Schüler*innen-Zahlen (s. Tabelle 5). Für die weiterführende Analyse kann JGST jedoch als eindeutig schwächster Faktor ausgeschlossen werden. Zur Frage nach dem Eigentempo bei fortgeschrittenen Schüler*innen kann als grobe Orientierungslinie die Mittelwertreihe für S nach JGST festgehalten werden:

| JGST | 3 (n = 1) | 4 (n = 3) | 5 (n = 3) | 6 (n = 10) | 7 (n = 9) |
|--------------|-----------|-----------|-----------|------------|-----------|
| S | 3,28 | 3,73 | 3,91 | 4,08 | 4,37 |
| WPM (S*60)/2 | 98 | 112 | 117 | 122 | 131 |

Tab. 5: Jahrgangsstufen-Mittelwerte auf Basis der n = 26 Aufnahmen in S (Silben pro Sekunde) und Doppsilben pro Minute (hypothet. BPM bzw. WPM).

Abbildung 7 zeigt die Kennwerte der einzelnen Aufnahmen. Die bunte vertikale „Perlenkonstellation“ zu einer Aufnahme repräsentiert das differenzierte Vorlese-Profil eines Kindes zu einem bestimmten Zeitpunkt, diagnostiziert anhand einer Prima-vista-Aufnahme des Textes „Dolmetscher“ von ca. einer halben Minute Dauer.

Explorative Aufschlüsse zum Zusammenwirken der Faktoren sollen nun über lineare Regressionsanalysen gewonnen werden. Der Faktor SPASS erwies sich dabei als schwächster Faktor, weshalb er für die folgenden Analysen ausgeschlossen wurde. Vor diesem Hintergrund soll die bisherige Evidenz noch einmal zusammengefasst werden. Ausgangspunkt dafür ist die zentrale Hypothese (s. Einführung), dass sich mit den erhobenen Daten ein Modell nach dem Muster $C \sim B + A + \text{andere Faktoren}$ bestätigen lässt. Was die Rangfolge der Faktoren angeht, korreliert B stets so stark mit C, dass der Verdacht gerechtfertigt ist, dass mit B in hohem Maße dasselbe gemessen worden ist wie mit C (Trennschärfe). Hinsichtlich der verbleibenden Faktoren sind unterschiedliche Rangfolgen festzustellen gewesen.



Anmerkungen: Rating-Mittelwerte C = „Vorlesewettbewerb“, B = „Stimmführung“, A = „Hervorhebung“, E = - Fehlersumme („Errors“), S = Silben/sek („Speed“), SPASS („Engagement“, s. Sappok, Linnemann & Stephany 2020).

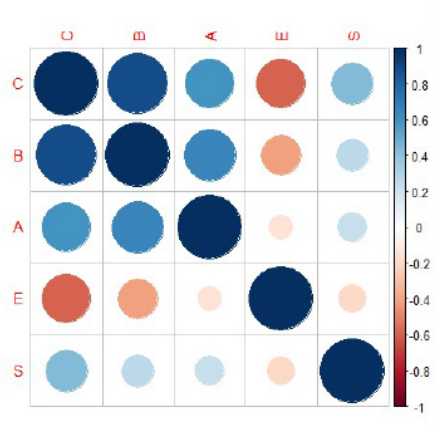
Abb. 7: Aufnahmespezifische „Advanced prosody“-Register für die Diagnostik von Vorlesekompetenz. Erfasste Merkmale zu $n = 26$ Aufnahmen fortgeschrittener Schüler*innen: Die Aussagekraft der Parameter im Spiegel von r -Werten ist durch die Perlengröße angedeutet.

Die A-Daten haben sich bzgl. Phrasen 4 und 5 als zu wenig reliabel gezeigt und wurden dort durch Mittelwerte ersetzt. Um das Zusammenwirken der Faktoren und hier vor allem die Rolle von Faktor A ausschließlich anhand vollständiger Datensätze zu klären, wurden diese Phrasen für die Erstellung von linearen Modellen ganz ausgeschlossen. So wurden die verbleibenden vier Audiophrasen (1, 2-1, 2-2 und 3, s. Tabelle 2) zu Aufnahme-Mittelwerten zusammengefasst. Diese repräsentieren die via C diagnostizierte Vorlesekompetenz und ihre potenziellen Komponenten. Der Unterschied schlägt sich in leichten Abweichungen der Kennwerte im Vergleich von Abbildung 6 b) und 8 a) nieder.

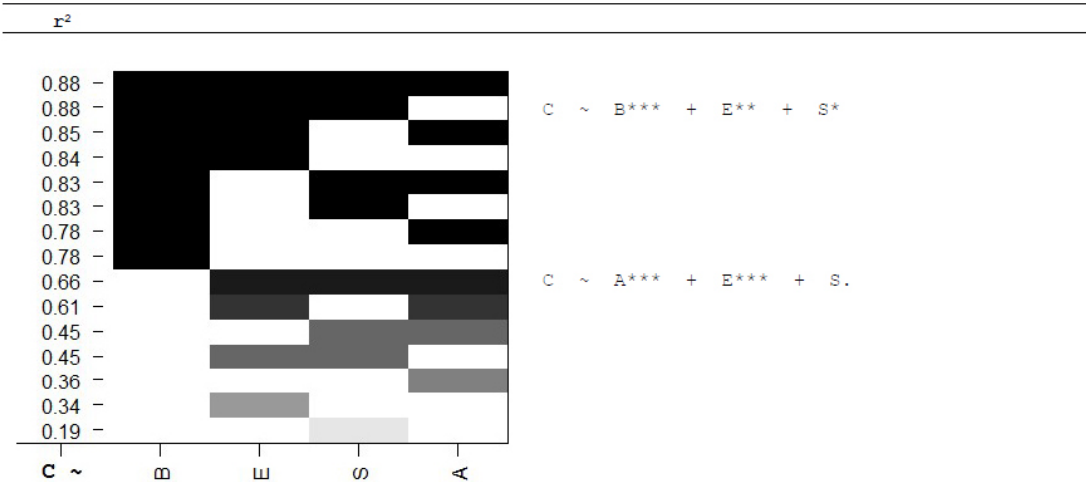
Im Hinblick auf die Korrelationsverhältnisse zeigt sich nun die Rangfolge (Abbildung 8 a): $B > A=E > S$.

a) Korrelationsmatrix Aufnahmen

| | C | B | A | E | S |
|---|------|------|------|------|---|
| C | 1 | | | | |
| B | .88 | 1 | | | |
| A | .60 | .65 | 1 | | |
| E | -.58 | -.40 | -.15 | 1 | |
| S | .44 | .26 | .22 | -.19 | 1 |



b) lineare Modelle



Anmerkungen: Das Ranking bezieht sich auf die leaps-Analyse (Lumley 2020) geordnet nach r^2 ; *** = $p < .001$, ** = $p < .01$, * = $p < .05$, . = $p < .1$ (s. Tabelle 6)

Abb. 8: Ebene der Aufnahmen ohne Phrasen 4 u. 5:

a) Korrelationskoeffizienten r und Visualisierung (vgl. Abbildung 6).

b) Ranking linearer Modelle zur Aufklärung von Varianz in C. Oben mit allen Faktoren, unten ohne Faktor B. Mit $B + E + S$ und $A + E + S$ ergeben sich zwei Modelle, wodurch die Rollen von Faktor B und Faktor A vergleichbar werden (s. Tabelle 4).

Die Exploration besteht in der Generierung und Gewichtung aller linearen Regressionsanalysen zur Prädiktion von C, die mit den Faktoren B, E, S und A möglich sind. Verwendet wurde hierfür das „leaps“ Package in R (Lumley 2020). Das Ergebnis besteht in einem Ranking der Faktorkombinationen nach in C aufgeklärter Varianz anhand von r^2 .

Zusammenfassend sollen zunächst alle Modelle betrachtet werden, die unter Beteiligung von B möglich sind (= „Pistolengriff“-Block in Abbildung 8 b) links oben). Es zeigt sich die Rangfolge $B > E > S > A$.

Der untere Teil von Abbildung 8 b) bezieht sich auf die Modelle ohne B. Hier zeigt sich eine Teilumkehrung der Ränge der verbleibenden Faktoren: $A = E > S$.

Die angebotenen Modelle wurden mit der Funktion $\text{lm}()$ in R einzeln neu gerechnet und bestätigt. Angegeben sind die zwei Modelle, die als maximal aufschlussreich erachtet werden. Beim ersten dieser Modelle (8 b) oben) wurde Faktor A ausgeschlossen, da er keinen Beitrag leistet ($C \sim B + E + S$). Beim zweiten Modell wurde Faktor B durch Faktor A ersetzt ($C \sim A + E + S$).

Bei Modell 1 ist der adj. r^2 -Wert von .86 noch einmal höher als der für das einfache Modell $C \sim B^{***}$, wo adj. r^2 bei .77 liegt. Dies spricht gegen einen kategorischen Mangel an Trennschärfe zwischen C und B. Die damit angesprochene Problematik wird in der Diskussion aufgegriffen.

Nur aus der Perspektive der beiden Modelle lässt sich eine aufschlussreiche Gegenüberstellung von Faktor B und Faktor A erreichen. Faktor A und B korrelieren nach Abbildung 8 a) mit $r = .65$ (adj $r^2 = .40$, $p < 0,001$). Vor diesem Hintergrund könnte A als eine Art „schlechter Ersatz“ für B angesehen werden. Auch dieser Befund wird in der Diskussion aufgegriffen. Ein summativer Einfluss im Sinne von $C \sim B + A$, wie er im Fujisaki-Modell (Abbildung 2 links) mit dem Prinzip Superimposition vorausgesetzt wird, konnte hier nicht belegt werden. Auch dies wird in der Diskussion aufgegriffen.

5 | Diskussion und Ausblick

Untersucht wurde das diagnostische Potenzial von 26 Audioaufnahmen eines Sachtextes von ca. 30 Sekunden Dauer. Es handelte sich um genau diejenigen Aufnahmen aus einer Längsschnitt-Stichprobe von $N = 99$, die auf der NAEP-Fluency-Skala einhellig Level 4 zugeordnet bekamen (Sappok, Linnemann & Stephany 2020). Jede Aufnahme wurde in 6 Phrasen von max. 22 Textsilben Länge geschnitten und es ergab sich eine Stichprobe von 156 Audiophrasen, auf der die Untersuchung beruht. Hierfür wurde eine dreizügige Ratingprozedur entwickelt und mit einer Stichprobe von 39 studentischen Rater*innen im Rahmen eines E-Seminars erprobt. Alle Rater*innen bewerteten alle 156 Audiophrasen am Computer als Hörstimuli in individuell teilrandomisierter Form. Team A bewertete informationsstrukturell motivierte phonetische Wortprominenz auf der binären Skala A („Zielwort hervorgehoben vs. Gegenbeispiel“). Team B benotete die Phrasen-Intonationskontur auf der vierstufigen Skala B („Stimmführung insgesamt satzzeichengemäß?“) mit den Antwortmöglichkeiten: sehr gut, gut, befriedigend, ausreichend oder weniger. Team C benotete Vorlesekompetenz global durch die Vergabe von 1 bis 10 Punkten in einem Casting-Szenario.

In der vorliegenden Untersuchung wurde somit weiterführende Grundlagenforschung betrieben, deren Befunde nun im Hinblick auf ihre Perspektiven diskutiert werden sollen. Zu nennen ist hier einmal die Perspektive, wie die Grundlagenforschung selber weitergeführt wird (z. B. Optimierung von Skalenkonstruktion). Hierzu wurden an den Anfang zwei Heurismen gestellt. Diese sollen nun im Unterkapitel Diagnostik diskutiert werden. Außerdem ist die Perspektive zu diskutieren, wie die in vorliegender Untersuchung generierte Evidenz zur Praxis weitergeführt werden kann. Dazu wurden konkretere Hypothesen aufgestellt. Diese Hypothesen sollen im Unterkapitel Didaktik aufgegriffen werden. Zum Bezug zum Textverstehen wurden zwei spekulative Axiome formuliert. Auch wenn der Bezug zum Textverstehen in vorliegender Untersuchung nicht im Zentrum stand, sollen diese Axiome in der Zusammenfassung noch einmal angesprochen werden.

5.1 | Diagnostik

Zunächst wird die Weiterarbeit aus Grundlagenperspektive fokussiert. Die Heuristiken I und II haben sich grundsätzlich bestätigt. Als Ausgangspunkt für die konkrete Weiterarbeit lassen sich die folgenden Anknüpfungspunkte für eine Optimierung des behandelten Diagnostikansatzes definieren:

- bessere theoretische Fundierung (Konstruktebene)
- besserer Lesetext (Auswahl oder Konstruktion von Textmaterial für die Diagnostik)
- bessere diagnostische Instrumente (Skalenkonstruktion, Benutzeroberfläche)
- „bessere“ Rater*innen (Expert*innen, z. B. aufwändiger geschulte Studierende oder Lehrkräfte)

Bei der Untersuchung der Inter-Rater-Reliabilität zeigte sich klar $C \gg B > A$. Was den DOL-Text angeht, so können die Phrasen 1 bis 4 als geeignet für die Erhebung von A gelten, Phrasen 4 und 5 nicht. Als Ansatz für eine Verbesserung der Skala kann gelten, dass die Hörbedingungen verbessert werden müssen. Als Orientierung kann dabei weiterhin die Akzent-Komponente nach Fujisaki (1988, S. Abb. 3) gelten.

Auch Skala B ist verbesserungsbedürftig. Auf Konstruktebene ist mit dem Bezug zur Zeichensetzung und damit den grobsyntaktischen intonationsrelevanten Verhältnissen ein prosodischer Aspekt angesprochen, der gegenüber A weniger gehobene Aspekte von Textverstehen voraussetzt. Die untersuchten fortgeschrittenen Schüler*innen lassen in puncto syntaxgemäßer Stimmführung häufig nichts zu wünschen übrig. Es hat aber den untersuchten Rater*innen womöglich gewisse Schwierigkeiten bereitet, B zu fokussieren und andere Eigenschaften von gutem Vorlesen dabei auszublenden (die Korrelation $B \sim C$ liegt bei $r = .9$, vgl. Abb. 6 b).

In Hinblick auf diese Probleme sind bzgl. zukünftiger Rater*innen-Auswahl verschiedene Maßnahmen zu erwägen, z. B. gratifizierte Schulungsmaßnahmen mit Beispiel-Audiophrasen, Rekrutierung aus anderen Populationen (z. B. Lehrkräfte). Außerdem sollte bzgl. der Skalen A und B über Modifikationen der Konzeption und der Instruktion in den Skalen selbst nachgedacht werden. Demgegenüber kann aber festgehalten werden, dass es mit Skala C offenbar gelingt, ein latentes Personenmerkmal "Vorlesekompetenz" reliabel zu erheben.

5.2 | Didaktik

Abschließend wird vor diesem Hintergrund die Weiterarbeit aus Praxisperspektive fokussiert. Im Mittelpunkt des vorliegenden Beitrags stand folgende

zentrale Hypothese: Die erhobenen Ratingdaten können zu einem explorativen Modell nach dem Muster $C \sim B + A + \text{weitere Faktoren}$ integriert und methodisch und didaktisch weiterführend interpretiert werden.

Der Befund, dass im schlimmsten Fall mit B etwas sehr Ähnliches wie mit C gemessen wird, ist ernüchternd und zeigt auf, dass Skala B konzeptionell überarbeitet werden muss. Der Befund, dass mit A etwas Ähnliches wie mit C gemessen wird, ist hingegen in unerwarteter Weise aufschlussreich. Dass ein minimales Kontingent von bereits vier binären Items genügt, um Aspekte

von Vorlesekompetenz in nennenswertem Maße widerzuspiegeln, zeigt, dass Akzent- bzw. Prominenzverhältnisse mit verbesserten Methoden weiter untersucht werden sollten, zumal hier Textverständnis recht eindeutig als Vorbedingung angesehen werden kann. Hervorzuheben ist, dass die Analyse von Hervorhebung auf Wortebene oder ein vergleichbares, optimiertes Konstrukt mit signalphonetischen Mitteln erfolgen könnte, da es sich eher als andere Merkmale automatisch lokalisieren lässt.

Für die Gültigkeit der Hypothesen, die zum Abschluss von Unterkapitel 2 aufgeführt wurden, Hypothese 1: Aus 50% aller Grundschüler*innen werden bis Mitte Sekundarstufe I keine flüssigen Leser*innen,

Hypothese 2 a) Das Überschreiten der Schwelle zu NAEP-Level 4 bedeutet, lediglich early prosody erworben zu haben,

Hypothese 2 b) Innerhalb von NAEP-Level 4 (advanced prosody) ist ein breites Spektrum von Binnenabstufung auszudifferenzieren,

konnten bereits in Sappok & Fay (2018) und Sappok, Linnemann & Stephany (2020) Hinweise gesammelt werden. Hinsichtlich Hypothese 1 lassen die geringen Stichproben allerdings keine Generalisierung zu. Zu Hypothese 2 a) und b) kann hingegen verallgemeinert werden, dass die Pinnell-Skala sehr anfällig für Deckeneffekte ist (ebd.). Hier setzen die untersuchten Skalen an. Skala C kann als voll einsatzfähig für die Diagnose von advanced prosody gelten. Dies bedeutet jedoch eine Abkehr vom Begriff Leseflüssigkeit im engeren Sinne, denn es spielen auch Faktoren wie Stimme-Verstellen eine Rolle, während Geschwindigkeit kaum mehr eine Rolle spielt (s. u.). Die Skalen B und A können aus unterschiedlichen Gründen noch nicht als einsatzfähig gelten. Die Untersuchung hat aber konkrete Anknüpfungspunkte für die Weiterarbeit ergeben.

| Modell | adj. R ² | Faktor | b | SE | t | df | p |
|--------|---------------------|--------|-------|------|-------|----|---------------|
| 1 | .86 | B | 0,96 | 0,11 | -8,89 | 22 | < .0001 |
| | | E | -0,77 | 0,25 | -3,14 | 22 | .0048 |
| | | S | 0,46 | 0,18 | 2,57 | 22 | .0173 |
| 2 | .62 | A | 1,79 | 0,48 | 3,73 | 22 | .0012 |
| | | E | -1,43 | 0,39 | -3,70 | 22 | .0013 |
| | | S | 0,57 | 0,30 | 1,89 | 22 | .0727 (n. s.) |

Anmerkungen: abhängige Variable: C = Vorlesewettbewerb, unabhängige Variablen: B = Stimmführung, E = Errors, S = Speed, A = Hervorhebung

Tab. 6: Lineare Modelle zur Gegenüberstellung von Faktor A und B als Komponenten von C (s. Abb. 9). Geschwindigkeit (S) spielt die geringste Rolle.

Aus schulpraktischer Perspektive steht die Frage im Vordergrund, wie didaktische Ressourcen optimal investiert werden können. Dazu lassen sich in einem letzten Schritt die in den Modellen aufgezeigten Steigungsverhältnisse auswerten (Spalte b in Tabelle 6). Dabei werden fünf „Advanced-Prosody-Faustregeln“ formuliert, die für den Unterricht Orientierung bieten könnten. Modell 1 (Spalte b) ist so zu lesen: Wenn jew. die beiden anderen Faktoren konstant bleiben,

gelten folgende Gesetzmäßigkeiten: für + 1 B (hochskaliert auf 1 bis 10) gibt es + 0,96 C-Punkte; für + 1 Error pro Phrase gibt es – 0,77 C-Punkte; für + 1 Silbe pro Sekunde gibt es + 0,46 C-Punkte.

Der Befund, dass B und C nicht trennscharf genug sind, tut der potenziellen Relevanz von Stimmführung keinen Abbruch. Für die didaktische Arbeit kann hierzu folgende Faustregel eingesetzt werden:

(1) *Beachte deine Stimmführung auf dem Weg zum Satzzeichen.*

Auch Faktor E spielt eine nennenswerte Rolle. Hinsichtlich der Fehler kann wie in der Grundschule nach wie vor die Faustregel ausgegeben werden:

(2) *Mache keine Flüchtigkeitsfehler, finde deine Problemfelder.*

Der Einfluss von S, einem der wichtigsten Faktoren bei der Diagnose von Leseflüssigkeit im Primärkontext, ist gering. Didaktischer Handlungsbedarf in Hinblick auf eine Steigerung besteht höchstens im Andante-Bereich (bis $S = 3,6$ bzw. 108 BPM). Die Steigung ist bei S vordergründig dahingehend zu interpretieren, dass + 1 S fast so viele C-Pluspunkte bringt wie - 1 E, also ein Fehler weniger pro Phrase. Hält man sich vor Augen, dass + 1S etwa + 30 BPM entspricht, erweist sich eine Investition in Geschwindigkeit im Spiegel des Modells allerdings klar als uninteressant. Wahrscheinlich profitieren manche Schüler*innen eher von einer beschwichtigenden Faustregel:

(3) *Finde dein Eigentempo. So schnell wie möglich zu lesen war nur in der Grundschule wichtig.*

Die Steigungsbefunde des zweiten Modells (Tabelle 6) können folgendermaßen interpretiert werden: Wenn jew. die beiden anderen Faktoren konstant bleiben, gelten folgende Gesetzmäßigkeiten: für + 1 A gibt es + 1,79 C-Punkte; für + 1 Error pro Phrase gibt es – 1,43 C-Punkte; für + 1 Silbe pro Sekunde gibt es + 0,57 C-Punkte.

Die Relevanzverhältnisse lassen sich bzgl. E und S recht gut mit denen zu Modell 1 vergleichen. Für A kann aus didaktischer Perspektive festgehalten werden, dass mit knapp 2 Pluspunkten ein hoher C-Profit winkt, wenn man lernt, einer informationsstrukturell gebotenen Hervorhebung Rechnung zu tragen. Als Faustregel lässt sich das z. B. so formulieren:

(4) *Wörter, auf die es ankommt: erkenne und betone sie.*

5.3 | Zusammenfassung

Eine Zusammenfassung zur vorliegenden Untersuchung sollte mit dem Begriff *advanced prosody* beginnen. Dieser Begriff wurde gewählt, um die Notwendigkeit einer Ausdifferenzierung des Konstrukts Leseflüssigkeit hervorzuheben. Die herkömmlichen Verwendungskontexte von *reading fluency* beziehen sich auch international fast ausschließlich auf den Primärbereich. *Advanced prosody* in der weiterführenden Schule zeichnet sich vor allem dadurch aus, dass Geschwindigkeit kaum noch relevant ist. Vor diesem Hintergrund muss der Begriff *reading fluency* als *genus proximum* für *reading prosody* hinterfragt werden, denn für Forschung und Praxis gehen mit *advanced prosody* andere Probleme einher als mit *early prosody*. Diese Probleme wurden mit vorliegender Untersuchung vor dem Hintergrund folgender „spekulativer Axiome“ exploriert (siehe Einführung):

Axiom I: Kompetenz beim lauten Lesen ist ein Indikator für Aspekte von Textverstehen.

Axiom II: Umgekehrt können solche Aspekte über die Förderung von lautem Lesen mitgefördert werden („bootstrapping“).

Hinsichtlich Axiom I haben sich trotz verbleibender Desiderate neue Hinweise auf die Relevanz von Informationsstruktur als Indikator für Textverständnis auf hohem Niveau und damit diagnostisches Potenzial ergeben.

Hinsichtlich Axiom II kann zum Abschluss als didaktischer Ansatz das „Verstehen durch Sichselber-Vorlesen“ angeregt werden. Eine solche Art der Förderung von Textverstehen über eine Förderung von lautem Lesen ist wahrscheinlich erst angebracht und erfolgreich, wenn die Gefilde von advanced fluency erreicht sind. Unterschiedlich gezielt trainiert werden könnte neben einem *publikumswirksamen* lauten Lesen demnach auch ein *epistemisches* lautes Lesen, das beim leisen Lesen „im Notfall“ zu Rate gezogen wird. Eine entsprechende Faustregel wäre:

(5) Zusammenhänge kann man hören: *Wenn du einen Satz nicht verstehst, lies ihn dir mal selber vor.*

6 | Literatur

- Beuschel-Menze, H., Ferber, M., Grimm, H., Grimm, I., Menze, F. & Wiebel, K. H. (2009). *Das große Schülerlexikon von A-Z*. cbj.
- Blaauw, E. (1995). On the perceptual classification of spontaneous and read speech. Om Dissertations Series. *Research Institute for Language and Speech*, Utrecht University.
- Boersma, P. & Weenink, D. (2020). *Praat: doing phonetics by computer* [Computer program]. Version 6.1.16. <http://www.praat.org/> [16.09.2021].
- Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading*. U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. Government Printing Office.
- Dahan, D. (2015). Prosody and language comprehension. *WIREs Cognitive Science*, 6(5), 441-452.
- Fisseni, B. (2011): *Focus: Interpretation? Empirical Investigations on Focus Interpretation*. Universitätsverlag Rhein-Ruhr KG an der Universität Duisburg-Essen.
- Fujisaki, H. (1988). A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. *Vocal physiology: Voice production, mechanisms and functions* 2, 347-355.
- Funke, R. (2018). Lautes Lesen – Was ist das und wozu dient es? *Leseräume* 5, 88-95.
- Gamer, M., Lemon, J. & Singh, I. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement*. R package version 0.84.1. <https://CRAN.R-project.org/package=irr> [16.09.2021].
- Godde, E., Bosse, M.-L. & Bailly, G. (2020). *A review of reading prosody acquisition and development*. University Grenoble Alpes.
- Groen, M. A., Veenendaal, N. J. & Verhoeven, L. (2018). The role of prosody in reading comprehension: evidence from poor comprehenders. *Journal of Research in Reading*, 42(1), 37-57.
- Hallgren, Kevin A. (2012): Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol*, 8(1), 23-34.

- Holle, K. (2006). Flüssiges und phrasiertes Lesen (fluency). In: S. Weinhold (Hrsg.), *Schriftspracherwerb empirisch. Konzepte – Diagnostik – Entwicklung*. Schneider Verlag Hohengehren.
- Kegel, G. (1990): Sprach- und Zeitverarbeitung bei sprachauffälligen und sprachunauffälligen Kindern. In: G. Kegel u.a. (Hg.): *Sprechwissenschaft und Psycholinguistik 4. Beiträge aus Forschung und Praxis* (S. 229–255). Westdeutscher Verlag.
- KMK (2004). *Bildungsstandards im Fach Deutsch für den Primarbereich*. Beschluss vom 15.10.2004. Bonn: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.
- Kuhn, M. R., Schwanenflugel, P. J. & Meisinger, E. B. (2010). Aligning Theory and Assessment of Reading Fluency: Automaticity, Prosody, and Definitions of Fluency. *Reading Research Quarterly*, 45, 230-251.
- Laver, J. (1994). *Principles of Phonetics*. Cambridge University Press.
- Lenhard, W. & Schneider, W. (2006). *ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler*. Hogrefe.
- Lenhard, W. & Lenhard, A. (2014-2017). *Berechnung des Lesbarkeitsindex LIX nach Björnson*. <http://www.psychometrica.de/lix.html> [16.09.2021]. Psychometrica.
- Lumley T. (based on Fortran code by Alan Miller) (2020). *leaps: Regression Subset Selection. R package version 3.1*. <https://CRAN.R-project.org/package=leaps> [16.09.2021].
- Miller, G. A. (1956): The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review* 63, 81–97.
- Möbius, B. (1993). *Ein quantitatives Modell der deutschen Intonation. Analyse und Synthese von Grundfrequenzverläufen*. Narr.
- Nix, D. (2011). *Förderung von Leseflüssigkeit. Theoretische Fundierung und empirische Überprüfung eines kooperativen Lautlese-Verfahrens im Deutschunterricht*. Juventa.
- Pinnell, G. S., Pikulski, J. J, Wixson, K. K., Campbell, J. R., Gough, P. B. & Beatty, A. S. (1995). *Listening to children read aloud. Data from NAEP's Integrated Reading Performance Record (IRPR) at Grade 4*. Office of Educational Research and Improvement, U.S. Department of Education.
- Pompino-Marschall, B. (2009). *Einführung in die Phonetik*. De Gruyter.
- Pöppel, E. (1997): A hierarchical model of temporal perception. *Trends in Cognitive Sciences* 2, 56–61.
- R Core Team (2014). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. <http://www.R-project.org/> [16.09.2021].
- Rietveld, T. & Chen, A. (2006). How to Obtain and Process Perceptual Judgments of Intonational Meaning. In: S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter & J. Schließer (Hg.): *Methods in Empirical Prosody Research* (S. 283–320). De Gruyter.
- Rosebrock, C. & Nix, D. (2006). Forschungsüberblick: Leseflüssigkeit (Fluency) in der amerikanischen Leseforschung und -didaktik. *Didaktik Deutsch*, 20, 90–112.
- Rosebrock, C. & Nix, D. (2015). *Grundlagen der Lesedidaktik und der systematischen Leseförderung*. Schneider.
- Rosebrock, C., Nix, D., Rieckmann, C. & Gold, A. (2016). *Leseflüssigkeit fördern. Lautleseverfahren für die Primar- und Sekundarstufe*. Kallmeyer.
- Röttig, S., Schwerkolt, C. & Nottbusch, G. (2021). 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 (in diesem Band).

- Sappok, C. (2002). *Metrum, Füllung und Signal. Eine Untersuchung der Betonungsintervalldauern bei 32 Aufnahmen einer Strophe aus dem Gedicht „Bimini“ von Heinrich Heine*. Magisterarbeit, Technische Universität Berlin.
- Sappok, C. (2010): The quantitative organization of speech. *Proc. Speech Prosody 2010*, paper 102
- Sappok, C. & Arnold, D. (2012a): On the Normalization of Syllable Prominence Ratings. In: *Proceedings of Speech Prosody, 6th International Conference of the International Speech Communication Association*.
- Sappok, C. & Arnold, D. (2012b): More on the Normalization of Syllable Prominence Ratings. In: *Proceedings of Interspeech, 13th Annual Conference of the International Speech Communication Association*.
- Sappok, C. & Fay, J. (2018): Prosodische Aspekte von Leseflüssigkeit messen. Evaluation einer Rating-prozedur mit Audioaufnahmen von DrittklässlerInnen. *Didaktik Deutsch*, 44, 61-83.
- Sappok, C., Linnemann, M. & Stephany, S. (2020): Leseflüssigkeit – Prosodie – Leseverstehen. Eine Longitudinalstudie zur Entwicklung der Leseflüssigkeit von Jahrgangstufe 3 bis 7. In: I. Rautenberg (Hrsg.): *Evidenzbasierte Forschung zum Schriftspracherwerb* (S. 175-209). Schneider Verlag Hohengehren.
- Schwanenflugel, P. J., Westmoreland, M. R. & Benjamin, R. G. (2013): Reading fluency skill and the prosodic marking of linguistic focus. *Reading and Writing*, 26(6), 9-30.
- Stephany, S., Linnemann, M., Goltsev, E. & Becker-Mrotzek, M. (2021). Prosodische Aspekte der Leseflüssigkeit als Indikator für Lesekompetenz. Analysen mithilfe prosodischer Lupenstellen (in diesem Band).
- Tillmann H. G. & Mansell, P. (1980): *Phonetik*. Klett-Cotta.
- Weaver, W. & Shannon, C. E. (1949): *The Mathematical Theory of Communication*. Urbana Ill.
- Wimmer, H. & Mayringer, H. (2014). *Salzburger Lese-Screening für die Schulstufen 2-9*. Hogrefe.

Gerrit Helm

Das Ausmaß der Unterschiede in den Lese-Items des Salzburger Lese-Screenings 2–9

Eine linguistisch-empirische und Tablet-gestützte Analyse

1 | Einleitung

Der vorliegende Beitrag¹ verfolgt das Ziel, eine Auswahl von Lese-Items des *Salzburger Lese-Screenings 2-9* (Wimmer & Mayringer 2014) linguistisch-analytisch und empirisch zu untersuchen, um den tatsächlichen Umfang der im didaktischen Diskurs vermuteten Ungleichgewichten der (Lese-)Anforderungen dieser festzustellen. Hierzu wurden exemplarisch 30 Items der Testform A1 des *Salzburger Lese-Screenings 2-9* in einem ersten Schritt linguistisch analysiert und anschließend mit Hilfe eines computergestützten Verfahrens die individuellen Lese- und Bearbeitungszeiten einer Kohorte von 208 Schüler*innen aus fünf Grundschulen exakt erfasst. Möglich wurde dadurch, Unterschiede in den durchschnittlichen Bearbeitungszeiten einzelner Lese-Items des SLS aufzuzeigen. Ein Vergleich der linguistischen Analyse mit den empirischen Daten lässt einschätzen, ob und inwieweit ausgewählte linguistische Merkmale der Lese-Items final Unterschiede in den Bearbeitungszeiten evozieren.

Hierbei zeigt sich, dass die Lese-Items des *Salzburger Lese-Screenings* zwar durchaus geeignet sind, zwischen Klassenstufen zu differenzieren, dennoch rufen zahlreiche der Items signifikant abweichende Bearbeitungszeiten hervor, was ein Ungleichgewicht zwischen diesen anzeigt. Bemerkenswert ist, dass das (z.T. merkliche) Ausmaß der Ungleichheiten nicht unbedingt linguistisch vorauszusagen ist und sich so die Frage erhärtet, was das SLS eigentlich genau 'misst'.

Das genannte SLS 2–9, welches ursprünglich lediglich als *Screening* für „basale Lesefertigkeit“ (vgl. Wimmer & Mayringer 2014: 11) entwickelt wurde, nimmt gemeinsam mit dem ebenfalls als *Screening* entworfenen *Stolperwörter Lese-Test* (Metze 2009) und dem umfangreicheren *ELFE II Leseverständnistest* (Lenhard, Lenhard & Schneider 2018) eine mittlerweile fast exklusive Stellung in der Schulpraxis und ebenso in der (deutsch-)didaktischen Forschung ein. Zugleich übt die Fachdidaktik Kritik an diesen Verfahren, insbesondere dem SLS, und sogar Zweifel an der Validität des Vorgehens werden geäußert: Vermutet wird, dass einige der konstruierten Lese-Items (semantisch und kulturell) voraussetzungsreicher seien als andere (vgl. Gailberger & Nix 2013: 19; Esslinger 2015: 132). Darüber hinaus mangelt es den Items „i.d.R. [an] konzeptionell systematische[m] Wortmaterial“ (Zepnik 2016: 90), sodass höchstens eine Sondierung ('stärkere' vs. 'schwächere' Leser*innen) erreichbar wird (vgl. Bredel, Fuhrhop & Noack 2017).

¹ Ich bedanke mich herzlich für die Beratung und Unterstützung bei Prof. Dr. Nanna Fuhrhop.

Vor diesem Hintergrund ist es erstaunlich, dass *Screenings* und Lesediagnose-Instrumente im Allgemeinen nur selten selbst zum Objekt empirischer Forschung werden. Verlassen wird sich hier (vermutlich) auf die umfangreichen Normierungsstudien (z.B. Gärtner 2010 bzgl. SLS), die die Instrumente 'mitliefern'. Im Kontext meiner Dissertationsarbeit ergab sich die Möglichkeit, dies zumindest für einen Ausschnitt des SLS 2–9 empirisch in einer Satellitenstudie aufzuarbeiten.

Mein Beitrag beginnt mit einer Kurzzvorstellung des *Salzburger Lese-Screenings* und der daran geübten Kritik (Abschnitt 2). Generiert werden hieraus leitende Fragestellungen (Abschnitt 3). Anschließend wird eine (einfache) Form einer linguistischen Kodierung entworfen (Abschnitt 4) und auf das SLS angewandt (Abschnitt 5). Die Satellitenstudie wird vorgestellt (Abschnitt 6) und die resultierenden Ergebnisse werden aufgearbeitet (Abschnitt 7). Eine abschließende Diskussion beendet diesen Beitrag.

2 | Das Salzburger Lese-Screening 2-9: Merkmale und Kritikpunkte

Das *Salzburger Lese-Screening 2–9* (Wimmer & Mayringer 2014) ist, wie schon seine Vorgänger SLS 1–4 (Wimmer & Mayringer 2003) und SLS 5–8 (Auer et al. 2005), ein standardisiertes Verfahren mit dem Anspruch, „ökonomisch [...] Schüler mit Schwächen in basaler Lesefertigkeit“ (Wimmer & Mayringer 2014: 11) zu identifizieren, welches sowohl in Form eines Einzel- als auch eines Klassentests durchführbar ist. Im Schulkontext gehört das SLS 2–9, neben dem nicht-standardisierten *Stolperwörter Lesetest* (Metze 2009) und dem in seiner Konzeption umfangreicheren *ELFE II Leseverständnistest* (Lenhard, Lenhard & Schneider 2018), zu den verbreiteteren Verfahren². In der deutschdidaktischen (Lese-)Forschung werden diese z.B. herangezogen, um den Ertrag von Leseförderprogrammen aufzuzeigen (vgl. z.B. Belgrad et al. 2010; Lauer-Schmaltz, Rosebrock & Gold 2014) oder um Referenzwerte für weiterführende (Lese-)Diagnosen zu generieren (vgl. Sappok; Röttig et al. - beide in diesem Band). Inwieweit der schulpraktische wie auch der deutschdidaktische Umgang mit dem SLS 2–9 das mögliche Leistungsspektrum dieses bloßen *Lese-Screenings* übersteigt, möchte ich abschließend diskutieren (Abschnitt 8).

Obgleich sich die genannten drei Verfahren sicherlich in ihren Konzeptionen und hinsichtlich der Ziele voneinander unterscheiden, sind die Umsetzungen in ihren Funktionen und Item-Bögen, also das, was den Schüler*innen tatsächlich begegnet, vergleichbar: Den Proband*innen werden (i.d.R. in Paper-and-Pencil-Form) mehrseitige Item-Bögen vorgelegt, die jeweils Bilder (*ELFE* Untertest 1) oder kurze Sätze (aus je einem Verb und Ergänzungen) (*ELFE* Untertest 2; *Stolperwörter Lesetest*; SLS 2–9) enthalten, mit dem Auftrag, deren 'Plausibilität' (SLS 2-9) oder Grammatikalität (*ELFE*; *Stolperwörter*) einzuschätzen oder herzustellen. Die Anzahl der in einem festen Zeitrahmen bearbeiteten Items ergibt dann einen Leistungs(roh-)wert der Proband*innen, welcher zumeist mit Hilfe einer Normtabelle in eine Aussage über altersgemäße

² Wie mir bekannte Lehrer*innen berichten, wird das SLS nicht selten im Vorfeld von Elternsprechtagen, Zeugnis-Konferenzen, etc. eingesetzt, um gewissermaßen eine 'gute Begründung' für Notenentscheidungen an der Hand zu haben. Zum Beispiel die Grundschule am Buntzelberg (Berlin), die Alexander-von-Humboldt Mittelschule (Marktredwitz) und weitere haben (laut Internetauftritt) in ihr Schulprogramm aufgenommen, dass das SLS 2–9 stets am Ende eines Schuljahres durchgeführt wird.

Lesefertigkeit überführt werden kann (vgl. etwa Wimmer & Mayringer 2014: 16; Lenhard, Lenhard & Schneider 2018: 63).

Im Fall des *Salzburger Lese-Screenings 2–9* haben Proband*innen ein Zeitfenster von drei Minuten, um möglichst viele aus insgesamt 100 „inhaltlich relativ einfachen Sätzen“ (Wimmer & Mayringer 2014: 11) durch Ankreuzen als ‘richtig’ oder ‘falsch’ zu markieren. Die Anzahl der in drei Minuten erwartungsgemäß beantworteten Sätze liefert den Roh-Wert, der durch Normtabellen (welche für die Klassenstufen 2 bis 9 vorliegen) in einen *Lesequotienten* umgerechnet werden kann. Zur Verdeutlichung ziehe ich fünf Beispielitems aus dem SLS 2–9 hier als (1)–(5) heran:

- (1) SLS Form A1, Item 2: *Wasser ist nass.*
- (2) SLS Form A1, Item 9: *In der Nacht scheint die Sonne.*
- (3) SLS Form A1, Item 12: *Honig schmeckt sauer.*
- (4) SLS Form A1, Item 24: *Die Schale der Orange ist violett.*
- (5) SLS Form A1, Item 25: *Schuhe trägt man an den Füßen.*

Auf den ersten Blick scheint die Bearbeitung dieser Items in der Tat einfach zu sein. Erst der zweite Blick verdeutlicht, dass die ‘richtige’ Beantwortung dann doch nicht vollends unproblematisch ist (vgl. auch Esslinger 2015: 132f.). Folgende Aspekte fallen auf:

- Item 9 („In der Nacht scheint die Sonne“) soll als ‘falsch’ markiert werden, was jedoch voraussetzt, dass keinerlei astrologisch-physikalisches (Vor-)Wissen herangezogen wird (etwa: Sie scheint, doch wir sehen sie nicht).
- Item 12 („Honig schmeckt sauer.“) und Item 24 („Die Schale der Orange ist violett.“) können nur von Test-Teilnehmer*innen als ‘falsch’ markiert werden, zu deren Wissens- und Erfahrungsbasis der Geschmack von Honig bzw. das Aussehen von Orangen gehört.
- Item 24 enthält mit „Orange“ einen Gallizismus (vgl. Eisenberg 2012: 187).
- Item 24 enthält ein Kopulaverb („ist“) gegenüber einem transitiven Verb („trägt) in Item 25³.
- Item 24 enthält eine komplexe Nominalgruppe („Die Schale der Orange“) gegenüber zwei einfachen Nominalgruppen („Schuhe“+ „man“) plus einer Präpositionalgruppe („an den Füßen“) in Item 25.

Gailberger und Nix (2013) kommen bei der Betrachtung der Items der Vorgängerfassung SLS 5–8 zu dem Schluss, dass einige der Items als „kulturell bedingt voraussetzungsreich gelten müssen“ (ebd.: 19), eine Feststellung, die hier auch auf die aktualisierte Fassung des *Screenings* übertragbar wird. Eine strenge linguistische Betrachtungsweise legt ferner die Vermutung nahe, dass unterschiedliche (syntaktische) Strukturen in den Items durchaus unterschiedliche Lese-Anforderungen (noch abseits der semantischen Anforderungen) stellen.

³ Hinzu kommt, dass in einer topologischen Betrachtung (z.B. Wöllstein 2014; Gallmann 2015) das Subjekt („man“) in Item 24 ins Mittelfeld rückt, während z.B. Item 25 das Subjekt („Die Schale der Orange“) im Vorfeld des Verbes behält.

In der (schulischen) Anwendung des SLS kann dies aber bedeuten, dass es Items im SLS-Set gibt, bei denen teilnehmende Schüler*innen Zeit durch nicht unmittelbar leserelevante kognitive Prozesse verlieren, was sich dann negativ auf den erreichten Roh-Wert und so die Ergebnisse auswirkt. Auslöser wären hier nicht (mangelnde) Lesefertigkeiten, sondern (semantische) Aspekte von Vor- und Weltwissen oder Unterschiede im syntaktischen *Parsen* der Sätze.

Die offensichtlich vorliegende mangelnde Kontrolle potentiell schwierigkeiterzeugender (linguistischer) Merkmale in den Items ist für ein *Screening*-Verfahren gleich zweifach hinderlich:

- Es führt zu Zweifeln an der Validität der Instrumente. Zwar erhebt z.B. das SLS 2–9 lediglich den Anspruch, ein *Screening* zu sein, dennoch behaupten die Autoren: „Nun könnte gegen das Testkonzept eingewendet werden, dass die Beurteilung von Sätzen mehr involviert als Wortverarbeitung [...]. Alternativ kann Ineffizienz der sprachlichen Satzverarbeitung und/oder der inhaltlichen Beurteilungen zu niedrigen Testleistungen führen. [...] Das ist alles möglich, trifft aber eher nicht zu“ (Wimmer & Mayringer 2014: 16, Markierung G.H.). Eine solche Behauptung suggeriert Validitäten, die empirisch zu fundieren wären⁴. Darüber hinaus gilt der Einfluss linguistischer Faktoren auf den Lese-Prozess als Konsens (vgl. Rautenberg & Reißig 2015) und selbst die Prämisse der Ökonomie in einem *Screening* darf darüber nicht hinwegtäuschen.
- Zum Zweiten verhindert eine mangelnde Kontrolle der linguistischen Faktoren in der Testentwicklung, diese dann in die Auswertung der Testwerte mit einzubeziehen, um z.B. Aussagen zum systematischen Wissen über Buchstaben (vgl. Zepnik 2016), zum morphologischen Bewusstsein (z.B. Carlisle 2000) oder zum syntaktischen Lesen (Esslinger 2015) der getesteten Proband*innen tätigen zu können. Testergebnisse verbleiben somit auf rein quantitativer Ebene (Schüler*in X erreichte Y Punkte). Eine Antwort auf die (qualitative) Frage, an welcher Stelle Punkte gewonnen oder verloren wurden (d.h., wo sich ‘Stärken’ und/oder ‘Schwächen’ zielgerichtet fördern ließen), bleibt offen (vgl. dazu auch Bredel, Fuhrhop & Noack 2017). Sicherlich erhebt ein auf Ökonomie bedachtes *Screening* nicht den Anspruch an eine derart umfassende Diagnose, doch wird damit die Frage umso sensibler, *was genau* (dann) eigentlich erfasst wird und v.a. wie die Ergebnisse des *Screenings* gedeutet und weiterverwendet werden können⁵.

Eingeräumt werden muss, dass das *Salzburger Lese-Screening* und ebenso auch der *Stolperwörter Lesetest* und der *ELFE II* nicht den Anspruch erheben, deutschdidaktische Instrumente zu

⁴ Wimmer und Mayringer (2014) ziehen hierzu eine Studie von Hawelka und Wimmer (2008 und 2010) heran, in welcher jedoch a) erwachsene Leser*innen mit Dyslexie untersucht und b) Items eingesetzt wurden, die „vergleichbar mit denen des SLS“ (ebd.), aber nicht exakt besagte Items waren. Es verbleibt m.E. eine Hypothese, dass dort generierten Erkenntnisse auf das Lesen von Schüler*innen bei „vergleichbaren Items“ übertragbar sind.

⁵ Darüber hinaus zeigt sich, dass z.B. der durchaus umfangreichere und detaillierte *ELFE II Leseverständnistest* hier kaum unterschiedlich verfährt: Zwar gibt es hier drei Untertests zum Wort-, Satz- und Textlesen, doch sind diese auf den zweiten Blick nicht linguistisch-strukturiert konzipiert: So stehen im Wortlesetest durchaus morphologisch komplexe neben einfachen Worten, Einsilber neben Zwei- und Dreisilbern und Fremdwörter (z.B. „*Laser*“) neben Wörtern des Kernwortschatzes (z.B. „*Lampe*“) und solchen, die heute seltener gebraucht werden (z.B. „*Lumpen*“), (alle drei Item 16; Lenhard, Lenhard & Schneider 2018: 2), womit ebenfalls wenig differenzierte Aussagen über graphematisches, morphologisches oder syntaktisches Bewusstsein der Schüler*innen möglich werden.

sein. Alle drei Verfahren stammen aus psychologischen bzw. allgemeindidaktischen Kontexten⁶. Zu beobachten ist jedoch, dass sie sowohl schulisch als auch in der empirischen (Lese-)Forschung z.T. die Funktion/Aufgabe einer (umfassenden) Lesediagnose übernehmen bzw. dort jeweils als alleinige Instrumente herangezogen werden (z.B. in Belgrad et al. 2010). Obgleich die Kritik an den Instrumenten nur in diesen (nicht-intendierten) Kontexten auch angebracht wirkt, erhält die Überprüfung der Kritikpunkte so doch eine Relevanz.

3 | Forschungsvorhaben und Forschungsfrage

Die zuvor herausgestellten (Kritik-)Punkte am *Salzburger Lese-Screening*, im Speziellen an der Neu-Fassung *Salzburger Lese-Screening 2–9* (Wimmer & Mayringer 2014) lassen sich m.E. in zwei Vermutungen kondensieren:

- Vermutung 1: Einzelne Lese-Items des SLS 2–9 sind durch semantische (Vor-/ Weltwissen), v.a. jedoch linguistische Faktoren 'schwerer' resp. 'leichter' (als andere).
- Vermutung 2: Das Ungleichgewicht in den Anforderungen der Lese-Items erzeugt wiederum Unterschiede in den Lese- und Bearbeitungszeiten bei einzelnen Items (wodurch die Bearbeitung spezifischer Einzelitems einen Zeitvorteil bzw. -nachteil erzeugt).

Bereits die zuvor präsentierten Beispiel-Items aus dem SLS 2–9 zeigen auf, dass es Unterschiede z.B. hinsichtlich der linguistischen Faktoren zwischen Einzel-Items gibt. Allein die unterschiedlichen Satz-Längen bei den Items machen dies mehr als offensichtlich. Die Funktionsweise des *Salzburger Lese-Screenings* setzt jedoch eigentlich eine Konstruktion voraus, bei der alle Items 'ungefähr' vergleichbar sein sollten. Vor allem relevant ist dies, da ein jeder Item-Bogen stets in unterschiedlichen Varianten (Form A1 und A2; vgl. Wimmer & Mayringer 2014) vorliegt, in denen die Abfolge der Items variiert. Proband*innen unterschiedliche Reihungen von Items darzubieten, generiert jedoch nur dann keine Vor- oder Nachteile, wenn der Schwierigkeitsgrad der Items vergleichbar ist. Schon der erste Eindruck des SLS 2–9 (s.o.) kann jedoch Grund zur Skepsis geben. Die leitende Forschungsfrage sei deshalb nicht die nach dem *ob*, sondern stets die nach dem *Ausmaß*, dem *Scope*, der Abweichungen und Unterschiede und (im zweiten Schritt) nach deren Einfluss, dem *Impact*. In eine Forschungsfrage zusammengeführt formuliert: *In welchem Ausmaß unterscheiden sich Lese-Items im SLS 2–9 und welchen Einfluss hat dies auf die Lese-/Bearbeitungszeiten?*

Die Beantwortung kann nicht in einem Schritt erfolgen, sondern macht es notwendig, die Forschungsfrage in drei kleinere Forschungsfragen aufzuteilen, die dann nacheinander in den Fokus gerückt werden:

- Forschungsfrage 1: In welchem Ausmaß unterscheiden sich Lese-Items des SLS 2–9 hinsichtlich linguistischer Aspekte voneinander?
- Forschungsfrage 2: In welchem Ausmaß unterscheiden sich tatsächliche Lese- und Bearbeitungszeiten von einzelnen Items des SLS 2–9 voneinander?

⁶ So sind sowohl die verstorbenen SLS-Autoren Heinz Mayringer und Heinz Wimmer wie auch Wolfgang und Alexandra Lenhard (*ELFE*) Lehrende der Psychologie (gewesen) und bezeichnen sich an keiner Stelle als Deutschdidaktiker*innen.

- Forschungsfrage 3: Inwieweit hängen Unterschiede in den Lese- und Bearbeitungszeiten mit Unterschieden in den linguistischen Aspekten einzelner Items zusammen.

Für die Bearbeitung der ersten Forschungsfrage ist bereits eine theoretische Analyse der Frage-Items des SLS 2–9 unter linguistischen Gesichtspunkten erkenntnisgenerierend. Hier gilt es, eine geeignete Kodierung zu finden, die auch strukturell verschiedene Lese-Items in zueinander vergleich- und relativierbare Zahlenwerte fasst. Der hier verfolgte heuristische Ansatz besteht darin, Quantitäten und Vorkommen linguistischer Zugriffsgrößen (einfach) auszuzählen und die Items hierzu auf Segmentebene (Buchstaben und graphemische Verarbeitungseinheiten), segmentübergreifender Ebene (graphematische Silben und Worte) und wortübergreifender Ebene (Konstituentenlänge) in den Blick zu nehmen und abschließend miteinander zu vergleichen.

Die Bearbeitung der zweiten Forschungsfrage hingegen setzt ein empirisches Erhebungsverfahren voraus, welches erlaubt, Lese- und Bearbeitungszeiten einzelner Lese-Items des SLS bei allen Teilnehmer*innen Millisekunden-genau zu erfassen. Hierzu bietet sich ein computer- bzw. *Tablet*-gestütztes Verfahren an, wie es im Abschnitt 5 vorgestellt wird. *Tablets* ermöglichen eine exakte Aufzeichnung des Zeitintervalls zwischen Darbietung eines Lese-Items und der Eingabe einer Antwort durch die User*in und liefern so automatisiert Aussagen sowohl über Antwort-Richtigkeit wie auch die damit verbundene Bearbeitungszeit.

Ist eine linguistische Kodierung und Analyse der Items vorgenommen und sind empirische Messwerte zu allen Einzel-Items aufgezeichnet, ist die Beantwortung der dritten Forschungsfrage vor allem eine statistische Fragestellung: Mit Hilfe einer multiplen Regressionsanalyse soll ermittelt werden, welches Modell – und auf diese Weise, welche(r) der (linguistischen) Einflussfaktor(en) – am geeignetsten ist, Unterschiede in den Item-bezogenen Bearbeitungszeiten zu erklären. Beantwortet werden kann damit zwar nicht die Frage, *was genau* Lese- und Bearbeitungszeiten bei einzelnen Items im SLS 2–9 ‘kürzer’ oder ‘länger’ macht (und damit auch nicht die Frage, *was genau das Salzburger Lese-Screening misst*), doch wird auf dieser Basis die Aussage möglich, *welche* und v.a. *wie viele* Faktoren und Variablen hierzu als potentielle Einflussgrößen in Frage kommen. Im Idealfall wäre das Ergebnis ein Modell mit möglichst wenigen Variablen, sodass zumindest eine Eingrenzung der Faktoren z.B. auf (hauptsächlich) die segmentale oder segmentübergreifende Ebene möglich wird.

4 | Linguistische Kodierung des Salzburger Lese-Screenings 2–9

In den Fokus der linguistischen Kodierung und Analyse genommen werden die Lese-Items 1–30 der Test-Form A1 des SLS 2–9 in aktuellster Fassung (Wimmer & Mayringer 2014). Diese Auswahl ist lesedidaktisch sinnvoll, da es sich hier um die Items handelt, die i.d.R. von allen Schüler*innen (aller Klassenstufen) ‘geschafft’ werden (vgl. ebd.: 45ff.), wenn das *Screening* schulisch eingesetzt wird. Darüber hinaus sorgt diese Einschränkung für einen ‘handhabbaren’ Umfang, der nicht zu sehr das Potential hat, den Rahmen eines Sammelbandbeitrages zu sprengen. Die zu betrachtenden Items seien damit folgende:

- | | |
|--------------------------------------|---|
| [1] Bäume können sprechen. | [16] Menschen essen gerne Gras. |
| [2] Wasser ist nass. | [17] Der Gärtner pflanzt Blumen. |
| [3] Der Löwe brüllt. | [18] Ameisen können tanzen. |
| [4] Flugzeuge können fliegen | [19] Elefanten sind sehr leichte Tiere. |
| [5] Schweine haben Flügel. | [20] Schwarz ist eine helle Farbe. |
| [6] Im Winter ist es heiß. | [21] Die Armbanduhr zeigt die Zeit an. |
| [7] Pflanzen brauchen Wasser. | [22] Das Gras ist rosarot. |
| [8] Vögel bauen Nester. | [23] Schiffe fahren auf Schienen. |
| [9] In der Nacht scheint die Sonne. | [24] Die Schale der Orange ist violett. |
| [10] Steine sind hart. | [25] Die Schuhe trägt man an den Füßen. |
| [11] Ein Riese ist groß. | [26] Ein Picknick macht man im Haus. |
| [12] Honig schmeckt sauer. | [27] In einem Krankenhaus arbeiten viele Ärzte. |
| [13] Bäume haben Wurzeln. | [28] Am Nordpol gibt es viel Eis und Schnee. |
| [14] Ein Kreis ist rund. | [29] Die Kuh frisst Gras von einem Teller. |
| [15] Mit der Angel fängt man Fische. | [30] Das Besteck braucht man zum Essen. |

(Wimmer & Mayringer 2014)

In der Sprachwissenschaft existiert m.W. gegenwärtig kein normiertes Verfahren, mit dessen Hilfe ein Item-Set wie das vorliegende hinsichtlich schwierigkeiterzeugender Merkmale kodiert und ausgewertet werden kann. Verwiesen wird an einigen Stellen auf den *Lesbarkeitsindex (LIX)* nach Björnson (1968). Dieser ist jedoch sowohl durch seine Anlage als Text-Analyse-Werkzeug und durch seine Funktionsweise, den Grenzwert von sechs Buchstaben als Maßstab für Leseschwierigkeit zu nutzen (jedes Wort mit mehr als sechs Buchstaben gilt als 'schwer', jedes mit weniger Buchstaben als 'leicht'), für diese Analyse ungeeignet.

Zum Einsatz kommen soll daher ein eigenes heuristisches Kodierungs-Raster. Dieses Raster ist ein lesedidaktisches, das keinerlei Anspruch auf Vollständigkeit aus linguistischer Perspektive erheben kann: Zu beobachten ist, dass im sprachwissenschaftlichen und sprachdidaktischen Diskurs Definitionen und Auffassungen spezifischer Bezugsgrößen zuweilen nebeneinanderstehen. Dies kann jeder Kodierung durchaus Komplexität verleihen (als prominentes Beispiel sei hier auf den kontrovers geführten Diskurs zum *Graphem*-Begriff verwiesen (Gallmann 1985; Günther 1988; Neef 2005; Rezec 2009; Primus 2004, 2006, 2010 uvm.)). Vorgeschlagen wird deshalb eine Form der Kodierung, die auf basale Zugriffsgrößen auf Segmentebene, segmentübergreifender Ebene und wortübergreifender Ebene abzielt.

4.1 | Kodierung und Analyse auf Segment-Ebene

Ein erster (auf der Hand liegender) Faktor für die Kodierung und Analyse ist sicherlich, die Anzahl der ‘Buchstaben’ pro Lese-Item zu ermitteln⁷. Rückbinden lässt sich dies z.B. an die ‘Buchstaben’-Auffassung von Neef (2005): Dort werden neben dem Inventar des Modernen Römischen Alphabets (MRA) auch die Elemente |ä|, |ö|, |ü| sowie |ß| zu den ‘Buchstaben’ gezählt. Erreicht wird so quasi die Zählung der Tastatur-Anschläge ohne Leerzeichen. Konkret werden z.B. |sch| und |ch| als drei bzw. zwei ‘Buchstaben’ gezählt, das Wort „Schienen“ besteht in der Analyse also aus acht ‘Buchstaben’ (|s|+|c|+|h|+|i|+|e|+|n|+|e|+|n|). Operiert wird mit dem Inventar:

‘Buchstaben’: a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, r, s, t, u, v, w, x, y, z, ä, ö, ü, ß

Jeder Versuch, einen Faktor oberhalb (also größer) der benannten ‘Buchstaben’ für die Kodierung und Analyse heranzuziehen, führt unumgänglich in den umfangreichen und teilweise (stark) kontrovers geführten Diskurs um den Terminus *Graphem*. Dieser kann hier freilich nur angedeutet werden (vgl. etwa: Gallmann 1985; Kohrt 1985; Günther 1988; Rezec 2009; Primus 2004, 2006, 2010; und zur Frage, ob ein solcher überhaupt sinnvoll anzusetzen ist, vgl. z.B. Neef 2005). Auf einer Meta-Ebene eint alle Diskurspositionen der Grundgedanke, dass es im Umfeld der oben herausgestellten ‘Buchstaben’ Einheiten gibt, die (beim Lesen) gemeinsam (d.h. zusammen) verarbeitet werden (können). Abgezielt wird hierbei häufig auf Kombinationen wie |sch|, |ch|, |sp|, |st|, aber auch auf |ie|, |ei|, |äu|, |ng| bzw. auf |en|, |el| und andere mehr. Im Rahmen meiner Dissertationsarbeit entwickle ich deshalb den Vorschlag, lesedidaktisch keiner spezifischen Leitposition den Vorrang zu gewähren und zugleich auch keine auszuschließen. Es soll also weder eine laut-geleitete Auffassung (z.B. Bloomfield 1933; Trost 1939), eine schriftfunktional-geleitete Auffassung (z.B. Gallmann 1985; Günther 1988; Eisenberg 2013) noch eine silbenstrukturell-geleiteten Auffassung (z.B. Primus 2004, 2006, 2010; Primus, Berg & Wagner 2016) uneingeschränkt gelten. Stattdessen wird als *graphemische Verarbeitungseinheit* eine Größe angesetzt, die alle ‘potentiellen Grapheme’ kodiert. Die Argumentationskette ist hier folgende⁸:

Zwei oder drei ‘Buchstaben’ bilden gemeinsam eine *graphemische Verarbeitungseinheit*, wenn sie...

1. gemeinsam mit einem Laut korrespondieren (vgl. Trost 1939; Bloomfield 1933; Rezec 2009).
2. gemeinsam eine Bedeutungsunterscheidung in schriftlicher Sprache evozieren (vgl. Gallmann 1985; Günther 1988; Eisenberg 2013a; Fuhrhop 2013).
3. gemeinsam eine Skelettposition der graphematischen Silbe besetzen (vgl. Primus 2004, 2010; Berg, Primus & Wagner 2016; Berg 2019).

⁷ Ich setze den Terminus ‘Buchstaben’ hier in einfache Anführungszeichen, um zu markieren, dass mein Beitrag an dieser Stelle eher das Alltagsverständnis dieses Terminus meint und sich nicht konkret auf eine spezifische Definition des Begriffes (die es durchaus gibt) beruft. Aus eben diesem Grund setze ich ferner das Inventar dieser Bezugsgröße in |_|.

⁸ Für eine ausführliche Erläuterung möchte ich auf meine in der Vorbereitung befindliche Dissertationsarbeit: „Über die Entwicklung eines (Tablet-)Lese-Screenings zur Diagnose hierarchieniedriger Lese-Fertigkeiten aus sprachdidaktischer Perspektive“ verweisen.

4. gemeinsam in hoher Frequenz in schriftlicher Sprache vorkommen (nach Berg 2019).

In allen weiteren Fällen gelten 'Buchstaben' dann isoliert als *graphemische Verarbeitungseinheiten*. In Kombination der Argumente entsteht folgendes Inventar:

| | |
|---|---|
| Graphemische Verarbeitungseinheiten: | a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, r, s, t, u, v, w, x, y, z, ä, ö, ü, ß, sc, sch, ch, ck, qu, ng, au, äu, eu, ei, ie, ih, eh, uh, oh, ah, ee, st, sp, er, el, te, en, in, an, le, tz |
|---|---|

Ein Wort wie „Schienen“ besteht somit aus 4 Einheiten (|sch|+|ie|+|n|+|en|), das Item „Bäume können sprechen“ aus 14 Einheiten (|B|+|äu|+|m|+|e|+|k|+|ö|+|n|+|n|+|en|+|sp|+|r|+|e|+|ch|+|en|).

4.2 | Kodierung und Analyse auf segmentübergreifender Ebene

Unabhängig davon, welches Verständnis und welcher Faktor graphematisch als Segment angesetzt werden, ist die nächstgrößere Zugriffsgröße die Silbe. Orientiert werden kann sich hier an der Auffassung der graphematischen Silbe nach Fuhrhop/Buchmann (2016; ebenso Fuhrhop 2013: 216ff.; Eisenberg 2013a: 296ff.). Obgleich die exakte Position von Silbengrenzen in Einzelitems Anlass zur Diskussion bieten kann, führt dies jedoch kaum zu einer Veränderung der Quantitäten der vorzufindenden Silben. Es ließen sich z.B. Argumente finden, ein graphematisches Wort wie „sprechen“ als „sprech-en“ oder „spre-chen“ aufzufassen, doch bleibt es in beiden Analysen ein Zweisilber und die Silben-Quantität ist mit „2“ zu beziffern. Auf diese Weise führt die Kodierung/Analyse der graphematischen Silben in den Lese-Items m.E. stets zu eindeutigen Zahlenwerten: Ein Item wie „Bäume können sprechen“ besteht in dieser Logik aus sechs graphematischen Silben (drei Worte à zwei Silben).

Ein weiterer vergleichbar intuitiver Faktor für die Kodierung auf segmentübergreifender Ebene ist die Anzahl der Worte pro Lese-Item. Zurückgegriffen werden kann hier auf die Definition des graphematischen Wortes nach Fuhrhop (2013). Diese generiert eine eindeutige Möglichkeit der Kodierung: „Das graphematische Wort ist eine Graphemfolge, die durch Leerzeichen (Spalten) begrenzt ist und selbst keine internen Leerzeichen enthält“ (ebd.: 251). Ein Lese-Item wie „Bäume können sprechen“ enthält in der Konsequenz drei graphematische Worte, die durch Leerzeichen begrenzt sind, ein Item wie „In einem Krankenhaus arbeiten viele Ärzte“ derer sechs. Komposita, wie das genannte „Krankenhaus“ aber auch „Armbanduhr“, gelten stets als ein graphematisches Wort (da sie selbst keine internen Leerzeichen enthalten).

Eine dritte segmentübergreifende Größe (noch unterhalb des graphematischen Wortes) ist das Morphem. Morpheme und die morphologische Struktur als Einflussfaktoren auf das Lesen sind an vielen Stellen betrachtet und nachgewiesen worden (z.B. Carlisle & Stone 2005; Bredel, Noack & Plag 2013; Bangel & Müller 2014, 2015; Fix 2015). Aus den genannten Arbeiten jedoch ein Werkzeug zur Kodierung und Analyse herzuleiten, verbleibt eine Herausforderung: v.a. bleibt ungeklärt, in welchen Fällen das Auftreten morphologischer Komplexität als Lesehilfe und in welchen als Herausforderung fungiert. Zwar gilt die morphologische Konstanthaltung bei Stämmen und Affixen („Bäcker“ nicht „*Becker“ wegen „backen“) als Lesehilfe (z.B. Bangel

& Müller 2015:21), doch zeigt Fix (2015) auch, dass das Auftreten genauso wie das Nicht-Auftreten morphologischer Komplexität als Leseherausforderung gelten kann. Im anglo-amerikanischen Raum weisen Carlisle & Stone (2005) bzw. Carlisle & Katz (2006) nach, dass Proband*innen Zweisilber aus zwei Morphemen (z.B. „sadly“, „shady“) schneller lesen als Zweisilber aus nur einem Morphem („lady“, „pretty“). Vor diesem Hintergrund soll in der vorliegenden Kodierung auf eine Betrachtung der morphologischen Strukturen verzichtet werden.

4.3 | Kodierung und Analyse auf wortübergreifender Ebene

Wie bei den Morphemen ist auch der Einfluss der syntaktischen Satz-Struktur auf das Lesen an vielen Stellen nachgewiesen worden (z.B. Esslinger 2015; Zepnik & Zepter 2015; zuvor Huggins 1977; Bentin et al. 1990). Es gilt, dass steigende syntaktische Komplexität eine Leseherausforderung darstellt (vgl. z.B. Marthohardjono et al. 2005). Während der Testbogen A1 des SLS 2–9, insgesamt betrachtet, eine große Diversität in den syntaktischen Strukturen aufweist (angefangen von den Stelligkeiten der Verben bis hin zur Verteilung der Konstituenten auf die topologischen Felder), unterscheiden sich die Items 1–30 in erster Linie in der Komplexität (der Länge) der Wortgruppen bzw. Konstituenten (vgl. Eisenberg 2013b: 27). Hier soll die Annahme leitend sein, dass Leser*innen alle Bestandteile einer syntaktischen Konstituente (z.B. „sehr leichte Tiere“) hierarchisch integrieren (müssen) (vgl. Esslinger 2015: 122). Eine erhöhte Anzahl an graphematischen Worten in einer Konstituente lässt damit eine Zunahme der Lese-Zeit erwarten. Kodiert werden soll deshalb die durchschnittliche Länge der syntaktischen Konstituenten pro Lese-Item. Ein Item wie „Wasser ist nass“ erhält in dieser Logik den Durchschnittswert „1“ (Zwei Konstituenten „Wasser“ und „nass“ aus jeweils einem Wort), das Item „In der Nacht scheint die Sonne“ erhält hingegen den Durchschnittswert „2,5“ (eine Konstituente aus drei „In der Nacht“ und eine Konstituente aus zwei Worten „die Sonne“).

5 | Ergebnisse der linguistischen Kodierung und Analyse

Die Abbildungen 1 und 2 liefern eine grafische Aufarbeitung der Quantitäten der segmentalen Zugriffsgrößen der ersten 30 Items der Form A1 des SLS 2–9. Die x-Achse kodiert jeweils die Nummer des betrachteten Lese-Items (in der Reihenfolge, wie diese in der Form A1 vorkommen; s. oben), während die y-Achse die Quantität des jeweiligen Faktors zeigt. Die schwarze Waagerechte markiert den Mittelwert, die gestrichelten Linien in blau und rot jeweils eine (blau) bzw. 1,5 Standardabweichungen (rot):

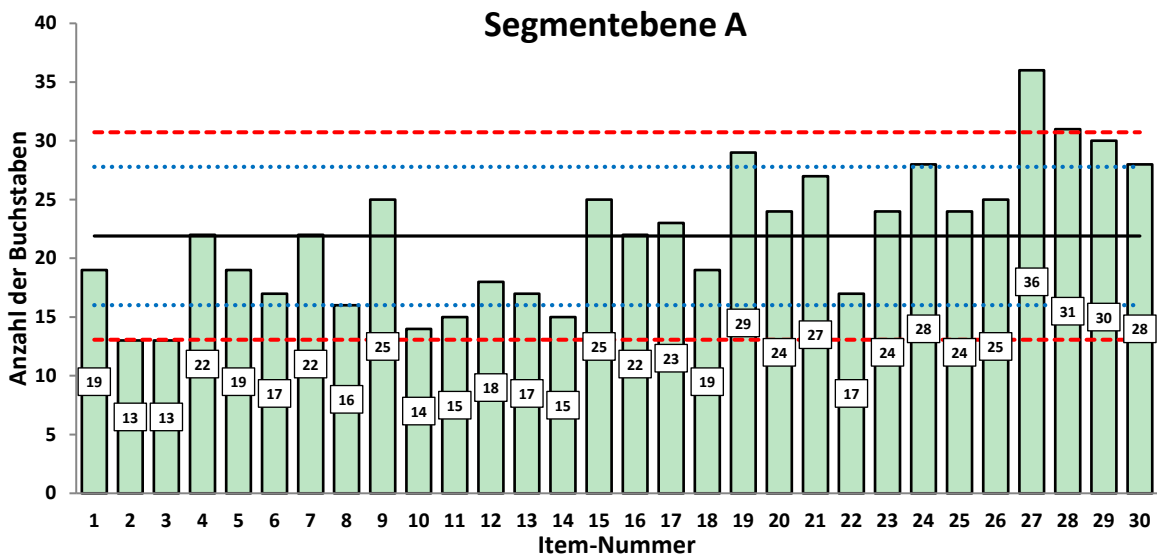


Abb. 1: Die Anzahl der Buchstaben in den 30 Lese-Items des SLS 2–9.

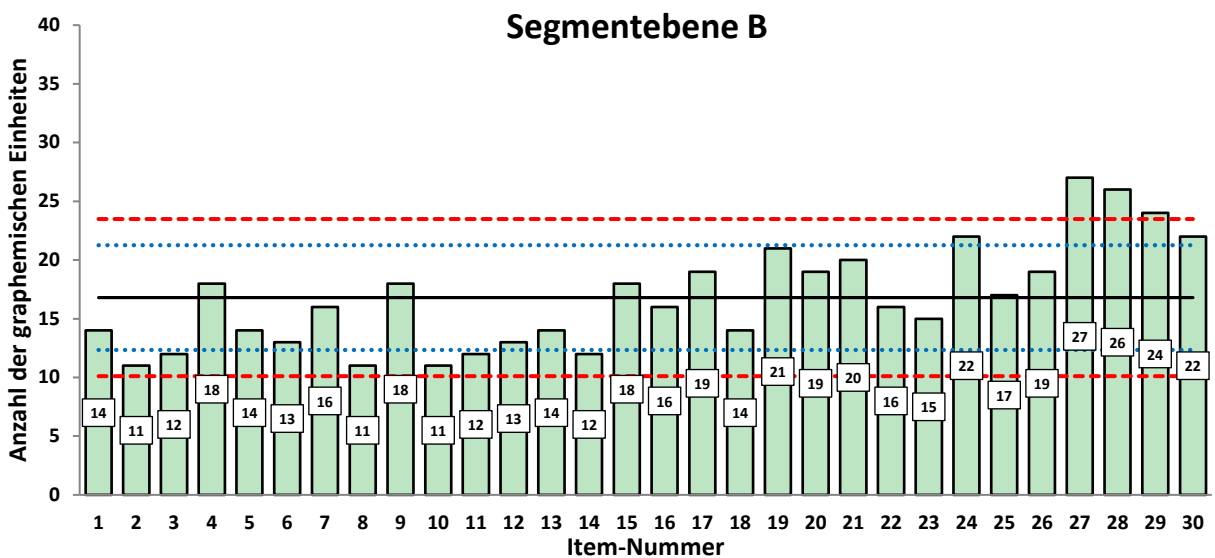


Abb. 2: Die Anzahl der graphemischen Einheiten in den 30 Lese-Items des SLS 2–9 .

Die linguistische Analyse auf Segmentebene legt offen, was zuvor (oben) bereits vermutet wurde: Es gibt eine starke Varianz in den Quantitäten der betrachteten linguistischen Zugriffsgrößen. Zieht man den Grenzwert von 1,5 Standardabweichungen als Markierung für eine relevante Abweichung heran, so kann bei einigen Lese-Items davon gesprochen werden, dass sich diese auf Segmentebene merklich vom Durchschnitt der Lese-Items abheben. Insgesamt ist zu beobachten:

- Die Items 27 („In einem Krankenhaus arbeiten viele Ärzte“) und 28 („Am Nordpol gibt es viel Eis und Schnee“) weichen sowohl hinsichtlich der Buchstaben als auch hinsichtlich der graphemischen Verarbeitungseinheiten deutlich (nach oben) ab.

- Die Items 2 („Wasser ist nass“) und 3 („Der Löwe brüllt“) weichen hinsichtlich der Buchstaben um mehr als 1,5 Standardabweichungen und hinsichtlich der graphemischen Einheiten um mehr als eine Standardabweichung (nach unten) ab.
- Weitere Items (8, 10, 11, 14, 24 und 29) weichen hinsichtlich beider Kodierungen um eine Standardabweichung vom Durchschnitt ab.

Denkbar ist, dass insbesondere die genannten Items bei der Durchführung des SLS einen Zeitvor- oder -nachteil generieren. Hinsichtlich der Forschungsfrage (1) nach dem Ausmaß der Unterschiede zwischen Items kann bereits an dieser Stelle festgestellt werden, dass dies durchaus erheblich ist, wenn vier aus 30 Items bereits auf Segmentebene derart differieren.

Die Abbildungen 3 und 4 zeigen die Quantitäten der linguistischen Zugriffsgrößen auf segmentübergreifender Ebene:

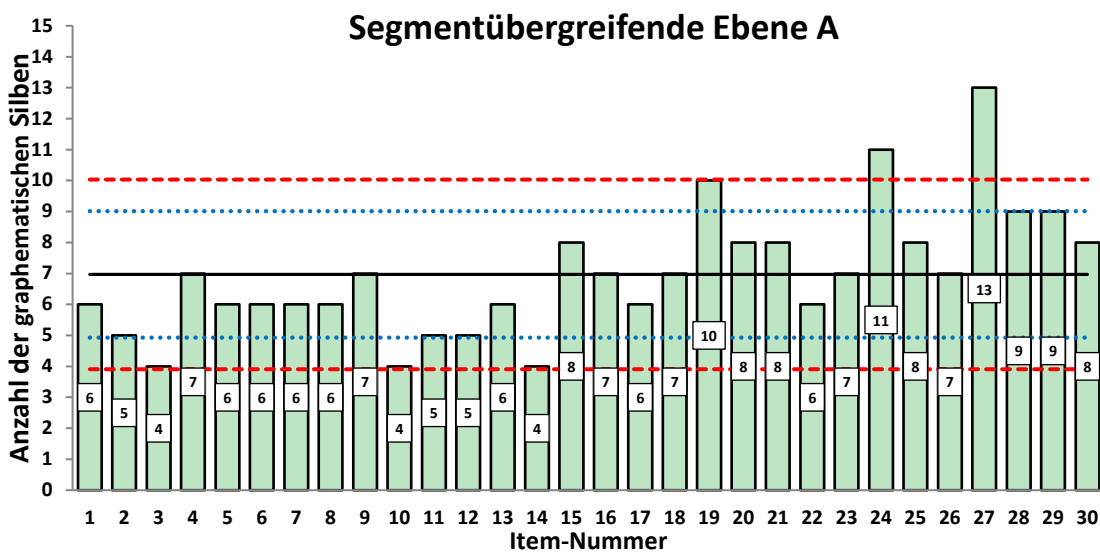


Abb. 3: Die Anzahl der graphematischen Silben in den 30 Lese-Items des SLS 2-9

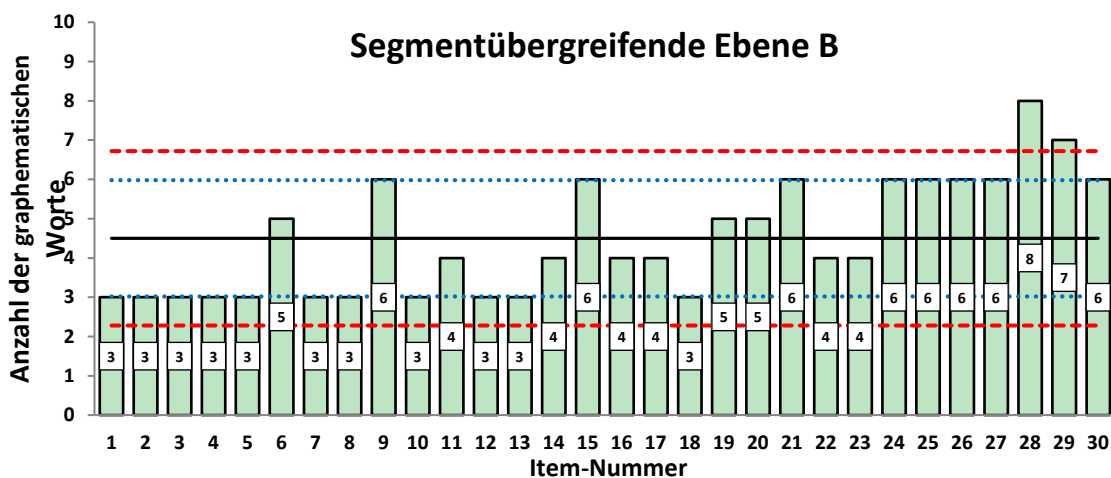


Abb. 4: Die Anzahl der graphematischen Worte in den 30 Lese-Items des SLS 2-9

Auf der segmentübergreifenden Ebene setzen sich die bereits auf Segmentebene getätigten Beobachtungen weitgehend fort. Festzustellen ist:

- Die Items 27 („In einem Krankenhaus arbeiten viele Ärzte“) und 24 („Die Schale der Orange ist violett“) weichen hinsichtlich der Anzahl der Silben signifikant (nach oben) ab, das Item 19 („Elefanten sind sehr leichte Tiere“) 1,49 Standardabweichungen.
- Auf der Ebene der Silben weicht kein Item mehr als 1,5 Standardabweichungen nach unten ab, einige Items (3, 10 und 14) jedoch um mehr als eine Standardabweichung.
- Die Items 28 („Am Nordpol gibt es viel Eis und Schnee“) und 29 („Die Kuh frisst Gras von einem Teller“) weichen hinsichtlich der Anzahl der graphematischen Worte signifikant vom Durchschnitt ab.
- Kein Item weicht auf Wortebene stärker nach unten ab als eine Standardabweichung.

Die Abbildung 5 zeigt die grafische Aufarbeitung der Quantitäten auf wortübergreifender Ebene:

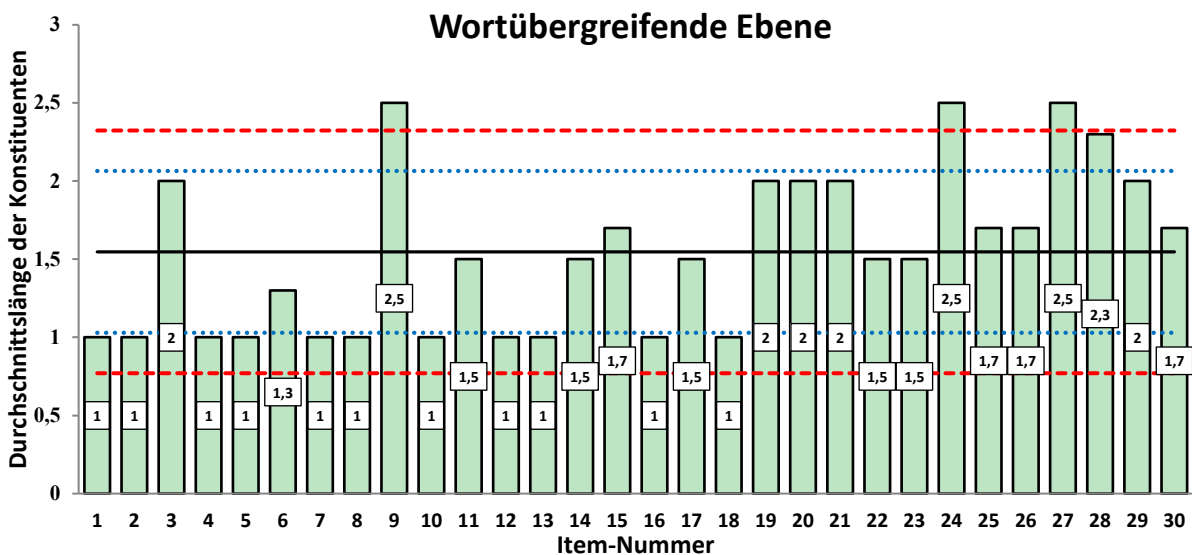


Abb. 5: Die durchschnittliche Länge der syntaktischen Konstituenten in den 30 Lese-Items

Einige der Lese-Items, die auf wortübergreifender Ebene bzw. hinsichtlich der syntaktischen Konstituenten auffällig werden, sind jene, die bereits auf Segment- und segmentübergreifender Ebene auffällig wurden:

- Item 27 („In einem Krankenhaus arbeiten viele Ärzte“) und 24 („Die Schale der Orange ist violett“) weichen hinsichtlich der durchschnittlichen Konstituentenlänge mehr als 1,5 Standardabweichungen vom Item-Durchschnitt ab.
- Item 9 („In der Nacht scheint die Sonne“) weicht ebenso um mehr als 1,5 Standardabweichungen in der Konstituentenstruktur ab.
- Das zuvor ebenso häufig auffällige Item 28 („Am Nordpol gibt es viel Eis und Schnee“) weicht auch auf Konstituentenebene um mehr als eine Standardabweichung nach oben ab

- Kein Item weicht mehr als eine Standardabweichung nach unten ab, wenn die Konstituentenstruktur betrachtet wird.

Als kurzes Zwischenfazit zur linguistischen Kodierung und Analyse der ausgewählten SLS-Items lässt sich festhalten: Auf allen betrachteten Analyseebenen finden sich Lese-Items, deren Abweichung 1,5 Standardabweichungen vom Mittelwert übersteigt. Einzelitems wie Item 24, 27 und 28 werden sogar gleich auf mehreren Ebenen auffällig. Abweichungen erfolgen häufiger ‘nach oben’ als ‘nach unten’ und lassen damit verlängerte Lese- und Bearbeitungszeiten erwarten. Die Tabelle 1 liefert eine Zusammenfassung der linguistischen Analyse und präsentiert vor allem, wie viele Lese-Items ‘leicht’ (mehr als eine Standardabweichung (nach oben/unten)) oder ‘stark’ (mehr als 1,5 Standardabweichungen (nach oben/unten)) differieren:

| Linguistischer Parameter | MW | Std-Abw. | Anzahl der Items mit leichten Abw. (+/- 1 Std.Abw) | Anzahl der Items mit starken Abw. (+/- 1,5 Std.Abw) | Korrelation zw. Item-Position und Parameter. |
|-----------------------------------|------|----------|--|---|--|
| Anzahl der Buchstaben | 21,9 | 5,88 | 8 (8,10,11,14,19,24,29,30) | 4 (2,3,27,28) | $p < 0.01$ (Spearman-Rho: .73) |
| Anzahl der graph. Einheiten | 16,8 | 4,46 | 8 (2,3,8,10,11,14,24,30) | 3 (27,28,29) | $p < 0.01$ (Spearman-Rho: .74) |
| Anzahl der graph. Silben | 6,97 | 2,04 | 4 (3,10,14,19) | 2 (24,27) | $p < 0.01$ (Spearman-Rho: .72) |
| Anzahl der graph. Worte | 4,5 | 1,48 | 9 (9,15,22,24,25,26,27,28,30) | 1 (29) | $p < 0.01$ (Spearman-Rho: .77) |
| Durchschnittl. Konstituentenlänge | 1,55 | 0,51 | 1 (28) | 3 (9,24,27) | $p < 0.01$ (Spearman-Rho: .61) |

Tab. 1: Ergebnis der linguistischen Auswertung der Lese-Items 1–30

Insgesamt stellt die Analyse folgende Items als *auffällig abweichend* und damit relevant/interessant für die weitere Analyse heraus:

Abweichung nach unten (→ Zeitvorteil?): Item 2, Item 3, Item 10

Abweichung nach oben (→ Zeitnachteil?): Item 9, Item 24, Item 27, Item 28

Bereits im Vorfeld des empirischen Teils der Studie wird damit interessant, wie sich Lese- und Bearbeitungszeiten der teilnehmenden Schüler*innen insbesondere bei den genannten sieben Items verhalten. Die Erwartung ist, dass auch die empirischen Zeit-Messwerte zu diesen Items abweichen, während die verbleibenden Werte nahe dem Durchschnitt liegen (sollten).

6 | Empirische Analyse des Salzburger Lese-Screenings 2–9

Auf Android-Tablets (konkret: Lenovo Tab M10) mit kontraststarkem 10,1 Zoll FullHD-Display wurde eine Variante einer (im Rahmen meines Dissertationsprojektes entwickelten) Lese-App mit den 30 Lese-Items aus der Form A1 des SLS 2–9 aufgespielt. Die App präsentiert jeweils ein einzelnes Lese-Item des Sets in schwarzer Fibel-Druckschrift (Fibel Nord) in konstanter Schriftgröße vor gleichbleibend weißem Hintergrund. Ein grüner Haken („dieser Satz stimmt“) oder

ein rotes Kreuz (=„dieser Satz stimmt nicht“) im unteren Teil des Bildschirms dienen zur Bearbeitung/Beantwortung der Items (durch Antippen) vergleichbar zur Print-Version des SLS. Ferner können Items mit einem Tippen auf ein gelbes Fragezeichen übersprungen werden. Jedes Item wird durch einen Countdown („...“ „:“ „:“) angekündigt. Die *App* erfasst das Zeitintervall von der Darbietung des Items (Stimulus) am Ende des Countdowns bis zur Eingabe einer Antwort durch die Proband*in⁹. Die Items werden ferner durch die *App* vor jeder Durchführung in der Reihenfolge randomisiert, sodass jede Proband*in eine individuelle Abfolge der Items bearbeitet.

Vor der Hauptuntersuchung präsentiert die *App* jeweils ein ‘Tutorial’, in welchem ihre Abläufe und Funktionsweise durch eine eingesprochene Stimme erklärt und fünf Übungs-Items bearbeitet werden.

Insgesamt nahmen 363 Schüler*innen der ersten bis vierten Klasse aus insgesamt fünf Grundschulen im Grenzgebiet von Niedersachsen und NRW an der Studie teil: Davon 98 Erstklässler*innen, 93 Zweitklässler*innen, 96 Drittklässler*innen und 76 Viertklässler*innen. Obgleich das SLS 2–9, wie schon der Name andeutet, nicht für die Klassenstufe 1 vorgesehen ist, legt das Manual des SLS 2–9 doch nahe, dass die Testung dieser Schulstufe nicht ausgeschlossen sein sollte: „Die ersten 50 Sätze des vorliegenden SLS entsprechen bezogen auf Länge und Inhalt weitgehend denjenigen des SLS 1–4“ (Wimmer & Mayringer 2014: 36)¹⁰. Da das gewählte *Tablet*-gestützte Vorgehen es nicht notwendig macht, dass für eine getestete Klassenstufe eine Normtabelle verfügbar ist, sprach also mehr für als gegen den Einbezug der 1. Klassen in dieser Testung.

Die *Tablet*-Stationen, bestehend aus einem *Tablet* mit Standhalterung und einem kind-gerechten Kopfhörer, wurden jeweils in Neben- oder Gruppenräumen in den Schulen aufgebaut, wo Kleingruppen von Schüler*innen unter Aufsicht der Testdurchführenden an diesen arbeiteten. Die Testleitenden setzten den Schüler*innen die Kopfhörer auf und starteten die *App*, die das weitere Vorgehen dann selbst steuerte. Nach Abschluss jeder Testung beantworteten die Schüler*innen kurze Fragen zu ihrer Mediennutzung im Alltag sowie zu den zu Hause und im Bekanntenkreis gesprochenen Sprachen, welches auf einem Protokollbogen notiert wurde.

91 der Schüler*innen gaben an, zu Hause noch eine weitere oder eine andere Sprache zu sprechen als in der Schule und wurden somit aus der Kerngruppe herausgenommen. Es verblieben so 272 monolinguale Schüler*innen in der Haupt-Testung.

7 | Ergebnisse

In einem ersten Schritt sollen Datensätze von Schüler*innen aus dem Gesamtdatensatz herausgefiltert werden, die die Items nicht gelesen, sondern die Antworten ‘geraten’ haben. Wie auch in der Print-Fassung des SLS sind Ergebniswerte solcher Schüler*innen daran zu erkennen,

⁹ Konkret wird die jeweilige System-Zeit der Tablets zum Zeitpunkt der Darbietung und der Eingabe gespeichert und die Differenz berechnet. Damit ist auszuschließen, dass die z.B. System-Auslastung die Messwerte beeinflussen kann.

¹⁰ Aus linguistischer Perspektive hat auch diese Äußerung der SLS-Autoren lediglich den Status einer Hypothese: Kein einziges Item der Formen A1 oder A2 des SLS 1–4 (Wimmer & Mayringer 2003) findet eine Wiederverwendung im SLS 2–9 (Wimmer & Mayringer 2014), es handelt sich um zwei völlig einzigartige Item-Sets.

dass ca. 50% der Items, in jedem Fall aber weniger als 75% der Items (23 von 30) korrekt bearbeitet wurden¹¹. Solche Datensätze werden in der hier dargebotenen Auswertung nicht berücksichtigt. Konkret umfasst der Datensatz damit 272 Ergebnissätze aus 17 Schulklassen:

| | Summe | Mädchen | Jungen |
|----------|--|---------|--------|
| Klasse 1 | 20 Schüler*innen (aus 3 ersten Klassen) | 10 | 10 |
| Klasse 2 | 65 Schüler*innen (aus 5 zweiten Klassen) | 32 | 33 |
| Klasse 3 | 73 Schüler*innen (aus 5 dritten Klassen) | 35 | 38 |
| Klasse 4 | 50 Schüler*innen (aus 4 vierten Klassen) | 24 | 26 |

Tab. 2: Zusammensetzung des Datensatzes

Die Abbildung 7 zeigt die durchschnittlichen Lese- und Bearbeitungszeiten der Lese-Items, die Proband*innen als 'stimmt' markieren mussten. Die Abbildung 8 (auf der Folgeseite) zeigt die Ergebniswerte der Items, die als 'stimmt nicht' markiert werden mussten.

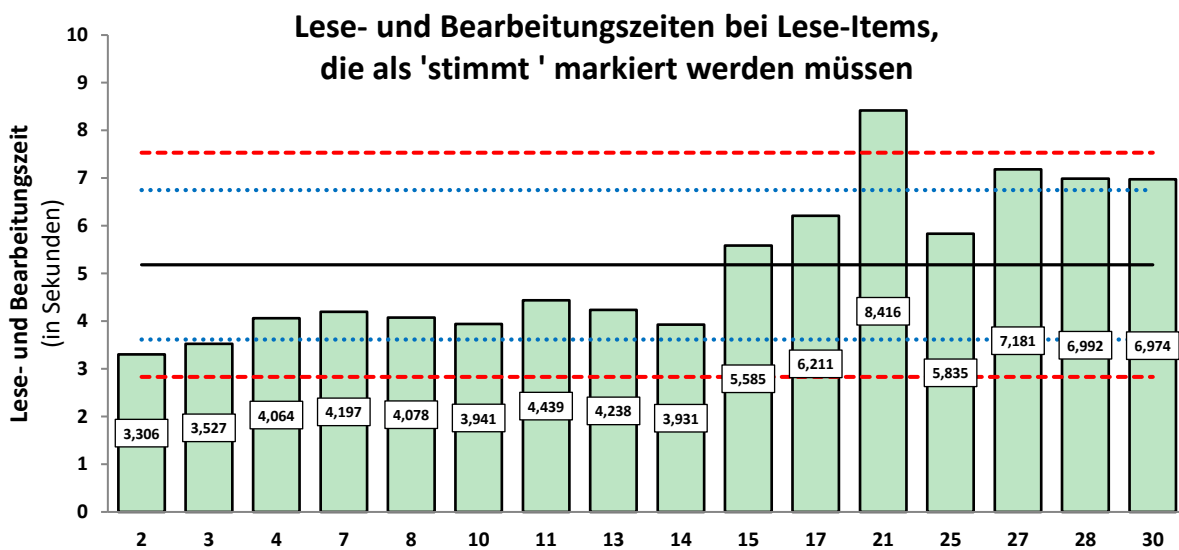


Abb. 7: Durchschnittliche Lese- und Bearbeitungszeiten aller Schüler*innen bei Items, die als 'stimmt' markiert werden mussten.

Erkennbar wird, dass die durchschnittlichen Lese- und Bearbeitungszeiten Item-bezogen durchaus variieren: Sie bewegen sich zwischen 3,3 und 8,9 s. Die Bearbeitung von Items, die als 'stimmt nicht' markiert werden sollen, erfordert durchschnittlich 0,34 Sekunden zusätzliche Zeit gegenüber solchen, die als 'stimmt' markiert werden müssen. Dieser Wert liegt sehr nahe

¹¹ Unter den ersten 30 Items müssen 14 Items als 'falsch' und 16 als 'richtig' markiert werden. Ein stetiges Tippen auf eine Schaltfläche sorgt also für 14 oder 16 korrekte Reaktionen. Für die hier präsentierte Auswertung wurden lediglich Daten von Schüler*innen herangezogen, die mehr als 23 korrekte Antworten (75%) lieferten.

an der von Krueger (1973) mit Tachistoskop-Verfahren festgestellten durchschnittlichen Verzögerung für *false*-Antworten in Klassenstufe 4 von 0,32 Sekunden (vgl. ebd.: 432). Die Datenwerte liegen hier also im Erwartungshorizont.

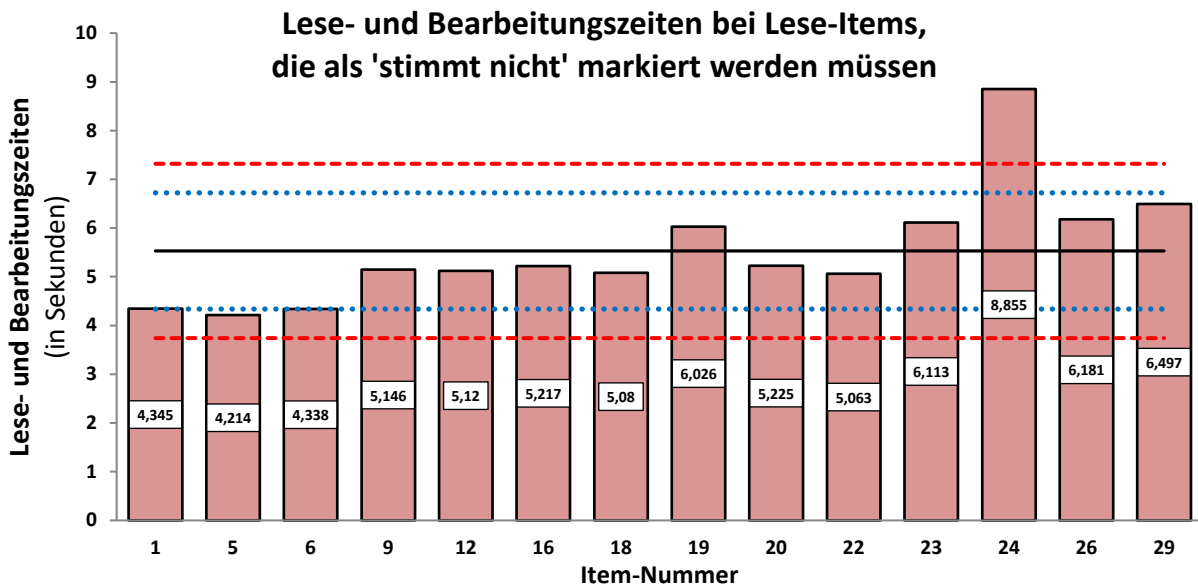


Abb: 8: Durchschnittliche Lese- und Bearbeitungszeiten aller Schüler*innen bei Items, die als 'stimmt nicht' markiert werden mussten.

Neun der 30 Items des SLS weichen in der gemessenen Bearbeitungszeit um mehr als eine Standardabweichung vom jeweiligen Mittelwert ab. Die Items 21 („Die Armbanduhr zeigt die Zeit an“) und 24 („Die Schale der Orange ist violett“) sogar um mehr als zwei (!) Standardabweichungen. Bereits auf dieser Basis kann die einleitend entwickelte Forschungsfrage (2) nach dem tatsächlichen Ausmaß der Abweichungen in den Lese- und Bearbeitungszeiten in der Art beantwortet werden, dass auch dieses Ausmaß erheblich ist: Mit einer Differenz von fast 5,6 Sekunden vom 'schnellsten' zum 'langsamsten' Item ist die Streuung größer als der Mittelwert der Bearbeitungszeiten insgesamt und auch die Relation von neun Items mit starken (signifikanten) Abweichungen zu lediglich 21 Items mit vergleichbaren Zeitwerten (1/3 zu 2/3) ist bemerkenswert.

Die nachfolgenden Abbildungen 9 und 10 unterscheiden die Lese- und Bearbeitungszeiten jeweils klassenstufenspezifisch. Die Werte für Klassenstufe 1 werden hier zunächst ausgespart, da diese eine derart große Differenz zu denen der Klassenstufen 2-4 aufweisen, dass alle vier Klassenstufen nicht sinnvoll in ein gemeinsames Diagramm überführt werden können.

Sichtbar wird, dass deutliche Differenzen in den Messwerten zwischen allen Klassenstufen bestehen. Hierbei gilt bei allen Items ausnahmslos: (Klasse 1 >) Klasse 2 > Klasse 3 > Klasse 4. Die Klassenstufe korrelieren signifikant ($p < 0.01$) mit den Lese- und Bearbeitungszeiten. Es zeigt sich hinsichtlich der Items, die als 'stimmt' markiert werden sollten, eine mittlere negative Effektstärke (Eff. -.491) und hinsichtlich der Items, die als 'stimmt nicht' markiert werden sollten, sogar eine hohe negative Effektstärke (Eff. -.528) nach Cohen (1988). Dies heißt (erwartungsgemäß): Je höher die Klassenstufe, desto geringer die Lese- und Bearbeitungszeit bei den Items.

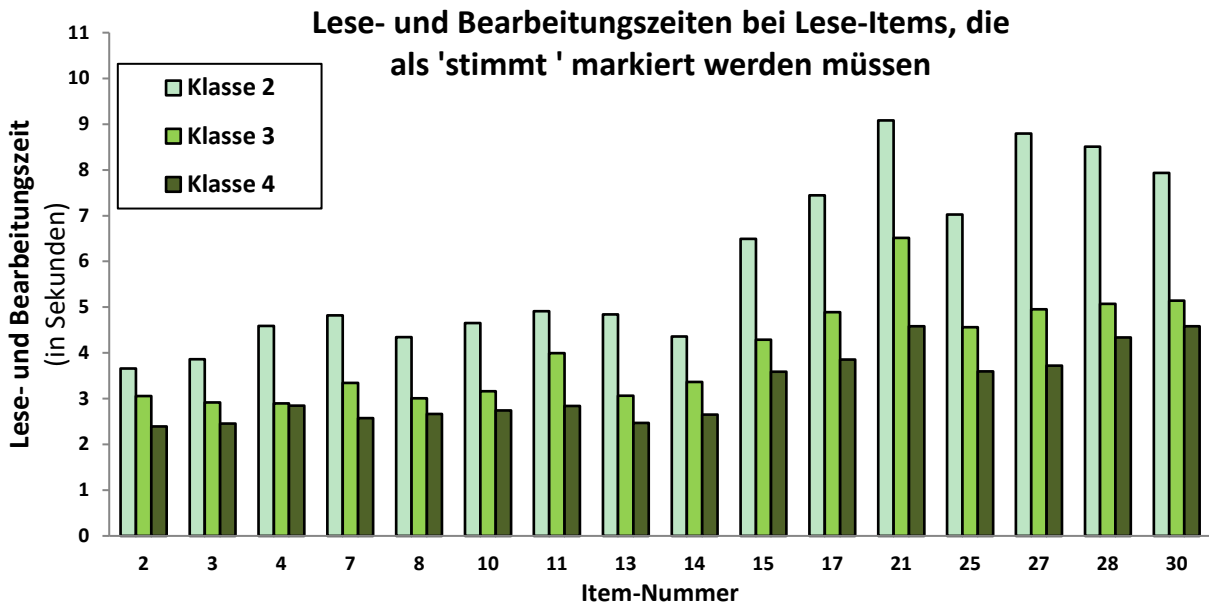


Abb: 9: Durchschnittliche Lese- und Bearbeitungszeiten bei Items, die als 'stimmt' markiert werden mussten nach Klassenstufe (2-4).

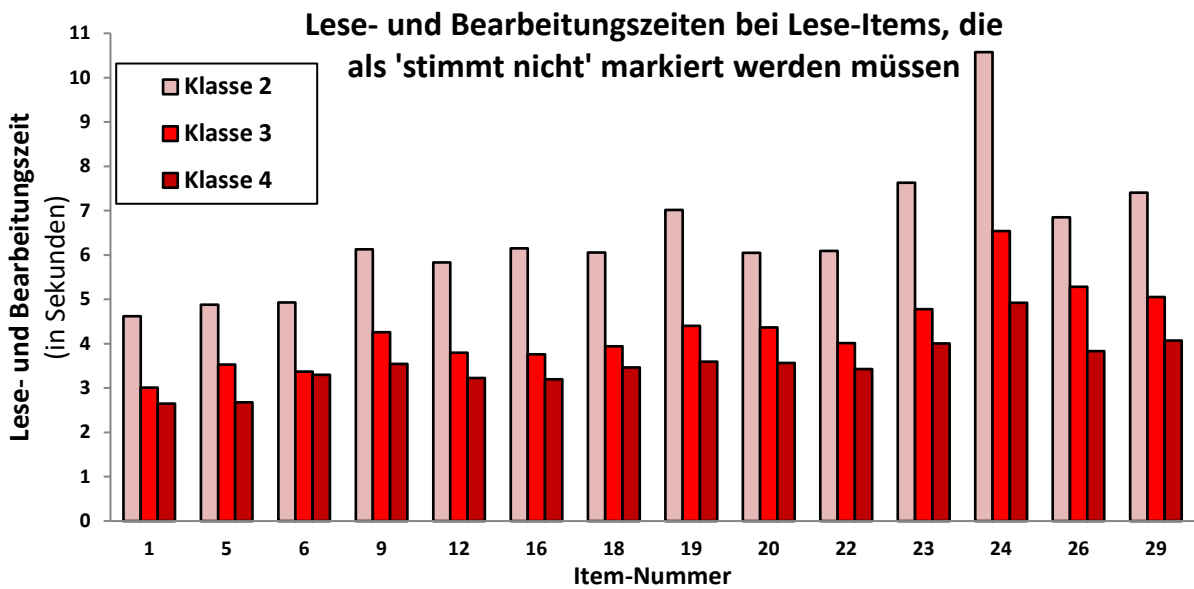


Abb: 10: Durchschnittliche Lese- und Bearbeitungszeiten bei Items, die als 'stimmt nicht' markiert werden mussten nach Klassenstufe (2-4).

Im Durchschnitt entstehen folgende Differenzwerte zwischen den getesteten Klassenstufen:

Zustimmung: Klasse 4 (+0,8) Klasse 3 (+1,9) Klasse 2 (+5,7) Klasse 1
 3,2 Sek →→→→ 4,0 Sek →→→→ 5,9 Sek →→→→ 11,6 Sek

Ablehnung: Klasse 4 (+0,7) Klasse 3 (+2,2) Klasse 2 (+5,4) Klasse 1
 3,5 Sek →→→→ 4,2 Sek →→→→ 6,4 Sek →→→→ 11,8 Sek

Es gelingt also durchaus, mit dem ausgewählten Item-Set des SLS zwischen den Klassenstufen zu differenzieren. Doch zeigt auch hier die grafische Darstellung, dass die Item-bezogenen Lese- und Bearbeitungszeiten innerhalb einer Klassenstufe durchaus stark variieren können: In der zweiten Klassenstufe schwanken die erfassten Werte zwischen 3,6 und 10,6 Sekunden (3657 ms und 10577 ms). Diese (große) Spanne der Werte von 6,9 Sekunden innerhalb der Klassenstufe ist damit um ein vielfaches größer als die mittlere Differenz zwischen den Werten von Klasse 2 zu Klasse 4 (2,7 Sek) und entspricht eher der mittleren Differenz der Werte zwischen den Klasse 1 und 4 (8,3 Sek). Auch in Klasse 3 schwanken die Messwerte noch zwischen 2,8 und 6,5 Sekunden und liegen über der mittleren Differenz zu Klasse 4. Selbst in Klasse 4 finden sich noch Item-bezogene Streuungen zwischen 2,3 und 4,9 Sekunden, also etwa so groß wie die mittlere Differenz zur Klasse 3.

Ins Verhältnis gesetzt, nimmt die Differenz der Item-bezogenen Bearbeitungszeiten von Klassenstufe zu Klassenstufe sogar eher zu als ab: In Klasse 4 weichen drei Items (Item 21, 24 und 30) um mehr als 1,5 Standardabweichungen vom Durchschnitt der Klassenstufenwerte ab, fünf weitere (Item 1, 2, 3, 5 und 28) um mehr als eine Standardabweichung. In Klasse 3 sind es 'nur' noch zwei Items (Item 21 und 24), die mehr als 1,5 Standardabweichungen, und vier (Item 3, 4, 26 und 30), die um mehr als eine Standardabweichung differieren, in Klasse 1 sogar zwei (Item 21 und 24) mit 1,5-facher und 'nur' noch drei Items (Item 2, 27 und 30) mit einfacher Abweichung. Überspitzt ausgedrückt, liefert das betrachtete Item-Set des SLS 2–9 somit in Klassenstufe 1 noch die verhältnismäßig konstantesten Werte.

Die Tabelle 3 (auf der Folgeseite) liefert eine Übersicht der Items, die in der empirischen Erhebung auffällig in jeglicher Form geworden sind.

Den empirischen Teil zusammenfassend werden 11 der 30 betrachteten Lese-Items durch die erfassten Lese- und Bearbeitungszeiten der Schüler*innen auffällig. Hierbei können Items unterschieden werden, die grundsätzlich in allen Klassenstufen auffällig werden, und solche, die lediglich 'lokal' in einer Klassenstufe hervortreten. Bemerkenswert ist hierbei, dass obgleich die Klassenstufe 1 nicht im Fokus der Konzeption des SLS 2–9 ist, kein einziges Item ausschließlich in dieser Klassenstufe auffällig wird. Lese-Items, die in Klasse 1 eine stärkere Abweichung der Bearbeitungszeit evozieren, tun dies mindestens auch (und relational stärker) in einer weiteren Klassenstufe. Items, die nur klassenstufenspezifisch (und nicht einmal hinsichtlich des Gesamtkorpus) größere Abweichungen hervorrufen, kommen ausschließlich in Klasse 3 und 4 vor, nicht jedoch in Klasse 1 und 2.

| Item | Gesamt | Klasse 1 | Klasse 2 | Klasse 3 | Klasse 4 |
|--|----------------|----------------|----------------|----------------|----------------|
| Bäume können sprechen. | unauffällig | unauffällig | unauffällig | unauffällig | -1 Std.-Abw. |
| Wasser ist nass. | -1 Std.-Abw. | -1 Std.-Abw. | -1 Std.-Abw. | unauffällig | -1 Std.-Abw. |
| Der Löwe brüllt. | -1 Std.-Abw. | unauffällig | -1 Std.-Abw. | -1 Std.-Abw. | -1 Std.-Abw. |
| Flugzeuge können fliegen. | unauffällig | unauffällig | unauffällig | -1 Std.-Abw. | unauffällig |
| Schweine haben Flügel. | -1 Std.-Abw. | unauffällig | unauffällig | unauffällig | -1 Std.-Abw. |
| Die Armbanduhr zeigt die Zeit an. | +1,5 Std.-Abw. | +1,5 Std.-Abw. | +1,5 Std.-Abw. | +1,5 Std.-Abw. | +1,5 Std.-Abw. |
| Die Schale der Orange ist violett. | +1,5 Std.-Abw. | +1,5 Std.-Abw. | +1,5 Std.-Abw. | +1,5 Std.-Abw. | +1,5 Std.-Abw. |
| Ein Picknick macht man im Haus. | unauffällig | unauffällig | unauffällig | +1 Std.-Abw. | unauffällig |
| In einem Krankenhaus arbeiten viele Ärzte. | +1 Std.-Abw. | +1 Std.-Abw. | +1 Std.-Abw. | unauffällig | unauffällig |
| Am Nordpol gibt es viel Eis und Schnee. | +1 Std.-Abw. | unauffällig | +1 Std.-Abw. | unauffällig | +1 Std.-Abw. |
| Das Besteck braucht man zum Essen. | +1 Std.-Abw. | +1 Std.-Abw. | +1 Std.-Abw. | +1 Std.-Abw. | +1,5 Std.-Abw. |

Tab. 3: Auflistung der im Rahmen der empirischen Analyse auffällig gewordenen Lese-Items

8 | Zusammenführung der theoretischen und der empirischen Analysen

Die bisherigen Analysen haben aufgedeckt, dass das Ausmaß des Ungleichgewichtes hinsichtlich der linguistischen Zugriffsgrößen/Faktoren in den Lese-Items des SLS 2–9 einigermaßen umfangreich ist. Das Ausmaß der Ungleichheit der empirisch erfassten Lese- und Bearbeitungszeiten zu individuellen Items ist jedoch noch größer. Die Anschlussfrage wird damit, ob ein Zusammenhang zwischen diesen Ungleichgewichten besteht und das zweite (auch) eine Konsequenz des ersten sein kann.

Gegen einen (vollständigen) kausalen Zusammenhang der Ungleichgewichte spricht, dass eine Reihe von Items in der empirischen Datenerhebung auffällig wurde, die hinsichtlich der Quantitäten der linguistischen Zugriffsgrößen zuvor kaum bemerkenswert schienen. Hierzu gehören das Item 21 („Die Armbanduhr zeigt die Zeit an“), das Item 5 („Schweine haben Flügel“) sowie das Item 26 („Ein Picknick macht man im Haus“). Das Item 9 („In der Nacht scheint die Sonne“) auf der anderen Seite, welches anfangs (sowohl) hinsichtlich der Konstituentenstruktur (als auch hinsichtlich der Semantik) als potentiell herausfordernd eingestuft wurde, zeigt im Datensatz (überraschenderweise) keinerlei merkliche Abweichungen. Die statistische Analyse der Zusammenhänge zwischen den linguistischen Merkmalen und den erfassten Bearbeitungszeiten der SLS-Items 1–30 erfolgt zunächst mit Hilfe der Statistiksoftware SPSS. In der Analyse der Korrelationen nach *Spearman* zeigen alle kodierten linguistischen Merkmale eine signifikante ($p < 0.01$) Korrelation mit den erfassten Bearbeitungszeit-Messwerten. Dies legt nahe, dass ein Zusammenhang zwischen Lesezeiten und linguistischen Merkmalen besteht.

Mit Hilfe einer multiplen Regressionsanalyse (in der Software R-Studio) kann ermittelt werden, welches Modell (und damit welche potentiellen Einflussfaktoren) am geeignetesten ist, Unterschiede in erfassten Lese- und Bearbeitungszeiten zu erklären. Als potentielle Faktoren zur Vorhersage der Lesezeiten kommen neben den kodierten linguistischen Merkmalen der Items zu-

sätzlich die (im Anschluss an jeweiligen Testungen) im Protokollbogen erfassten personenbezogenen Merkmale der Schüler*innen in Frage. Hier wurden die Schüler*innen gebeten, Aussagen über ihre Mediennutzung zu tätigen. Testdurchführende fragten die Schüler*innen jeweils, welche technischen Geräte zu Hause *existieren* (*Wir haben...*), welche die Schüler*innen selbst *nutzen* dürfen (*Ich darf...*) und welche sie selbst *besitzen* (*Nur mir gehört...*). Gezielt gefragt wurde hierbei nach den folgenden fünf Kategorien, die durch Protokollierende mit (+) oder (-) markiert wurden:

- PC/Laptop
- Tablet/Ipad
- Smartphone
- Spielkonsole¹²
- portable Spielkonsole¹³

Darüber hinaus wurde anhand einer vierstufigen Skala erfragt, ob die Schüler*innen (*sehr gern, gern, ungern, sehr ungern*) und (*sehr viel, viel, wenig, sehr wenig*) lesen.

Die Tabelle (4) listet die potentiellen Einflussfaktoren auf die erfassten Lese- und Bearbeitungszeiten auf, die für die Regressionsanalyse herangezogen wurden:

| Parameter/Faktor | Erhebungsart | Messniveau |
|--|-----------------------|---------------------------------------|
| Nummer des Lese-Items | SLS Bogen A1 | (metrisch) 1-30 |
| Quantität der Buchstaben | linguistische Analyse | (metrisch) min. 13 max. 36 |
| Quantität der graph. Einheiten | linguistische Analyse | (metrisch) min. 11 max. 27 |
| Quantität der graph. Silben | linguistische Analyse | (metrisch) min. 4 max. 13 |
| Quantität der graph. Worte | linguistische Analyse | (metrisch) min. 3 max. 8 |
| Durchschnittliche Länge der Konstituenten | linguistische Analyse | (metrisch) min. 1,0 max. 2,5 |
| Schule (eine der fünf Grundschulen) | Protokollbogen | (nominal) 1,2,3,4,5 |
| Klassenstufe | Protokollbogen | (metrisch) 1,2,3,4 |
| Händigkeit (beim Schreiben) | Protokollbogen | (nominal) <i>rechts, links, beide</i> |
| Geschlecht | Protokollbogen | (nominal) <i>Mädchen, Junge</i> |
| Quantität der Medien zu Hause | Protokollbogen | (metrisch) min. 0 max. 5 (s.oben) |
| Quantität der Medien, die genutzt werden | Protokollbogen | (metrisch) min. 0 max. 5 (s.oben) |
| Quantität der Medien im (nur) eigenen Besitz | Protokollbogen | (metrisch) min. 0 max. 5 (s.oben) |
| Eigene Lesezeit in der Freizeit | Protokollbogen | (metrisch) min. 1 max. 4 (s.oben) |
| Eigene Lesemotivation | Protokollbogen | (metrisch) min. 1 max. 4 (s.oben) |

Tab. 4: Potentielle Einflussfaktoren für die Bildung eines linearen Regressionsmodells

In der linearen Regressionsanalyse gilt ein Modell als *ideal*, wenn es möglichst wenige (unabhängige) Variablen/Faktoren enthält, doch zugleich eine möglichst hohe Vorhersagekraft (*adjustiertes R²*) hinsichtlich der abhängigen Variable (hier: die gemessenen Lesezeit-Werte) zeigt. Weniger treffende Modelle enthalten entweder zu viele Variablen oder zeigen einen insgesamt

¹² z.B.: Gibt es bei euch zu Hause/besitzt du selbst eine Playstation, eine Xbox, eine Switch, eine Wii oder etwas Ähnliches? Darfst du selbst damit spielen?

¹³ z.B.: Gibt es bei euch zu Hause einen DS3, DS lite, DS Portable, eine Playstation Portable oder etwas Ähnliches? Darfst du selbst...

(zu) niedrigen Einfluss auf die abhängige Variable. Im Fall der vorliegenden Daten hat sich folgendes Modell mit insgesamt zehn Variablen/ Einflussfaktoren als *ideal* erwiesen:

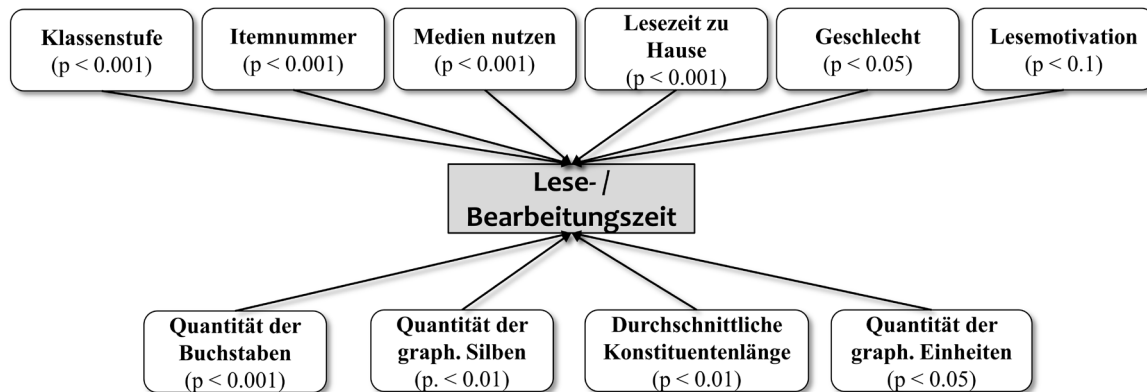


Abb. 11: Lineares Regressionsmodell zu Unterschieden in den erfassten Lesezeit-Werten

Die Analyse liefert damit folgende Ergebnisse:

- Mit zehn relevanten Faktoren liegt eine verhältnismäßig große Anzahl vor. Die Lese- und Bearbeitungszeiten der ersten 30 Items des SLS 2–9 werden demnach von vielen (unterschiedlichen) Faktoren beeinflusst (und nicht etwa von einigen wenigen).
- Auf Seiten der linguistischen Merkmale ist einzig die *Anzahl der graphematischen Worte* kein notwendiger (relevanter) Einflussfaktor in der Modellbildung (obgleich das SLS 2–9 zuerst als Wort- (und Satz-)Lese-Screening verstanden werden soll (vgl. Wimmer & Mayringer 2014: 16)).
- Die *Klassenstufe*, die *Item-Nummer* (Position des Items in der Originalvorlage) und die *Quantität der Buchstaben* sind die Faktoren mit dem größten Einfluss auf die Lesezeit.

Mit notwendiger Vorsicht lässt sich durch die Fokussierung der genannten drei größten Einflussfaktoren eine Art *Vorhersagemodell* ableiten (Abbildung 12), welches einschätzen lässt, wie stark die Lese- und Bearbeitungszeiten von den Hauptfaktoren abhängen bzw. durch diese verfälscht werden.

| | adj. R ² | Faktor | B | Std. Er. | t-Wert | p |
|--------|---------------------|------------------|----------|----------|---------|--------|
| Modell | .25 | Klassenstufe | -2232,32 | 80,47 | -27,749 | < .001 |
| | | Buchstabenanzahl | 107,94 | 18,68 | 5,777 | < .001 |
| | | Item-Nummer | 81,25 | 14,16 | 5,737 | < .001 |

Residual Standard Error: 4349 on 338 degrees of freedom

Abb. 12: Vorhersagemodell zu Einflussfaktoren auf die Lese- und Bearbeitungszeiten

Zu betonen ist zuerst, dass dieses Modell – auch in Konsequenz der Reduktion auf lediglich drei Faktoren – einen vergleichsweise hohen Freiheitsgrad (resp. *degree of freedom*) aufweist. Dies meint, dass weitere (größere) Abweichung in den Lese- und Bearbeitungszeiten auch durch Faktoren außerhalb des Modells ausgelöst werden können. Gemäß der Analyse zuvor (Abbildung 11) kommen hier v.a. die Mediennutzung und die eigene Lesezeit (zu Hause) als Einflüsse in Frage.

Davon abgesehen lassen sich aus dem Vorhersagemodell drei Abhängigkeitsverhältnisse entnehmen, die auf Grund des explorativen und nicht-exhaustiven Charakters des Modells in erster Linie als Daumenformeln verstanden werden dürfen.

Es gilt: Verbleiben alle weiteren Faktoren konstant...

- verkürzt sich die Lese- und Bearbeitungszeit jedes Lese-Items um etwa 2,2 Sekunden (-2232,99 gemäß Modell) von Klassenstufe zu Klassenstufe. Drittklässler*innen lesen je Item 2,2 Sekunden 'kürzer' als Zweitklässler*innen. Viertklässler*innen lesen jeweils 4,4 Sekunden 'schneller' als Zweitklässler*innen.
- verzögert sich die Lese- und Bearbeitungszeit in jedem Lese-Item um 0,1 Sekunden (107,94 gemäß Modell) mit jedem zusätzlichen Buchstaben¹⁴ im Item-Text.
- verzögert sich die Lese- und Bearbeitungszeit um 0,08 Sekunden (81,25 gemäß Modell) mit zunehmender Item-Nummer im Salzburger Lese-Screening 2–9. Ein Lese-Item B, welches 10 Items 'später' als ein Item A im Test-Bogen abgedruckt ist, bedarf in der Theorie 0,8 Sekunden mehr Bearbeitungszeit als Item A. Über den Gesamtfragebogen aus 100 Lese-Items entsteht ein Zeit-Zuwachs von acht Sekunden vom ersten zum letzten Lese-Item.

Noch einmal betont werden soll der theoretische Charakter dieser Daumenformeln. Allein die oben präsentierten Ergebnisse verdeutlichen, dass eine tatsächliche Vorhersage und umgekehrt Erläuterung von Lese- und Bearbeitungszeiten in den ersten 30 Items des SLS nur erschwert möglich ist.

9 | Zusammenfassung und Diskussion

Ausgangspunkt dieses Beitrages war die im deutschdidaktischen Diskurs mehrfach getätigte Beobachtung und Vermutung, dass sich Lese-Items des *Salzburger Lese-Screenings 2–9* (Wimmer & Mayringer 2014) hinsichtlich ihrer Anforderungen voneinander unterscheiden und auf diese Weise ein Ungleichgewicht zwischen den Items entsteht (vgl. Gailberger & Nix 2013; Esslinger 2015; Zepnik 2016 uvm.). Vor dem Hintergrund der hohen schulpraktischen wie auch deutschdidaktisch-empirischen Relevanz des Erhebungsinstrumentes (z.B. in Belgrad et al. 2010; Lauer-Schmaltz, Rosebrock & Gold 2014; uvm.) sollte linguistisch-theoretisch und empi-

¹⁴ Bereits an dieser Stelle wird deutlich, dass es sich hierbei um ausschließlich theoretische Überlegungen handeln kann: In der Praxis lässt das Deutsche kaum zu, dass zusätzliche Buchstaben ergänzt werden, ohne dass hierdurch eine Veränderung irgendeiner Art angezeigt wird (andernfalls wäre der Buchstabe redundant). Wird beispielsweise 'nur' ein <e> am Ende des Wortes *Hund* angefügt (*Hunde*), so kommt nicht nur ein weiterer Buchstabe hinzu, sondern es ändert sich auch die Flexionsform (zu Plural), womit ein weiterer Einflussfaktor variiert wäre.

risch untersucht werden, welches Ausmaß diese Unterschiede und das Ungleichgewicht tatsächlich annimmt. Dass ein Ungleichgewicht festgestellt werden wird, galt bereits im Vorfeld als gesetzt, überraschend ist dennoch das festgestellte Ausmaß.

Die betrachteten ersten 30 Lese-Items der Fragebogenform A1 des SLS 2–9 sind nicht nur jene, die alle Schüler*innen aller Klassenstufen bei der Bearbeitung (sicher) durchlaufen, es sind auch die, die optisch (z.B. durch Satzlängen, Wortlängen und Anzahl der Textzeilen) zueinander maximal vergleichbar scheinen¹⁵. Umso bemerkenswerter ist, wie strukturell und (empirisch) ergebnisverschieden die noch vergleichbareren Items des SLS-Sets ausfallen: Sieben der 30 (ersten) Items zeigen bereits in der theoretischen linguistischen Analyse signifikante Abweichungen hinsichtlich der Quantitäten der linguistischen Zugriffsgrößen, sodass bereits auf dieser Ebene von stark verschiedenen Leseanforderungen ausgegangen werden muss.

Die *Tablet*-gestützt erfassten Lese- und Bearbeitungszeiten von insgesamt 208 Schüler*innen lassen empirisch 11 der 30 Items hervortreten, die durch starke (signifikante) Abweichungen von den durchschnittlichen Lese- und Bearbeitungszeiten gekennzeichnet sind.

Werden linguistische und empirischen Daten zueinander ins Verhältnis gesetzt, zeigt sich zwar, dass v.a. die Quantität der ‘Buchstaben’ (z.B. Neef 2005) und auch der graphematischen Silben (Fuhrhop 2013) die Lese- und Bearbeitungszeiten der Schüler*innen beeinflussen, doch gibt es noch eine große Zahl (mindestens acht) weiterer und vergleichbar starker Faktoren, die diese linguistischen Faktoren in (zahlreichen) Einzelfällen zu überschreiben scheinen. So ist zwar das empirisch (zeit-)schnellste Item 2 („Wasser ist nass“) auch jenes, das hinsichtlich der segmentalen, segmentübergreifenden und wortübergreifenden linguistischen Analyse als anforderungsärmstes bestimmt wurde, die (mit) empirisch längeren Items 21 („Die Armbanduhr zeigt die Zeit an“) und 30 („Das Besteck braucht man zum Essen“) hingegen scheinen auf allen linguistischen Analyseebenen unauffällig. Zuletzt generiert das in der linguistischen Analyse auffällige Item 9 („In der Nacht scheint die Sonne“) empirisch Zeitwerte, die vollends im Durchschnitt liegen. Die Vermutung liegt nahe, dass hier v.a. semantische Aspekte des Vor- und Weltwissens die Zeitwerte beeinflussen (etwa: Schüler*innen *wissen* im digitalen Zeitalter nicht mehr, was eine „Armbanduhr“ ist...) ¹⁶.

Die empirisch erfassten Lese- und Bearbeitungszeiten unterscheiden sich zwischen den Klassenstufen 1–4 signifikant und zeigen darüber hinaus eine konstante Staffelung nach Klassenstufe bei jedem einzelnen Item. Auf einen ersten Blick scheint das SLS 2–9 damit geeignet, generelle Unterschiede zwischen Klassenstufen zu erfassen. Gegen eine solche Auffassung spricht jedoch, dass individuelle und Item-bezogene Unterschiede in erfassten Messwerten i.d.R. um ein Vielfaches größer ausfallen als die durchschnittliche Differenz der Bearbeitungszeiten zwischen Klassenstufen: Ob eine Schüler*in ‘liest wie in Klasse 4 oder 2’ wird damit z.T. stärker vom individuellen Item abhängig als von der tatsächlichen Klassenstufe.

¹⁵ Dass ein Vergleich von Lese-Items wie 68 („Wenn viel Schnee liegt, kann man einen Schneemann bauen“) zu 69 („In einer Tintenpatrone, die man in einen Füllfeder gibt, befindet sich Heidelbeersaft“) zahlreiche theoretische wie empirische (Lesezeit-)Unterschiede zu Tage fördern würde, bedarf keiner Forschungsarbeit.

¹⁶ Aus platztechnischen Gründen konnte die ‘Richtigkeit’ der Antworten der Schüler*innen in diesem Beitrag nicht betrachtet werden. Zusammenfassend ist hier (nicht überraschend) das Item 24 („Die Schale der Orange ist violett“) das Item, das in der Studie am häufigsten ‘falsch’ bearbeitet wurde. Den ‘zweiten Platz’ teilen sich die Items 21 („Die Armbanduhr...“) und Item 12 („Honig schmeckt sauer“), was die einleitend formulierten Bedenken hinsichtlich der semantischen Anforderungen insbesondere des zweiten Items bestätigt.

Auffallend ist darüber hinaus, dass mit zunehmender Klassenstufe mehr Lese-Items das Potential bekommen, Lese- und Bearbeitungszeiten im Verhältnis zum Jahrgangsstufendurchschnitt stärker (verlangsamend) zu beeinflussen. Dies lässt zwei Interpretationen zu: Es deutet sich an, dass:

- entweder fortschreitender Lese-Erwerb von Klassenstufe zu Klassenstufe Unterschiede in den Lesefähigkeiten und -Fertigkeiten zwischen individuellen Schüler*innen verhältnismäßig eher verschärft als nivelliert
- oder dass fortschreitender Erwerb der Lesefähigkeiten und -Fertigkeiten die Lesezeit-Messungen für weitere ('Stör-')Faktoren eher anfälliger als unanfälliger macht.

In jedem Fall jedoch wird die Auswahl geeigneter Test-Items mit steigender Klassenstufe relevanter und sensibler, was das offensichtliche Ungleichgewicht im Item-Set des SLS 2–9 zu einem Problem macht, das mit zunehmender Klassenstufe (noch) an Tragkraft gewinnt¹⁷.

Am Ende bleiben als Eindrücke m.E. zwei Aspekte zurück: Zum Ersten das tatsächliche Itembezogene Ungleichgewicht der Lesezeiten und zum Zweiten das Eingeständnis, dass die teilweise erheblichen Unterschiede bei einigen Lese-Items linguistisch/lesedidaktisch schwer vorauszusehen/zu erklären sind.

Jetzt kann diesem Beitrag vorgeworfen werden, das *Salzburger Lese-Screening 2–9* einer Qualitätskontrolle zu unterziehen, welche es selbst nie zu bestehen behauptet hat. Eine kritische Prüfung des Manuals, z.B. in der bereits dargebotenen Aussage, Messwerte seien nicht durch Faktoren der Satzverarbeitung oder der inhaltlichen Beurteilung beeinflusst (Wimmer & Mayringer 2014: 16), zeigen zum Ersten, dass dieser Vorwurf nur z.T. berechtigt wäre: Wimmer & Mayringer (2014) vermitteln (schon) den Eindruck, selbst einen solchen Anspruch zu erheben. Zum Zweiten erfordern die schulpraktischen und deutschdidaktisch-empirischen Settings, in denen das SLS (z.B. 2–9) gegenwärtig tatsächlich eingesetzt wird, eigentlich ein Erhebungsinstrument, das genau einer solchen Qualitätsprüfung standhält. Im Zweifelsfall behaupteten die Autoren des SLS nie, ein solches Instrument zu liefern. Ob man es *nichtsdestotrotz* als ein solches Instrument auffasst und nutzt – und darauf möchte dieser Beitrag hinweisen – sollte mehrfach überdacht werden.

¹⁷ Eine interessante Anschlussfrage wird damit, wie sich Lese- und Bearbeitungszeiten bei Schüler*innen der Klassenstufen 5–9 entwickeln.

10 | Literatur

- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz- Testkonzeption und Ergebnisse. In: J. Baumert & E. Klieme (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69–140). Leske und Burdich.
- Auer, M., Gruber, G., Mayringer, H. & Wimmer, H. (2005). *SLS 5–8. Salzburger Lese-Screening für die Schulstufen 5–8*. Hogrefe Verlag GmbH & Co. KG.
- Bangel, M. & Müller, A. (2014). Zur Entwicklung morphologischer Bewusstheit und basaler Lesefähigkeit durch die Arbeit an Wort(bildungs)strukturen. Erste Ergebnisse einer Interventionsstudie. *Didaktik Deutsch*, 36, 42–63.
- Bangel, M. & Müller, A. (2015). Einsichten in morphologische Strukturen als Grundlage für die Entwicklung der Dekodierfähigkeit. In: I. Rautenberg & T. Reißig (Hrsg.), *Lesen und Lesedidaktik aus linguistischer Perspektive* (S. 17–53). Peter Lang GmbH.
- Belgrad, J., Knapp, W. & Lindel, M. (2010). *Leseförderung durch Vorlesen – ein Forschungsprojekt an der Pädagogischen Hochschule Weingarten*. www.lesefoerderung-durch-vorlesen.de/ [16.09.2021]
- Bentin, S., Deutsch, A. & Liberman, I. (1990). Syntactic Competence and Reading Ability in Children. *Journal of Experimental Child Psychology*, 49(1), 147–172.
- Berg, K. (2013). *Morphosyntax nominaler Einheiten im Niederdeutschen*. Winter.
- Berg, K., Primus, B. & Wagner, L. (2016). Buchstabenmerkmal, Buchstabe, Graphem. In: U. Domahs, & B. Primus (Hg.), *Handbuch Laut, Gebärde, Buchstabe* (S. 337–355). Walter de Gruyter GmbH & Co. KG.
- Bloomfield, L. (1933). *Language*. George Allen & Unwin Ltd.
- Bredel, U., Fuhrhop, N. & Noack, C. (2017). *Wie Kinder lesen und schreiben lernen*. 2. Aufl. Narr Francke Attempto.
- Bredel, U., Noack, C. & Plag, I. (2013). Morphologie lesen. Stammkonstanzschreibung und Leseverstehen bei starken und schwachen Lesern. In: M. Neef & C. Scherer (Hrsg.), *Die Schnittstelle von Morphologie und geschriebener Sprache*. (S. 211–249). Walter de Gruyter GmbH & Co. KG.
- Bredel, U. & Maaß, C. (2016). *Duden Leichte Sprache. Theoretische Grundlagen. Orientierung für die Praxis*. Duden.
- Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words. Impact on reading. *Reading and Writing: An Interdisciplinary Journal*, 12, 169–190.
- Carlisle, J. F. & Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly*, 40(4), 428–449.
- Carlisle, J. F. & Katz, L. A. (2006). Effects of word and morpheme familiarity on reading derived words. *Reading and Writing*, 19, 669–693.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Science*. Lawrence Erlbaum Associates Publishers.
- Eisenberg, P. (2012). *Das Fremdwort im Deutschen*. Walter de Gruyter GmbH & Co. KG.
- Eisenberg, P. (2013a). *Das Wort. Grundriss der deutschen Grammatik*. Verlag J.B. Metzler.
- Eisenberg, P. (2013b). *Der Satz. Grundriss der deutschen Grammatik*. Verlag J.B. Metzler.

- Esslinger, G. (2015). Syntaktisches Lesen unter besonderer Berücksichtigung der Interpunktion – Theorie, Testkonzeption und empirische Befunde. In: I. Rautenberg & T. Reißig, (Hrsg.), *Lesen und Lesedidaktik aus linguistischer Perspektive* (S. 117–152). Peter Lang GmbH.
- Fix, M. (2015). Morphologie als Lesehilfe. Eine empirische Untersuchung mit Grundschulern. In: I. Rautenberg & T. Reißig (Hrsg.), *Lesen und Lesedidaktik aus linguistischer Perspektive* (S.55–90). Peter Lang GmbH.
- Fuhrhop, N. (2013). Graphematik. In: N. Fuhrhop & J. Peters (Hrsg.), *Einführung in die Phonologie und Graphematik* (S.180–294). Verlag J.B. Metzler.
- Fuhrhop, N. & Buchmann, F. (2016). Graphematische Silbe. In: U. Domahs & B. Primus (Hrsg.), *Handbuch Laut, Gebärde, Buchstabe* (S. 357–376). Walter de Gruyter GmbH & Co. KG.
- Gailberger, S. & Nix, D. (2013). Lesen und Leseförderung in der Primar- und Sekundarstufe. In: S. Gailberger & F. Wietzke (Hrsg.), *Handbuch Kompetenzorientierter Deutschunterricht* (S. 32–69). Beltz.
- Gallmann, P. (1985). *Graphematische Elemente geschriebener Sprache*. Max Niemeyer Verlag.
- Gallmann, P. (2015). Das topologische Modell. Basisartikel. In: A. Wöllstein (Hrsg.), *Das topologische Modell für die Schule* (S. 1–36). Schneider Verlag Hohengehren.
- Gärtner, K. (2010). Normierungsstudie zum Salzburger Lese-Screening. *Statistische Nachrichten*, 5, 383–389.
- Günther, H. (1988). *Schriftliche Sprache. Strukturen geschriebener Wörter und ihre Verarbeitung beim Lesen*. Max Niemeyer Verlag.
- Hawelka, S., Gagl, B. & Wimmer, H. (2010). A dual-route perspective on eye movement of dyslexic readers. *Cognition*, 115(3), 367–379.
- Huggins, A. W. F. (1977). *Syntactic Aspects of Reading Comprehension*. Bolt Beranek and Newman Inc.
- Kohrt, M. (1985). *Problemgeschichte des Graphembegriffs und des frühen Phonembegriffs*. Walter de Gruyter.
- Krueger, L. (1973). Effects of letter-pair frequency and orientation on speed of ‘same’-‘different’ judgments by children and adults. *Bulletin of the Psychonomic Society*, 2, 431–433.
- Lauer-Schmaltz, M., Rosebrock, C. & Gold, A. (2014). Lautlesetandems in der Grundschule. Bedingungen und Grenzen ihrer Wirksamkeit. *Didaktik Deutsch*, 37, 45–61.
- Lenhard, W., Lenhard, A. & Schneider, W. (2018). *ELFE II. Ein Leseverständnistest für Erst- bis Siebtklässler – Version II*. Hogrefe Verlag GmbH & Co. KG.
- Levshina, N. (2015). *How to do Linguistics with R. Data exploration and statistic analysis*. John Benjamins Publishing Company.
- Martohardjono, G., Otheguy, R., Gabriele, A., de Goeas-Malone, M., Szupica-Pyranowaski, M., Troseth, E., Rivero, S. & Schutzman, Z. (2005). The Role of Syntax in Reading Comprehension. A Study of Bilingual Readers. In: J. Cohen, K. T. McAlister, K. Rolstad & J. MacSwan (Hrsg.), *Proceedings of the 4th International Symposium on Bilingualism* (S. 1522–1544). Cascadilla Press.
- Metze, W. (2009). *Stolperwörter Lesetest*. www.wilfriedmetze.de/Handweisungen_2009.pdf [16.09.2021].
- Neef, M. (2005). *Die Graphematik des Deutschen*. Max Niemeyer Verlag.
- Primus, B. (2004). A featural analysis of the Modern Roman Alphabet. *Written Language & Literacy*, 7, 235–274.

- Primus, B. (2006). Buchstabenkomponenten und ihre Grammatik. In: U. Bredel & H. Günther (Hrsg.), *Orthographietheorie und Rechtschreibunterricht* (S. 1–34). Max Niemeyer Verlag.
- Primus, B. (2010). Strukturelle Grundlagen des deutschen Schriftsystems. In: U. Bredel, A. Müller & G. Hinney (Hrsg.), *Schriftsystem und Schriffterwerb: linguistisch – didaktisch – empirisch*. Max Niemeyer Verlag.
- Rautenberg, I. & Reißig, T. (2015). *Lesen und Lesedidaktik aus linguistischer Perspektive*. Peter Lang GmbH.
- Rezec, O. (2009). *Zur Struktur des deutschen Schriftsystems*. [Dissertation].
- Rosebrock, C. & Nix, D. (2014). Ein Modell zur Lesekompetenz aus didaktischer Perspektive. In: C. Rosebrock & D. Nix (Hrsg.): *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung* (S. 13–32). Schneider Verlag Hohengehren.
- Röttig, S., Schwerkolt, C. & Nottbusch, G. (2020). Die Entwicklung der Leseflüssigkeit in der Grundschule. (in diesem Band).
- Sappok, C. (2020). Exploring Advanced Prosody – eine Best-Practice-Untersuchung in der weiterführenden Schule. (in diesem Band).
- Trost, P. (1939). Bemerkungen zum deutschen Vokalsystem. *Travaux du Cercle Linguistique de Prague*, 8, 319–327.
- Wimmer, H. & Mayringer, H. (2003). *SLS 1–4. Salzburger Lese-Screening für die Schulstufen 1–4*. Hogrefe Verlag GmbH & Co. KG.
- Wimmer, H. & Mayringer, H. (2014). *SLS 2–9. Salzburger Lese-Screening für die Schulstufen 2–9*. Hogrefe Verlag GmbH & Co. KG.
- Wöllstein, A. (2014). *Topologisches Satzmodell*. 2. Aufl. Universitätsverlag Winter.
- Zepnik, S. (2016). *Linguistische Leseprofile von Grundschülerinnen und Grundschulern zur Ermittlung basaler Lesestrategien*. [Dissertation].
- Zepnik, S. & Zepter, A. (2015). Die Syntax für das Lesen nutzbar machen. In: I. Rautenberg & T. Reißig (Hrsg.), *Lesen und Lesedidaktik aus linguistischer Perspektive* (S. 91–116). Peter Lang GmbH.

Anke Schmitz, Fabiana Karstens & Jörg Jost

Vermittlung von Lesestrategien im Umgang mit Sachtexten im Deutschunterricht

Vergleich von Unterrichtsbeobachtung und Lehrerbefragung

1 | Einleitung

Lesestrategien, um deren Vermittlung im Deutschunterricht es in diesem Beitrag geht, sind kognitive Werkzeuge (Rosebrock & Nix 2017), die es Leser*innen ermöglichen, selbstregulativ bei komplexen und schwierigen Texten mentale Kohärenz herzustellen. Da selbstreguliertes Lesen eine vertiefte Auseinandersetzung mit Texten befördert (Kleinbub 2016; Lankes & Carstensen 2007), kann dieses in besonderem Maße zu kognitiver Aktivierung beitragen – einem Qualitätsmerkmal guten Unterrichts, das als prädiktiv für den Kompetenzerwerb gilt (Klieme & Rakoczy 2008). Im Unterricht sollten Lehrkräfte daher vielfältige Lerngelegenheiten anbieten, damit ihre Schüler*innen Techniken und Strategien für den selbstregulierten Umgang mit Texten kennenlernen.

Die Vermittlung von Lesestrategien ist explizit als Kompetenzziel des Deutschunterrichts verankert (KMK 2003). Über die Auseinandersetzung mit auch formbezogenen und sprachlichen Merkmalen von Texten ist er hierfür der angemessene, wenn auch vor dem Hintergrund der Forschung zu fachspezifischen Lesestrategien nicht notwendigerweise der alleinige Ort. Wenngleich Lesestrategien prinzipiell auch Zugangsmöglichkeiten für literarische Texte bieten können (Bremerich-Vos 2001), wird ihre Vermittlung und Anwendung im Deutschunterricht insbesondere an die Arbeit mit Sachtexten geknüpft. Ein Grund dafür liegt vermutlich im psychologischen Ursprung des Konzepts der Lernstrategien (Mandl & Friedrich 2006), welches auf einer überfachlichen Modellierung von Selbstregulation (Boekaerts 1999) beruht. Deren Wirkung kann im Umgang mit Sachtexten besonders gut betrachtet und erforscht werden, während sich fiktionale Texte dem Zugriff unspezifischer Lern- und Lesestrategien entziehen (Frederking 2013; Leubner & Saupe 2014). Diesbezüglich zeigen Interventionsstudien zur Förderung des Leseverstehens mithilfe von Sachtexten, dass selbstreguliertes Lesen effektiv trainiert werden kann, bspw. durch den Einsatz von Lesestrategien oder die metakognitive Steuerung des Leseprozesses (z. B. Dignath & Büttner 2008; Souvignier & Mokhlesgerami 2006).

Auf dem Weg von der wissenschaftlichen Entwicklung von evidenzbasierten Strategietrainings, Förderkonzepten und Prinzipien wirksamer Strategievermittlung hin zur Implementation in den Unterricht liegt jedoch das sogenannte *research to practice-gap* (Souvignier & Phillip 2016). Die mit dem Begriff benannte Lücke zwischen Theorie und Praxis veranschaulicht, dass die Forschungsbemühungen nur bedingt im Unterricht ankommen. Die Gründe für misslingende oder

gar nicht erst erfolgreiche Implementationen können vielfältig sein und bspw. in einer zu geringen unterrichtlichen Praktikabilität von Förderkonzepten (ebd.) oder in einer unzureichenden Passung auf die Lernbereiche des Deutschunterrichts liegen. Folgen sind, dass Lehrer*innen Konzepte nicht nutzen oder sie undifferenziert an ihre unterrichtlichen Gegebenheiten anpassen und damit möglicherweise entscheidende Wirkmechanismen verändern. Jene Forschungsarbeiten, die auf den *research to practice-gap* hinweisen, evaluieren insbesondere die Wiedergabetreue von im Unterricht eingesetzten Strategietrainings, verlieren jedoch unterrichtstypische Instruktionen und Anwendungspraktiken von selbstreguliertem Lesen des regulären Deutschunterrichts aus dem Blick – obgleich dies wichtige Anhaltspunkte für die Reflexion über Förderprinzipien und das Handeln der Lehrpersonen bieten würde.

Für eine bessere Implementation evidenzbasierter Prinzipien einer gelingenden Strategievermittlung in die Unterrichtspraxis – wie bspw. die explizite Instruktion (Duke & Pearson 2002) – ist aktuell zu wenig darüber bekannt, wie Lehrer*innen im regulären Unterricht vor allem in der Sekundarstufe I bei der Arbeit mit Sachtexten tatsächlich vorgehen (Philipp 2014) und wie sie ihr Handeln einschätzen.

Mit dem vorliegenden Beitrag wird ein mehrperspektivischer Einblick in die Praxis der Förderung selbstregulierten Lesens in ausgewählten Deutschstunden in fünften Klassen geboten und nach Optimierungspotenzial gefragt. Mittels ratingbasierter Unterrichtsbeobachtungen und einer Lehrkräftebefragung werden spezifische Aspekte der Strategievermittlung fokussiert, die sich in empirischen Studien als wirksam erwiesen haben, und es werden allgemeine Rahmenbedingungen des Leseunterrichts betrachtet. Anders als Studien aus der bildungswissenschaftlichen Unterrichtsforschung, in denen allgemein Leseunterricht betrachtet wird (Stahns, Rieser & Lankes 2017), liegt der Fokus dieses Beitrags auf Unterrichtsstunden, in denen das selbstregulierte Lesen in all seinen Facetten und eine Auseinandersetzung mit Sachtexten über strategiebasierte Zugänge erfolgt.

Nachfolgend werden das selbstregulierte Lesen und Prinzipien einer unterrichtlichen Vermittlung dargestellt, um daran anschließend den Erhebungskontext sowie das Design der Unterrichtsbeobachtung und der Befragung zu erläutern. Auf die Ergebnisdarstellung der Unterrichtsbeobachtungen und der Lehrerbefragung folgt ein Perspektivabgleich der Befunde. Die Diskussion widmet sich einerseits methodologischen Aspekten der durchgeführten Erhebungen, andererseits werden Entwicklungspotenziale für das Handeln von Lehrpersonen im Deutschunterricht, die weitere Beforschung des selbstregulierten Lesens in der Deutschdidaktik und die universitäre Lehrerbildung aufgezeigt.

2 | Zum Konzept des selbstregulierten Lesens im Umgang mit Sachtexten und dessen Vermittlung im Unterricht

2.1 | Dimensionen des selbstregulierten Lesens

Das Verstehen von Sachtexten erfordert aufgrund ihrer zuweilen komplexen, differenzierten Struktur ein zielgerichtetes Lesen. Hierarchiehohe Verstehensprozesse, wie die globale Kohärenzbildung und die Konstruktion eines mentalen Modells (Kintsch 1998), die beim Verstehen von Sachtexten in der Sekundarstufe in besonderem Maße aktiv von den Leser*innen prozessiert werden müssen, lassen sich durch selbstreguliertes Lesen steuern (Rosebrock & Nix 2017).

Es wurde mehrfach nachgewiesen, dass sich der Einsatz von Strategien im Umgang mit Sachtexten förderlich auf das Textverstehen von Lernenden auswirkt (u.a. Dignath & Büttner 2008).

Selbstreguliertes Lesen beruht auf verschiedenen Fähigkeiten, darunter die Nutzung von Lesetechniken, die durch eine bewusste und zielgerichtete Auswahl in einer konkreten Handlungssituation eine strategische Funktion erhalten (Bräuer 2015). Dabei wird z.B. seit Weinstein und Mayer (1986) unterschieden in kognitive Organisations-, Elaborations- und Wiederholungsstrategien. Dieser Unterscheidung nach reduzieren oder strukturieren Organisationsstrategien Informationen, beispielsweise durch Markierung oder Visualisierung von Inhalten (Philipp 2012: 43). Elaborationsstrategien unterstützen eine tiefere Textverarbeitung mittels Aktivierung von Vorwissen, Generierung von eigenen Beispielen sowie der weiterführenden Nutzung des Gelesenen (ebd.: 44). Um Inhalte langfristig zu speichern, sind schließlich Wiederholungsstrategien hilfreich, beispielsweise durch Abschreiben des Textes, wiederholtes Lesen oder mündliches Wiedergeben des Inhalts (ebd.: 43).

Der Einsatz kognitiver Strategien sollte dem naturgemäßen Ablauf eines Verstehensprozesses entsprechen (Stichwort *vor, während und nach dem Lesen*), um so die spezifischen Anforderungen in den einzelnen Phasen auch gezielt zu unterstützen (Philipp 2012: 51 f.; Willenberg 2004: 10, 15). Darüber hinaus kommen beim selbstregulierten Lesen aber auch metakognitive Lesestrategien sowie ressourcenbezogene Strategien zum Tragen. Dabei umfassen metakognitive Lesestrategien Selbststeuerungsaktivitäten der Planung, Überwachung und Regulation des Leseprozesses (Differenzierung für den Lernprozess: vgl. Zimmerman 2002). In der Planungsphase vor dem Lesen geht es z. B. um das Fokussieren eines Leseziels und die Auswahl möglicher Lesestrategien. Während der Überwachung werden das erlangte Textverständnis sowie der Strategieeinsatz beobachtet. Reflexionsstrategien sind nach dem Lesen dienlich, um den Leseprozess hinsichtlich des erzielten Textverständnisses, des Strategieeinsatzes und des Umgangs mit Textschwierigkeiten zu reflektieren. Ressourcenbezogene Strategien rahmen den Einsatz kognitiver und metakognitiver Lesestrategien. Sie richten sich auf die Aktivierung und Regulation motivationaler und emotionaler Bedingungen (z.B. die Aktivierung und Aufrechterhaltung von Lesemotivation, Interesse und Anstrengungsbereitschaft) sowie die Einrichtung einer unterstützenden, lernförderlichen Leseumgebung (Schiefele & Pekrun 1996).

2.2 | Unterrichtliche Vermittlung des selbstregulierten Lesens (zwischen Anspruch und Wirklichkeit)

Da selbstreguliertes Lesen kognitive Aktivierung begünstigt (Kleinbub 2016; Lankes & Carstensen 2007) und sich bei vielen Schüler*innen in der Regel nicht von allein entwickelt (Dole, Nokes & Drits 2009), sollten im Deutschunterricht wiederholt Gelegenheiten zum Üben, Anwenden und Reflektieren von kognitiven und metakognitiven Lesestrategien geschaffen werden (Willenberg 2004). Dabei kann durch Interventionsstudien als gesichert gelten, dass sich eine kombinierte Vermittlung eben von kognitiven und metakognitiven Strategien als besonders förderlich erweist (Souvignier & Mokhlesgerami 2006). Ein weiteres wichtiges Vermittlungsprinzip besteht darin, dass Lehrkräfte Lesestrategien explizit vermitteln, indem sie ihren Schüler*innen den Nutzen sowie die Anwendung vergegenwärtigen (Duke & Pearson 2002; Pressley, Wharton-McDonalds, Mistretta-Hampston & Echevarria 1998). Zudem sollten die Lernenden (wie oben bereits angedeutet) verschiedene kognitive Strategiearten der Organisation, Elaboration und Wiederholung erproben können, um während des Leseprozesses eine individuelle Auswahl

aus einem Strategierepertoire vornehmen zu können. Flankierend sollten Lehrkräfte im Unterricht schließlich ressourcenbezogene Strategien adressieren, indem sie die Schüler*innen beispielsweise zum Lesen eines bestimmten Themas motivieren, an ihre Lebenswelt anknüpfen, mit Lob und Bestärkung beim Lesen oder bei auftretenden Schwierigkeiten agieren und die Lernenden durch Maßnahmen der Differenzierung unterstützen (z.B. bezogen auf das Lesematerial, die Aufgabenstellung oder die Wahl der Sozialform). Mit diesen auf die Ressourcen der Lernenden ausgerichteten Handlungen lässt sich ein unterstützendes Sozialklima im (Lese-)Unterricht erzeugen, was ein weiteres Merkmal guten Unterrichts darstellt (Klieme & Rakoczy 2008).

Werden bei der unterrichtlichen Leseförderung diese Vermittlungsaspekte berücksichtigt, kann von einem auf diesen Lernbereich bezogenen qualitätvollen Unterrichts*angebot* gesprochen werden (vgl. das Angebots-Nutzungs-Modell, Helmke 2015). Einschränkend ist jedoch darauf hinzuweisen, dass die entwickelten Prinzipien eines strategieorientierten Unterrichts und die nachgewiesenen Effekte auf das Textverstehen auf Befunden kontrollierter, i.d.R. pädagogisch-psychologischer, zeitintensiver Interventionsforschung basieren und nur bedingt auf natürliche Unterrichtssettings übertragen werden können. Meta-Analysen zeigen, dass die Effekte von Strategien auf das Textverstehen geringer ausfallen, wenn Lehrkräfte anstatt Forschende die Vermittlung durchführen und wenn sie im regulären Klassenkontext anstatt unter randomisierten Bedingungen stattfindet (Dignath & Büttner 2008; Okkinga et al. 2018).

Aufschlussreich sind vor allem Beobachtungsstudien zum Unterrichtshandeln im Regelunterricht des Faches Deutsch. In diesen wird wiederholt deutlich, dass die Auseinandersetzung mit Sachtexten durch eine weitgehend fehlende Strategievermittlung und seltene Strategieanwendung gekennzeichnet ist, wobei hier vor allem die metakognitiven Strategien der ‚Überwachung‘ anzusprechen sind (Anmarkrud & Bråten 2012; Kleinbub 2010). Dieser Befund spiegelt sich auch in Designs der Lehrerbefragungen, in denen Lehrkräfte bekunden, nur wenige kognitive Lesestrategien zu vermitteln (Lankes & Carstensen 2007; Schmitz 2019). Ferner legen die Auskünfte der Lehrpersonen in der neusten IGLU-Studie nahe, dass „nur eine Minderheit der Kinder von Lehrkräften unterrichtet wird, die in nennenswertem Umfang explizit lehren, wie und wann Lesestrategien einzusetzen sind“ (Bremerich-Vos et al. 2017: 293). Zudem werden kognitive Lesestrategien, die eine hohe Aktivität auf Seiten der Schüler*innen erfordern (bspw. das Formulieren von Teilüberschriften und/oder die Transformation von Textinhalten in eine andere Darstellungsform) selten im Unterricht gefordert (ebd., Schmitz & Jost 2019). Primär reduktiv-organisierende Strategien, wie das Zusammenfassen und das Unterstreichen, werden nach Angaben der Lehrkräfte im Unterricht praktiziert (Bremerich-Vos et al. 2017; Schmitz 2019; Schmitz & Jost 2019). Ergänzend konstatierte bereits Ness (2008), dass (Fach-)Lehrkräfte Kompensationsmaßnahmen wie die Anpassung eines Textes an die Lernausgangslagen der Schüler*innen unterrichtlicher Instruktion von Lesestrategien vorziehen (vgl. ferner Pressley et al. 1998). Da die Aktivierung von ressourcenbezogenen Strategien mit Blick auf das Textverstehen im Leseunterricht bislang nicht als solches erforscht wurde, lassen sich nur Vergleiche mit Studien herstellen, die allgemeine Dimensionen wie das Sozialklima erhoben haben. Indikatoren zur Erfassung des Sozialklimas im Deutsch- bzw. Leseunterricht, z.B. unterrichtliche Fehlerkultur, Schüler-Lehrer-Beziehung oder Motivierung mittels Lob, werden von Lehrkräften sowie externen Beobachter*innen als ausgeprägt beurteilt (Lotz 2016; Stahns et al. 2017).

Obgleich die Studien aufgrund ihres methodischen Zugangs, der Stichproben und Operationalisierungen nicht stringent zu vergleichen sind, scheint die Gestaltung eines auf Selbstregulation beim Lesen ausgerichteten Deutschunterrichts Lehrkräfte vor besondere Herausforderungen zu stellen, worauf zuvor bereits mit dem *research to practice-gap* hingewiesen wurde (Souvignier & Philipp 2016). Dabei ist die Befundlage für den Unterricht mit Blick auf die Primarstufe gewichtiger als mit Blick auf die Sekundarstufe I, sodass insbesondere für Letztere empirische Einblicke dringend erforderlich sind.

3 | Fragestellungen

Um den aufgeführten Desideraten zu begegnen, wurden in der vorliegenden Untersuchung das unterrichtliche Handeln von Lehrkräften bei der Vermittlung von selbstreguliertem Lesen im Umgang mit Sachtexten in fünften Klassen mittels Unterrichtsbeobachtungen und Lehrerbefragung erhoben. Die Fragestellungen lauten:

- (1) Wie gestalten Lehrkräfte ihren Deutschunterricht, wenn sie selbstreguliertes Lesen am Beispiel eines Sachtextes vermitteln?
- (2) Wie beurteilen die Lehrkräfte ihr eigenes Unterrichtshandeln?
- (3) In welcher Relation stehen die Unterrichtsbeobachtungen zur Selbsteinschätzung der Lehrpersonen und an welchen Stellen zeigt sich Optimierungspotenzial angesichts der Vermittlung von selbstreguliertem Lesen im Deutschunterricht?

4 | Methode

4.1 | Erhebungskontext und Stichprobe

Die Unterrichtsbeobachtungen erfolgten im Rahmen der Evaluationsstudie EILe¹, in der die Implementierung von Leseförderkonzepten in den Regelunterricht formativ evaluiert und begleitet wurde (Schmitz et al. 2021). An EILe beteiligten sich insgesamt 23 Projektschulen mit 54 Klassen, die sich zum Ziel setzten, ein vorgegebenes Leseförderkonzept zum selbstregulierten Lesen in den Deutschunterricht zu integrieren.

Die nachfolgend vorgestellten Unterrichtsbeobachtungen wurden im Jahr 2017 in der zweiten Schuljahreshälfte in fünften Klassen durchgeführt. Beobachtet wurden 31 Deutschstunden, davon in 9 Schulen jeweils drei Klassen und in 2 Schulen zwei Klassen. Es handelt sich um 31 Lehrpersonen (davon 5 männlich), die im Schnitt 39 Jahre alt ($SD = 10,82$) waren und 13 Jahre Berufserfahrung ($SD = 10,23$) hatten. Die Auswahl der Beobachtungsstunden richtete sich nach der Freiwilligkeit der Lehrpersonen und der Zustimmung der Erziehungsberechtigten.

Es handelte sich um ratingbasierte Beobachtungen zweier geschulter Projektmitarbeiterinnen mithilfe von Audioaufnahmen und dem Beobachtungsbogen Lestra-BD (Schmitz, Karstens & Jost 2020). Um in den Klassen eine vergleichbare Situierung zu erzeugen, erhielten die Lehrkräfte den Auftrag, einen für die Lernenden noch unbekanntem Sachtext ihrer Wahl strategieorientiert zu erarbeiten. Die Sachtexte umfassten i.d.R. etwa eine Seite und behandelten

¹ Das Projekt EILe ist ein vom BMBF gefördertes Projekt (Souvignier & Jost, FK: 01J11506A/B; Laufzeit: 01/16-12/18).

Sachthemen zu Tierwelten, Sportarten, Geschichtlichem oder Umweltpolitischem. Um über die punktuellen Unterrichtsbeobachtungen hinaus das übliche Handeln der Deutschlehrer*innen zu erfassen, wurde eine größere Stichprobe von 135 Deutschlehrpersonen, darunter auch die 31 beobachteten Lehrkräfte, befragt (Schmitz 2019). Sie waren durchschnittlich 41 Jahre alt ($SD = 10,19$) und hatten 14 Jahre Berufserfahrung ($SD = 9,82$).

4.2 | Dimensionen der Beobachtungsstudie und der Befragung

In der Beobachtungsstudie wurden folgende Prinzipien eines strategieorientierten Unterrichts erhoben:

- die Unterstützung des Leseprozesses (vor dem Lesen, während des Lesens, nach dem Lesen) durch kognitive und metakognitive Lesestrategien (mit Blick auf Sachtexte),
- die Explizitheit der Instruktion,
- die Aktivierung von ressourcenbezogenen Strategien durch Motivierung, Involvierung und Berücksichtigung von Heterogenität.
- Ergänzend wurden darüber hinaus die Aktivierung der Schüler*innen, die Lernformen während der Textarbeit, die Arten des Lesens sowie die Zeitnutzung betrachtet.

Tabelle 1 enthält die Beobachtungsdimensionen, Beispiele und Skalierungen.

| Beobachtungsdimension | Beschreibung | Einschätzung |
|---|---|---|
| Instruktion kognitiver und metakognitiver Lesestrategien | | |
| Instruktion kognitiver Strategien auf Prozessebene | Erhebung, ob kognitive Lesestrategien in den Lesephasen angeleitet werden: <ul style="list-style-type: none"> • Hinführung bevor der Text gelesen wird (vor) • Während des Lesens und Bearbeitens (während) • Anschlussphase, nachdem der Text bearbeitet ist (nach) | Rating jeweils: ja / nein |
| Explizitheit der Instruktion in den Phasen des Leseprozesses | <ul style="list-style-type: none"> • Implizit • Explizit | Rating jeweils: ja / nein |
| Kognitive Strategiearten in den Phasen des Leseprozesses | Differenzierung von kognitiven Strategiearten in den Phasen (vor, während, nach) | Rating: Organisation / Elaboration / Wiederholung |
| Grad der Explizitheit kognitiver Strategieinstruktion in der Gesamtstunde | <ul style="list-style-type: none"> • Sachkompetenz: Lesestrategien spielen keine Rolle in der Stunde • Implizit: Text wird strategisch erarbeitet, keine Vermittlung der Lesestrategie, lediglich Anwendung einer Technik. • Implizit prozedural: Text wird strategisch erarbeitet, die Anwendung der Technik wird erklärt, aber ohne Verweis auf jene als Lesestrategie • Explizit deklarativ: Technik wird konkret als Lesestrategie bezeichnet • Explizit konditional: Nutzen von Lesestrategie wird erklärt • Explizit prozedural: Anwendung von Lesestrategie wird erklärt | Rating: jeweils ja / nein |
| Modellierung der Vermittlung | Visualisierung der Strategievermittlung | Rating: ja / nein |
| Instruktion metakognitiver Strategien auf Prozessebene | Erfassung von metakognitiven Strategiearten: <ul style="list-style-type: none"> • Festlegung/Erarbeitung eines Leseziels • Planung des Lesens • Reflexion über das strategische Vorgehen • Reflexion über das Textverständnis • Reflexion über Textschwierigkeiten | Rating: jeweils ja / nein |

| | | |
|-----------------------------------|--|--|
| Lernformen während der Textarbeit | Plenum Einzelarbeit Kooperative Lernformen (Partner- oder Gruppenarbeit) | Relativer Anteil in der Gesamtstunde Rating jeweils: Likert 1-4 |
|-----------------------------------|--|--|

Quellen: Neukonstruktion unter Rückgriff auf Boekaerts (1999), Wagner, Helmke & Rösner (2009), Duke & Pearson (2002), Bos et al. (2006), Mandl & Friedrich (2006), Weinstein & Mayer (1986), Zimmerman (2002)

Ressourcenbezogene Unterstützung des Leseprozesses

| | | |
|-------------|---|---|
| Motivierung | <ul style="list-style-type: none"> • Thematische Motivierung: Erhebung, inwiefern in der Stunde das Interesse der Lernenden für das Thema geweckt wird. • Berücksichtigung der Lebenswelt: Erhebung, ob ein Bezug zur Lebenswelt der Lernenden hergestellt wird (Beispiele, Materialien, Erfahrungen, Alltagsrelevanz). | Einschätzung der Gesamtstunde Rating jeweils: Likert 1-4 |
|-------------|---|---|

| | | |
|---------------|--|---|
| Unterstützung | <ul style="list-style-type: none"> • Lob und Bestärkung: Erhebung, ob die Lehrperson hinsichtlich des Verstehens von Texten, der Anwendung von Strategien mit Lob/Bestärkung ausspricht. • Unterstützung bei Leseschwierigkeiten: Erhebung, ob die Lernenden ermuntert werden, sich bei Verstehensschwierigkeiten oder der Anwendung von Lesestrategien anzustrengen, es nochmal zu probieren, nachzudenken etc. | Einschätzung der Gesamtstunde Rating jeweils: Likert 1-4 |
|---------------|--|---|

| | | |
|--------------------------|---|---------------------------|
| Umgang mit Heterogenität | Erhebung von Differenzierungsmaßnahmen: <ul style="list-style-type: none"> • Lesestrategien • Lesematerial • Aufgaben • Lernformen | Rating jeweils: ja / nein |
|--------------------------|---|---------------------------|

Quellen: Neukonstruktion unter Rückgriff auf Wagner et al. (2009), Bos et al. (2006), Frey et al. (2009), Pressley et al. (1998)

Aktive Lernzeit für Erwerb von selbstreguliertem Lesen

| | | |
|-------------------------------|---|---|
| Aktivierung der Schüler/innen | Erhebung der Schüleraktivierung während der Textarbeit <ul style="list-style-type: none"> • Frage-Antwort-Muster im Plenum und stark lehrergeleitete Textarbeit | Einschätzung der Gesamtstunde Rating jeweils: Likert 1-4 |
|-------------------------------|---|---|

| | | |
|----------------|---|---|
| Art des Lesens | <ul style="list-style-type: none">• Frage-Antwort-Muster im Plenum unter reger Beteiligung der Lernenden• Individuelle oder kooperative Textarbeit Formen des Lesens <ul style="list-style-type: none">• Lautes Lesen• Individuelles, leises Lesen | Rating jeweils: ja / nein |
| Zeitnutzung | Einschätzung von Zeiträumen , in welchem die Lernenden den Text lesen und bearbeiten können und Lesestrategien erklärt/angeleitet oder thematisiert werden. <ul style="list-style-type: none">• Zeit zur Vermittlung von Lesestrategien• Zeit zur Anwendung von Lesestrategien• Zeit zur Reflexion von Lesestrategien und Textverstehen• Zeit zum eigenständigen Lesen des Textes | Einschätzung der Gesamtstunde Rating jeweils: Likert 1-4 |

Quellen: Neukonstruktion unter Rückgriff auf Bos et al. (2006)

Tab. 1: Dimensionen der Beobachtungsstudie, Beispiele und Skalierungen.

Anmerkung: Likert-Skala: 1 = trifft nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu; 4 = trifft zu.

Die Güte der Beobachtungen wurde durch die Interrater-Reliabilität Kappa zweier Beobachterinnen, die als wissenschaftliche Mitarbeiterinnen im Projekt arbeiteten, ermittelt. Die Werte ($.57 \leq \kappa \leq 1.00$) legen eine mäßige bis vollständige Übereinstimmung der Beobachterinnen nahe (Schmitz et al. 2020).

In der Lehrerbefragung wurden ausgewählte Aspekte der in Tabelle 1 dargestellten Beobachtungsdimensionen durch Likert-Skalen erfasst, um Häufigkeiten oder das Ausmaß persönlicher Zustimmung für das übliche Unterrichtshandeln abzubilden (Schmitz 2019). Erfasst wurde die Instruktion von kognitiven Lesestrategien, metakognitiven Lesestrategien, die Aktivierung der Schüler*innen und die Arten des Lesens durch Häufigkeiten (1 = *in fast keiner Unterrichtsstunde*, 2 = *in einigen Unterrichtsstunden*, 3 = *in den meisten Unterrichtsstunden*, 4 = *in (fast) jeder Unterrichtsstunde*). Der Grad der Explizitheit, die Motivierung und Unterstützung sowie der Umgang mit Heterogenität wurden durch die Skalierung 1 = *stimmt überhaupt nicht*, 2 = *stimmt eher nicht*, 3 = *stimmt eher*, 4 = *stimmt völlig* erfasst. Diese Skalierung orientiert sich an den Skalen großer Schulleistungsstudien wie PISA und IGLU. Skalen, bei welchen Mittelwerte gebildet wurden, erzielten eine gute interne Konsistenz ($\alpha = .79$ bis $.81$) (Schmitz 2019).

5 | Ergebnisse

5.1 | Wie gestalten Lehrkräfte ihren Deutschunterricht, wenn sie selbstreguliertes Lesen am Beispiel eines Sachtextes vermitteln?

Betrachtet werden zunächst die kognitiven Strategien. Bei der Instruktion auf der Prozessebene wurde beobachtet, dass die Lehrkräfte ihre Lernenden in fast jeder Stunde mit kognitiven Strategien auf den Sachtext vorbereiten (Abb. 1). Differenziert nach ‚vor, während und nach dem Lesen‘ lässt sich dies wie folgt beschreiben: *Vor* dem Lesen bevorzugen die Lehrkräfte eine implizite Vorbereitung (67 Prozent der Stunden), in nur 3 Prozent der Stunden erfolgt keine Vorbereitung. *Während* des Lesens überwiegen hingegen explizite Vermittlungen (52 Prozent), implizit ist die Vermittlung hier in 45 Prozent der Stunden, keine Vermittlung erfolgte in 3 Prozent der Stunden. *Nach* dem Lesen werden überwiegend implizite Vermittlungen sichtbar (47 Prozent), gefolgt von expliziten Vermittlungen (40 Prozent) und nicht erfolgter Nachbereitung (13 Prozent).

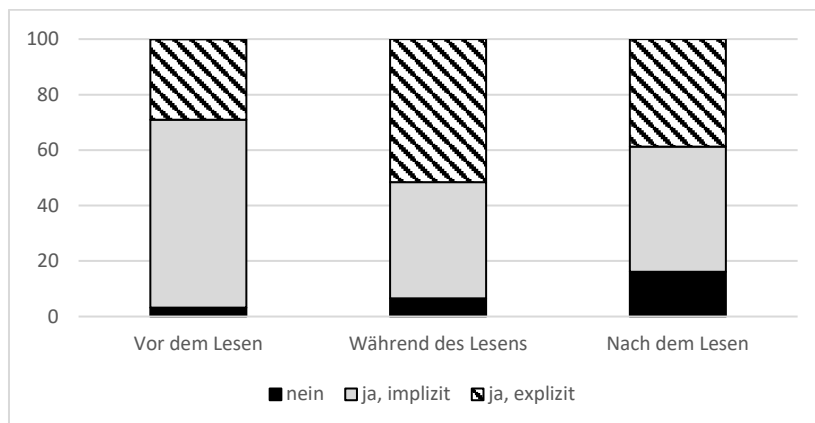


Abb. 1: Explizitheit der Instruktion kognitiver Strategien in den Phasen des Leseprozesses.

Welche Strategiearten und Kombinationen die Deutschlehrer*innen während der Textbearbeitung instruieren, ist Abb. 2 zu entnehmen.

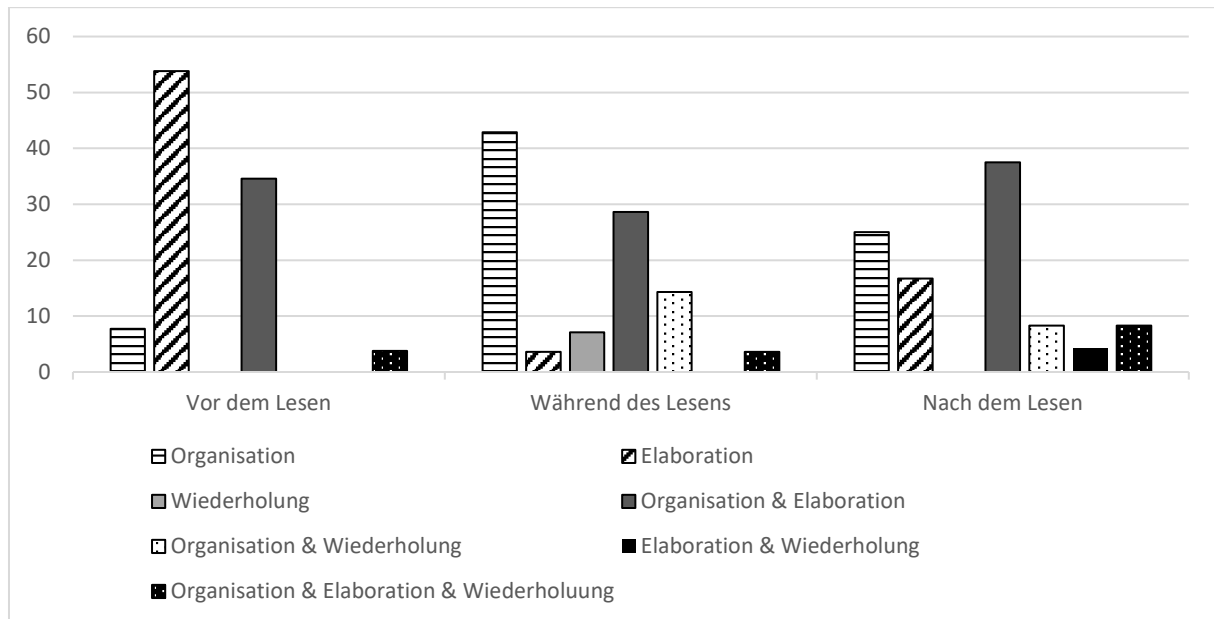


Abb. 2: Kognitive Strategiearten in den Phasen des Leseprozesses.

Die bereits bekannten Strategiearten und -kombinationen betreffend, werden in der Hälfte der Stunden (54 Prozent) *vor* dem Lesen des Textes ausschließlich Elaborationsstrategien aktiviert. In anderen Stunden (35 Prozent) werden sowohl Organisations- als auch Elaborationsstrategien von den Lehrkräften vermittelt und eingesetzt. *Während* des Lesens fokussieren die Lehrkräfte in 43 Prozent der Stunden auf Organisationsstrategien und in 29 Prozent werden sowohl Organisations- als auch Elaborationsstrategien beobachtet. Zur *Nach*bereitung des Lesens werden in 38 Prozent Organisations- und Elaborationsstrategien beobachtet, gefolgt von 25 Prozent der Stunden mit ausschließlich Organisationsstrategien und 17 Prozent mit ausschließlich Elaborationsstrategien. Wiederholungsstrategien nehmen während und nach dem Lesen nur eine Randstellung ein.

Unabhängig von der Phase des Lesens handelt es sich bei den Elaborationsstrategien vorwiegend um die Aktivierung von inhaltlichem Vorwissen und das Generieren inhaltlicher Hypothesen. Bei den Organisationsstrategien werden das überfliegende Lesen, das Markieren von Schlüsselbegriffen und die Visualisierung des Textinhaltes praktiziert. Wiederholende Strategien werden durch mehrfaches Lesen realisiert. Kaum beobachtet werden Aktivitäten wie das Generieren von eigenen Beispielen, das selbstständige Formulieren von Fragen an den Textinhalt, das Zusammenfassen desselben oder ein z.B. paraphrasierendes Notieren.

Auf die Gesamtstunde bezogen stellt Abbildung 3 dar, welche Formen von Implizitheit oder Explizitheit in den Stunden vorliegen. Es handelt sich dabei um die Gesamtheit der 31 Stunden, da in jeder Stunde Strategien vermittelt wurden, wenngleich nicht in jeder Phase des Lesens (vgl. Abb.1).

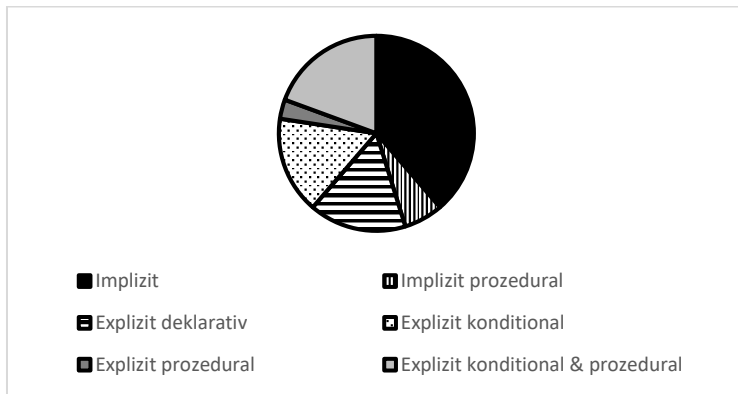


Abb. 3: Grad der Explizitheit kognitiver Strategieinstruktion in der Gesamtstunde.

In keiner beobachteten Stunde wird auf die Vermittlung von kognitiven Lesestrategien verzichtet. Diese Beobachtung lässt sich wie folgt differenzieren:

- 39 Prozent der Deutschstunden sind implizit,
- 7 Prozent implizit prozedural.
- In 16 Prozent der Stunden wird explizit deklarativ instruiert, indem beispielsweise erwähnt wird, dass das Markieren eine Lesestrategie ist.
- Der Nutzen wird in weiteren 16 Prozent der Stunden erläutert (d.h. explizit konditional)
- und weiter erfolgt in 20 Prozent der Stunden eine kombinierte Instruktion von Nutzen und Vorgehen (explizit konditional und prozedural).
- Unterricht, in dem Lesestrategien auf explizit prozedurale Weise vermittelt werden, bildet mit lediglich (3 Prozent) eine Randerscheinung.

In etwa einem Drittel der Stunden (32 Prozent) wird der Strategieeinsatz visualisiert, z. B. mittels Verdeutlichung der Strategieanwendung an einer Dokumentenkamera oder an einem Overheadprojektor. Die auf einer vierstufigen Skala bewerteten Lernformen während der Textarbeit erfolgen häufig im Plenum ($M = 2,42$, Modus = 3) und Einzelarbeit ($M = 2,84$, Modus = 3), aber nie bzw. selten in Gruppen- oder Partnerarbeit ($M = 1,48$, Modus = 1).

Hinsichtlich der Instruktion metakognitiver Strategien auf Prozessebene zeigt sich, dass die verschiedenen Strategiearten unterschiedlich etabliert zu sein scheinen. Die Lehrkräfte geben in jeder zweiten Stunde ein Leseziel vor und planen in 65 Prozent der Stunden den Strategieeinsatz. Die Reflexion des Textverständnisses wird in 52 Prozent der Stunden beobachtet. Die Reflexion über Textschwierigkeiten erfolgt in 49 Prozent aller Stunden, wohingegen über das strategische Vorgehen in 39 Prozent der Stunden reflektiert wird.

Die Ausprägungen der ressourcenbezogenen Unterstützung des Leseprozesses finden sich in Tabelle 2.

| Items | N | M | SD |
|---------------------------------------|----|------|------|
| Motivierung | | | |
| Thematische Motivierung | 31 | 2,26 | 0,97 |
| Berücksichtigung der Lebenswelt | 31 | 2,29 | 1,10 |
| Unterstützung | | | |
| Lob und Bestärkung | 31 | 2,61 | 1,02 |
| Unterstützung bei Leseschwierigkeiten | 30 | 2,00 | 1,02 |

Tab. 2: Mittelwerte und Standardabweichungen für die Motivierung und Unterstützung.

Anmerkung: Likert-Skala: 1 = trifft nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu; 4 = trifft zu.

Beobachtet wird, dass die Lehrkräfte ihre Schüler*innen nur wenig für das spezifische Thema des Textes motivieren sowie an ihre Lebenswelt anknüpfen. Hingegen ist die Unterstützung durch die Lehrkräfte höher ausgeprägt, da sie ihre Lernenden eher loben und bestärken. Eine Ermutigung der Lernenden bei auftretenden Leseschwierigkeiten wurde nicht beobachtet.

Dem Umgang mit Heterogenität wird in 20 Prozent aller Stunden durch differenzierte Aufgabenstellungen und Strategien entsprochen und in 13 Prozent durch unterschiedlich schwierige Textversionen. Wenig bzw. gar nicht etabliert sind eine differenzierte Nutzung von Lesestrategien (in nur einer Stunde beobachtet) oder die Arbeit in individuellen Lernformen zur Texterschließung.

Hinsichtlich der Aktivierung der Schüler*innen zeichnet sich insgesamt das Bild ab, dass in der Mehrheit der Stunden Schüler-Lehrer-Gespräche im Plenum unter reger Beteiligung der Lernenden stattfinden und schülerzentrierte, individuelle sowie kooperative Lernformen seltener ermöglicht werden.

In 40 Prozent der Stunden wird nur individuell und leise gelesen. Sowohl lautes Lesen im Klassenverband (z. B. Reihum-Lesen) als auch leises Lesen werden in jeder zweiten Stunde praktiziert. Das laute Lesen als alleinige Form der Texterschließung findet in 7 Prozent der Stunden statt.

Wie Tabelle 3 verdeutlicht, schaffen die beobachteten Lehrkräfte einen Zeitraum, um Lesestrategien zu vermitteln und diese anwenden zu lassen. Etwas geringer wird der Zeitraum eingeschätzt, um über Lesestrategien und das Textverstehen zu reflektieren. Zeit für das eigenständige Lesen wird lt. der Beobachterinnen durchaus geschaffen.

| Items | N | M | SD |
|---|----|------|------|
| Zeit zur Vermittlung von Lesestrategien | 31 | 2,94 | 0,99 |
| Zeit zur Anwendung von Lesestrategien | 31 | 3,74 | 0,63 |
| Zeit zur Reflexion von Lesestrategien und Textverstehen | 30 | 2,70 | 1,15 |
| Zeit zum eigenständigen Lesen des Textes | 31 | 3,58 | 0,72 |

Tab. 3: Mittelwerte und Standardabweichungen der Items Zeitnutzung.

Anmerkung: Likert-Skala: 1 = trifft nicht zu, 2 = trifft eher nicht zu, 3 = trifft eher zu; 4 = trifft zu.

5.2 | Wie beurteilen die Lehrkräfte ihr eigenes Unterrichtshandeln?

Die Befragung der Deutschlehrkräfte ($N = 135$) zu ihrem üblichen Vorgehen ergibt, dass sie kognitive Lesestrategien durchaus vermitteln ($M = 2,53$, $SD = 0,37$) (Schmitz 2019: 24). Hinsichtlich der kognitiven Strategiearten leiten sie unter den Organisationsstrategien vor allem das Markieren, Strukturieren und Zusammenfassen an. Seltener finden Visualisierungen und überfliegendes Lesen Anwendung im Unterricht. Unter den Elaborationsstrategien wird das Formulieren von Hypothesen am häufigsten instruiert, wohingegen das Formulieren von eigenen Beispielen und das Generieren von Fragen seltener erfolgen. Das wiederholte Lesen des Textes wird in einigen Stunden praktiziert (ebd.: 24 f.). Bezüglich der Instruktion stimmen die Lehrpersonen eher zu, Lesestrategien explizit zu vermitteln (ebd.). Sie bekunden, Lesestrategien zudem konditional und prozedural zu instruieren (ebd.: 25). Die Häufigkeitsanalyse zeigt diesbezüglich, dass 90 Prozent der Befragten angeben, deklaratives Strategiewissen zu vermitteln und den Nutzen einer Lesestrategie sowie die Anwendung zu erklären (konditional und prozedural). Die strategische Textarbeit und die Einübung der Strategien erfolgen überwiegend gemeinsam mit den Lernenden (90 Prozent), seltener in Einzelarbeit (60 Prozent) und noch seltener in kooperativen Lernformen (Partner- oder Gruppenarbeit insg. 40 Prozent).

Metakognitive Lesestrategien werden vergleichbar oft vermittelt wie kognitive Lesestrategien ($M = 2,50$, $SD = 0,55$), davon die Reflexion über das Textverständnis am häufigsten. In einigen Stunden wird ein Leseziel erarbeitet oder angezeigt, weniger häufig wird das strategische Vorgehen geplant und darüber reflektiert (ebd.: 26).

Die Aktivierung von ressourcenbezogenen Strategien durch Motivierung und Unterstützung wird von den Lehrpersonen als eher zutreffend eingeschätzt. Sie geben an, ihre Schüler*innen eher für das Thema des Textes zu motivieren ($M = 2,98$, $SD = 0,45$) und sie zu unterstützen ($M = 3,42$, $SD = 0,36$) (Schmitz 2019: 27). Die Frage nach dem Umgang mit Heterogenität durch den Einsatz von verschiedenen Textversionen und/oder Lesestrategien ergibt, dass die Hälfte der Lehrkräfte in einer Klasse einen einheitlichen Text austeilte und diesen mit denselben oder unterschiedlichen Lesestrategien erschließen lässt (jeweils 50 Prozent). 30 Prozent der Lehrpersonen teilen unterschiedliche Textversionen aus und lassen diese überwiegend mit den gleichen Lesestrategien erarbeiten (nur 20 Prozent geben an, den Lesestrategieeinsatz zu variieren).

Knapp 70 Prozent der Lehrpersonen unterrichten die Klasse in den meisten bis (fast) allen Stunden im Plenum. Einzel-, Partner oder Gruppenarbeit finden laut 43 Prozent der Befragten in den meisten bis (fast) allen Stunden statt. 80 Prozent geben ferner an, dass sie in den meisten

bis (fast) allen Stunden lautes Lesen im Klassenverband praktizieren. 55 Prozent stimmen zu, leises Lesen in den meisten bis (fast) allen Stunden zu ermöglichen.

5.3 | In welcher Relation stehen die Unterrichtsbeobachtungen zur Selbsteinschätzung der Lehrpersonen und an welchen Stellen zeigt sich Optimierungspotenzial angesichts der Vermittlung von selbstreguliertem Lesen im Deutschunterricht?

Ein Vergleich der Ergebnisse aus beiden Studien weist an einigen Stellen Übereinstimmungen auf, an manchen Stellen zeigen sich leichte Abweichungen und auch deutlichere Diskrepanzen.

Übereinstimmungen die Instruktion betreffend liegen dahingehend vor, welche kognitiven Strategiearten die Lehrkräfte bevorzugt instruieren (insb. klassische kognitive Organisations- und Elaborationsstrategien) und in welchen Lernformen die strategische Textarbeit in der Regel erfolgt (meist in Plenumsarbeit, teils in Einzelarbeit und seltener in kooperativen Lernformen). Konvergenzen zeichnen sich auch bei den metakognitiven Lesestrategien ab: Die Reflexion über das Textverständnis und die Setzung eines Leseziels werden als ähnlich häufig eingeschätzt, die Reflexion über das strategische Vorgehen als etwas seltenere Aktivität. Lediglich die Planung des strategischen Vorgehens unterscheidet sich, insofern dieses lt. Selbstauskunft eher selten stattfindet, wohingegen es in den Beobachtungen in 65 Prozent der Stunden ermittelt wurde. Beide Studien deuten auf aktivierte Schüler*innen während der Schüler-Lehrer-Gespräche im Plenum hin, wohingegen kooperative Lernformen weniger etabliert zu sein scheinen.

Eine leichte Diskrepanz zeigt sich in der beobachteten und selbst eingeschätzten Instruktionsweise: Sie wurde in den Beobachtungen in 46 Prozent der Stunden als implizit wahrgenommen und in 54 Prozent als explizit, was bei der Anzahl an beobachteten Stunden kein bedeutender Unterschied ist. In der Befragung stimmten hingegen 90 Prozent der Lehrpersonen eher der Aussage zu, explizite Instruktionen zu realisieren. Bezogen auf den Umgang mit Heterogenität geben die Befragten an, vor allem Textversionen zu variieren und weniger die Strategien. In den Deutschstunden wurden differenzierte Textvariationen und Lesestrategien jedoch selten beobachtet. Das Lesen wurde in den Beobachtungen etwa zur Hälfte leise und zur Hälfte laut praktiziert, bei den Befragungen geben sogar 80% der Lehrpersonen an, in den meisten bis (fast) allen Stunden lautes Lesen einzufordern. Die ressourcenbezogene Unterstützung bringt noch deutlichere Unterschiede hervor. Während sich die Lehrpersonen als eher motivierend und unterstützend einschätzen, wurde dies in den Beobachtungen als wenig ausgeprägt beurteilt.

6 | Diskussion

Dass die Studien einerseits Übereinstimmungen, andererseits Diskrepanzen aufweisen, war zu erwarten. Die Unterschiede können darin begründet sein, dass sich die Beobachtungen auf eine konkrete Unterrichtsstunde beziehen, bei der Befragung jedoch keine spezifische Stunde im Mittelpunkt stand, sondern das übliche Handeln der Lehrpersonen. Die erfassten Indikatoren können zum Zeitpunkt der Beobachtungen durchaus anders ausgeprägt sein, als es üblicherweise der Fall ist. Hinsichtlich der Adressierung von ressourcenbezogenen Strategien ist es bspw. denkbar, dass die Mittelwerte in den beobachteten Unterrichtsstunden im Durchschnitt geringfügiger ausgeprägt sind, wenngleich die Lehrkräfte im Allgemeinen motivierend agieren. Die Unterschiede können zudem dem Deutungsspielraum (Inferenz) während der Beobachtung

und bei der Selbsteinschätzung geschuldet sein. Kognitive und metakognitive Strategien wurden so operationalisiert, dass ihr Vorkommen oder ihre Abwesenheit eindeutig zu bestimmen ist (z. B. durch die Entscheidung zwischen ja oder nein). Gleiches gilt für die unterrichtlichen Sichtstrukturen wie die Lernformen während der Textarbeit und die aktive Lernzeit für den Erwerb von selbstreguliertem Lesen. Hingegen beruhen die Angaben zur ressourcenbezogenen Unterstützung auf Einschätzungen, deren Start- und Endpunkte einer Skala individuell interpretiert werden können. Nicht auszuschließen ist auch das Kriterium der sozialen Erwünschtheit, da sich die Lehrkräfte in den Fragebögen vor allem bei der Motivierung und Unterstützung besser einschätzen, als dies in den Beobachtungen ersichtlich wurde.

Wenngleich angenommen werden könnte, dass es sich um Unterricht in Schulen mit positiver Selbstselektion handelt und dieser daher nicht repräsentativ sei, zeigen identische Beobachtungen in fünf Kontrollschulen, dass sich der Deutschunterricht in den Projektschulen von dem in den Kontrollschulen, in denen kein Leseförderkonzept zur Verfügung stand, nicht unterscheidet (Schmitz et al. 2021). Die Befunde angesichts eines strategieorientierten Unterrichts vergleichend, lassen sich daher einige Entwicklungspotenziale aufzeigen – wenngleich an dieser Stelle einschränkend darauf hinzuweisen ist, dass für eine angemessene Beurteilung des Unterrichtsangebots die Lernvoraussetzungen der Klassen einbezogen werden müssten. Jedoch stellt sich ungeachtet dieser Einschränkung ein sich wiederholender Befund ein (Anmarkrud & Bråten 2012; Kleinbub 2016; Schmitz & Jost 2019), der uns zeigt, dass Lehrkräfte augenscheinlich wenige und – darüber hinaus – stets die gleichen Strategiearten vermitteln. Schüler*innen scheinen vor allem das Markieren, das Generieren von Hypothesen zum Inhalt mittels Überschrift sowie das überfliegende und wiederholte Lesen kennenzulernen. Dabei wäre eine höhere Varianz bei der Vermittlung von kognitiven Lesestrategien wichtig, womit vor allem die Instruktion kognitiv anspruchsvollerer Strategien wie das Formulieren von Fragen an den Text, das Generieren von Beispielen, Visualisieren und Zusammenfassen gemeint ist. Weiter wäre es angesichts der steigenden Heterogenität in Klassen wichtig, eine explizite Instruktion auch in kooperativen Lernsettings zu realisieren und die eher seltener angewendeten metakognitiven Strategien der Planung und der Reflexion des erfolgten Strategieeinsatzes zu integrieren (Schmitz 2020). Auch ließe sich beim Generieren von Fragen an den Text als kognitive Lesestrategie eine Verbindung zur metakognitiven Strategie der Reflexion über das bis dato erlangte Textverständnis erzielen.

Die Befunde zu den ressourcenbezogenen Strategien der Motivierung und Unterstützung deuten – wenn auch mit Blick auf die Befunde in den vorliegenden Studien widersprüchlich – darauf hin, dass Lehrpersonen über die unterrichtliche Aktivierung von ressourcenbezogenen Strategien reflektieren sollten bzw. diese Techniken bewusst thematisieren sollten, damit ihre Lernenden diese wichtigen Facetten der Selbstregulation erwerben. Angesichts heterogener Lernvoraussetzungen und des Prinzips der individuellen Strategieanwendung wäre es zudem wichtig, dass die Lernenden (wie zuvor erwähnt) nicht nur *unterschiedliche* Strategien kennenlernen, sondern auf einen *individuellen* Lesestrategieeinsatz (individuell bspw. hinsichtlich des persönlichen Leseziels oder der vorliegenden Textsorte etc.) hingewiesen werden. Dies könnte langfristig auch den Aufwand des augenscheinlich landläufigen Erstellens von verschiedenen Textversionen reduzieren (Ness 2008). Anstatt auf die Heterogenität der Schüler*innen mit vereinfachten Texten mit differenzierten Lesestrategien zu antworten, fordert diese eben nicht

heraus, sich auch mit für sie schwierigen Texten strategisch auseinanderzusetzen (Schmellentin & Gilg 2016).

Es zeigt sich folglich insgesamt, dass die hervorgebrachten Unterrichtsroutinen Bezugspunkte darstellen, um die Forschung zum selbstregulierten Lesen zu reflektieren und praxistaugliche(re) Maßnahmen sowie Trainings zu entwickeln (Schmitz 2019). Insbesondere von der Psychologie wurden in den vergangenen Jahren zahlreiche Konzepte, Materialien und Methodentrainings zum selbstregulierten Lesen entwickelt und zum großen Teil auch evaluiert. Da diese jedoch aufgrund ihrer vornehmlich psychologischen Provenienz an das Konzept des selbstregulierten Lernens angelehnt und daher fach- und gegenstandsbezogen wenig ausgeprägt sind, bedarf es einer deutschdidaktischen, empirischen Konkretisierung des Konstruktes. Bislang hat die Deutschdidaktik dazu jedoch vergleichsweise wenig anzubieten.

Dabei dürfte man sich von einer deutschdidaktischen Konkretisierung des selbstregulativen Lesens eine – neben fach- und gegenstandsspezifischen weiterhin auch stärker an der Textoberflächencharakteristik verschiedener Textsorten orientierte – Abgrenzung des selbstregulierten Lesens vom allgemeineren (psychologischen) Konzept des selbstregulierten Lernens versprechen. Konkret könnte es in deutschdidaktischer Perspektive weiterhin auch um eine stärkere Berücksichtigung schulischer Leseprozesse, um schulische Heterogenitätsanforderungen vor dem Hintergrund von Differenzierungsmöglichkeiten bei der Strategieauswahl und -kombination, aber auch um das Ableiten didaktischer Handlungsoptionen für Deutschlehrkräfte gehen (vgl. erste Arbeiten in Willenberg 2004; Praxis Deutsch Heft 31; Karstens, Schmitz & Jost 2019; Schmitz 2020).

Die Ergebnisse können aber auch Anhaltspunkte dafür geben, Lehrer*innen die Potenziale der Strategienutzung im Leseprozess verstärkt transparent zu machen. An diesem Punkt ist die universitäre deutschdidaktische Ausbildung gefragt, in der nach einer ersten (unsystematischen) Erhebung von Lehrinhalten deutschdidaktischer Seminare an verschiedenen Universitäten die Förderung selbstregulierten Lesens kaum thematisiert und Begriffe wie Selbstregulation oder Lesestrategien in Ankündigungen zu Lehrveranstaltungen – auch solchen zum Lesen – nur wenig vorkommen. Gerade in der universitären Lehrerbildung liegt aber, ebenso wie in Lehrerfortbildungen, die Möglichkeit, Lehrer*innen für die förderlichen Maßnahmen selbstregulierten Lesens zu sensibilisieren. Da das Wissen der Lehrpersonen einen wesentlichen Einflussfaktor für das Gelingen der Förderung selbstregulierten Lesens darstellt (Souvignier & Philipp 2016), wäre eine prominentere Verankerung der Thematik in geeigneten Modulen des Lehramtsstudiums für alle Schulformen wünschenswert sowie eine Implementation in übergreifenden Beratungs- und Unterstützungsangeboten zentraler universitärer Einrichtungen sinnvoll.

7 | Literatur

- Anmarkrud, Ø. & Bråten, I. (2012). Naturally-occurring comprehension strategies instruction in 9th-grade language arts classroom. *Scandinavian Journal of Educational Research*, 56(6), 591–623.
- Boekaerts, M. (1999). Self-regulated learning: where we are today. *International Journal of Educational Research*, 31, 445–457.
- Bos, W., Strietholt, R., Goy, M., Stubbe, T. C., Tarelli, I. & Hornberg, S. (2010). *IGLU 2006. Dokumentation der Erhebungsinstrumente*. Waxmann.
- Bräuer, C. (2015). Lesestrategien erlernen, Lesestrategien entwickeln im Unterricht. In: M. Kämper-van den Boogaart & K.-H. Spinner (Hrsg.), *Lese- und Literaturunterricht, DTP 11(3)* (S. 153-196). Schneider.
- Bremerich-Vos, A. (2001). Zum Lehren von Lernstrategien im Umgang mit Texten und mit Lyrik im Besonderen. In: C. Köppert & K. Metzger (Hrsg.), *Entfaltung innerer Kräfte. Blickpunkte der Deutschdidaktik* (S. 149-162). Friedrich.
- Bremerich-Vos, A., Stahns, R., Hußmann, A. & Schurig, M. (2017). Förderung von Leseflüssigkeit und Leseverstehen im Leseunterricht. In: A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valentin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern im internationalen Vergleich* (S. 279-296). Waxmann.
- Dignath, C. & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, 3, 231–264.
- Dole, J. A., Nokes, J. D., & Driets, D. (2009). Cognitive strategy instruction. In: G. G. Duffy & S. E. Israel (Hrsg.), *Handbook on research in reading comprehension* (S. 347–372). Routledge.
- Duke, N. K. & Pearson, P. D. (2002). Effective practices for developing reading comprehension. In: A. E. Farstrup & S. J. Samuels (Hrsg.), *What research has to say about reading instruction* (S. 205–242). International Reading Association.
- Frederking, V. (2013). Literarische Verstehenskompetenz erfassen und fördern. In: S. Gailberger & F. Wietzke (Hrsg.), *Handbuch kompetenzorientierter Deutschunterricht* (S. 117-144). Beltz.
- Frey, A., Taskinen, P., Schütte, K., Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hamman, M., Klieme, E., Pekrun, R. (2009) (Hrsg.). *PISA 2006 Skalenhandbuch. Dokumentation der Erhebungsinstrumente*. Waxmann.
- Helmke, A. (2015). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Kallmeyer.
- Karstens, F., Schmitz, A. & Jost, J. (2019). Gutes Material zur Förderung von selbstreguliertem Lesen von Sachtexten im Deutschunterricht? Analyse exemplarischen Fördermaterials für die Sekundarstufe I unter Einbezug pädagogisch-psychologischer und deutschdidaktischer Perspektiven. *Leseforum*, 1, 1–13.
- Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. University Press.
- Kleinbub, I. (2010). *Unterrichtsqualität im Leseunterricht. Eine videobasierte Analyse in vierten Klassen*. Wissenschaftlicher Verlag.
- Kleinbub, I. (2016). Ein Blick ins Klassenzimmer: deutschdidaktische Unterrichtsforschung am Beispiel der Videostudie VERA – Gute Unterrichtspraxis. In: M. Krelle & W. Senn (Hrsg.), *Qualitäten von Deutschunterricht* (S. 15–36). Klett.

- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54(2), 222–237.
- Kultusministerkonferenz der Länder der Bundesrepublik Deutschland (KMK) (2003). *Bildungsstandards für den mittleren Schulabschluss*. KMK.
- Lankes, E.-M. & Carstensen, C. H. (2007). Der Leseunterricht aus der Sicht der Lehrkräfte. In: W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin. (Hrsg.), *IGLU 2006. Lesekompetenz von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 161-193). Waxmann.
- Leubner, M. & Saupe, A. (2014). *Lesestrategien für die Hypothesenbildung und die Erschließung von Handlungen*. Schneider Verlag Hohengehren.
- Mandl, H. & Friedrich, H. F. (Hrsg.) (2006). *Handbuch Lernstrategien*. Hogrefe.
- Ness, M. (2008). Supporting secondary readers: When teachers provide the “what” and not the “how”. *American Secondary Education*, 37(1), 80–95.
- Okkinga, M., van Steensel, R., van Gelderen, A. J. S., van Schooten, E., Slegers, P. J. C. & Arends, L. R. (2018). Effectiveness of Reading-Strategy Interventions in Whole Classrooms: A Meta-Analysis. *Educational Psychological Review*, 1215–1239.
- Philipp, M. (2012). Einige theoretische und begriffliche Grundlagen. In: M. Philipp & A. Schilcher (Hrsg.), *Selbstreguliertes Lesen. Ein Überblick über wirksame Leseförderansätze* (S. 38–58). Kallmeyer.
- Philipp, M. (2014). Leseunterricht in der Grundschule - vom Ist-Zustand und vom Soll-Zustand. Was Beobachtungsstudien lehren. In: R. Valtin & I. Tarelli (Hrsg.), *Lesekompetenz nachhaltig stärken. Evidenzbasierte Maßnahmen und Programme* (S. 122–165). Deutsche Gesellschaft für Lesen und Schreiben.
- Pressley, M., Wharton-McDonalds, R., Mistretta-Hampston, J., & Echevarria, M. (1998). Literacy instruction in 10 fourth-grade classrooms in upstate New York. *Scientific Studies of Reading*, 2(2), 159–194.
- Rosebrock, C. & Nix, D. (2017). *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung*. Schneider Verlag Hohengehren.
- Schiefele, U. & Pekrun, R. (1996). Psychologische Modelle des fremdgesteuerten und selbstgesteuerten Lernens. In: F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie. Pädagogische Psychologie. Psychologie des Lernens und der Instruktion* (S. 249–278). Hogrefe.
- Schmellentin, C. & Gilg, E. (2016). Lesen im schulischen Fachunterricht. In: A. Bertschi-Kaufmann & T. Graber (Hrsg.), *Lesekompetenz – Leseleistung – Leseförderung. Grundlagen, Modelle und Materialien* (S. 198–212). Klett und Balmer Verlag.
- Schmitz, A. (2020). *Prozesse der Vermittlung von selbstreguliertem Lesen in der regulären Praxis des Deutschunterrichts*. Leseforum.
- Schmitz, A. (2019). Reading instruction in 5th grade. Teachers' perspectives on promoting self-regulated reading in language and content area teaching. *Research in Subject Matter Teaching and Learning*, 2, 16–31.
- Schmitz, A. & Jost, J. (2019). Selbstreguliertes Lesen in der Sekundarstufe I aus der Sicht von Deutschlehrkräften. Ergebnisse einer Unterrichtstagebuch-Studie. In: E. Christophel, M. Hemmer, F. Korneck, T. Leuders & P. Labudde (Hrsg.), *Fachdidaktische Forschung zur Lehrerbildung. Beiträge der fachdidaktischen Forschung* (S. 205-214). Waxmann.
- Schmitz, A., Karstens, F. & Jost, J. (2020). Beobachtung der Vermittlung von selbstreguliertem Lesen im Deutschunterricht mit dem Beobachtungsbogen Lestra-BD. In: K. Mackowiak, C. Beckerle, C. Titz &

- S. Gentrup (Hrsg.), *Forschungsinstrumente im Kontext (schrift-)sprachlicher Bildung* (S. 103–119). Klinkhardt.
- Schmitz, A., Zeuch, N., Karstens, F., Meudt, S.-I., Souvignier, E. & Jost, J. (2021). Leseförderung im Schul- und Unterrichtsalltag implementieren. Erste Erkenntnisse des Evaluationsprojektes BiSS-ELLE. In: S. Gentrup, S. Henschel, K. Schotte, L. Beck & P. Stanat (Hrsg.), *Sprach- und Schriftsprachförderung gestalten: Evaluation von Qualität und Wirksamkeit umgesetzter Konzepte*. Kohlhammer.
- Stahns, R., Rieser, S. & Lankes, E.-M. (2017). Unterrichtsführung, Sozialklima und kognitive Aktivierung im Deutschunterricht in vierten Klassen. In: A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valentin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern im internationalen Vergleich* (S. 251–277). Waxmann.
- Souvignier, E. & Mokhlesgerami, J. (2006). *Using self-regulation as a framework for implementing strategy instruction to foster reading comprehension*. *Learning and Instruction*, 16, 57–71.
- Souvignier, E. & Philipp, M. (2016). Implementation – Begrifflichkeiten, Befunde und Herausforderungen. In: M. Philipp & E. Souvignier (Hrsg.), *Implementation von Lesefördermaßnahmen. Perspektiven auf Gelingensbedingungen und Hindernisse* (S. 9–22). Waxmann.
- Wagner, W., Helmke, A. & Rösner, E. (2009). *DESI. Deutsch Englisch Schülerleistungen International. Dokumentation der Erhebungsinstrumente für Schülerinnen und Schüler, Eltern und Lehrkräfte. Materialien zur Bildungsforschung*. DIPF.
- Weinstein, C. E. & Mayer, R. E. (1986). The teaching of learning strategies. In: M. C. Wittrock (Ed.), *Handbook of research on teaching* (S. 315–325). Macmillan Publishing.
- Willenberg, H. (2004). Lesestrategien. Vermittlung zwischen Eigenständigkeit und Wissen. *Praxis Deutsch*, 31, 6–15.
- Zimmerman (2002). Becoming a self-regulated learner: an overview. *Theory Into Practice*, 41(2), 64–70.

Ruven Stahns, Svenja Rieser & Anke Hußmann

Zur Passung von Leseförderung und Schülerkompetenzen im Leseunterricht in Deutschland

Ergebnisse der Lehrkräftebefragung der Internationalen Grundschul-Lese-Untersuchung 2016

1 | Einleitung

Im Jahr 2016 haben Viertklässler*innen in Deutschland zum vierten Mal an der Internationalen Grundschul-Lese-Untersuchung (IGLU) teilgenommen (Bremerich-Vos, Wendt & Bos 2017). Das Ziel dieser Studie ist ein internationaler Vergleich der Lesekompetenzen von Schüler*innen im Primarstufenalter. Begleitet wird ein zu diesem Zweck durchgeführter Lesetest u.a. durch schriftliche Befragungen der Lehrkräfte, deren Klassen an IGLU teilnehmen (Hußmann et al. 2017). Die gewonnenen Daten liefern u.a. Informationen darüber, welche Maßnahmen zur Förderung von Lesekompetenz im Leseunterricht realisiert werden.

Im vorliegenden Beitrag werden auf der Grundlage dieser Befragung gewonnene Befunde dazu präsentiert, wie häufig im Leseunterricht am Ende der Grundschulzeit in Deutschland ausgewählte Lesestrategien instruiert werden und wie Stillesephasen gestaltet werden. Dabei wird überprüft, ob zwischen den Angaben der Lehrkräfte hierzu und dem Anteil an schwachen Leser*innen in den Schulklassen Zusammenhänge bestehen.

Für eine effektive Leseförderung müssen Lehrkräfte den Lernstand ihrer Schüler*innen einschätzen, um fundierte Entscheidungen für Verfahren zur Leseförderung treffen zu können. Dabei orientieren sie sich im besten Fall an empirischen Befunden zu deren Wirksamkeit (Rosebrock & Nix 2017: 9f.). Für Kinder, die noch nicht flüssig lesen können, wird es z.B. wenig förderlich sein, wenn Lehrkräfte ihnen anspruchsvolle Lesestrategien vorstellen oder sie im Unterricht ohne Unterstützung Kinder- oder Jugendbücher lesen lassen (Abschnitt 3). Die Ergebnisse der vorliegenden Untersuchung sollen dazu beitragen, die Frage zu klären, inwiefern Lehrkräfte in Deutschland bei der Gestaltung des Leseunterrichts den Lernstand ihrer Schüler*innen berücksichtigen.

Im Folgenden werden einige Ergebnisse von IGLU 2016 skizziert (Abschnitt 2) und die in der Untersuchung berücksichtigten Lesefördermaßnahmen eingeführt (Abschnitt 3). Im Anschluss werden die Forschungsfragen (Abschnitt 4) sowie die Datengrundlage und das methodische Vorgehen (Abschnitt 5) vorgestellt. Schließlich werden die Ergebnisse der Untersuchung präsentiert (Abschnitt 6) und diskutiert (Abschnitt 7).

2 | Lesekompetenzen und Leseförderung am Ende der Grundschulzeit in Deutschland: ausgewählte Ergebnisse von IGLU 2016

IGLU ist eine Leseleistungsstudie, an der in der Erhebungsrunde 2016 Schüler*innen aus 57 Ländern und Regionen teilgenommen haben. In Deutschland wird die Studie am Ende des vierten Schuljahres durchgeführt. Im Jahr 2016 haben in Deutschland 3959 Schüler*innen aus 208 Klassen an IGLU teilgenommen (Hußmann et al. 2017). Im Mittelpunkt der Studie steht ein Lesetest, für den die Teilnehmer*innen eines von 16 Testheften bearbeiten, das einen Erzähltext, einen informierenden Text sowie Testitems mit Fragen zu den beiden Texten enthält (Bremerich-Vos, Wendt & Bos 2017). Steht zu Beginn der Grundschulzeit der Erwerb von Lesefertigkeiten auf der Wort- und Satzebene im Mittelpunkt, sollte im Verlauf der Grundschulzeit die Auseinandersetzung mit zunehmend längeren und komplexeren Texten möglich werden. Mitunter wird das als ein Weg vom „learning to read“ zum „reading to learn“ beschrieben (Rosebrock & Gold 2018: 13). Im Lesetest von IGLU, der auf die Lesekompetenz von Schüler*innen am Ende der Grundschulzeit ausgerichtet ist, werden basale Lesefertigkeiten vorausgesetzt, daher werden mehr oder weniger hierarchiehohe Prozesse des Textverstehens überprüft (Bremerich-Vos, Wendt & Bos 2017: 82ff.). Allerdings zeigen Forschungsbefunde der letzten Jahre, dass auch zum Ende der Grundschulzeit und darüber hinaus hierarchieniedrige Leseprozesse nicht von allen Schüler*innen flüssig realisiert werden (Rosebrock & Nix 2017: 34ff.; Gailberger et al. in diesem Band). Daher wurden in IGLU 2016 zusätzlich zu dem international eingesetzten textorientierten Lesetest der Wort- und Satzverständnistest von ELFE II (Lenhard, Lenhard & Schneider 2017) durchgeführt, um die Lesekompetenzen sehr schwacher Leser*innen beschreiben zu können.

Die Leistungswerte von IGLU ergeben sich aus der Berechnung von Plausible Values (Hußmann et al. 2017: 62ff.). Der Leistungsmittelwert der Schüler*innen in Deutschland liegt bei IGLU 2016 bei 537 Punkten (*SD*: 78) und damit signifikant oberhalb des internationalen Mittelwerts (*M*: 521; *SD*: 78) und nicht signifikant unterhalb des Mittelwerts der EU-Teilnehmerländer bzw. -regionen (*M*: 540; *SD*: 71) (vgl. dazu und zum Folgenden Bremerich-Vos, Wendt & Bos 2017: 91ff.). Im Vergleich zu den Ergebnissen der vorherigen IGLU-Erhebungen zeigt sich in IGLU 2016 eine leichte (nicht signifikante) Verringerung der mittleren Leseleistungen in Deutschland. Um die Leseleistungen der Viertklässler*innen, die an IGLU teilnehmen, inhaltlich zu bestimmen, wird in jeder Erhebungsrunde der Studie ein Kompetenzstufenmodell entwickelt. Die Beschreibung der Kompetenzstufen wird vorgenommen, indem die Leseanforderungen der Aufgaben ermittelt werden, die von Kindern, die Leseleistungen auf den jeweiligen Stufen erbringen, mit einer bestimmten Wahrscheinlichkeit gelöst werden. Etwa sechs Prozent der Schüler*innen in Deutschland erreichen im Jahr 2016 lediglich Stufe I des verwendeten Lesekompetenzstufenmodells. Auch die Ergebnisse von ELFE II weisen auf erhebliche Defizite dieser Schüler*innen bei den hierarchieniedrigen Leseprozessen hin: Im Mittel erreichen sie 41,8 von 75 Punkten auf der Wort- und 18,8 von 36 Punkten auf der Satzebene. Defizite auf einer immer noch eher hierarchieniedrigen Ebene haben auch die Schüler*innen auf Kompetenzstufe II (13 Prozent der Viertklässler*innen; Wortverständnis: 50,5 Punkte, Satzverständnis: 19,6 Punkte). Bedarf an Förderung zeigt sich für diese Kinder also bereits im Bereich hierarchieniedriger Leseprozesse. Die Schüler*innen, deren Leistungen auf Kompetenzstufe III einzuordnen sind, sind ebenfalls noch nicht dazu in der Lage, durchgängig stimmige mentale Modelle von altersgemäßen Texten auszubilden. Da ihre Schwierigkeiten aber eher auf einer hierarchiehöheren Ebene

des Leseprozesses liegen, müsste die Leseförderung anders aussehen als die Förderung der Kinder auf den Stufen I und II. Wiederum anders müsste die Förderung von Kindern auf den Stufen IV und V des Lesekompetenzstufenmodells gestaltet werden. Sie sind gute bzw. ausgesprochen gute Leser*innen, für die spezifische Förderangebote im Leseunterricht zu gestalten wären.

Die Befunde aus IGLU 2016 machen deutlich, dass auch noch am Ende der Grundschulzeit die Förderung der Leseflüssigkeit neben der Förderung des Textverstehens ein Ziel des Leseunterrichts ist. Ausgehend vom Lernstand der Schüler*innen muss eine effektive Förderung der unterschiedlichen Aspekte von Lesekompetenz mit unterschiedlichen Maßnahmen erfolgen (zusammenfassend NICHHD 2000; Rosebrock & Nix 2017). Wenn die Förderung hierarchieniedriger Lesekompetenzen resp. der Leseflüssigkeit notwendig ist, ist auf Grundlage des aktuellen Forschungsstands der Einsatz unterschiedlicher Varianten begleitenden und wiederholenden laut-lesens mit (Peer-)Feedback zu empfehlen – im deutschsprachigen Raum wurden in den letzten Jahren vor allem Effekte der Förderung in Form von Lautlesetandems untersucht (Rosebrock et al. 2010; Walter, die & Petersen 2012; Müller et al. 2013; Lauer-Schmaltz, Rosebrock & Gold 2014; Munser-Kiefer 2014). Befunde einer Befragung von Lehrkräften, die im Rahmen von IGLU 2016 durchgeführt wurde, deuten darauf hin, dass Lautlesetandems im Unterricht in vierten Klassen seltener realisiert werden als andere Verfahren, deren förderliche Wirkung fraglich ist, z.B. Reihumlesen (Bremerich-Vos et al. 2017: 283ff.). Demnach werden 44 Prozent der Schüler*innen in Deutschland von Lehrkräften unterrichtet, die angeben, in ihrem Leseunterricht niemals Tandems aus stärkeren und schwächeren Schüler*innen einen Text gemeinsam halblaut lesen zu lassen (ebd.: 287). Es gibt aber den Ergebnissen von IGLU 2016 zufolge in fast allen vierten Klassen in Deutschland Leser*innen, deren Leistungen auf den Stufen I oder II des Kompetenzstufenmodells von IGLU einzuordnen sind und die damit zur „Risikogruppe“ im Lesen gehören. Da Lautlesetandems sich vor allem für die Förderung von Schüler*innen eignen, die noch nicht flüssig lesen können, ist zu erwarten, dass sie vor allem im Unterricht von Lehrkräften eingesetzt werden, die in Klassen mit einem hohen Anteil schwacher Leser*innen arbeiten. Das aber zeigt sich in den Ergebnissen von IGLU 2016 nicht (ebd.: 289). Die betreffenden Befunde weisen darauf hin, dass Leselehrkräfte im Unterricht nicht unbedingt umsetzen, was auf Grundlage des Forschungsstands zur effektiven Leseförderung sinnvoll wäre (ebd.: 295).

3 | Förderung von Lesekompetenzen durch Vermittlung von Lesestrategien und stilles Lesen

Im Rahmen von IGLU 2016 wurden neben Daten zur Realisierung von Lautlesetandems auch solche zur Vermittlung von Lesestrategien und zur Gestaltung von Stillelesephasen im Leseunterricht erhoben.

Schüler*innen den Einsatz von Lesestrategien zu vermitteln, wird in der einschlägigen Forschung empfohlen, wenn es das Ziel ist, hierarchiehöhere Leseprozesse zu fördern (z.B. Munser-Kiefer 2012; Munser-Kiefer 2014; Rosebrock & Nix 2017; Souvignier 2014). Allerdings gibt es durchaus Lesestrategien, die das Verstehen auf einer relativ hierarchieniedrigen Ebene wie der lokalen Kohärenzbildung unterstützen (Rosebrock & Nix 2017: 73).

Verstehendes Lesen setzt eine kognitive Auseinandersetzung mit einem Text voraus: Die „komplexeren mentalen Handlungen, die zum Textverstehen führen“ (Rosebrock & Nix 2017: 74), können als Lesestrategien bezeichnet werden. Kompetente Leser*innen nutzen Lesestrategien häufig unbewusst, sie sind dem bewussten Zugriff allerdings zugänglich und ihr Einsatz gilt als lehr- und lernbar (z.B. Palincsar & Brown 1984; Kruse et al. 2013; Schulte & Souvignier 2013). Dabei ist von den Lehrkräften zu vermitteln, wie aus einer Lesetechnik (wie dem Unterstreichen) eine Lesestrategie wird. Mit Rosebrock und Nix (2017: 85ff.) dürfen Lehrkräfte Schüler*innen nicht nur Lesetechniken vorstellen, wenn diese zum kompetenten Umgang mit Lesestrategien befähigt werden sollen. Die Lernenden müssten die Auswahl und den Einsatz geeigneter Lesetechniken für die Auseinandersetzung mit bestimmten Texten und Arbeitsaufträgen erlernen. Dabei dürfe nicht nur die Entwicklung kognitiver Strategien vorangetrieben werden: Um abzuschätzen, ob der Einsatz einer Strategie die Textrezeption verbessert, seien metakognitive Strategien zu entwickeln, die ein Monitoring des Strategieeinsatzes ermöglichen (vgl. dazu und zum Folgenden z.B. Artelt, Naumann & Schneider 2010: 78; Rosebrock & Nix 2017: 80ff.).

- Zu den kognitiven Strategien gehören Wiederholungsstrategien (z.B. mehrmalige Lektüre eines Textes, wörtliche Zusammenfassungen des Textes), Elaborationsstrategien (z.B. Vorausagen über den Inhalt, Textvergleiche, die Zusammenfassung von Textabschnitten in eigenen Worten) und Organisations- bzw. Reduktionsstrategien (Markieren von als zentral erkannten Textteilen, Zusammenfassungen von Textabschnitten in wenigen Sätzen).
- Metakognitive Strategien schließlich beziehen sich auf den Einsatz der kognitiven Lesestrategien im Leseprozess selbst, indem sie diesen quasi zielorientiert überwachen (z.B. in der Planung, Durchführung und Sicherung des Leseprozesses und seines mentalen Ergebnisses).

Neben der Vermittlung von Lesestrategien wird auch die förderliche Wirkung von Formen ‚offenen‘ Unterrichts auf die Lesekompetenz diskutiert. Häufig wird dieser Unterricht in Form von Stillesephasen gedacht, in der z.B. Kinder- und Jugendliteratur gelesen wird (Bertschi-Kaufmann & Schneider 2006; Schneider & Bertschi-Kaufmann 2007; Rosebrock et al. 2010; Rieckmann 2020). Dabei können Lesemotivation oder Leseselbstkonzept im Fokus stehen – und damit Aspekte der „Subjektebene“ (Rosebrock & Nix 2017: 16) von Lesekompetenz. Es kann aber auch ein Zusammenhang zwischen der stillen Lektüre und der Verbesserung der Leseflüssigkeit und des Textverstehens hergestellt werden (Rosebrock et al. 2010). Dabei wird von einem Einfluss der gelesenen Textmenge auf die Lesekompetenzen auf der Prozessebene ausgegangen (ebd.: 36). Hinsichtlich der förderlichen Effekte der Stilleseverfahren stehen häufig schwache Leser*innen im Mittelpunkt, die kaum außerschulische Erfahrungen mit Buchlektüre gemacht haben (Rosebrock & Nix 2014: 58ff.). Potenziell sei die Wirksamkeit häufigen stillen Lesens im Unterricht allerdings nicht auf diese Schüler*innen beschränkt, da stärkere ihre Kompetenzen durch die Lektüre anspruchsvoller Texte ebenfalls weiterentwickeln könnten (ebd.).

Einen Vorschlag zur Gestaltung von Stillesephasen im Unterricht stellt das Kilometer-Lesen dar (Rosebrock & Nix 2017: 60f.). Im Mittelpunkt steht dabei stilles Lesen im Unterricht, für das die Schüler*innen selbstständig eine Auswahl von Kinder- und Jugendbüchern aus einer Bücherkiste vornehmen. Die gelesene Textmenge wird in Meter umgerechnet und addiert; die gelesenen Texte werden in einen „Lese-Reise-Pass“ (ebd.: 61) eingetragen. Die Texte werden nicht im Unterrichtsgespräch aufgearbeitet und es sind auch keine Aufgaben zum Textverständnis zu

bearbeiten. Rosebrock, Rieckmann, Nix und Gold (2010) haben die Effekte des Kilometer-Lesens auf die Entwicklung unterschiedlicher Aspekte der Lesekompetenz bei 12-jährigen Hauptschüler*innen untersucht. Hinsichtlich der Entwicklung von Leseflüssigkeit und Textverstehen bleiben die Schüler*innen, die am Kilometer-Lesen teilgenommen haben, hinter denen zurück, die in Form der Lautlesetandems gefördert wurden. Hinsichtlich der Entwicklung der Lesemotivation und des Leseselbstkonzepts zeigen sich sogar negative Veränderungen zwischen Prätest und Follow-Up-Test. Die Autor*innen gehen davon aus, dass das Kilometer-Lesen voraussetzungsreicher als die Arbeit in Lautlesetandems ist: Während für die Arbeit in Lautlesetandems maximal lokale Kohärenz herzustellen sei, setze effektive Förderung beim Kilometer-Lesen globale Kohärenzbildung „als Bedingung seiner Wirksamkeit“ (ebd.: 51) voraus, weil sich nur dann „Lesegenuss“ (ebd.: 50) einstellen könne. Der Effekt auf der Subjektebene sei wiederum Voraussetzung dafür, dass es zu einer Verbesserung auf der Prozessebene komme.

Weitere Befunde zum stillen Lesen stammen von Bertschi-Kaufmann und Schneider (2006), die Ergebnisse einer Interventionsstudie vorgelegt haben, in der u.a. Effekte der Leseförderung im Rahmen offenen, interessen geleiteten Leseunterrichts überprüft werden. Im Mittelpunkt dieses Unterrichts steht die individuelle Auseinandersetzung von Schüler*innen mit Kinder- und Jugendliteratur, wobei die Kinder frei in ihren Entscheidungen hinsichtlich der Auswahl von Texten für Lesestunden und der Anschlusskommunikation sind. Für die Primarstufe (3. bzw. 4. Klasse) lässt sich ein positiver Effekt des offenen Leseunterrichts auf die „Entwicklung der Erinnerung an Gelesenes“ (ebd.: 415) nachweisen. Darin sehen Bertschi-Kaufmann und Schneider einen Beleg dafür, dass die „Habitualisierung des Lesens“ (ebd.: 402) sowie das „emotionale Involvement“ (ebd.) im offenen Unterricht tendenziell stärker gefördert werden als im Unterricht einer Kontrollgruppe. Ein Effekt auf unterschiedliche Aspekte von Lesemotivation zeigt sich im Vergleich von Treatment- und Kontrollgruppe in einer späteren Publikation aus dem Projekt nicht (Schneider & Bertschi-Kaufmann 2007). Für weitere Aspekte von Lesekompetenz (Leseflüssigkeit, Leseverstehen) zeigen sich im Vergleich mit dem Entwicklungsverlauf der Kontrollgruppe nicht nur keine positiven, sondern auch hier z.T. negative Effekte (Bertschi-Kaufmann & Schneider 2006). Zudem weisen Schneider und Bertschi-Kaufmann (2007) eine positivere Entwicklung der Leseflüssigkeit in der Kontrollgruppe als in der Treatmentgruppe nach.

4 | Forschungsfragen

Lehrkräfte stehen vor der Herausforderung, unterschiedliche Leistungsniveaus von Schüler*innen einer Klasse einzuschätzen und durch passende Angebote zu fördern. In den folgenden Analysen wird der Frage nachgegangen, ob die Gestaltung des Leseunterrichts mit dem Anteil an schwachen Leser*innen in einer Klasse zusammenhängt. Wir knüpfen an die Vorarbeit von Bremerich-Vos, Stahns, Hußmann und Schurig (2017) an, die überprüft haben, ob ein Zusammenhang zwischen der Realisierung von Lautlesetandems im Unterricht und dem Anteil schwacher Leser*innen in einer Klasse besteht. In der vorliegenden Untersuchung wird geprüft, ob sich für die Instruktion von Lesestrategien und die Gestaltung von Stillelesephasen im Unterricht Unterschiede zwischen Klassen mit einem hohen und einem mittleren bzw. niedrigen Anteil schwacher Leser*innen zeigen.

Die Vermittlung von Lesestrategien wird meist zur Förderung hierarchiehöherer Leseprozesse empfohlen. Daher könnte davon ausgegangen werden, dass die Vermittlung von Lesestrategien

seltener in Klassen mit vielen (sehr) schwachen Leser*innen erfolgt, die sich weniger mit dem Textverstehen auseinandersetzen müssen bzw. können als mit der Leseflüssigkeit. Da es Lesestrategien gibt, die den Verstehensprozess auf einer relativ hierarchieniedrigen Ebene unterstützen, ist es nicht entscheidend, ob sich der Unterricht in Klassen mit vielen und wenigen schwachen Leser*innen hinsichtlich der Instruktion von Lesestrategien insgesamt unterscheidet. Von Bedeutung ist, ob den Schüler*innen die Lesestrategien vermittelt werden, die aufgrund ihrer Voraussetzungen angemessen erscheinen, und ob sich das in Unterschieden zwischen den Angaben der Lehrkräfte zur Vermittlung von Strategien widerspiegelt. Dazu gehört auch, dass in Klassen mit einem niedrigen Anteil schwacher Leser*innen intensiver anspruchsvolle Lesestrategien vermittelt werden, die auf die Unterstützung hierarchiehöherer Verstehensprozesse zielen.

Die Befunde zur Wirksamkeit der Stillleseverfahren stimmen insgesamt skeptisch, was deren förderliche Wirkung auf die unterschiedlichen Ebenen und Aspekte von Lesekompetenz angeht. Vermutlich bringen eher die starken Leser*innen die Voraussetzungen mit, um vom stillen Lesen im Unterricht zu profitieren. Rosebrock und Nix (2017: 65ff.) plädieren für ein Verständnis von Wirksamkeit, das auch einbezieht, dass leseschwache Schüler*innen an die Buchlektüre herangeführt werden. Wenn Lehrkräfte diese Einschätzung teilen, könnten sie Stillleseverfahren wie das Kilometer-Lesen in Klassen mit vielen schwachen Leser*innen einsetzen.

5 | Daten und Methoden

Die Analysen basieren auf Daten aus dem Leistungstest und der Lehrkräftebefragung von IGLU 2016. In einem ersten Auswertungsschritt werden die Angaben zur Umsetzung von Leseförderverfahren von Lehrkräften mit einem hohen Anteil schwacher Leser*innen einerseits und einem mittleren bzw. niedrigen Anteil schwacher Leser*innen andererseits gegenübergestellt. Im zweiten Schritt werden Korrelationen zwischen den Angaben zum Einsatz von Leseförderverfahren und dem Anteil schwacher Leser*innen in der Klasse berechnet. Der Anteil schwacher Leser*innen wird auf Grundlage der Ergebnisse des Lesetestes aus IGLU 2016 bestimmt.¹ Zudem werden die Analysen jeweils mit Angaben zum Anteil der Schüler*innen mit Bedarf an Förderunterricht im Lesen wiederholt, die von den Lehrkräften stammen, die an der Lehrkräftebefragung von IGLU 2016 teilgenommen haben.² Lehrkräfte planen ihren Unterricht vermutlich eher auf der Grundlage ihrer Einschätzung des Leistungsniveaus der Schüler*innen als auf der Grundlage der Ergebnisse von standardisierten Leistungstests. Da Lehrkräfteeinschätzungen und Testergebnisse nicht übereinstimmen müssen, werden die Analysen in den beiden genannten Varianten durchgeführt.

*Zum ermittelten Anteil schwacher Leser*innen*

Als schwache Leser*innen werden in diesem Sinne die Kinder betrachtet, die im IGLU-Lesetest 475 Punkte oder weniger erreicht haben und deren Leistungen somit höchstens auf der Kompetenzstufe II des Stufenmodells von IGLU 2016 zu verorten sind (Bremerich-Vos, Wendt & Bos 2017: 92). Für jede Klasse wird bestimmt, welcher Anteil an Schüler*innen dieser Gruppe

¹ Im Folgenden wird vom „ermittelten“ Anteil schwacher Leser*innen gesprochen, auch wenn die Lesekompetenzwerte der einzelnen Schüler*innen auf einer Schätzung von Plausible Values beruhen.

² Im Folgenden wird vom „geschätzten“ Anteil schwacher Leser*innen gesprochen, um die Unterscheidung von den „ermittelten“ Werten zu erleichtern.

angehört. Der durchschnittliche Anteil schwacher Leser*innen in den Klassen liegt bei $M = 20,15$ Prozent ($SD: 21,42$ Prozent). Der Anteil rangiert zwischen den Klassen von 0 bis 100 Prozent (Abb. 1). Um die Einteilung der Klassen in Gruppen mit einem hohen bzw. mittleren bis niedrigen Anteil schwacher Leser*innen für die vorliegende Untersuchung vorzunehmen, werden die Leistungsdaten der Stichprobe anhand der dritten Quartilsgrenze in zwei Gruppen geteilt. Eine Gruppe umfasst alle Klassen, in denen der aufgrund des IGLU-Tests ermittelte Anteil schwacher Leser*innen mindestens bei 26,36 Prozent liegt (dritte Quartilsgrenze). Hier wird von einem hohen Anteil schwacher Leser*innen ausgegangen. Die anderen Klassen gehören zur Gruppe mit einem mittleren oder niedrigen Anteil schwacher Leser*innen.

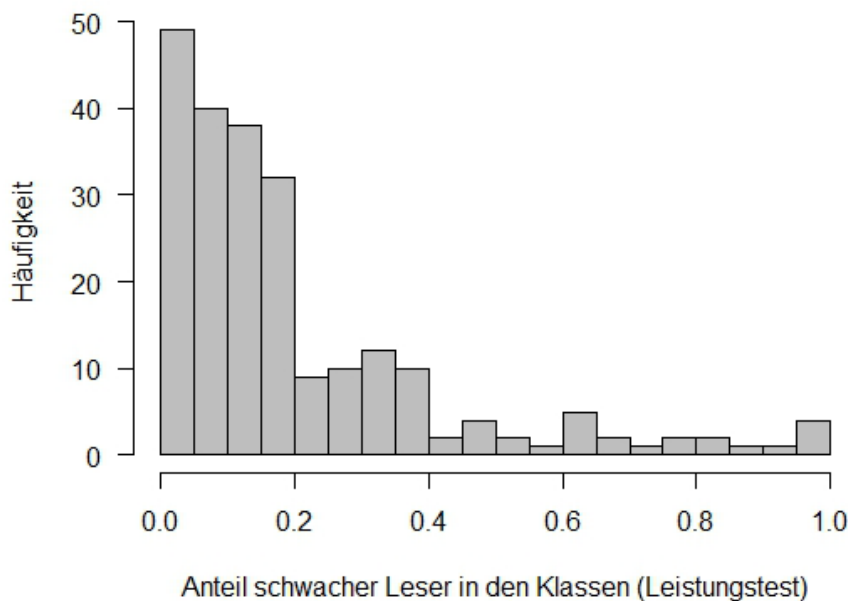


Abb. 1: Häufigkeitsverteilung des Anteils schwacher Leser*innen in einer Klasse auf Grundlage des Leistungstests (Histogramm)

*Zum geschätzten Anteil schwacher Leser*innen*

Insgesamt haben in Deutschland 227 Lehrkräfte an IGLU 2016 teilgenommen. Im Lehrkräftefragebogen wurden sie um Angaben zur Klassengröße und um eine Einschätzung zur Anzahl der Kinder in ihren Klassen gebeten, die Förderunterricht im Lesen benötigen („*Wie viele Schüler benötigen Förderunterricht im Lesen?* Anmerkung: *bezieht sich auf die an IGLU teilnehmende Klasse*“). Auf Grundlage dieser Angaben wird der geschätzte Anteil schwacher Leser*innen berechnet. Dazu wird die von den Lehrkräften geschätzte Zahl der Kinder mit Förderbedarf durch die Klassengröße dividiert. Die Werte liegen zwischen 0 und 100 Prozent (Abb. 2). Der Mittelwert liegt mit $M = 17,5$ Prozent ($SD: 16,37$) etwas niedriger als der ermittelte Anteil leseschwacher Kinder. Um zu prüfen, ob die Gestaltung des Unterrichts in Abhängigkeit von dem durch die Lehrkräfte geschätzten Anteil schwacher Leser*innen variiert, wird ebenfalls eine Einteilung in Klassen mit einem hohen bzw. mittleren bis niedrigen geschätzten Anteil schwacher Leser*innen anhand der dritten Quartilsgrenze vorgenommen. Diese liegt für den geschätzten Anteil an Leser*innen mit Förderbedarf bei 24,0 Prozent. Daher werden alle Klassen mit einem

Anteil ab 24 Prozent an Schüler*innen mit Bedarf an Förderunterricht im Lesen der Gruppe mit einem hohen geschätzten Anteil schwacher Leser*innen zugeordnet, die übrigen Klassen der Gruppe mit einem mittleren bzw. niedrigen Anteil.

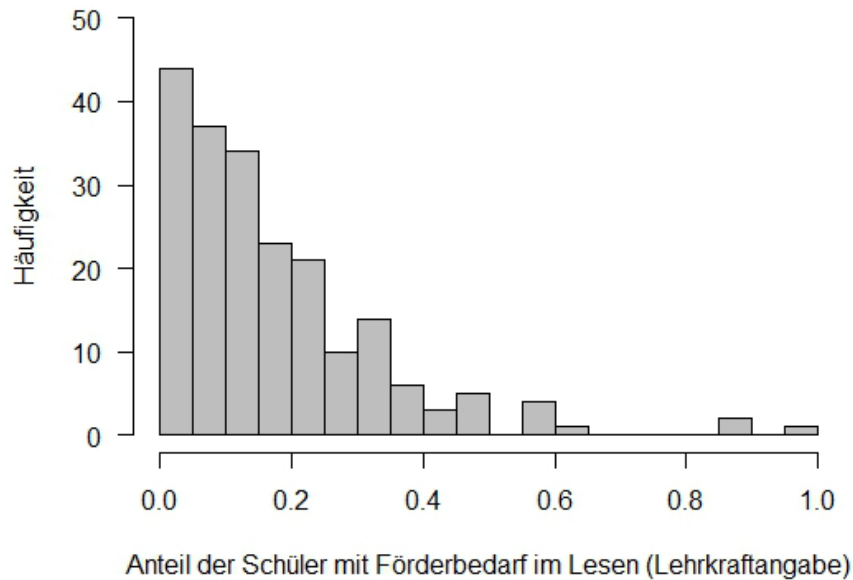


Abb. 2: Häufigkeitsverteilung des Anteils der Schüler*innen in einer Klasse, die nach Einschätzung der Lehrkräfte Förderunterricht im Lesen benötigen (Histogramm)

Die von den Lehrkräften geschätzten Anteile schwacher Leser*innen und die aufgrund des IGLU-Tests ermittelten Anteile korrelieren zu $r = .54^*$ miteinander.

Der Lehrkräftefragebogen von IGLU 2016 enthält 13 Items, mit denen erfragt wird, wie häufig die Lehrkräfte verschiedene Lesestrategien erklären und zeigen bzw. ob sie vermitteln, wie und wann Lesestrategien eingesetzt werden sollten. Um zu erfassen, wie Stillesephasen im Unterricht gestaltet werden, werden die Lehrkräfte im Lehrkräftefragebogen gebeten anzugeben, wie häufig sie den Kindern im Unterricht Zeit zum stillen Lesen einräumen. Von den im Rahmen von IGLU 2016 befragten Lehrkräften haben 22 diese Frage nicht beantwortet. Ein geringer Anteil an Lehrkräften räumt den Schüler*innen niemals Zeit zum stillen Lesen ein ($n = 3$). Die übrigen Lehrkräfte ($n=202$) haben vier weitere Items bearbeitet, in denen sie um Angaben zur Auswahl und dem Umgang mit den Texten gebeten werden, die in den Stillesephasen gelesen werden. Alle Items sind auf einer vierstufigen Skala (von „jeden oder fast jeden Tag“ bis „nie oder fast nie“) zu beantworten. Sie werden für die Analysen so kodiert, dass ein hoher Wert für eine häufige Umsetzung steht.

Im Folgenden werden getrennt für die Gruppen die Mittelwerte und Standardabweichungen für die einzelnen Items berichtet. Die Analysen dienen dazu, zu zeigen, wie häufig die Strategien instruiert werden bzw. wie Stillesephasen in den Klassen gestaltet werden. Die statistische Signifikanz der Unterschiede wird mit T-Tests für unabhängige Stichproben geprüft. In einem zwei-

ten Auswertungsschritt werden Korrelationen zwischen den Angaben zur Unterrichtsgestaltung und dem (ermittelten und geschätzten) Anteil schwacher Leser*innen in der Klasse berechnet. Alle Analysen werden mit SPSS 25 durchgeführt.

6 | Ergebnisse

6.1 | Unterrichtsgestaltung in Abhängigkeit vom ermittelten Anteil schwacher Leser*innen

Vermittlung von Lesestrategien

In Tabelle 1 sind die deskriptiven Ergebnisse für die Instruktion von Lesestrategien zu finden. Die Lehrkräfte nutzen die vierstufige Antwortskala für alle Items aus. In beiden Gruppen liegen die Mittelwerte für die Mehrheit der Items nahe am theoretischen Mittelwert von 2,5. Allerdings werden einige relativ anspruchsvolle Lesestrategien, die sich auf hierarchiehöhere Verstehensleistungen beziehen, in beiden Gruppen seltener thematisiert (Vergleich des Textaufbaus, Verknüpfung von Informationen im Text über Abschnittsgrenzen hinweg oder Übertragen wichtiger Informationen in ein anderes Darstellungsformat). In beiden Gruppen erklären und zeigen die Lehrkräfte besonders häufig, wie die Schüler*innen wichtige Stellen in einem Text markieren können. Das passiert etwas häufiger in Klassen mit einem mittleren bzw. niedrigen Anteil schwacher Leser*innen.

Zudem wird im Unterricht in beiden Gruppen häufig erklärt und gezeigt, wie die Schüler*innen prüfen können, ob ihnen alle Wortbedeutungen in einem Text bekannt sind. Diese Strategie wird etwas häufiger in Klassen mit vielen schwachen Leser*innen thematisiert. Den Schüler*innen beider Gruppen wird relativ häufig vermittelt, wie sie die zentralen Ideen von Texten zusammenfassen können. Zudem wird in Klassen mit einem hohen Anteil schwacher Leser*innen häufiger darüber gesprochen, wann Lesestrategien eingesetzt werden sollten. Die Unterschiede zwischen den Mittelwerten der beiden Gruppen fallen aber überwiegend gering aus. Der Vergleich der Mittelwerte durch T-Tests zeigt keine signifikanten Unterschiede zwischen den Gruppen.

| Wie oft tun Sie im Leseunterricht die folgenden Dinge? | Klassen mit einem hohen Anteil leistungsschwacher Leser*innen | | Klassen mit einem mittleren/niedrigen Anteil leistungsschwacher Leser*innen | | Mittelwertvergleich (T-Test bei unabhängigen Stichproben) | | |
|--|---|------|---|------|---|-----|------|
| | M | SD | M | SD | t | df | p |
| Den Schüler*innen erklären und zeigen, ... | | | | | | | |
| wie man verschiedene Lesestrategien einsetzt. | 2,35 | 0,74 | 2,42 | 0,76 | 0,589 | 196 | .557 |
| wann man verschiedene Lesestrategien einsetzt. | 2,19 | 0,69 | 2,34 | 0,73 | 1,238 | 196 | .217 |
| wie man Fragen zum Text entwickelt, während man liest. | 2,12 | 0,76 | 2,16 | 0,76 | 0,334 | 197 | .739 |
| wie man zentrale Ideen eines Textes zusammenfasst. | 2,54 | 0,70 | 2,63 | 0,71 | 0,787 | 198 | .432 |
| wie sie den Textinhalt mit ihrem Vorwissen verbinden können. | 2,42 | 0,73 | 2,50 | 0,74 | 0,661 | 196 | .509 |
| wie man wichtige Textstellen markiert (z. B. durch Unterstreichen). | 2,85 | 0,72 | 2,99 | 0,68 | 1,326 | 198 | .186 |
| dass man wichtige Textstellen manchmal wörtlich abschreiben sollte. | 2,15 | 0,89 | 2,09 | 0,78 | -0,504 | 198 | .615 |
| wie sie während des Lesens Voraussagen über den weiteren Verlauf des Textes anstellen. | 2,13 | 0,71 | 2,07 | 0,71 | -0,564 | 198 | .573 |
| wie sie überprüfen, ob ihnen alle Wortbedeutungen bekannt sind. | 2,69 | 0,96 | 2,55 | 0,94 | -,923 | 197 | .357 |
| wie sie den Textaufbau mit dem von anderen Texten vergleichen. | 1,71 | 0,78 | 1,80 | 0,68 | 0,816 | 198 | .416 |
| wie sie Informationen im Text über Abschnittsgrenzen hinweg verknüpfen. | 1,76 | 0,84 | 1,87 | 0,80 | 0,780 | 193 | .436 |
| wie sie Überschriften zu Textabschnitten formulieren. | 2,27 | 0,66 | 2,34 | 0,69 | 0,665 | 196 | .507 |
| wie sie wichtige Informationen in ein anderes Darstellungsformat überführen (z. B. Mind-Map, Tabelle, Diagramm). | 1,85 | 0,74 | 1,97 | 0,66 | 1,134 | 199 | .258 |

Tab. 1: Deskriptive Werte zur Instruktion von Lesestrategien in Klassen mit einem hohen bzw. mittleren bis niedrigen Anteil schwacher Leser*innen sowie die Ergebnisse der Überprüfung der statistischen Bedeutsamkeit der Mittelwertsunterschiede mittels T-Tests bei unabhängigen Stichproben.

Gestaltung von Stillesephasen

In Tabelle 2 finden sich die deskriptiven Werte zur Gestaltung von Stillesephasen. Die Texte, die in den Stillesephasen von allen gelesen werden, werden in allen untersuchten Klassen im Anschluss an die Lektüre relativ häufig Gegenstand des Unterrichtsgesprächs. Seltener müssen die Kinder nachhalten, wie viel sie tatsächlich gelesen haben. Lehrkräfte in Klassen mit einem hohen Anteil schwacher Leser*innen geben an, häufiger Texte für die Stillesephasen auszuwählen. In den anderen Klassen übernehmen das Auswählen häufiger die Kinder. Auch hier weisen die T-Tests keine signifikanten Unterschiede zwischen den Gruppen aus.

| In meinem Unterricht haben Schüler*innen Zeit, still zu lesen. Falls ja: Wie häufig kommt Folgendes beim stillen Lesen vor? | Klassen mit einem hohen Anteil leistungsschwacher Leser*innen | | Klassen mit einem mittleren/niedrigen Anteil leistungsschwacher Leser*innen | | Mittelwertvergleich (T-Test bei unabhängigen Stichproben) | | |
|---|---|------|---|------|---|-----|------|
| | M | SD | M | SD | t | df | p |
| Ich übernehme die Auswahl der Texte für Stillesephasen. | 2,60 | 0,96 | 2,44 | 0,95 | -0,958 | 174 | .340 |
| Die Schüler*innen können die Texte nach eigenen Interessen auswählen. | 2,53 | 0,86 | 2,71 | 0,94 | 1,118 | 178 | .265 |
| Texte, die alle Schüler*innen gelesen haben, werden nach der Lektüre Gegenstand des Unterrichtsgesprächs. | 2,81 | 0,94 | 2,77 | 0,88 | -0,252 | 179 | .802 |
| Die Schüler*innen müssen nachhalten, wie viel sie gelesen haben (z. B. durch ein Lesetagebuch). | 1,59 | 0,90 | 1,59 | 0,93 | 0,018 | 170 | .986 |

Tab. 2: Deskriptive Werte zur Häufigkeit von Gestaltungsmerkmalen von Stillesephasen in Klassen mit einem hohen bzw. mittleren bis niedrigen Anteil schwacher Leser*innen sowie die Ergebnisse der Überprüfung der statistischen Bedeutsamkeit der Mittelwertsunterschiede mittels T-Tests bei unabhängigen Stichproben.

6.2 | Unterrichtsgestaltung in Abhängigkeit vom geschätzten Anteil schwacher Leser*innen

Vermittlung von Lesestrategien

Legt man der Gruppierung den geschätzten Anteil schwacher Leser*innen zugrunde, finden sich nur vereinzelt größere Unterschiede zwischen den Gruppenmittelwerten zur Instruktion von Lesestrategien (Tabelle 3). Die Unterschiede fallen jedoch deutlicher aus als bei den Analysen, die auf dem auf der Grundlage des IGLU-Tests ermittelten Anteil schwacher Leser*innen beruhen.

| Wie oft tun Sie im Leseunterricht die folgenden Dinge? | Klassen mit einem geschätzt hohen Anteil leistungsschwacher Leser*innen | | Klassen mit einem geschätzt mittleren/niedrigen Anteil leistungsschwacher Leser*innen | | Mittelwertvergleich (T-Test bei unabhängigen Stichproben) | | |
|--|---|------|---|------|---|--------|------|
| | M | SD | M | SD | t | df | p |
| Den Schüler*innen erklären und zeigen, ... | | | | | | | |
| wie man verschiedene Lesestrategien einsetzt. | 2,45 | 0,76 | 2,38 | 0,75 | -0,572 | 196 | .568 |
| wann man verschiedene Lesestrategien einsetzt. | 2,33 | 0,74 | 2,29 | 0,71 | -0,408 | 196 | .684 |
| wie man Fragen zum Text entwickelt, während man liest. | 2,18 | 0,82 | 2,14 | 0,74 | -0,334 | 197 | .739 |
| wie man zentrale Ideen eines Textes zusammenfasst. | 2,71 | 0,76 | 2,57 | 0,69 | -1,180 | 198 | .239 |
| wie sie den Textinhalt mit ihrem Vorwissen verbinden können. | 2,59 | 0,83 | 2,44 | 0,70 | -1,219 | 196 | .224 |
| wie man wichtige Textstellen markiert (z. B. durch Unterstreichen). | 2,98 | 0,71 | 2,95 | 0,69 | -0,304 | 198 | .761 |
| dass man wichtige Textstellen manchmal wörtlich abschreiben sollte. | 2,12 | 0,97 | 2,10 | 0,75 | -0,114 | 71,491 | .910 |
| wie sie während des Lesens Voraussagen über den weiteren Verlauf des Textes anstellen. | 2,22 | 0,78 | 2,04 | 0,68 | -1,534 | 198 | .127 |
| wie sie überprüfen, ob ihnen alle Wortbedeutungen bekannt sind. | 2,80 | 0,96 | 2,51 | 0,94 | -1,898 | 197 | .059 |
| wie sie den Textaufbau mit dem von anderen Texten vergleichen. | 1,80 | 0,89 | 1,77 | 0,63 | -0,237 | 67,589 | .813 |
| wie sie Informationen im Text über Abschnittsgrenzen hinweg verknüpfen. | 1,84 | 0,89 | 1,84 | 0,79 | 0,010 | 193 | .992 |
| wie sie Überschriften zu Textabschnitten formulieren. | 2,34 | 0,69 | 2,32 | 0,68 | -0,201 | 196 | .841 |
| wie sie wichtige Informationen in ein anderes Darstellungsformat überführen (z. B. Mind-Map, Tabelle, Diagramm). | 1,86 | 0,83 | 1,97 | 0,63 | 0,822 | 70,741 | .414 |

Tab 3: Deskriptive Werte der Items zur Instruktion von Lesestrategien in Klassen mit einem hohen bzw. mittleren bis niedrigen Anteil schwacher Leser*innen sowie die Ergebnisse der Überprüfung der statistischen Bedeutsamkeit der Mittelwertsunterschiede mittels T-Tests bei unabhängigen Stichproben.

Fast alle Lesestrategien werden den Angaben der Lehrkräfte zufolge in Klassen mit einem hohen Anteil schwacher Leser*innen häufiger thematisiert als in den anderen Klassen. Die Unter-

schiede fallen insgesamt aber gering aus. In Klassen mit einem hohen Anteil schwacher Leser*innen wird z.B. häufiger thematisiert, wie überprüft werden kann, ob alle Wortbedeutungen in einem Text bekannt sind. Gleiches gilt für die Strategie, Voraussagen über den weiteren Verlauf eines Textes vorzunehmen. Die T-Tests zeigen keine signifikanten Unterschiede zwischen den Gruppen.

Gestaltung von Stillesephasen

Hinsichtlich der Gestaltung von Stillesephasen zeigen sich zwischen den Gruppen deutlichere Unterschiede als bei den Analysen zum ermittelten Anteil schwacher Leser*innen (Tabelle 4). So geben Lehrkräfte, die den Anteil an schwachen Leser*innen in ihrer Klasse als hoch einschätzen, an, häufiger Texte zum stillen Lesen für die Kinder auszuwählen. Die Lehrkräfte der anderen Gruppen überlassen diese Aufgabe häufiger den Kindern. Außerdem geben Lehrkräfte von Klassen mit einem hohen Anteil schwacher Leser*innen an, häufiger mit den Kindern nach der Lektüre über die Texte zu sprechen. Bei dem Vergleich der Mittelwerte durch T-Tests zeigen sich signifikante Unterschiede zwischen den Gruppen: In Klassen mit einem geringen Anteil schwacher Leser*innen können die Schüler*innen Texte häufiger nach eigenen Interessen auswählen und in Klassen mit einem hohen Anteil schwacher Leser*innen werden die Texte häufiger im Anschluss an die Lektüre im Unterrichtsgespräch besprochen.

| In meinem Unterricht haben Schüler*innen Zeit, still zu lesen. Falls ja: Wie häufig kommt Folgendes beim stillen Lesen vor? | Klassen mit einem geschätzt hohen Anteil leistungs-schwacher Leser*innen | | Klassen mit einem geschätzt mittleren/niedrigen Anteil leistungs-schwacher Leser*innen | | Mittelwertvergleich (T-Test bei unabhängigen Stichproben) | | |
|---|--|------|--|------|---|-----|------|
| | M | SD | M | SD | t | df | p |
| Ich übernehme die Auswahl der Texte für Stillesephasen. | 2,71 | 0,93 | 2,41 | 0,95 | -1,737 | 174 | .084 |
| Die Schüler*innen können die Texte nach eigenen Interessen auswählen. | 2,40 | 0,91 | 2,74 | 0,91 | 2,076 | 178 | .039 |
| Texte, die alle Schüler*innen gelesen haben, werden nach der Lektüre Gegenstand des Unterrichtsgesprächs. | 3,02 | 0,85 | 2,71 | 0,90 | -2,044 | 179 | .042 |
| Die Schüler*innen müssen nachhalten, wie viel sie gelesen haben (z. B. durch ein Lesetagebuch). | 1,54 | 0,79 | 1,61 | 0,95 | 0,422 | 170 | .674 |

Tab. 4: Deskriptive Werte zur Häufigkeit von Gestaltungsmerkmalen von Stillesephasen in Klassen mit einem hohen bzw. einem mittleren bis niedrigen Anteil schwacher Leser*innen sowie die Ergebnisse der Überprüfung der statistischen Bedeutsamkeit der Mittelwertsunterschiede mittels T-Tests bei unabhängigen Stichproben.

6.3 | Korrelationen zwischen der Unterrichtsgestaltung und dem ermittelten sowie dem geschätzten Anteil schwacher Leser*innen

| Wie oft tun Sie im Leseunterricht die folgenden Dinge? | Korrelationskoeffizient (r) | |
|--|--|---|
| | Anteil schwacher Leser*innen berechnet auf Grundlage der Ergebnisse des Leistungstests | Anteil an Schüler*innen mit Förderbedarf im Lesen nach Einschätzung der Lehrkraft |
| wie man verschiedene Lesestrategien einsetzt. | .02 | -.01 |
| wann man verschiedene Lesestrategien einsetzt. | -.05 | -.02 |
| wie man Fragen zum Text entwickelt, während man liest. | .01 | .08 |
| wie man zentrale Ideen eines Textes zusammenfasst. | -.05 | .09 |
| wie sie den Textinhalt mit ihrem Vorwissen verbinden können. | -.06 | .12 |
| wie man wichtige Textstellen markiert (z. B. durch Unterstreichen). | -.11 | .01 |
| dass man wichtige Textstellen manchmal wörtlich abschreiben sollte. | -.04 | .03 |
| wie sie während des Lesens Voraussetzungen über den weiteren Verlauf des Textes anstellen. | .04 | .12 |
| wie sie überprüfen, ob ihnen alle Wortbedeutungen bekannt sind. | .13 | .14* |
| wie sie den Textaufbau mit dem von anderen Texten vergleichen. | -.06 | -.01 |
| wie sie Informationen im Text über Abschnittsgrenzen hinweg verknüpfen. | .01 | .05 |
| wie sie Überschriften zu Textabschnitten formulieren. | -.03 | .06 |
| wie sie wichtige Informationen in ein anderes Darstellungsformat überführen (z. B. Mind-Map, Tabelle, Diagramm). | -.10 | -.03 |

Tab. 5: Korrelationen zwischen dem Anteil an schwachen Leser*innen (basierend auf dem Leistungstest bzw. den Lehrereinschätzungen) und den Lehrerangaben zur Häufigkeit, mit der Strategien im Unterricht instruiert werden. * $p < .05$.

Vermittlung von Lesestrategien

Um zu prüfen, ob ein Zusammenhang zwischen der Unterrichtsgestaltung und dem (ermittelten und geschätzten) Anteil schwacher Leser*innen besteht, werden bivariate Korrelationen berechnet. Die Ergebnisse für die Instruktion von Strategien finden sich in Tabelle 5. Keine der Korrelationen mit dem ermittelten Anteil schwacher Leser*innen wird statistisch bedeutsam. Es gibt keinen Zusammenhang zwischen der Häufigkeit der Instruktion von Lesestrategien und dem auf Grundlage des IGLU-Tests ermittelten Anteil schwacher Leser*innen in einer Klasse. Für den durch die Lehrkräfte geschätzten Anteil schwacher Leser*innen findet sich ein signifikanter Zusammenhang mit der Unterrichtsgestaltung: Je höher der geschätzte Anteil an schwachen Leser*innen in einer Klasse ist, desto häufiger instruieren die Lehrkräfte Lesestrategien, die darauf zielen, dass die Schüler*innen prüfen können, ob ihnen alle Wortbedeutungen in einem Text bekannt sind ($r = .14^*$).

Gestaltung von Stillesephasen

In Tabelle 6 werden die Korrelationen zwischen dem Anteil schwacher Leser*innen und den Gestaltungsmerkmalen von Stillesephasen dargestellt. Es ist kein statistisch bedeutsamer Zusammenhang zwischen dem ermittelten Anteil an schwachen Leser*innen in einer Klasse und der Gestaltung von Stilleseverfahren nachzuweisen. Ein signifikanter Zusammenhang findet sich in den Ergebnissen auf Grundlage des geschätzten Anteils schwacher Leser*innen: Je höher dieser ist, desto seltener dürfen die Kinder Texte, die sie während der Stillesephasen lesen, nach eigenen Interessen auswählen ($r = -.20^*$).

| | Korrelationskoeffizient (r) | |
|---|--|---|
| | Anteil schwacher Leser*innen berechnet auf Grundlage der Ergebnisse des Leistungstests | Anteil an Schüler*innen mit Förderbedarf im Lesen nach Einschätzung der Lehrkraft |
| Ich übernehme die Auswahl der Texte für Stillesephasen. | .10 | .11 |
| Die Schüler*innen können die Texte nach eigenen Interessen auswählen. | -.08 | -.20* |
| Texte, die alle Schüler*innen gelesen haben, werden nach der Lektüre Gegenstand des Unterrichtsgesprächs. | .03 | .13 |
| Die Schüler*innen müssen nachhalten, wie viel sie gelesen haben (z. B. durch ein Lesetagebuch). | -.06 | -.04 |

Tab. 6: Korrelationen zwischen dem Anteil an schwachen Leser*innen (basierend auf dem Leistungstest bzw. den Lehrereinschätzungen) und den Lehrerangaben zur Häufigkeit bestimmter Gestaltungsmerkmale von Stillesephasen. * $p < .05$.

7 | Diskussion

Es zeigt sich, dass viele der elaborativen und organisierenden Lesestrategien, deren Einsatz die Bearbeitung anspruchsvoller Leseaufgaben unterstützen kann (zu denken wären hier z.B. an Textvergleiche, Übertragen von Informationen in andere Formate u.a.m.), im Leseunterricht am Ende der Grundschulzeit insgesamt seltener thematisiert werden als Oberflächenstrategien wie das Unterstreichen wichtiger Textstellen (vgl. dazu auch Bremerich-Vos et al. 2017: 289ff.). In Klassen mit wenigen schwachen Leser*innen wäre zu erwarten, dass die Vermittlung anspruchsvoller Lesestrategien eine bedeutendere Rolle spielt als in Klassen mit vielen Kinder, deren Schwierigkeiten mit dem Lesen auf einer hierarchieniedrigen Ebene beginnen. Den Ergebnissen von IGLU 2016 zufolge spielen die Lesestrategien, die auf eine Unterstützung hierarchiehöherer Prozesse zielen, in Klassen mit einem mittleren bzw. niedrigen Anteil schwacher Leser*innen keine bedeutendere Rolle als in den Klassen mit vielen schwachen Leser*innen.

Insgesamt finden sich zwischen den Gruppen kaum Unterschiede in den Angaben zur Unterrichtsgestaltung – weder bei den Angaben dazu, wann und wie Lesestrategien eingesetzt werden sollten, noch bei der Instruktion der meisten Strategien. Eine statistisch signifikante Korrelation mit dem Anteil schwacher Leser*innen ist lediglich bei einem Item nachzuweisen: Lehrkräfte, die nach eigener Einschätzung in Klassen mit vielen förderbedürftigen Leser*innen arbeiten, vermitteln nach eigenen Angaben häufiger, wie die Schüler*innen überprüfen können, ob ihnen alle Wortbedeutungen eines Textes bekannt sind. Kleinbub (2010) hat auf Grundlage von Unterrichtsaufzeichnungen darauf aufmerksam gemacht, dass im Leseunterricht in vierten Klassen vor der Lektüre von Texten häufig Wortbedeutungen geklärt werden und dass andere kognitive und metakognitive Strategien im Unterricht von untergeordneter Bedeutung sind. Mit dem in der vorliegenden Untersuchung berücksichtigten Item wird nicht erfragt, ob im Unterricht Wortbedeutungen geklärt werden, um das Verständnis zu sichern; das Item zielt darauf, wie man prüfen kann, ob alle Wortbedeutungen bekannt sind. Die Validität der Lehrerangaben vorausgesetzt, wird damit eine metakognitive Strategie vermittelt. Weitere signifikante Zusammenhänge lassen sich hinsichtlich der Instruktion von Lesestrategien nicht nachweisen. Weitere Strategien, die auf die Unterstützung relativ hierarchieniedriger Leseprozesse zielen, werden in Klassen mit vielen schwachen Leser*innen demnach auch nicht häufiger instruiert als in Klassen mit weniger leseschwachen Kindern.

Phasen stillen Lesens im Unterricht bieten sich an, um Schüler*innen die Lektüre ganzer Texte zu ermöglichen. In der Hoffnung auf Effekte für Kinder, deren familiäre Lesesozialisation nicht ideal verlaufen ist, sehen Rosebrock und Nix (2017: 71) einen Grund, Stillesephasen in den Unterricht zu integrieren, auch wenn positive Effekte eines auf stilles Lesen ausgerichteten Unterrichts in einschlägigen Studien nicht immer nachgewiesen werden können. Phasen stillen Lesens sind Bestandteil des Unterrichts eines großen Teils der Lehrkräfte der Stichprobe aus IGLU 2016. Der Anteil der Lehrkräfte, die angeben, Schüler*innen im Unterricht niemals Zeit für stilles Lesen einzuräumen, ist deutlich kleiner als der Anteil der Lehrkräfte, die angeben, dass Lautlesetandems in ihrem Unterricht nie eine Rolle spielen (Abschnitt 2).

Gruppenunterschiede zeigen sich bei den deskriptiven Befunden vor allem bei der Auswahl der Texte. Diese überlassen die Lehrkräfte in Klassen mit vielen schwachen Leser*innen (ermittelter und geschätzter Anteil) seltener den Kindern. Auch die einzige signifikante Korrelation zur Gestaltung der Stillesephasen betrifft diesen Aspekt der Unterrichtsgestaltung: Je höher der berichtete Anteil an schwachen Leser*innen in der Klasse ist, desto weniger Einfluss haben die

Schüler*innen auf die Auswahl der Texte, die still gelesen werden. Das könnte ein Aspekt „kompetente[r] Begleitung und Kontrolle“ (Rosebrock & Nix 2017: 68) von Stillelesephasen sein, der sich positiv auf die Lesekompetenz auswirken könnte, wie Rosebrock und Nix (2017) auf Grundlage der Forschungslage zum stillen Lesen im Unterricht annehmen: „Offensichtlich fördert ein Setting, in dem selbstbestimmt im Rahmen des Unterrichts gelesen wird, nicht per se messbare Lesekompetenz.“ (ebd.). Es müsse sichergestellt sein, dass die Schüler*innen, die Texte auch tatsächlich verstehend lesen und die Lektüre nicht nur „als ein ‚Durchkommen‘ durch den Text“ (ebd.) begreifen. Möglicherweise sei es sinnvoll, die Fördersituation so zu gestalten, „dass die Lektüre zwar ein weitgehend eigenständiger und in vielen Komponenten selbstorganisierter Akt bleibt, zugleich aber durch eine kompetente Begleitung und Kontrolle stabilisiert wird“ (ebd.) (vgl. dazu auch Rieckmann 2020).

Zudem zeigt sich bei den deskriptiven Befunden, dass die Texte, die in den Stillelesephasen gelesen werden, in beiden Gruppen und unabhängig von der Frage, ob der ermittelte oder der geschätzte Anteil schwacher Leser*innen zugrunde gelegt wird, im Anschluss an die Lektüre relativ häufig Gegenstand des Unterrichtsgesprächs werden. Das ist z.B. im Kilometer-Lesen nach Rosebrock und Kolleg*innen nicht so vorgesehen (Rosebrock & Nix 2017: 60f.). Dass die Lehrkräfte sich in ihrem Unterricht mit den Ergebnissen der Lektüre der Schüler*innen auseinandersetzen, kann ebenfalls als Aspekt von Begleitung und Kontrolle gedeutet werden. Es liegt nahe, das auch mit dem Befund zu verbinden, dass die Lehrkräfte in beiden Gruppen relativ selten mithilfe von Lesetagebüchern o.Ä. nachzuhalten scheinen, was und wie viel gelesen wurde; in vielen Klassen scheint das „Nachhalten“ ein Teil des Unterrichts zu sein.

Vor dem Hintergrund der Erkenntnisse empirischer Forschung zur Leseförderung im Unterricht scheinen in Deutschland nicht alle Bedürfnisse der unterschiedlich leistungsstarken Schüler*innen hinreichend berücksichtigt zu werden. Darauf haben schon die Befunde von Bremerich-Vos et al. (2017) zur Realisierung von Lautlesetandems im Unterricht hingewiesen. Dass sich in der vorliegenden Untersuchung kaum signifikante Zusammenhänge zwischen dem Anteil schwacher Leser*innen und dem Einsatz der beiden Leseförderverfahren zeigen, deutet darauf hin, dass eine noch deutlichere Vermittlung von Informationen darüber notwendig ist, für welche Zielgruppe die unterschiedlichen Maßnahmen zur Leseförderung geeignet sind.

Effektive Leseförderung im Unterricht setzt diagnostische Kompetenz aufseiten von Lehrkräften voraus (Rosebrock & Nix 2017; Hußmann et al. i.V.). IGLU 2016 zeigt, dass der von den Lehrkräften geschätzte Anteil an Leser*innen mit Förderbedarf im Lesen in etwa dem auf Grundlage des IGLU-Tests ermittelten Anteil an schwachen Leser*innen entspricht (Bremerich-Vos et al. 2017: 281). Kompetenzwerte, die auf einer Schätzung von Lesekompetenzen für die einzelnen Schüler*innen beruhen, wie sie im Rahmen von IGLU generiert werden, dienen nicht dazu, die individuelle Leseförderung vorzubereiten. In der alltäglichen Unterrichtsplanung von Lehrkräften dürften standardisierte Tests wie ELFE II sowie Einschätzungen von Schülerkompetenzen durch Lehrkräfte von Bedeutung sein. In einschlägigen Publikationen zur Leseförderung finden sich praxisorientierte Vorschläge zur Diagnose von Lesekompetenzen (z.B. Rosebrock et al. 2013: 81ff.; Gailberger & Nix 2013). Lehrkräfte damit vertraut(er) zu machen, könnte dazu beitragen, dass die Leseförderung gezielter gestaltet wird. Denn: Die Mittelwerte zwischen den Angaben zum Leseunterricht zwischen den Gruppen (hoher bzw. mittlerer und niedriger Anteil

schwacher Leser*innen) weisen deutlichere Unterschiede auf, wenn die Gruppen auf Grundlage der Einschätzung der Lehrkräfte gebildet werden. Das könnte man als Hinweis darauf deuten, dass die Lehrkräfte auf erkannte Leseschwierigkeiten durchaus reagieren.

Zum Abschluss zwei Anmerkungen zu der vorliegenden Studie.

Angaben von Lehrkräften zu ihrem eigenen Unterricht werden mitunter kritisch gesehen (Philipp 2014: 126). Die Frage, „inwiefern Fragebogen-Daten zur faktischen Unterrichtspraxis zu trauen ist“ (ebd.), ist Gegenstand zahlreicher Studien (Clausen 2002). Angaben von Lehrkräften zum eigenen Unterricht haben Vor- und Nachteile: Lehrkräfte könnten im Hinblick auf die Einschätzung ihrer Unterrichtspraxis zu selbstwertdienlichen Verzerrungen neigen; das kann in einer einseitig positiven Einschätzung des eigenen Unterrichts zum Ausdruck kommen (Stahns, Rieser & Lankes 2017). Allerdings zeigen Ergebnisse aus IGLU auch, dass Zusammenhänge zwischen Angaben von Lehrkräften zum Leseunterricht und Lesekompetenzen bestehen (Stahns & Rieser 2018). Zudem ist es möglich, Lehrkräfte Einschätzungen von Unterrichtsmerkmalen vornehmen zu lassen, wenn dafür fachliches und fachdidaktisches Wissen gefordert ist. Schüler*innen, auf deren Einschätzungen viele Studien zur Unterrichtsgestaltung beruhen, können das, Ergebnissen zahlreicher Studien zufolge, eher nicht. Zudem können Lehrkräfte Angaben zur Gestaltung ihres Unterrichts über einen längeren Zeitraum vornehmen, während in Videostudien wie der von Kleinbub (2010) einzelne oder wenige Stunden aufgezeichnet werden. Insofern haben Angaben von Lehrkräften zu ihrem eigenen Unterricht ihren Platz in der empirischen Unterrichtsforschung (vgl. dazu Clausen 2002).

Des Weiteren ist eine Besonderheit der vorliegenden Untersuchung zu diskutieren. Systematische Leseförderung zeichnet sich durch die Abfolge von (individueller) Diagnostik und anschließender Förderung aus, der Leistungstest von IGLU wurde aber durchgeführt, nachdem der Unterricht durchgeführt wurde, auf den sich die Lehrkräfte in ihren Angaben beziehen. In der vorliegenden Untersuchung werden die u.a. auf Grundlage der Testergebnisse geschätzten Plausible Values genutzt, um den Anteil schwacher Leser*innen in der Klasse zu bestimmen. Es ist u.E. davon auszugehen, dass Schüler*innen, die zum Testzeitpunkt zur Gruppe der schwachen Leser*innen gehören, in dem Zeitraum vorher nicht zur Gruppe der starken Leser*innen gehört haben. Möglich wäre natürlich, dass (einige) Schüler*innen zum Testzeitpunkt nicht mehr zur Gruppe der schwachen Leser*innen gehört haben, während das für den Zeitraum noch galt, auf den sich die Lehrkräfte in ihren Angaben zum Unterricht beziehen. Mit größeren Veränderungen des Anteils schwacher Leser*innen innerhalb der Klassen ist aber zumindest kurzfristig nicht zu rechnen. Des Weiteren sollten solche Veränderungen dann auch zu Anpassungen im Unterricht führen. Das wiederum sollte sich in den Angaben der Lehrkräfte zum Unterricht widerspiegeln. Da der Erwerb von Lesekompetenz sich über längere Zeiträume erstreckt und dabei immer wieder eine Passung des Unterrichts stattfindet, ist davon auszugehen, dass die Daten der vorliegenden Studie die Beantwortung der Fragestellung erlauben.

8 | Literatur

- Artelt, C., Naumann, J. & Schneider, W. (2010). Lesemotivation und Lernstrategien. In: E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider, P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 73–112). Waxmann.
- Bertschi-Kaufmann, A. & Schneider, H. (2006). Entwicklung von Lesefähigkeit: Massnahmen – Messungen – Effekte. Ergebnisse und Konsequenzen aus dem Forschungsprojekt „Lese- und Schreibkompetenzen fördern“. *Schweizerische Zeitschrift für Bildungswissenschaften*, 28 (3), 393–424.
- Bremerich-Vos, A., Stahns, R., Hußmann, A. & Schurig, M. (2017). Förderung von Leseflüssigkeit und Leseverstehen im Leseunterricht. In: A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T.C. Stubbe, R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 279–296). Waxmann.
- Bremerich-Vos, A., Wendt, H. & Bos, W. (2017). Lesekompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In: A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T.C. Stubbe, R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 79–142). Waxmann.
- Clausen, M. (2002): *Unterrichtsqualität: Eine Frage der Perspektive?* Waxmann.
- Gailberger, S. & Nix, D. (2013). Lesen und Leseförderung in Primar- und Sekundarstufe. In: S. Gailberger, F. Wietzke (Hrsg.): *Handbuch kompetenzorientierter Deutschunterricht* (S. 36–69). Beltz.
- Hußmann, A., Hüninghake, R., Hartung, N., Euker, N. & Kuhl, J. (i.V.). Förderung im Lernbereich Schriftsprache. In: J. Kuhl, A. Vossen, N. Hartung & C. Wittich (Hrsg.), *Evidenzbasierte Förderung bei Lernschwierigkeiten in der inklusiven Schule*. Ernst Reinhardt Verlag.
- Hußmann, A., Wendt, H., Kasper, D., Bos, W. & Goy, M. (2017). Ziele, Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU 2016). In: A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T.C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016 Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S.29–78). Waxmann.
- Kleinbub, I. D. (2010). *Unterrichtsqualität im Leseunterricht. Eine videobasierte Analyse in vierten Klassen*. Wissenschaftlicher Verlag.
- Kruse, G., Rickli, U. Riss, M. & Sommer, T. (2013). *Lesen. Das Training*. Schulbuchverlag plus AG.
- Lauer-Schmaltz, M., Rosebrock, C. & Gold, A. (2014). Lautlesetandems in der Grundschule. *Didaktik Deutsch* 19 (37), 45–61.
- Lenhard, W., Lenhard, A. & Schneider, W. (2017). *ELFE II. Ein Leseverständnistest für Erst- bis Siebtklässler – Version II*. Hogrefe.
- Munser-Kiefer, M. (2012). Leseförderung mit (meta-)kognitiven Strategien in der Grundschule? Befunde und Schlussfolgerungen aus einer Interventionsstudie. *Zeitschrift für Grundschulforschung*, 5 (2), 122–135.
- Munser-Kiefer, M. (2014). *Leseförderung im Leseteam in der Grundschule. Eine Interventionsstudie zur Förderung von basaler Lesefertigkeit und (meta-)kognitiven Lesestrategien*. Waxmann.
- Müller, B., Križan, A., Hecht, T., Richter, T. & Ennemoser, M. (2013). Leseflüssigkeit im Grundschulalter: Entwicklungsverlauf und Effekte systematischer Leseförderung. *Lernen und Lernstörungen*, 2 (3), 131–146.

- NICHD – National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel: Teaching children to read – An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. NICHD.
- Palincsar, A. & Brown, A.L. (1984). Reciprocal Teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117–175.
- Philipp, M. (2014). Leseunterricht in der Grundschule – vom Ist-Zustand und vom Soll-Zustand. Was Beobachtungsstudien lehren. In: R. Valtin, I. Tarelli (Hrsg.), *Lesekompetenz stärken – evidenzbasierte Maßnahmen und Programme* (S. 122–165). Deutsche Gesellschaft für Lesen und Schreiben.
- Rieckmann, C. (2020). *Grundlage der Lesedidaktik. Band 2: „Eigenständiges Lesen“*. 3. Auflage. Schneider Verlag Hohengehren.
- Rosebrock, C. & Gold, A. (2018). Flüssigkeit als Kategorie für die Diagnose und Förderung von Lesefertigkeit. In: S. Kutzelmann, C. Rosebrock (Hrsg.), *Praxis der Lautleseverfahren* (S.7–19). Schneider Verlag Hohengehren.
- Rosebrock, C. & Nix, D. (2017). *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung*. 8. korrigierte Auflage. Schneider Verlag Hohengehren.
- Rosebrock, C., Nix, D., Rieckmann, C. & Gold, A. (2013). *Leseflüssigkeit fördern. Lautleseverfahren für die Primar- und Sekundarstufe*. 5. Auflage. Klett Kallmeyer.
- Rosebrock, C., Rieckmann, C., Nix, D. & Gold, A. (2010). Förderung der Leseflüssigkeit bei leseschwachen Zwölfjährigen. *Didaktik Deutsch* 28, 33–58.
- Schneider, H. & Bertschi-Kaufmann, A. (2007). Selbstgestaltete Lektüre oder ein angeleitetes Lesetraining? Zu den Effekten zweier Methoden der Leseförderung. *Bulletin suisse de linguistique appliquée*, 29–52.
- Schulte, E. & Souvignier, E. (2013). Der Lese-Sportler. Ein Programm für individuelle Leseförderung. *Die Grundschulzeitschrift*, 27 (7), 58–61.
- Souvignier, E. (2014). Lesestrategien zur Förderung des Leseverständnisses und der Ansatz einer diagnosebasierten, individualisierten Leseförderung. In: R. Valtin, I. Tarelli (Hrsg.), *Lesekompetenz nachhaltig stärken. Evidenzbasierte Maßnahmen und Programme* (S. 167–182). Deutsche Gesellschaft für Lesen und Schreiben.
- Stahns, R., Rieser, S. & Lankes, E.-M. (2017). Unterrichtsführung, Sozialklima und kognitive Aktivierung im Deutschunterricht in vierten Klassen. In: A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T.C. Stubbe, R. Valtin (Hrsg.), *IGLU 2016 Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 251–277). Waxmann.
- Stahns, R. & Rieser, S. (2018). Qualität des Leseunterrichts in vierten Klassen unter den Bedingungen von Mehrsprachigkeit. Ergebnisse der Lehrkräftebefragung von IGLU 2011. *Zeitschrift für Grundschulforschung*, 11 (1), 131–145.
- Walter, J., Ide, S. & Petersen, A. (2012). Kooperatives Lernen auf der Basis von Lesetandems. Entwicklung und Evaluation eines tutoriellen Lesetrainings zur Steigerung der Leseflüssigkeit. *Zeitschrift für Heilpädagogik*, 63, 448–464.

Steffen Gailberger, Britta Pohlmann, Laura Reichenbach, Franziska Thonke & Jonna Wolters

Zum nachhaltigen Einfluss von Lautleseverfahren auf Leseflüssigkeit, Leseverstehen, Rechtschreibung sowie Kompetenzen jenseits des Deutschunterrichts

Bericht zur Longitudinalstudie des Hamburger BiSS-Verbunds Primarstufe aus der Perspektive von Lesedidaktik und empirischer Bildungswissenschaft für schulische Praxis und Leseförderung

1 | Durch „Bildung in Sprache und Schrift“ zu mehr Chancengerechtigkeit und Bildungsgleichheit? Aufriss und Einführung ins Projekt

Im folgenden Beitrag¹ wird eine empirische Longitudinalstudie zur systematischen Leseförderung vorgestellt, die das Hamburger BiSS-Projekt für die Primarstufe bzw. die darin geförderten strukturell benachteiligten Schüler*innen von 2015 bis 2018 über insgesamt sieben Testzeitpunkte² von Jg. 2 bis Jg. 5 begleitete. Die Studie liefert Hinweise, wie Jungen und Mädchen mit sozio-ökonomischer und/oder sprachlicher Benachteiligung im Bemühen um gleiche Bildungschancen und Bildungsgerechtigkeit nachhaltig gefördert werden können. Unter Verwendung eines ganzheitlichen Lesebegriffs wurden im Rahmen des Projekts *erstens* das Lesen aus kognitiver Perspektive, *zweitens* die Einstellung zu Literatur und Schriftlichkeit im Allgemeinen, *drittens* die motivationale Einstellung gegenüber dem eigens für das Projekt initiierten Leseunterricht und *viertens*, da die Leseförderung in Form eines stark an Partner- oder Gruppenarbeit orientierten „Lesebandes“ (Gailberger & Hauschild 2018) organisiert wird, auch soziale Kompetenzen gefördert.³ Somit standen die Entwicklung und Förderung von Lesegeläufigkeit und Lesefähigkeit ebenso im Fokus wie der damit verbundene Auf- und Ausbau der Lesemotivation, des Leseinteresses und der Entwicklung von positiv lesebegleitenden Emotionen. Die Förderung dieser verschiedenen Komponenten sollte sich schließlich – so die erkenntnisleitende These des Projekts – über eine Förderung der Lesekompetenz hinaus ebenso positiv auf die Leistungen in anderen Kompetenzbereichen (auch jenseits des Deutschunterrichts) auswirken.

¹ Die Autor*innen danken Nadja Frehe (Hamburger Landesinstitut für Lehrerbildung und Schulentwicklung/ Grundschule Lutterothstraße), Andreas Heintze und Eric Vaccaro (Hamburger Behörde für Schule und Berufsbildung) sowie allen am Projekt beteiligten Lehrkräften und Schulleitungen für ihre tatkräftige Unterstützung des Projekts. Nützliche Literaturhinweise erhielten sie außerdem von Anke Schmitz und Gerrit Helm – auch ihnen sei herzlich gedankt.

² Aus schulpragmatischen Gründen wurde die Kontrollgruppe in derselben Zeit nur zu drei Testzeitpunkten getestet (s. Abbildung 1).

³ Aus Platzgründen beschränken wir uns im vorliegenden Bericht auf die Aspekte 1, 2 und 3.

Der vorliegende Bericht soll informieren, inwiefern es dem Projekt und den in ihm beteiligten Lehrerinnen und Lehrern gelungen ist, den zum größten Teil stark benachteiligten Kindern nachhaltig zu verbesserten Kompetenzen zu verhelfen.

2 | Grundlegendes zum Hamburger BiSS-Projekt für die Primarstufe

Unter dem in der Hamburger Behörde für Schule und Berufsbildung offiziell geführten Titel *Systematische Leseförderung in der Grundschule* verfolgte das Hamburger primarstufenspezifische BiSS-Projekt das Ziel, die (im Sinne des in der deutschsprachigen Praxis vorherrschenden Lesebegriffs aus der Perspektive der Lesesozialisationsforschung modellierte) Lesekompetenz der beteiligten Schüler*innen von Jahrgang zwei bis Ende Jahrgang vier zu fördern (vgl. im Überblick Garbe, Ohlsen & Philipp 2009; Pieper 2019 oder Rosebrock & Nix 2017). Das Lesetraining richtete sich also insbesondere an Schüler*innen mit ungünstigeren Ausgangslagen, für die eine frühzeitige gezielte Förderung von besonderer Bedeutung ist, was wiederum unmittelbaren Einfluss auf die Auswahl der im Projekt verwendeten Methoden zur Förderung der Lesegeläufigkeit, des lesenden Textverstehens und Textverarbeitens sowie zur Förderung der Lesemotivation hatte.

Zur Lesegeläufigkeit

Lesegeläufigkeit wird seit nahezu vier Jahrzehnten empirisch erforscht und auf dieser Basis gemeinhin als Brücke zur Lesekompetenz angesehen (vgl. hierzu etwa LaBerge & Samuels 1974, Pikulski & Chard 2005, Klauda & Guthrie 2008 oder Fuchs et al. 2001 sowie die weiterführenden Beiträge von Röttig et al., Stephany et al., Juska-Bacher et al. und von Sappok im vorliegenden Band). Dabei werden bekanntlich vier Aspekte von Lesegeläufigkeit differenziert, die im Rahmen einer erfolgsversprechenden Lesekompetenzausbildung in einem Mindestmaß gegeben sein oder aber gefördert werden müssen und die mit Blick auf die in den vorherigen Klammern erwähnten Bandbeiträge im Folgenden knapp skizziert werden sollen:

Zunächst sind disfluent lesende Mädchen und Jungen meist leicht an einer mangelnden *Dekodiergenauigkeit* zu erkennen, da ihr Anteil an sinnentstellenden Verlesungen sowie ihr Aufwand, dies selbst fortwährend korrigieren zu müssen, sehr hoch sind. Im Sinne einer auszubildenden ausreichenden Lesegenauigkeit muss sich also der Zugriff auf das semantische Lexikon fehlerfrei vollziehen lassen, so dass dem dekodierten Schriftbild des Wortes eine passende Bedeutung zugewiesen werden kann. Flüssig zu lesen bedeutet zweitens aber auch, *automatisiert* zu lesen. Versierte Leser*innen sind dazu befähigt, Wörter oder ganze Wortgruppen ‚auf einen Blick‘ zu erkennen, also auf den sogenannten Sichtwortschatz zurückzugreifen, wohingegen nicht automatisiert lesenden Schüler*innen diese Möglichkeit nicht gegeben ist. Diese müssen beim Entziffern eines Wortes mehrfach ansetzen oder sind beim Dekodieren längerer Wörter auffällig langsam. Automatisiertes Lesen bedeutet also konkret, nicht mehr den Lesevorgang an sich dekodierend oder lautierend fokussieren zu müssen, sondern die Dekodierprozesse ohne gesonderte Aufmerksamkeit vollziehen zu können. Dies setzt wiederum kognitive Ressourcen frei, die dann beispielsweise zur Verarbeitung des Textinhalts dienen können. Als dritte Komponente des flüssigen Lesens und quasi aus dem Zusammenspiel von Lesegenauigkeit und Automatisierung ergibt sich die *Lesegeschwindigkeit*, die in Wörtern pro Minute (WpM) gemessen wird. Kommen Lesegenauigkeit und *fehlende* Automatisierung zusammen, so bewirkt dies, dass Bezüge innerhalb von Sätzen als lokale Kohärenzen nicht angemessen hergestellt

werden können. Schwach lesende Schüler*innen vergessen im Laufe eines Satzes, weiterhin auf dessen Inhalt zu achten oder sie sind einfach nicht dazu in der Lage, so dass sie am Satzende nicht mehr erinnern, was soeben noch am Satzanfang zu lesen stand. Diese Schüler*innen, denen es an Lesegenauigkeit und Automatisierung mangelt, lesen somit deutlich langsamer als lesekompetentere Gleichaltrige. Den vierten die Lesegeläufigkeit beeinflussenden Faktor stellen *Betonung und Phrasierung* während des Leseprozesses dar, also inwiefern ein*e Schüler*in die Sätze eines Textes im Sinne der Grammatik der Geschichte angemessen lesen oder vorlesen kann. Dabei gibt die stimmige Betonung Aufschluss darüber, ob die Textvorgaben berücksichtigt und somit die semantischen *und* syntaktischen Sinnabschnitte bewusst wahrgenommen werden konnten. Disfluent lesenden Schüler*innen fehlt diese Wahrnehmung. Sie verharren häufig noch auf der Ebene von Ein- oder Zweiwortschritten und vermögen es demnach nicht, Sätze im Rezeptionsprozess prosodisch in sinnvolle, grammatisch zusammengehörige Teilabschnitte zu segmentieren. Außerdem schaffen sie es noch nicht, das eigene Textverstehen des Gelesenen unter Verwendung von impliziten oder expliziten Textsignalen mit einer sinnvollen Betonung, einer flexiblen und sinngebenden Intonation, Pausengestaltungen und angemessenen Rhythmisierungen zu unterstützen. Dies gilt dabei nicht nur für laut- oder vorlesende Settings, sondern ebenso für Situationen, in denen leise gelesen wird (vgl. Nix 2011).

Zu den weiteren kognitiven Teilprozessen des Lesens

Ihrer „Brückenfunktion“ (Pikulski & Chard 2005) hin zur Lesekompetenz kann die Lesegeläufigkeit nur nachkommen, wenn sie ein erfolgreich arbeitendes und v.a. *interdependentes Zusammenspiel* aller vier genannten Aspekte aufweist und sich dieses erfolgreich auch auf das lesende *Textverständnis* im Sinne hierarchiehöherer Verstehensprozesse auswirkt. Da das limitierte menschliche Arbeitsgedächtnis darauf angewiesen ist, dass alle beim Lesen aktivierten Teilprozesse so reibungslos wie möglich verlaufen, sind *fluent* lesende Schüler*innen den *disfluent* lesenden Schüler*innen gegenüber klar im Vorteil. Denn sie müssen keine gesonderte kognitive Aufmerksamkeit mehr auf die zuvor beschriebenen basalen Leseprozesse richten, sondern sind bereits dazu in der Lage, ihre kognitiven Ressourcen uneingeschränkt für den eigentlichen Lese- und Verstehensvorgang zu nutzen. Demgegenüber blockiert der bloße Dekodiervorgang bei noch disfluent lesenden Mädchen und Jungen das Arbeitsgedächtnis auf eine Weise, dass sie sich neben dem Lesen selbst nur schwer auf den Inhalt oder die sprachliche Gestaltung des Gelesenen konzentrieren können. Schwach lesende Schüler*innen benötigen demnach also zusätzliche mentale Kapazitäten für die Entzifferung der Schrift, was nicht nur ein bloß mangelhaftes Verstehen des Textes, sondern auch frühe Erschöpfungserscheinungen und damit einhergehend eine nur gering ausgebildete Lesemotivation zur Folge hat.

Flüssiges Lesen gilt daher als eine sehr wichtige Vorläuferfähigkeit des *verstehenden Lesens*, wenn im Laufe und nach Beendigung des Rezeptionsprozesses Mentale Modelle (respektive Situationsmodelle) des Textes generiert werden sollen. Nach der einschlägigen kognitionspsychologischen/psycholinguistischen Beschreibung (wie bei Christmann & Groeben 2006, Christmann 2019, Schnotz 2006 oder Kintsch 1998) kann dies dann gelingen,

- wenn es der bzw. dem Rezipierenden gelingt, auf der Basis der vornehmlich semantischen Prädikat-Argument-Struktur und über sogenannte Segmentierungsstrategien unter Zuhilfenahme der Syntax (auf *Satzebene*),

- über Koreferenzstrategien, Thema-Rhema-Struktur-Erkennungen und Inferenzbildungen (auf Absatzebene),
- und schließlich über die zyklische Verarbeitung der im Arbeitsgedächtnis gesammelten und abgespeicherten Propositionen sowie über die schemaorientierte Verwendung von vorwissensbasierten Scripts und Story Grammars (als *Text-Leser-Interaktion*)

eben jene Mentalen bzw. Situationsmodelle zu prozessieren, die es der bzw. dem Rezipierenden erlauben, mehr oder weniger *ad hoc* über den zuvor gelesenen Text – das heißt über dessen Figur(en), Ort(e), Zeit(en), Handlungskern(e) sowie Intention(en) – in anschlusskommunikativen Settings engagiert und/oder kritisch zu reflektieren.

Zum Lesen aus der Perspektive von Emotions- und Motivationspsychologie

Im Sinne des dem Hamburger BiSS-Projekts zugrunde gelegten weiten, didaktischen Lesebegriffs geht es neben der Förderung von Lesegeläufigkeit und lesendem Textverstehen außerdem um den Auf- und Ausbau von *Lesemotivation*, von *Interesse am* sowie von positiven *Emotionen beim* Lesen.

Das diachron ausgelegte Lesemotivations-Lesekompetenz-Modell von Möller und Schiefele (2004), das dem in der pädagogisch-psychologischen Forschung einschlägigen „Erwartungsmal-Wert-Modell“ folgt (vgl. etwa Hasselhorn & Gold 2013), bietet hierbei *die* Grundlage, auf der Lesemotivation auch beim Hamburger BiSS-Projekt verstanden und gefördert wird. Stark vereinfacht gesprochen unterzieht der Kern des „Erwartungsmal-Wert-Modells“ jegliche Tätigkeit zwei Fragen, die es aus motivationaler Sicht zu beantworten gilt: Die Antwort auf die Frage „*Mag ich das, was ich da tun soll/tun werde? Hat es für mich einen Nutzen und sind die ‚Kosten‘ dafür überschaubar und okay?*“ repräsentiert dabei die sogenannte *Wert-Komponente*, wohingegen die Antwort auf die Frage „*Schaffe ich das, was ich da tun soll/tun werde?*“ die Erwartungskomponente darstellt. Der motivationale Grad, mit dem schließlich eine Tätigkeit (wie bspw. das Lesen) durchgeführt werden soll, wird aus der ‚Verrechnung‘ beider Antworten miteinander ermittelt. Logisch schlussfolgernd werden Handlungen nach dem „Erwartungsmal-Wert-Modell“ also dann gerne und/oder erfolgreich ausgeführt, wenn sie entweder auf der *Wertebene* hochgradig positiv aufgeladen sind, und somit die Furcht vor einem etwaigen Scheitern verdrängen oder relativieren, oder wenn sie extrinsisch motiviert sind und somit als eine zwingende Notwendigkeit gelten, ein Blick auf den *Wert* der Handlung also mehr oder weniger irrelevant wird. Besonders gerne oder erfolgreich werden Handlungen jedoch dann ausgeführt, wenn sie beide Komponenten stimulieren und somit zu nachhaltigen ‚Erfolgen‘ mit hohem Gratifikationspotenzial führen.

Lesefördermaßnahmen werden von jungen Schüler*innen vor allem dann qualitativ nachhaltig positiv empfunden, wenn es ihnen im Rahmen des jeweiligen Lesesettings ermöglicht wird, beide oben erwähnten Komponenten miteinander zu verbinden. Mit Blick auf die lesespezifische ‚Situation Deutschunterricht‘ gilt es also, sowohl subjektiv unterschiedliche Empfindungen wie Lust, Interesse und Freude am Lesen (die *Wert-Komponente*) mit der eigenen Überzeugung, den Lesevorgang an sich erfolgreich meistern zu können (die *Erwartungs-Komponente*), zu verbinden. Auch wenn dies mitunter ein nicht allzu leichtes Unterfangen darstellt, erscheint es umso wichtiger, die Förderung der Lesemotivation zu systematisieren und damit nachhaltig zu gestalten, damit auch sie neben der Lesegeläufigkeit als eine zweite Brücke hin zur Lesekompetenz dienen kann.

Nun gibt (mit Groeben & Vorderer 1988: 12) die Lesemotivation als *quantitative* Komponente Auskunft darüber, wie intensiv und ausdauernd Schüler*innen in einer bestimmten Situation lesen. Da dies aus lesedidaktischer Sicht alleine aber nicht ausreicht, stellt das *Leseinteresse* als eine dritte Brücke hin zur Lesekompetenz die *qualitative* Komponente dar, die darüber Auskunft gibt, welche Autor*innen, Textsorten, Genres, Themen etc. gerne von den Schüler*innen gelesen werden und welche Lesestoffe hingegen nur widerwillig angegangen oder gar abgelehnt werden (ebd.). Von einer nachhaltigen Leseförderung kann also nur dann gesprochen werden, wenn sie neben der Motivation (= *Quantität* des Lesevorgangs, gemessen in Länge und Dauer) auch das Interesse (= *Qualität* des Lesevorgangs, gemessen in Intensität und Tiefe) der lesenden Jungen und Mädchen berücksichtigt. Unterrichtsarrangements im Sinne einer interessenssensiblen Leseförderung werden daher darauf achten, möglichst *interessante* Lektüren aus der Sicht der Schüler*innen auszuwählen bzw. diesen die Möglichkeit zu bieten, diese Auswahl selbst treffen zu können, da diese wiederum die lernbegleitenden positiven Emotionen bei den Lernenden fördern (vgl. Richter & Plath 2005, 2012; Garbe 2014a, 2014b).

Dies wiederum ist wichtig, wenn nun noch die Emotionspsychologie hinsichtlich ihres Beitrags zu einer nachhaltigen Leseförderung befragt wird. Bereits Willenberg (1999: 127) nannte die *Emotionen* die „heimliche Basis“ allen Lernens und verwies darauf, dass Emotionen nicht nur das Handeln allgemein determinieren (vgl. auch Lewis, Haviland-Jones & Barrett 2010), sondern v.a. auch aus lesedidaktischer Sicht eine entscheidende Rolle bei der schulischen Förderung des Lesens spielen. Dabei gilt es zu unterscheiden, ob es sich um *aktivierende positive* Emotionen wie Stolz oder Freude, um *aktivierende negative* Emotionen wie Angst vor schamvollen Situationen oder um *desaktivierende negative* Emotionen wie Panik, Hoffnungslosigkeit oder auch Langeweile z.B. hinsichtlich bestimmter Methoden (wie dem unvorbereiteten Reihumlesen) oder bestimmter unbeliebter Themen im Unterricht handelt (vgl. Richter & Plath 2005). Denn wo *positive* und begeisternde Emotionen kognitive Lernprozesse fördern, können *negative* Gefühle beim Lesen und Lernen genau das Gegenteil bewirken und sogar die Ausbildung von Kompetenzen gänzlich verhindern (vgl. Hasselhorn & Gold 2013). Die Berücksichtigung der emotionalen Befindlichkeiten beim Lesen von Schüler*innen ist vor diesem Hintergrund also stets notwendig, damit diese der eigentlichen Förderung mit Vorfreude entgegen- und im Nachhinein zufrieden auf diese zurückblicken können. Aus motivations- wie emotionspsychologischer Sicht muss bei erfolgsversprechenden Leseförderprojekten im schulischen Deutsch- oder im additiven Leseunterricht also für eine Methodenauswahl gesorgt werden, die es Schüler*innen allgemein, vor allem aber jenen schwach lesenden Mädchen und Jungen aus schriftfernen Welten und mit hoher sozialer Belastung, ermöglicht, diese Emotionen beim Lesen zu entwickeln, um sie dann wiederum als positive Wertzuschreibung einer Handlung für ihre eigene Lesemotivation nutzen zu können. Das Auslösen *aktivierender positiver* Emotionen⁴ ist in Folge dessen eines der Ziele, das sich das Hamburger BiSS-Projekt für die Primarstufe im Rahmen der Leseförderung gesetzt hat.

⁴ Vgl. hierzu Fußnote 3.

3 | Systematische Leseförderung und ihre Umsetzung in den teilnehmenden BiSS-Schulen

Nach Klärung der dem Hamburger BiSS-Projekt zugrundeliegenden Auffassung von Lesekompetenz und ihrer Förderung soll im Folgenden Anlage und Design der Studie in den Mittelpunkt gerückt werden. Hierzu wird zunächst die Umsetzung des Hamburger BiSS-Projekts erläutert, wobei zwischen organisatorisch-administrativen Anteilen auf der einen und konkret schulisch-methodischen Anteilen auf der anderen Seite unterschieden wird. Anschließend skizzieren wir die Hypothesen, denen wir im Rahmen dieser Studie nachgehen, und beschreiben das methodische Vorgehen zur deren Prüfung. Abschließend werden die Ergebnisse diskutiert und Implikationen für mögliche Nachfolgeprojekte aufgezeigt.

3.1 | Organisatorisch-administrative Voraussetzung: zur Leseförderung in verbindlichen Lesebändern

Die am Hamburger BiSS-Projekt teilnehmenden und durch Ausstattung mit Lektürematerial, kostenlosen Lehrer*innenfortbildungen und fortlaufenden Ergebnisevaluationen profitierenden Grundschulen mussten einige zentrale organisatorisch-administrative Voraussetzungen erfüllen und diese im Laufe des Projekts vertraglich bindend einhalten. So galt es, neben der obligatorischen Teilnahme an zwei Methodenworkshops pro Schuljahr für jede beteiligte Lehrkraft ein sogenanntes „Leseband“ (nach Gailberger & Hauschild 2018) fest in den schulischen Alltag der jeweiligen Schule zu implementieren, das je nach organisatorischen Möglichkeiten vor Ort drei bis fünf Mal in der Woche eine additive Lesezeit von mindestens zwanzig Minuten für die Schüler*innen garantierte. Alle Lehrkräfte der Schulkollegien waren an der im Leseband stattfindenden Förderung beteiligt, da diese ausdrücklich *nicht* dem Deutschunterricht zugeordnet, sondern – als Form zusätzlicher Sprachförderung – neutral als (verbindliche) Lesezeit organisiert und durchgeführt wurde. Zur Leseförderung im Rahmen eben dieser festen Lesezeit gehörte ferner das Ein- und Ausläuten des Lesebands, wodurch die Leseförderung zur festen Ausstattung des täglichen Unterrichts- und Schulalltags der Mädchen und Jungen avancierte. Das Leseband diente daher als verlässliche Größe nicht nur für die Einhaltung der zeitlichen Vorgaben, sondern auch als tägliche Instanz, der die Schüler*innen mit positiv aktivierenden Emotionen entgegenblicken konnten (vgl. hierzu zusammenfassend Hasselhorn & Gold 2013: 119-129).

Diese Form der organisierten, systematischen und eine gesamte Schule umspannenden Leseförderung hat nicht nur dazu geführt, dass die an ‚BiSS Hamburg‘ teilnehmende Grundschule Kirchdorf Preisträgerin des Deutschen Lesepreises 2017 wurde (vgl. Gailberger & Hauschild 2018), sie kann auch als Grundvoraussetzung für die Wirksamkeit der ausgewählten Lesefördermethoden gewertet werden, um die es im Folgenden gehen wird.

3.2 | Die in der täglichen Schulpraxis verwendeten Lesetrainingsmethoden

Wie in der Einleitung bereits angedeutet wurde, war dem Hamburger BiSS-Projekt für die Primarstufe, von dem hier aus theoretischer sowie im Folgenden aus empirischer wie praxis- und unterrichtsnaher Perspektive die Rede sein soll, eine einjährige Vorlauf- und Pilotierungsphase vorgeschaltet. Diese erachteten wir als notwendig, um *erstens* unter Einsatz empirisch-diag-

nostischer Tests sowie *zweitens* mithilfe von Austauschrunden, die zusammen mit den Sprachlernkoordinator*innen und den Deutschlehrkräften der beteiligten Schulen durchgeführt wurden, den Stand der Lesegeläufigkeit und der Lesekompetenz der Schüler*innen dieser Schulen zu ermitteln, um in einem nachfolgenden Schritt wiederum Entscheidungen über die einzusetzenden Lesefördermethoden entscheiden zu können. Aufgrund der intern gesammelten Erkenntnisse dieser Vorlauf- und Pilotierungsphase entschied sich das Hamburger Primarstufen-Projekt für die im Schuljahr 2015/2016 startende Kernkohorte dazu, im Rahmen des in den Schulen installierten Lesebandes niedrigschwellige Verfahren einzusetzen, die als Lautleseverfahren a) zuvörderst die Lesegeläufigkeit fördern, dabei b) aus organisatorischer wie didaktisch-methodischer Sicht sowohl bei den Lehrenden wie bei den Lernenden als beliebt gelten und somit c) neben dem Lesen aus kognitiver Perspektive ebenso die Lesemotivation wie die positiv aktivierenden Emotionen der Jungen und Mädchen beeinflussen. Dabei kamen sowohl Lautleseverfahren zum Einsatz, die bereits weitgehend Einzug in den landläufigen Grundschulalltag gefunden haben, als auch Methoden, die für das Projekt neu entwickelt oder spezifiziert wurden und die – im Sinne von Binnendifferenzierung und Umgang mit Heterogenität – die Bandbreite, die Variabilität und damit auch die Flexibilität der teilnehmenden Lehrer*innen in ihren individuellen Lerngruppen erhöhten.

Die folgenden Abschnitte skizzieren, nach aufsteigender Komplexität für die beteiligten Schüler*innen geordnet, jene Methoden, die im Rahmen des Projekts in den Lesebänden am häufigsten zur Anwendung kamen.

Vorlesen und Mitlesen

Das **Vorlesen und Mitlesen** ist ein lautes wie leises Lesen, bei dem ein zuvor ausgewähltes Buch bzw. eine zuvor ausgewählte Geschichte je nach Lesefertigkeit der Kinder zehn bis zwanzig Minuten von der Lehrkraft vor- und von den Schüler*innen mitgelesen wird. Die Kinder nutzen die gut hörbare Stimme als Lese-Modell, sie lehnen sich an dieses Modell an und kompensieren die eigenen Leseschwierigkeiten, indem sie simultan *leise* (das heißt entweder stumm oder halblaut) in den eigenen Texten mitlesen.

Wird mit den Kindern eine Ganzschrift gelesen, so hat es sich bewährt, die Bücher nach Beendigung der Leseinheit wieder einzusammeln. Diese Maßnahme ist in schwach lesenden Klassen deswegen von Vorteil, da somit keines der schon stärker lesenden Kinder dazu verleitet werden kann, den ausgewählten Text alleine weiterzulesen und ggf. den Klassenkameraden (z.B. aus Stolz oder Missgunst) den Ausgang der Geschichte zu verraten. Ein weiterer Vorteil dieser unbestritten einschränkenden Maßnahme liegt darin, das beliebte ‚Leseband‘ als *positiv* aufgeladene Lese-Situation zu verfestigen und es somit zu einem Ort zu machen, auf den sich die Kinder (bereits in der Pause/vor dem Einläuten) freuen können (vgl. Meyer 2011).

In den ersten zwei oder drei Wochen kann es vorkommen, dass nicht alle Kinder die Konzentration aufbringen, länger als zehn Minuten am Stück zu lesen. Vor allem die ganz schwach lesenden Kinder dürfen an dieser Stelle nicht überfordert werden, sonst verlieren sie schnell die Motivation, weiterhin an der Leseförderung teilzunehmen. In solchen Fällen sollte unbedingt vorzeitig gestoppt und die Textstelle markiert werden, um mit der Lektüre am Folgetag fortzufahren.

Das Vorlesetheater

Das **Vorlesetheater** (oder auch nur *Lesetheater*) stammt, ebenso wie die Methode *Vorlesen und Mitlesen*, aus dem Pool der Lautlese-Verfahren, kombiniert diese aber mit Aspekten der Wiederholung (vgl. ausführlich Nix 2006). Bei Lesetheater-Projekten werden vorzugsweise kurzepische Texte im narrativen Modus in einen quasi dramatischen Modus überführt; es können aber auch Teile bzw. Kapitel von Ganzschriften präpariert werden. Die so gewonnenen Lese-Scripts, die die Dialoge und Gedanken der Figuren und des Erzählers in direkter Rede enthalten, können je nach Leistungsniveau der Klasse entweder von den Schüler*innen selbst oder aber im Vorfeld von der Lehrkraft angefertigt werden. Leichter zu realisieren sind freilich Texte, die bereits dramatisiert vorliegen, z.B. in Form kurzer Theaterstücke oder lustiger Sketche.

Die Lehrkraft kann den Text zu Beginn einmal laut vorlesen. Da es allerdings auch Lerngruppen gibt, die sich die Geschichte selbst erlesen wollen, ist dies eine fakultative Vorgabe. Jedes Kind übernimmt nun die Rolle einer literarischen Figur, die Erzählinstanz in narrativen Texten sollte je nach quantitativer Dominanz unter mehreren Kindern aufgeteilt werden. Die Übungsroutine wird von Rosebrock und Nix (2017: 47f.) folgendermaßen beschrieben:

„Die Kinder üben mit diesen Scripts in Gruppenarbeit zum einen, eine von ihnen ausgewählte Figur des Textes durch das Vorlesen ‚zum Leben zu erwecken‘, also durch verschiedene sprecherische Mittel (vgl. dazu Ockel 2000) die charakterlichen Eigenschaften, Gefühle, Gedanken und Motive der Figur im Kontext der jeweiligen Handlungssituation möglichst angemessen stimmlich zu interpretieren. Andererseits müssen die Schülerinnen und Schüler ihren individuellen Vortrag mit den anderen Sprechern der Gruppe abstimmen, im Gruppenprozess die klangliche Inszenierung des Textes diskutieren und so eine Gesamtdeutung des Gelesenen zusammen erarbeiten. Die Zielsetzung des Lesetheaters besteht für die Schülerinnen und Schüler einer Gruppe darin, sich durch das wiederholte Lautlesen und den Austausch in der Gruppe für die abschließende Lese-Aufführung des Textes vorzubereiten: Im Lesetheater soll den Zuhörerinnen und Zuhörern der Text möglichst bildhaft und einprägsam, eben ‚lebendig‘, vor Augen geführt werden.“

Aufwerten und ergänzen lassen sich Lesetheater-Projekte, indem sie mit bekannten Routinen des Kasperle- bzw. Puppentheaters zusammengebracht werden. Hierbei kann es (durch Los, durch Freiwilligkeitsentscheidungen oder durch regelmäßige Abwechslung) zu einer Teilung der Gruppe in *Vorleser* auf der einen Seite und *Puppenspieler* auf der anderen Seite kommen.

Lesen mit dem Ich-Du-Wir-Würfel

Eigens für das Hamburger BiSS-Projekt wurde die Methode **Lesen mit dem Ich-Du-Wir-Würfel** entwickelt. Dabei handelt es sich um eine Art Gesellschaftsspiel, welches den Förderprinzipien des *wiederholten* Lautlesens folgt. Es setzt am Spaß in der Gruppe an und nutzt diesen lesedidaktisch aus. Das ‚Würfel-Lesen‘ wird am besten in Vierergruppen gespielt. Der vorher zu präparierende (oder zu erwerbende) Extra-Würfel wird jeweils zwei Mal mit den Seiten ICH, DU und WIR ausgestattet. Die ausgewählten Texte müssen in vier bis sechs etwa gleichgroßen Abschnitten vorliegen. Die Länge eines Abschnitts sollte in Abhängigkeit von der Leseflüssigkeit der Kinder 30 bis 50 Wörter nicht überschreiten. Sitzen die Kinder in ihren Gruppen im Kreis, beginnt das erste Auswürfeln, z.B. in alphabetischer Reihenfolge der Vornamen. Würfelt das erste Kind ein ICH, liest es den ersten Abschnitt halblaut vor. Bei DU darf ein Dritter als Vorleser bestimmt werden. Die anderen Kinder lesen jedes Mal still im eigenen Text simultan mit. Fällt die WIR-Seite, liest die gesamte Gruppe im Chor. Sollte es von den Mitgliedern der Gruppe erwünscht sein, dürfen die bereits besser lesenden Kinder immer dann leise einhelfen, wenn

ein lesendes Kind stockt oder sich verliert. In der zweiten Würfelrunde liest das nächste würfelnde Kind den vorherigen, also bereits bekannten Abschnitt erneut laut vor, danach dann den eigenen, neuen Absatz. In der dritten Runde wiederholt sich diese Routine mit dem nächsten Kind, erst danach liest dieses wiederum den eigenen Textabschnitt. Dieses Vorgehen wiederholt sich so oft, wie der Text Abschnitte hat. Zum Abschluss dürfen ein Kind oder alle Kinder im Chor den Text laut (und dann auch flüssig) vorlesen.

Lesen mit Hörbüchern

Das **Lesen mit Hörbüchern** ist eine denkbar einfache und zugleich sehr effektive Form der Förderung, die bereits für die Sekundarstufe 1 (vgl. Gailberger 2013) sowie in Ansätzen auch für die Primarstufe (Boll & Scholz 2012; Meyer 2011) empirisch erprobt wurde. Dabei wurde gezeigt, dass (in der Primarstufe ebenso wie in der Sekundarstufe 1) nach nur wenigen Wochen die Leseflüssigkeit (auf kognitiver Ebene), mehr noch aber die Lesemotivation, die Freude am Lesen im Deutschunterricht und weitere Determinanten des Lesens (auf subjektiver Ebene) gesteigert werden konnten. Darüber hinaus wertet das Lesen mit Hörbüchern das Lesen in der Schule an sich auf, indem das gemeinsame simultane Lesen und Hören eine ‚literarische Geselligkeit‘ erzeugt, die von den Kindern wie von ihren Lehrkräften zu gleichen Teilen genossen wird (vgl. Gailberger 2013).

Wie schon beim *Lesen und Mitlesen* läuft auch beim Hörbuchlesen für zehn bis zwanzig Minuten im Rahmen des Lesebandes ein Hörbuch in angenehmer Zimmerlautstärke. Auch hier sollte die Lesedauer in Abhängigkeit von der Übung und der Leseflüssigkeit der Kinder gewählt und im Laufe der Förderung gesteigert werden. Da das Lesen mit Hörbüchern bei ‚BiSS Hamburg‘ aber frühestens gegen Ende der dritten Klasse eingesetzt wird, sind grundsätzliche Probleme des Lesens wie bei der Methode *Vorlesen und Mitlesen* nicht zu erwarten. Dennoch ähnelt vieles der ersten Methode, so auch das Einsammeln der Texte nach Beendigung der täglichen Förderung (aus denselben Gründen).

3.2 | Probleme bei der Implementierung weiterer Fördermaßnahmen

Wie weiter oben angedeutet wurde, fördern die für das BiSS-Projekt vorgesehenen wiederholenden und/oder begleitenden Lautleseverfahren vornehmlich die Lesegeläufigkeit sowie einige weitere Determinanten des Lesens (wie bspw. die Lesemotivation), die aber ‚lediglich‘ als Vorbedingungen dafür gelten, dass Schüler*innen auch Lesekompetenz ausbilden können (vgl. aus psychometrischer Sicht Gailberger 2019, aus didaktischer Sicht vgl. Rosebrock & Nix 2017). Für eine systematische (Sprach)Förderung des Lesens über die (Lese)Geläufigkeit hinaus wurden den beteiligten Lehrkräften daher ebenso Förderansätze wie *Lesetagebücher*, einfache *Lesestrategien* sowie Kopiervorlagen zur *Wortschatzarbeit* zur Verfügung gestellt. Diese Förderinstrumente, namentlich die verwendeten Lesetagebücher „Das Lesetagebuch für Erzähltexte. 2.–4. Klasse“ (Wietzke 2011), „Das Lesetagebuch für Sachtexte. 3./4. Klasse“ (Wietzke 2013) sowie die verwendeten Lesestrategien aus „LESEN: Das Training“ (Kruse et al. 2011) folgen in ihrer Fördersystematik der grundlegenden Struktur ‚Vor dem Lesen‘, ‚Während des Lesens‘ sowie ‚Nach dem Lesen‘ und stützen sich in ihrer behutsamen Vorgehensweise auf lesetheoretische Annahmen wie lesedidaktische Prämissen, die mittlerweile als gesichert gelten und erfolgreich für die Primarstufe (wie für die Sekundarstufe) aufbereitet wurden (vgl. im Überblick Nix 2019).

Im Rahmen des in den oben beschriebenen Lesebänden organisierten BiSS-Projekts zeigte sich allerdings schnell und unmissverständlich, dass das Lesen der beteiligten Schüler*innen auch noch in den Jahrgängen 3 und 4 im Mittel zu schwach ausgebildet und daher die Dauer von 20 Minuten Lesezeit nicht ausreichend war, um neben den regelmäßig stattfindenden Lautleseverfahren auch noch (wie eigentlich geplant) Lesetagebücher, Lesestrategien oder Methoden der Wortschatzarbeit zu implementieren, weswegen diese im Laufe des Projekts gar nicht erst zum Einsatz kamen. Inwiefern sich dies (negativ?) auf die Veränderungen der teilnehmenden Kinder im Lesen ausgewirkt hat und ob daher von einem Scheitern des Projekts zur Förderung der Lesekompetenz von strukturell benachteiligten Schüler*innen der Jahrgangsstufe 2 bis 4 gesprochen werden muss, sei in den nachfolgenden Abschnitten erläutert.

4 | Hypothesen

Wie bereits betont, wird mit dem Hamburger BiSS-Projekt das Ziel verfolgt, die Lesekompetenz insbesondere von Schüler*innen aus sozial benachteiligten Milieus zu fördern, deren Familiensprache nicht Deutsch ist. Zusätzlich zur Förderung der Lesekompetenz, die im Wesentlichen über die Verbesserung der Leseflüssigkeit erfolgt, sollen Leseaktivität und Lesekompetenz durch motivierende Elemente des Trainings unterstützt werden. Durch damit einhergehende Transfereffekte sollen über erhöhte Lesekompetenz auch Kompetenzen in weiteren Domänen, wie im Bereich der Rechtschreibung, der Mathematik, den Naturwissenschaften und in Englisch Hörverstehen, die ebenfalls über zu erlesende Aufgaben erfasst werden, gestärkt werden.

Konkret werden folgende Hypothesen überprüft:

- (1) Die Teilnahme am BiSS-Lesetraining führt zu einer Steigerung der Leseflüssigkeit.
- (2) Die Steigerung der Leseflüssigkeit führt zu einem besseren Leseverstehen.
- (3) Die Teilnahme am BiSS-Lesetraining führt neben einer Verbesserung des Leseverstehens auch zur Verbesserung der Kompetenzen in weiteren Domänen (z.B. Rechtschreibung, Mathematik etc.).
- (4) Die Trainingseffekte zeigen sich insbesondere bei Schüler*innen mit schwächeren Ausgangslagen im Lesen.

5 | Methodisches Vorgehen

5.1 | Design der Studie

Die Studie ist als Längsschnittuntersuchung im Kontrollgruppen-Design angelegt. Die Entwicklung der Leseflüssigkeit der am Projekt beteiligten Kinder wird anhand eines standardisierten Tests zu sieben Testzeitpunkten untersucht und dabei zu drei Messzeitpunkten mit der Kontrollgruppe verglichen (**Hypothese 1**).

Um zu überprüfen, wie sich das Training auf die Lesekompetenz und die Kompetenzen in anderen Domänen auswirkt, werden vorliegende Daten aus den regelhaft in Hamburg durchgeführten Kompetenzmessungen KERMIT (= *Kompetenzen ermitteln*) herangezogen⁵. Dabei werden zunächst Trainings- und Kontrollgruppe hinsichtlich der Lesekompetenz verglichen und die Zusammenhänge zwischen Trainingsteilnahme, Leseflüssigkeit und Leseverstehen untersucht. Die Ergebnisse liefern Befunde, inwieweit sich das Training der Leseflüssigkeit auch auf das Leseverstehen auswirkt (**Hypothese 2**).

Zur Überprüfung der Hypothesen 3 und 4 wird eine neue Referenzgruppe gebildet, mit der die Trainingsgruppe hinsichtlich der KERMIT-Ergebnisse in den unterschiedlichen Domänen verglichen wird. Unter Nutzung schulstatistischer Daten wird über ein *Propensity Score Matching* eine Vergleichsgruppe sogenannter „statistischer Zwillinge“ generiert, die der Trainingsgruppe hinsichtlich verschiedener soziodemographischer Daten nahezu identisch ist. Berücksichtigt werden beim Matching-Verfahren die Merkmale Geschlecht, Migrationshintergrund, sozioökonomischer Status, Vorhandensein eines additiven Sprachförderbedarfs sowie soziale Belastung der Schule.⁶ Mit diesem Verfahren kann aus der gesamten Schüler*innenkohorte der Jahrgangsstufe eine Vergleichsgruppe gezogen werden, die sehr ähnliche Rahmenbedingungen aufweist wie die Trainingsgruppe und somit einen adäquaten Vergleich ermöglicht. Dies ist besonders wichtig, da für den Vergleich der Kompetenzstände in den unterschiedlichen Domänen (**Hypothese 3**) keine Ausgangsleistungen verfügbar sind. Für die Prüfung differentieller Effekte einzelner Teilgruppen (**Hypothese 4**) ist darüber hinaus eine annähernd gleiche Zusammensetzung der Gruppen erforderlich, die durch das Matching-Verfahren gewährleistet ist.

5.2 | Stichprobe

Da das Lesetraining insbesondere auf Schüler*innen aus bildungsbenachteiligten Familien abzielt, wurde die Beteiligung am Training vorrangig Schulen mit geringerem Sozialindex angeboten. Der Sozialindex ist ein Maß zur Bestimmung der sozialen Belastung von Schulen, der von 1 (stark belastet) bis 6 (kaum belastet) reicht. Über das gesamte Hamburger Stadtgebiet verteilt nahmen insgesamt sechs Schulen an der Förderung teil, die zuvor im Rahmen einer offenen Ausschreibung für das Projekt gewonnen werden konnten. Aus zwei weiteren vergleichbaren Schulen bildete sich die Kontrollgruppe. Alle Schulen weisen – mit zwei Ausnahmen – den Sozialindex 1 auf. Das Projekt wurde also an Standorten durchgeführt, an denen ein großer Teil der Schülerschaft im Sinne eines weiten Inklusionsbegriffs (vgl. Deutsche UNESCO-Kommission 2014) von unterschiedlichsten Marginalisierungsrisiken bedroht ist; sei es von Armut, von sprachlichen Aspekten wie Deutsch als Zweitsprache/DaZ, von Migration oder von Aspekten sozial-emotionaler Entwicklung.

⁵ Die Daten aus den verschiedenen Erhebungen wurden nach datenschutzrechtlicher Prüfung durch die Vertrauensstelle der Behörde für Schule und Berufsbildung (BSB) verknüpft und zur Durchführung der beantragten Analysen bereitgestellt.

⁶ Bei Interesse zu genaueren Angaben der Methodik und Güte der Matching-Verfahren können diese bei den Autor*innen erfragt werden.

Trainingsgruppe

Die Trainingsgruppe besteht insgesamt aus N = 284 Schüler*innen⁷ (137 Mädchen und 147 Jungen), die sich im Schuljahr 2015/16 in der zweiten Jahrgangsstufe befanden. 78 Prozent der Kinder weisen einen Migrationshintergrund und rund 36 Prozent einen additiven Sprachförderbedarf auf. Als Indikator für den sozioökonomischen Status wird der sogenannte RISE-Status⁸ genutzt, der in vier Stufen eingeteilt wird (sehr niedrig, niedrig, mittel und hoch). Aus der Trainingsgruppe werden 41 Prozent der Kinder der Gruppe mit dem sehr niedrigen RISE-Status zugeordnet, 29 Prozent der Kinder der Gruppe mit dem geringen RISE-Status und 29 Prozent der Gruppe mit dem mittleren Status. Lediglich 1 Prozent gehören der Gruppe mit dem hohen RISE-Status an.

Kontrollgruppe

Die Kontrollgruppe umfasst 90 Kinder (53 Mädchen und 37 Jungen). Es zeigt sich hinsichtlich des Migrationshintergrunds eine ähnliche Verteilung wie in der Trainingsgruppe, 76 Prozent weisen einen Migrationshintergrund auf. Bei rund einem Fünftel liegt ein additiver Sprachförderbedarf vor (19%). Bei der Zuordnung zum RISE-Status liegt eine leicht ungünstigere Verteilung als in der Trainingsgruppe vor: Die Mehrheit der Kinder (46%) wird dem sehr niedrigen RISE-Status zugeordnet, ein gutes Drittel (34%) dem niedrigen RISE-Status und ein Fünftel (20%) dem mittleren Status.

Matching-Gruppe

Aus der gesamten Alterskohorte der 18.659 Schüler*innen wurde für jedes Kind der Trainingsgruppe ein „statistischer Zwilling“ gezogen (Ziehen ohne Zurücklegen). Daher hat die Matching-Gruppe mit 284 Kindern denselben Umfang wie die Trainingsgruppe. Auch die Verteilung der zentralen soziodemographischen Daten entspricht aufgrund des Bildungsgrades der Matching-Gruppe mittels eines Propensity Score Matchings nahezu der Verteilung der Trainingsgruppe. Sie besteht aus 154 Jungen und 130 Mädchen, womit der Anteil der männlichen Gruppe etwas höher ist als in der Trainingsgruppe. Bei 80 Prozent von ihnen liegt ein Migrationshintergrund vor, 35 Prozent weisen einen additiven Sprachförderbedarf auf. Aus der Matching-Gruppe gehören 41 Prozent der Kinder der Gruppe mit sehr niedrigem RISE-Status an, 30 Prozent der Gruppe mit niedrigem RISE-Status und 29 Prozent der Gruppe zählen zum mittleren RISE-Status.

⁷ Für den Vergleich von Trainingsgruppe und Kontrollgruppe wurden nur die Daten von 215 Schüler*innen herangezogen, da nur für diese Kinder Ergebnisse zu den drei Messzeitpunkten im Salzburger Lese-Screening vorlagen.

⁸ Durch das „Rahmenprogramm Integrierte Stadtteilentwicklung“ (RISE) sollen Quartiere mit besonderem Förderungsbedarf identifiziert und unterstützt werden. Dazu werden verschiedene soziale Merkmale wie der Anteil von Schüler*innen ohne Schulabschluss, Empfänger*innen von Leistungen der Grundsicherung und der Anteil der Arbeitslosen erfasst. Die statistischen Gebiete werden anhand der Merkmale einem Statusindex mit den Kategorien sehr niedrig, niedrig, mittel und hoch zugeordnet (vgl. Behörde für Stadtentwicklung und Wohnen, www.hamburg.de).

5.3 | Eingesetzte Testverfahren

Salzburger Lese-Screening (SLS)

Zur Erfassung der Leseflüssigkeit wurde das Salzburger Lese-Screening 2-9 von Mayringer und Wimmer (2014) eingesetzt. Das Salzburger Lese-Screening ist ein häufig verwendetes und daher sehr anschlussfähiges Verfahren zur Feststellung von Defiziten in der basalen Lesefertigkeit von Schüler*innen, welche sich in einer langsamen Lesegeschwindigkeit zeigen (zur kritischen Beleuchtung des SLS aus linguistischer Perspektive vgl. Helm in diesem Band). Während der dreiminütigen Bearbeitungsdauer des Lese-Screenings sind die Schüler*innen dazu aufgefordert, eine Abfolge von möglichst vielen Sätzen leise zu lesen und hinter jedem Satz zu markieren, ob die inhaltliche Aussage des Satzes als richtig oder falsch zu bewerten ist. Über die Anzahl der fehlerfrei beurteilten Sätze, den Rohwert, lässt sich anhand einer Normtabelle der entsprechende Lesequotient ableiten. Der Lesequotient bietet Auskunft darüber, wie weit die basale Lesefertigkeit vom Durchschnitt der Normierungsstichprobe abweicht.

KERMIT

KERMIT (= *Kompetenzen ermitteln*) steht für ein verbindliches Testverfahren zur Erfassung der Kompetenzen von Schüler*innen in ausgewählten Jahrgangsstufen. KERMIT⁹ findet für Hamburger Schüler*innen in den Klassen 2, 3, 5, 7, 8 und 9 statt. In allen Jahrgangsstufen werden die Lesekompetenz und die mathematischen Kompetenzen der Schüler*innen, sowie zusätzlich Rechtschreibung (Jahrgang 3, 5 und 7) und in der weiterführenden Schule Englisch in Teilaspekten (Jahrgang 5, 7, 8 und 9) und Naturwissenschaften (5, 7 und 9) erhoben. Ziel dieses Verfahrens ist die Unterstützung der Schulen bei ihrer datengestützten Schul- und Unterrichtsentwicklung. KERMIT wird im Rahmen des BiSS-Projekts herangezogen, um über die Ergebnisse zur Leseflüssigkeit hinaus Informationen zur Lesekompetenz der Schüler*innen zu erhalten, sowie um mögliche Transfereffekte auf andere Domänen untersuchen zu können.

5.4 | Statistische Analysen

Zur Prüfung der Leistungsunterschiede zwischen der Trainingsgruppe und der Kontrollgruppe zu den drei Testzeitpunkten im Salzburger Lese-Screening wird eine einfaktorielle Varianzanalyse mit Messwiederholung berechnet. Für die Vergleiche zwischen Trainingsgruppe und Matching-Gruppe werden jeweils t-Tests für unabhängige Stichproben gerechnet. Da die Hypothesen eine Verbesserung der Leistungen zugunsten der Trainingsgruppe postulieren, wurden einseitige Testungen durchgeführt. Zur besseren Interpretierbarkeit der Effekte und Einschätzung der praktischen Relevanz werden bei statistisch signifikanten Unterschieden zusätzlich die Effektstärken berechnet (vgl. Cohen 1988). Die Zusammenhangsanalysen werden über Strukturgleichungsmodelle mittels der Statistiksoftware Mplus 7.0 (Muthén & Muthén 2012) vorgenommen. Um aus der Stichprobe aufgrund fehlender Werte nicht zu viele Fälle ausschließen zu müssen, werden durch multiple Imputation fehlende Werte geschätzt. Die soziodemographischen Daten und Leistungsdaten werden als Prädiktoren für die zu schätzenden Werte verwendet und die Analysen mit den gemittelten Werten aus 10 Imputationen berechnet.

⁹ Die in Hamburg zum Einsatz kommenden KERMIT 3- und 8-Testungen sind identisch mit den bundesweit üblichen VERA 3- und 8-Tests, ebenso wie auch alle weiteren KERMIT-Verfahren analoge Instrumente zu VERA darstellen. Auf ihre psychometrischen Unterschiede im Detail kann hier aus Platzgründen allerdings nicht eingegangen werden.

6 | Ergebnisse

6.1 | Auswirkungen der Fördermethoden auf die Leseflüssigkeit

In Abbildung 1 wird die Entwicklung der Lesegeschwindigkeit – gemessen über das SLS – für die Trainingsgruppe und die Kontrollgruppe über die drei Messzeitpunkte dargestellt (Anfang Jahrgangsstufe 2, Ende Jahrgangsstufe 3, Ende Jahrgangsstufe 4). Außerdem ist der Erwartungswert aus der Normstichprobe des SLS aufgeführt.

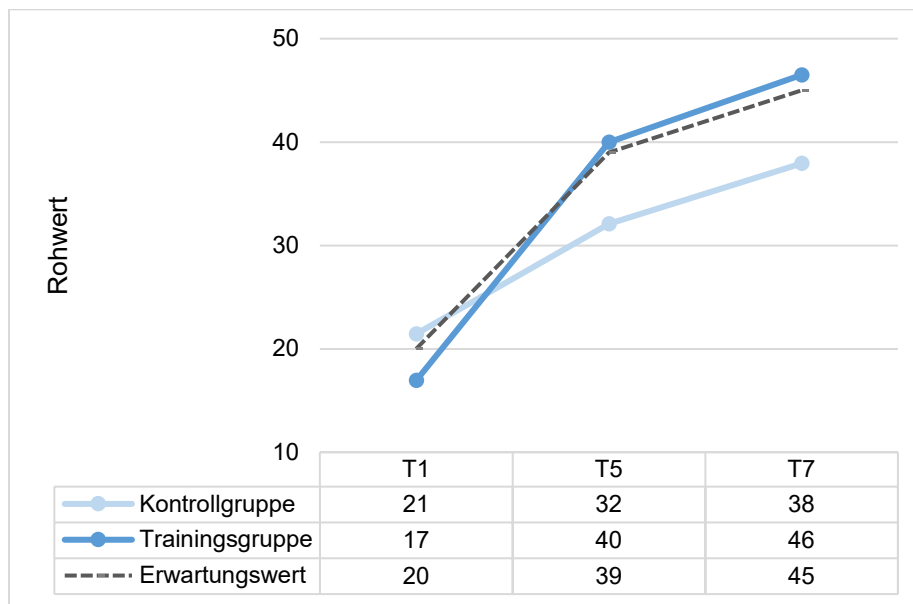


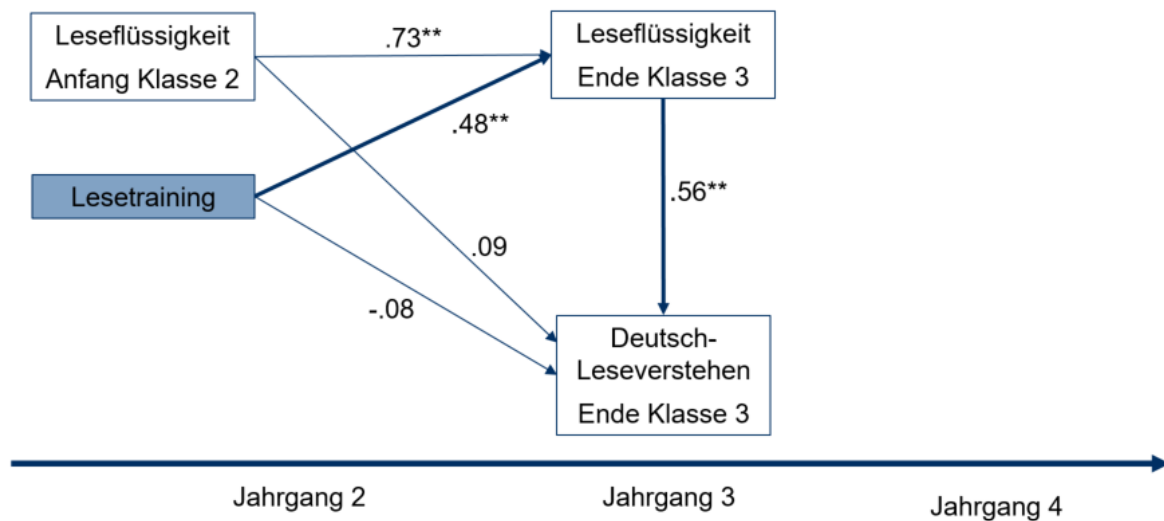
Abb. 1: Entwicklung der Lesegeschwindigkeit.

Die Ausgangswerte machen deutlich, dass es sich bei den am Projekt beteiligten Schüler*innen tatsächlich um eine Gruppe mit schwachen Lese-Ausgangslagen handelt. Die Ergebnisse der weiteren Messzeitpunkte zeigen jedoch auf, dass sich die Gruppe im Projektverlauf verbessert und den Erwartungswert am Ende sogar leicht übersteigt. Die Kontrollgruppe unterscheidet sich beim ersten Messzeitpunkt am Anfang der zweiten Jahrgangsstufe mit durchschnittlich 21 richtig beurteilten Sätzen von dem Mittelwert der Trainingsgruppe mit durchschnittlich 17 richtig beurteilten Sätzen. Diesen Rückstand kann die Trainingsgruppe am Ende der dritten Jahrgangsstufe aufholen und übertrifft mit einem Rohwert von 40 die Kontrollgruppe um 8 richtig beurteilte Sätze. Auch zum Ende der vierten Jahrgangsstufe liegt der Rohwert der Trainingsgruppe mit 46 richtig beurteilten Sätzen deutlich über dem Rohwert von 38 der Kontrollgruppe. Trotz einer schlechteren Ausgangslage am Anfang der zweiten Jahrgangsstufe im Vergleich zur Kontrollgruppe kann die Trainingsgruppe durch die Teilnahme am BiSS-Lesetraining ihre Defizite in der Leseflüssigkeit ausgleichen.¹⁰ Es ergibt sich eine signifikante Interaktion, $F(2,606) = 154.33, p < .01$.

¹⁰ In einem vorab veröffentlichten Zwischenbericht zum Hamburger BiSS-Projekt Primarstufe konnten die mithilfe des Salzburger Lesescreenings *psychometrisch* gewonnenen Ergebnisse zur signifikanten Steigerung der Lesegeschwindigkeit durch insgesamt 37 edumetrisch ausgewertete Lautleseprotokolle (auf den zusätzlichen Ebenen Dekodiergenauigkeit, sinnvolle Betonung und Lesegeschwindigkeit *WpM*) ergänzt werden (vgl. Gailberger 2018, 125f.).

6.2 | Wirkmechanismen zum Einfluss der Leseflüssigkeit

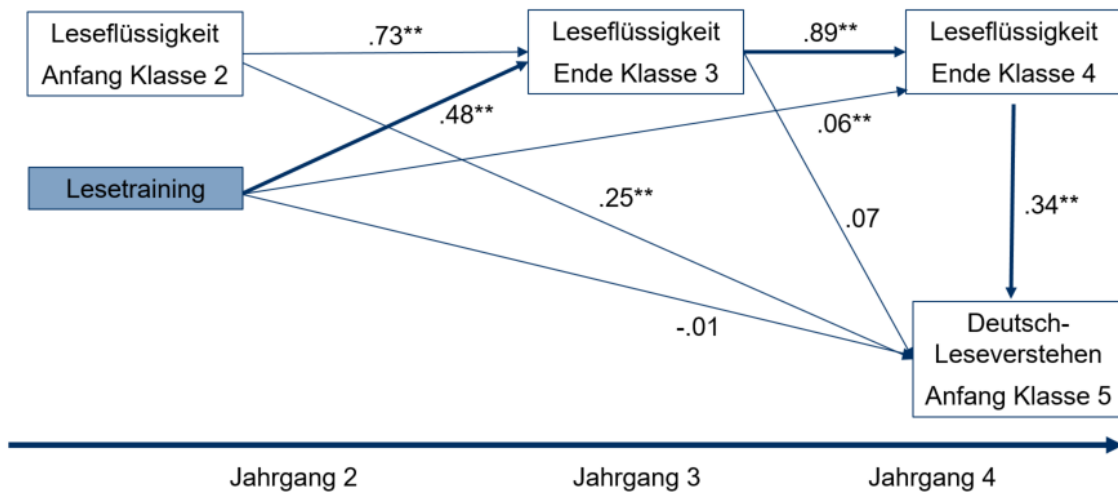
Anhand von Strukturgleichungsmodellen wird der Zusammenhang zwischen der Teilnahme am Lesetraining, der Leseflüssigkeit und dem Leseverstehen überprüft. Dabei wird zunächst der Einfluss auf das Leseverstehen am Ende von Jahrgangsstufe 3 (KERMIT 3) untersucht (siehe Abbildung 2). Wie unter Punkt 6.1 berichtet, zeigt sich ein positiver Einfluss der Trainingsteilnahme auf die Leseflüssigkeit am Ende von Jahrgangsstufe 3 ($\beta = .48^{**}$). Bei Kontrolle der Ausgangsleistung erzielen die Kinder der Trainingsgruppe demnach deutlich bessere Ergebnisse im Leseflüchtigkeits-Tests als die Kinder der Kontrollgruppe. Die Leseflüchtigkeit wiederum wirkt sich positiv auf das Leseverstehen aus ($\beta = .56^{**}$). Insgesamt ergibt sich damit ein bedeutsamer Effekt vom Lesetraining auf das Leseverstehen am Ende der Klassenstufe 3 ($\beta = .19^{**}$), der nahezu vollständig über die Leseflüchtigkeit vermittelt ist.



Anmerkungen: * = $p \leq .05$, ** = $p \leq .01$

Abb. 2: Einfluss auf das Leseverstehen Ende Klasse 3.

In einem zweiten Modell wird der Einfluss auf das Leseverstehen am Anfang von Klassenstufe 5 (KERMIT 5) untersucht. Dabei zeigt sich, dass sich der postulierte Wirkungszusammenhang auch über eine längere Zeitspanne nachweisen lässt und sich die Trainingseffekte auch auf die KERMIT-Ergebnisse am Anfang der Jahrgangsstufe 5 niederschlagen. Der Vorsprung in der Leseflüchtigkeit der Trainingsgruppe bleibt am Ende von Klassenstufe 4 bestehen und führt zu einem verbesserten Leseverstehen am Anfang der Klassenstufe 5 ($\beta = .34^{**}$). Der vermittelnde Einfluss der Leseflüchtigkeit lässt sich somit auch über die Primarstufe hinweg auf das Leseverstehen zu Beginn der Sekundarstufe I beobachten.



Anmerkungen: * = $p \leq .05$, ** = $p \leq .01$

Abb. 3: Einfluss auf das Leseverstehen Anfang Klasse 5

6.3 | Effekte auf verschiedene Kompetenzbereiche

Im Folgenden werden die Effekte des Trainings auf verschiedene Kompetenzbereiche untersucht. Dazu werden die jeweiligen Ergebnisse der KERMIT-Erhebungen der Trainingsgruppe mit denen der Matching-Gruppe verglichen. Grundsätzlich wird bei jeder KERMIT-Erhebung der Mittelwert auf 500 Punkte gesetzt. Dabei werden bei KERMIT 2 und 5 die Werte aller Hamburger Schüler*innen auf 500 Punkte gesetzt, bei KERMIT 3 handelt es sich hingegen um ein bundesweites Verfahren. Daher können die Mittelwerte verschiedener Erhebungen nicht unmittelbar miteinander verglichen werden.

| | Trainings-Gruppe | | Matching-Gruppe | | $t(df)$ | p | Cohen's d |
|----------|------------------|--------|-----------------|--------|------------|-----|-------------|
| | M | (S.E.) | M | (S.E.) | | | |
| KERMIT 2 | 458,79 | (5,10) | 442,98 | (5,96) | 2,03 (444) | .02 | 0.19 |
| KERMIT 3 | 428,44 | (7,76) | 400,21 | (7,99) | 2,49 (260) | .01 | 0.31 |
| KERMIT 5 | 459,93 | (6,18) | 443,66 | (6,18) | 1,80 (69) | .04 | 0.43 |

Anmerkungen: d = Effektgröße nach Cohen; $d \geq 0.2$ kleiner Effekt; $d \geq 0.5$ mittlerer Effekt; $d \geq 0.8$ großer Effekt; $p \leq .05$ signifikant; $p \leq .01$ hoch signifikant (1-seitig)

Tab. 1: Vergleich der Kompetenzwerte im Bereich Leseverstehen

Leseverstehen

Im ersten Schritt werden die Effekte des Trainings auf das Leseverstehen untersucht. Bereits am Ende von Jahrgangsstufe 2 sind signifikante Kompetenzunterschiede im Bereich Leseverstehen zugunsten der Trainingsgruppe vorhanden, ($t(444) = 2,03$, $p = .02$). Der Effekt ($d = 0.19$) ist allerdings gering. In Jahrgangsstufe 3 wird der Mittelwertunterschied zwischen den beiden Gruppen erneut signifikant, ($t(260) = 2,49$, $p = .01$). Mit einer Effektstärke von $d = 0.31$ hat

dieser Unterschied eine höhere Relevanz als in Jahrgangsstufe 2. Den Vorsprung in den Lesekompetenzwerten behält die Trainingsgruppe in Jahrgangsstufe 5 bei. Auch hier wird der Unterschied in den Kompetenzwerten der Trainingsgruppe und der Matching-Gruppe signifikant ($t(69) = 1,80, p = .04$). Vor allem ist aber darauf hinzuweisen, dass sich die Effektstärke nochmal erhöht hat ($d = 0.43$).

Rechtschreibung

Im Bereich Rechtschreibung erzielt die Trainingsgruppe zu beiden KERMIT-Testzeitpunkten bessere Ergebnisse als die Matching-Gruppe.

| | Trainings-Gruppe | | Matching-Gruppe | | $t(df)$ | p | Cohen's d |
|----------|------------------|------------|-----------------|------------|------------|-----|-------------|
| | M | ($S.E.$) | M | ($S.E.$) | | | |
| KERMIT 3 | 376,71 | (7,82) | 351,41 | (7,94) | 2,27 (489) | .01 | 0.21 |
| KERMIT 5 | 465,62 | (5,73) | 443,78 | (6,31) | 2,51 (139) | .01 | 0.43 |

Anmerkungen: d = Effektgröße nach Cohen; $d \geq 0.2$ kleiner Effekt; $d \geq 0.5$ mittlerer Effekt; $d \geq 0.8$ großer Effekt; $p \leq .05$ signifikant; $p \leq .01$ hoch signifikant (1-seitig)

Tab. 2: Vergleich der Kompetenzwerte im Bereich Rechtschreibung

Dieser Unterschied wird sowohl in Jahrgangsstufe 3, ($t(489) = 2,27, p = .01$) als auch in Jahrgangsstufe 5 signifikant, ($t(139) = 2,51, p = .01$). Der Effekt ist bei KERMIT 5 ($d = 0.43$) dabei größer als bei KERMIT 3 ($d = 0.21$).

Mathematik

| | Trainings-Gruppe | | Matching-Gruppe | | $t(df)$ | p | Cohen's d |
|----------|------------------|------------|-----------------|------------|------------|-----|-------------|
| | M | ($S.E.$) | M | ($S.E.$) | | | |
| KERMIT 2 | 453,37 | (5,72) | 441,16 | (6,14) | 1,44 (182) | .08 | |
| KERMIT 3 | 416,61 | (7,04) | 387,38 | (7,23) | 2,87 (169) | .01 | 0.44 |
| KERMIT 5 | 452,32 | (5,32) | 443,11 | (6,64) | 1,07 (72) | .14 | |

Anmerkungen: d = Effektgröße nach Cohen; $d \geq 0.2$ kleiner Effekt; $d \geq 0.5$ mittlerer Effekt; $d \geq 0.8$ großer Effekt; $p \leq .05$ signifikant; $p \leq .01$ hoch signifikant (1-seitig)

Tab. 3: Vergleich der Kompetenzwerte im Bereich Mathematik

Die Kompetenzmittelwerte der Trainingsgruppe liegen auch im Bereich Mathematik über den Werten der Matching-Gruppe.

Der Mittelwertsunterschied ist allerdings lediglich in Jahrgangsstufe 3 signifikant, ($t(169) = 2,87, p = .01$) und weist dabei eine Effektgröße von $d = 0.44$ auf.

Englisch Hörverstehen und Naturwissenschaften

Auch in den Bereichen Englisch-Hörverstehen und Naturwissenschaften verzeichnen die Schüler*innen der Trainingsgruppe höherer Kompetenzwerte als die Schüler*innen der Matching-Gruppe.

| | Trainings-Gruppe | | Matching-Gruppe | | t(df) | p |
|---------------------|------------------|--------|-----------------|--------|------------|-----|
| | M | (S.E.) | M | (S.E.) | | |
| Englisch | | | | | | |
| KERMIT 5 | 460,23 | (6,09) | 453,37 | (6,12) | 0,80 (74) | .21 |
| Naturwissenschaften | | | | | | |
| KERMIT 5 | 447,01 | (5,22) | 439,63 | (5,70) | 0,95 (243) | .17 |

d = Effektgröße nach Cohen; $d \geq 0.2$ kleiner Effekt; $d \geq 0.5$ mittlerer Effekt; $d \geq 0.8$ großer Effekt.
 $p \leq .05$ signifikant; $p \leq .01$ hoch signifikant (1-seitig)

Tab. 4: Vergleich der Kompetenzwerte in den Bereichen Englisch und Naturwissenschaften
 Anmerkungen:

Die Unterschiede sind jedoch statistisch nicht signifikant.

6.4 | Differentielle Effekte des Trainings für einzelne Schüler*innengruppen

Um zu überprüfen, ob sich die Trainingseffekte besonders bei Schüler*innengruppen mit – wie eingangs beschrieben – ungünstiger Ausgangslage zeigen, wurden verschiedene Teilgruppen miteinander verglichen. Die Vergleiche wurden dabei in Abhängigkeit von Geschlecht, Migrationshintergrund, Familiensprache und sozioökonomischem Hintergrund durchgeführt.

Geschlecht

Die Mittelwertvergleiche zeigen, dass sowohl Mädchen als auch Jungen aus der Trainingsgruppe in allen Jahrgangsstufen besser in den KERMIT-Erhebungen abschneiden als die Schüler*innen aus der Matching-Gruppe. Dieser Unterschied wird für die Jungen allerdings nur in Jahrgangsstufe 2 und 3, für die Mädchen nur in Jahrgangsstufe 3 signifikant. Der Unterschied in Jahrgangsstufe 2 zwischen den Jungen erreicht allerdings keine praktische Bedeutsamkeit. Lediglich bei den Unterschieden zum Zeitpunkt von KERMIT 3 kann von einem kleinen Effekt zugunsten der Mädchen ($d = 0.26$) und Jungen ($d = 0.25$) aus der Trainingsgruppe gesprochen werden.

| | Trainings-Gruppe | | Matching-Gruppe | | t(df) | p | Cohen's d |
|------------|------------------|---------|-----------------|---------|-------------|-----|-----------|
| | M | (S.E.) | M | (S.E.) | | | |
| Geschlecht | | | | | | | |
| männlich | | | | | | | |
| KERMIT 2 | 459,35 | (7,17) | 436,91 | (7,40) | 2,21 (1139) | .01 | 0.13 |
| KERMIT 3 | 412,66 | (10,69) | 386,92 | (10,43) | 1,72 (196) | .04 | 0.25 |
| KERMIT 5 | 457,28 | (8,71) | 439,24 | (8,37) | 1,49 (83) | .07 | |
| weiblich | | | | | | | |
| KERMIT 2 | 458,20 | (7,41) | 450,17 | (9,30) | 0,69 (622) | .25 | |
| KERMIT 3 | 445,38 | (11,81) | 415,95 | (13,02) | 1,69 (173) | .05 | 0.26 |
| KERMIT 5 | 462,78 | (8,51) | 448,89 | (9,36) | 1,09 (147) | .14 | |

Anmerkungen: d = Effektgröße nach Cohen; $d \geq 0.2$ kleiner Effekt; $d \geq 0.5$ mittlerer Effekt; $d \geq 0.8$ großer Effekt; $p \leq .05$ signifikant; $p \leq .01$ hoch signifikant (1-seitig)

Tab. 5: Vergleich der Kompetenzwerte im Bereich Leseverstehen nach Geschlecht

Migrationshintergrund

Die weiteren Analysen zeigen, dass die Trainingseffekte der Leseflüssigkeit besonders gewinnbringend für die Lesekompetenzen der Schüler*innen mit Migrationshintergrund sind.

| | Trainings-Gruppe | | Matching-Gruppe | | t(df) | p | Cohens's d |
|----------------------------|------------------|---------|-----------------|---------|------------|-----|------------|
| | M | (S.E.) | M | (S.E.) | | | |
| Migrationshintergrund | | | | | | | |
| mit Migrationshintergrund | | | | | | | |
| KERMIT 2 | 456,13 | (6,15) | 432,79 | (6,41) | 2,66 (611) | .01 | 0.22 |
| KERMIT 3 | 425,24 | (9,27) | 386,18 | (8,43) | 3,14 (810) | .01 | 0.22 |
| KERMIT 5 | 459,44 | (7,19) | 432,04 | (6,42) | 2,82 (156) | .01 | 0.45 |
| ohne Migrationshintergrund | | | | | | | |
| KERMIT 2 | 486,82 | (10,40) | 482,75 | (14,26) | 0,23 (440) | .41 | |
| KERMIT 3 | 467,79 | (16,96) | 446,82 | (19,63) | 0,80 (227) | .21 | |
| KERMIT 5 | 472,96 | (13,74) | 488,29 | (14,77) | 0,74 (86) | .23 | |

Anmerkungen: d = Effektgröße nach Cohen; $d \geq 0.2$ kleiner Effekt; $d \geq 0.5$ mittlerer Effekt; $d \geq 0.8$ großer Effekt; $p \leq .05$ signifikant; $p \leq .01$ hoch signifikant (1-seitig)

Tab. 6: Vergleich der Kompetenzwerte im Bereich Leseverstehen nach Migrationshintergrund

Sie erreichen sowohl in Klassenstufe 2 und 3, als auch noch in Klassenstufe 5 signifikant bessere KERMIT-Ergebnisse im Bereich Leseverstehen als die Kinder mit Migrationshintergrund aus der Matching-Gruppe. In Jahrgangsstufe 5 ist der Effekt am stärksten ($d = 0.45$).

Bei den Kindern ohne Migrationshintergrund zeigen sich keine signifikanten Unterschiede zwischen der Trainingsgruppe und der Matching-Gruppe.

Familiensprache

Die Analysen nach Familiensprache ergeben ein ähnliches Befundmuster.

| | Trainings-Gruppe | | Matching-Gruppe | | t(df) | p | Cohens's d |
|-------------------------|------------------|---------|-----------------|---------|------------|-----|------------|
| | M | (S.E.) | M | (S.E.) | | | |
| Familiensprache | | | | | | | |
| andere Familiensprache | | | | | | | |
| KERMIT 2 | 444,35 | (6,82) | 424,56 | (7,36) | 1,96 (163) | .03 | 0.31 |
| KERMIT 3 | 415,09 | (10,00) | 377,47 | (9,02) | 2,78 (702) | .01 | 0.21 |
| KERMIT 5 | 450,90 | (7,91) | 423,97 | (7,40) | 2,44 (106) | .01 | 0.47 |
| Familiensprache Deutsch | | | | | | | |
| KERMIT 2 | 480,65 | (7,22) | 470,04 | (9,82) | 0,86 (238) | .20 | |
| KERMIT 3 | 448,64 | (12,40) | 433,61 | (13,77) | 0,80 (175) | .21 | |
| KERMIT 5 | 473,61 | (9,11) | 472,59 | (10,24) | 0,07 (64) | .47 | |

Anmerkungen: d = Effektgröße nach Cohen; $d \geq 0.2$ kleiner Effekt; $d \geq 0.5$ mittlerer Effekt; $d \geq 0.8$ großer Effekt; $p \leq .05$ signifikant; $p \leq .01$ hoch signifikant (1-seitig)

Tab. 7: Vergleich der Kompetenzwerte im Bereich Leseverstehen nach Familiensprache

So lässt sich für die Kinder mit nicht-deutscher Familiensprache zu allen drei Testzeitpunkten ein signifikanter Unterschied im Leseverstehen zwischen der Trainingsgruppe und der Matching-Gruppe nachweisen. Ein weiteres Mal ist der Effekt in Jahrgangsstufe 5 am stärksten ($d = 0.47$).

Bei den Kindern mit deutscher Familiensprache schneidet ebenfalls die Trainingsgruppe besser ab als die Matching-Gruppe. Diese Unterschiede sind jedoch nicht signifikant.

RISE-Status

Bei der Betrachtung der Lesekompetenzwerte differenziert nach dem RISE-Status wird erneut sichtbar, dass vor allem die Schüler*innengruppen mit ungünstiger Ausgangslage von dem Training profitieren.

| | Trainings-Gruppe | | Matching-Gruppe | | t(df) | p | Cohens's d |
|--------------|------------------|---------|-----------------|---------|-------------|-----|------------|
| | M | (S.E.) | M | (S.E.) | | | |
| RISE-Status | | | | | | | |
| sehr niedrig | | | | | | | |
| KERMIT 2 | 437,42 | (7,82) | 412,98 | (9,66) | 2,00 (89) | .02 | 0.43 |
| KERMIT 3 | 409,94 | (11,99) | 368,87 | (10,97) | 2,50 (293) | .01 | 0.29 |
| KERMIT 5 | 454,96 | (9,12) | 419,41 | (8,58) | 2,77 (68) | .01 | 0.67 |
| niedrig | | | | | | | |
| KERMIT 2 | 466,11 | (9,24) | 463,84 | (10,25) | 0,16 (2027) | .43 | |
| KERMIT 3 | 426,55 | (13,84) | 415,02 | (15,12) | 0,56 (331) | .29 | |
| KERMIT 5 | 453,54 | (11,63) | 451,24 | (12,96) | 0,13 (93) | .45 | |
| mittel | | | | | | | |
| KERMIT 2 | 478,66 | (9,14) | 464,52 | (11,25) | 0,98 (945) | .16 | |
| KERMIT 3 | 453,30 | (15,49) | 431,21 | (15,03) | 1,02 (1260) | .15 | |
| KERMIT 5 | 470,17 | (10,84) | 470,95 | (11,03) | 0,05 (516) | .48 | |

Anmerkungen: d = Effektgröße nach Cohen; $d \geq 0.2$ kleiner Effekt; $d \geq 0.5$ mittlerer Effekt; $d \geq 0.8$ großer Effekt; $p \leq .05$ signifikant; $p \leq .01$ hoch signifikant (1-seitig)

Tab. 8: Vergleich der Kompetenzwerte im Bereich Leseverstehen nach RISE-Status

Die Schüler*innen mit sehr niedrigem RISE-Status aus der Trainingsgruppe erzielen im Bereich Leseverstehen zu allen drei Testzeitpunkten signifikant bessere Ergebnisse als die Kinder aus der Matching-Gruppe. Der Unterschied erreicht in Jahrgangsstufe fünf eine mittlere Effektgröße ($d = 0.67$).

Auch die Kompetenzwerte der Schüler*innen mit niedrigem und mittlerem RISE-Status aus der Trainingsgruppe liegen – bis auf eine Ausnahme der Kinder mit mittlerem RISE-Status in der fünften Jahrgangsstufe – über denen der Matching-Gruppe. Diese Unterschiede werden statistisch aber nicht signifikant.

7 | Diskussion der Ergebnisse

Abschließend werden die Ergebnisse auf die in Abschnitt 4 formulierten Hypothesen bezogen und hinsichtlich des Forschungsstandes und den sich daraus abzuleitenden schulpraktischen Konsequenzen diskutiert.

Zur Annahme 1: Die Teilnahme am BiSS-Lesetraining führt zu einer Steigerung der Leseflüssigkeit

Diese Hypothese kann bestätigt werden und war auch nach allem, was empirisch über die Förderung der Leseflüssigkeit bekannt ist, zu erwarten (vgl. etwa Rosebrock et al. 2011, Rosebrock & Nix 2017, Gailberger, Helm & Nix 2021): Der Vergleich der Ergebnisse zwischen der Kontroll- und der Trainingsgruppe zu den drei Testzeitpunkten im Salzburger Lese-Screening spricht dafür, dass die Teilnahme am BiSS-Lesetraining zu einer Steigerung der Leseflüssigkeit führt. Die Tatsache, dass die Schüler*innen des BiSS-Projekts signifikant profitieren, könnte als Anregung

dafür dienen, es der mit dem Deutschen Lesepreis 2017 ausgezeichneten Hamburger Grundschule Kirchdorf¹¹ gleich zu tun und ein verbindliches „Leseband“ (Gailberger & Hauschild 2018) einzuführen, das die Schüler*innen von der Vorschule an fünf Tagen in der Woche für 20 Minuten an das Lesen von (literarischen) Texten heranführt.

Ob der berichtete Effekt auch für Schüler*innen aus weniger benachteiligten Familien nachzuweisen ist, bedarf weiterer Forschung. Ebenso sollte geprüft werden, ob auch bei guten Leseausgangslagen eine weitere Verbesserung zu erzielen ist.

Zur Annahme 2: Die Steigerung der Leseflüssigkeit führt zu einem besseren Leseverstehen auf Textebene

In Übereinstimmung mit den vorangegangenen Forschungen bspw. von Pikulski & Chard (2005), Klauda & Guthrie (2008) oder Fuchs et al. (2001) zur bereits zitierten Brückenfunktion der Leseflüssigkeit zeigen die weiteren Ergebnisse, dass das Hamburger BiSS-Training, vermittelt über die Förderung der Leseflüssigkeit, wesentlich auch zu einem besseren Leseverstehen auf Textebene beiträgt. Dabei zeigt sich der positive Effekt aber nicht nur während des Trainings (gemessen mit KERMIT 3) und damit quasi ‚to the test‘, sondern auch nachhaltig und darüber hinaus, wie die **Follow-Up-Testung** zu Beginn der Jahrgangsstufe 5 (mittels KERMIT 5) gezeigt hat. Vor dem Hintergrund der in Abschnitt 3.2 skizzierten Schwierigkeiten, die die teilnehmenden Schulen mit der ursprünglich von BiSS Hamburg vorgesehenen Wortschatz- oder Lesestrategiearbeit *unisono* äußerten, und den Sorgen, die die Verfasser*innen dieses Beitrags damit verbanden, ist dieses Ergebnis auch aus schulpraktischer Sicht als besonders erfreulich einzustufen, zeigt es doch, dass sich das Lesen als zu erwerbende und als ausbaufähige Kompetenz über die Leseflüssigkeit hinaus auch bereits bei schwachen und/oder i.w.S. benachteiligten Schüler*innen fördern lässt, deren hierarchiehöhere Teilprozesse des Lesens hierfür eigentlich noch nicht ausreichend ausgebildet sind.

Zu prüfen wäre hier, ob dieser Effekt auch über einen noch längeren Zeitraum als drei bis vier Monate anhält, oder – was zu erwarten wäre – ob er sich ab einem bestimmten Zeitpunkt wieder abschwächt. Zusätzlich wäre zu prüfen, welchen Einfluss die Leseflüssigkeit auf die Entwicklung der Lesekompetenz von Grundschüler*innen behält, wenn weitere Faktoren, die die Lesekompetenz beeinflussen (wie Motivation, Interesse, Emotion, soziale Kompetenz etc.), in das Modell aufgenommen werden.

Zur Annahme 3: Die Teilnahme am BiSS-Lesetraining führt neben einer Verbesserung des Leseverstehens auch zu Verbesserungen in anderen Kompetenzen

Hinsichtlich der rechtschreibspezifischen Testungen der KERMIT-Erhebungen in Jahrgang 3 und 5 konnte gezeigt werden, dass die BiSS-Gruppe bessere Ergebnisse im Bereich **Rechtschreibung** als ihre Vergleichsgruppe erzielt. Dabei ist allerdings nicht alleine zu betonen, dass die gemessenen Unterschiede in Jahrgangsstufe 3 wie auch in Jahrgangsstufe 5 (als *Follow Up*) signifikant sind, sondern dass darüber hinaus die Effektstärke mit Dauer der Leseförderung (hier über drei Jahre) kontinuierlich zunimmt, so dass also (erneut) aus schulpraktischer Perspektive dafür plädiert werden kann, Sprach- und Leseförderungen schulübergreifend und dauerhaft verbindlich zu implementieren und nicht ‚auf halbem Weg‘ (bspw. nach wenigen Wochen oder Monaten) wieder abzubrechen.

¹¹ Vgl. <https://gskirchdorf.hamburg.de/unser-profil/deutsch/lesenlesenlesen/> [16.09.2021]

Für die Domäne **Mathematik** konnte ein signifikanter Unterschied zwischen Trainings- und Kontrollgruppe in der dritten Jahrgangsstufe gefunden werden, so dass anzunehmen ist, dass es den zuvor geförderten Schüler*innen leichter fiel, sich auf die mathematischen Prozeduren zu konzentrieren, da sie die gestellten schriftlichen Aufgaben besser lesen und verstehen konnten. Für **Englisch Hörverstehen** und **Naturwissenschaften** konnten hingegen die Effekte des Lesetrainings nicht nachgewiesen werden. Die Annahme, dass durch das Leseflüssigkeitstraining der Zugang zu den textlastigen Aufgaben bei KERMIT erleichtert wird, kann hier also nicht bestätigt werden. Möglicherweise sind die benötigten domänenspezifischen Kompetenzen so durchschlagend, dass auch schnelleres Lesen bzw. besseres Leseverstehen (s.o.) keinen Zugewinn mehr bringt. Die Leistungen in Englisch und den Naturwissenschaften wurden nach einer langen Phase ohne Training (Sommerferien und Schulbeginn in der weiterführenden Schule) erhoben, so dass es auch möglich ist, dass sich kleine Effekte bereits wieder verflüchtigt haben. Hier ist es angezeigt, parallel zum Lesetraining Leistungen in anderen Domänen zu erfassen, um mögliche Abschwächungseffekte über die Zeit ausschließen zu können.

In einem weiteren Schritt erscheint es auch sinnvoll, die eingesetzten Aufgaben zur Kompetenzmessung differenzierter zu betrachten. Möglichweise zeigen sich die Transfereffekte nur bei bestimmten Aufgabentypen.

Damit kann grundsätzlich ein fachspezifischer (**Rechtschreibung** – KERMIT 2 und 5) und ein zeitbezogener (**Rechtschreibung** – KERMIT 3 und 5 und **Mathematik** – KERMIT 3) Transfereffekt nachgewiesen werden, der jeweils nur während des Trainings auftritt.

Zur Annahme 4: Die Trainingseffekte zeigen sich insbesondere bei Schüler*innen mit schwächeren Ausgangslagen im Lesen

Es konnte nachgewiesen werden, dass besonders Schüler*innen mit Migrationshintergrund, mit nicht deutscher Familiensprache und/oder mit niedrigem RISE-Status von dem Training profitieren. Ein Training scheint für diese Gruppen daher besonders angezeigt. Zu untersuchen ist weiterhin, wie auch andere Lerner*innengruppen von dem angebotenen Training stärker profitieren können.

8 | Fazit und Ausblick: Der Hamburger BiSS-Verbund Primarstufe aus der Perspektive von Lesedidaktik, empirischer Bildungswissenschaft und schulischer Praxis

Das Akronym BiSS steht für „Bildung durch Sprache und Schrift“ und bezeichnet ein deutschlandweites Sprachförderprojekt, das gemeinsam von Bund und Ländern getragen wird und deutschlandweit in 101 Verbünde organisiert ist (vgl. Titz et al. 2017). Mithilfe wissenschaftlicher Begleitstudien unterstützt BiSS Bildungseinrichtungen von Flensburg bis München bei ihrer Arbeit in der sprachlichen Bildung, um bewährte Sprachfördermaßnahmen evaluieren und an Schulen bzw. Lehrer*innen weitervermitteln zu können. Im vorliegenden Beitrag wurde über die Langzeitstudie des Hamburger BiSS-Projekts für die Primarstufe berichtet, deren teilnehmende Schüler*innen zum größten Teil als sprachlich und/oder sozial stark benachteiligt gelten. Mithilfe verschiedener Testzeitpunkte konnten die Ausgangswerte der Trainings- und Kontrollgruppe, ihre Veränderung im Laufe des Projekts, die Zielwerte am Ende des Projekts sowie die Nachhaltigkeit der Intervention (nach ca. vier Monaten) dokumentiert werden. Dabei zeigte

sich, dass es mithilfe eines verbindlichen Lesebandes an (je nach organisatorischen Möglichkeiten vor Ort in den Schulen) drei bis fünf Tagen zu je 20 Minuten, klassenübergreifend und vom gesamten Kollegium getragen, gelingen kann, strukturell benachteiligte Mädchen und Jungen bereits zu Beginn ihrer Schulkarriere mit Lautleseverfahren so zu fördern, dass nicht nur ihre Leseflüssigkeit und ihr Leseverstehen auf Textebene, sondern darüber hinaus auch weitere Kompetenzen anderer Domänen (z.T. auch jenseits des Deutschunterrichts) davon profitieren. Keine Hinweise liefert die Studie indes zur Wirksamkeit einzelner Methoden. Hierzu bedarf es weiterer, differenzierterer Designs. Studien dieser Art könnten dann auch in den Blick nehmen, inwiefern sich Effektstärken einzelner Elemente des Lesetrainings ausfindig machen lassen. Auch gilt es, noch genauer herauszufinden, wie häufig und in welchem Umfang das Training stattfinden muss, damit es seine volle Wirkung entfalten kann. So waren nicht alle Schulen dazu in der Lage, ihr Leseband an allen fünf Tagen in der Woche stattfinden zu lassen und förderten so das Lesen ihrer Schüler*innen nur zu drei verbindlichen Terminen in der Woche. Als letzter Aspekt wären noch die Fragen nach der weiterführenden Nachhaltigkeit der Trainingseffekte und des Ausmaßes von möglichen Transfereffekten zu nennen, denen es noch weiter nachzugehen gelten wird.

9 | Literatur

- Behörde für Stadtentwicklung und Wohnen (o.J.). *Grundlagen der Stadtteilentwicklung. Sozialmonitoring*. <https://www.hamburg.de/sozialmonitoring/> [16.09.2021].
- Boll, V. & Scholz, J. (2012). *Leseförderung durch Hörbücher und durch Vorlesen im Vergleich. Eine empirische Studie in Grund- und Hauptschulklassen*. Lüneburg (unveröffentlichte Masterarbeit).
- Christmann, U. (2019). Lesepsychologie. In: M. Kämper-van den Boogaart & K. Spinner (Hrsg.). *Lesen und Literaturunterricht*. Band 1. (S. 150–201). Schneider Verlag Hohengehren.
- Christmann, U. & Groeben, N. (2006). Psychologie des Lesens. In: B. Franzmann, K. Hasemann, D. Löffler & E. Schön (Hrsg.), *Handbuch Lesen* (S. 145–223). Schneider Verlag Hohengehren.
- Cohen, J. (1988): *Statistical power analysis for the behavioral sciences*. Erlbaum.
- Deutsche UNESCO-Kommission e.V. (2014). *Inklusion: Leitlinien für die Bildungspolitik*. https://www.unesco.de/sites/default/files/2018-05/2014_Leitlinien_inklusive_Bildung.pdf [16.09.2021].
- Fuchs, L. S., Fuchs, D., Hosp, M. K. & Jenkins, J. (2001). Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis, *Scientific Studies of Reading*.
- Gailberger, S. (2013). *Systematische Leseförderung für schwach lesende Schüler. Zur Wirkung von lektürebegleitenden Hörbüchern und Lesebewusstmachungsstrategien*. Beltz-Juventa.
- Gailberger, S. & Hauschild, K. (2018). BiSS zum Deutschen Lesepreis. Das Hamburger BiSS-Projekt „Systematische Leseförderung für die Primarstufe“ – dargestellt am ‚Leseband‘ der Grundschule Kirchdorf. In S. Kutzelnann & C. Rosebrock (Hrsg.), *Lautleseverfahren in der Praxis*. Schneider Verlag Hohengehren.
- Gailberger, S., Helm, G. & Nix, D. (2021). Lesen und Leseförderung in der Primar- und Sekundarstufe 1. In: S. Gailberger & F. Wietzke (Hrsg.), *Handbuch kompetenzorientierter Deutschunterricht*. Beltz (i.Dr.).

- Garbe, C. (2014a). Mädchen lesen gern – Jungen auch?! Unterschiede im Leseverhalten und in der Leseleistung von Jungen und Mädchen erkennen und berücksichtigen. *Grundschulunterricht* 4(2), 4-7.
- Garbe, C. (2014b). Lesefreude – Lesemotivation – stabile Lesegewohnheiten. *Grundschulunterricht* 4(2), 8-11.
- Garbe, C., Olsen, N. & Philipp, M. (2009). *Lesesozialisation: Ein Arbeitsbuch für Lehramtsstudierende*. UTB Schöningh.
- Groebe, N. & Vorderer, P. (1988). *Leserpsychologie: Lesemotivation – Lektürewirkung*. Aschendorff.
- Hasselhorn, M. & Gold, A. (2013). *Pädagogische Psychologie. Erfolgreiches Lehren und Lernen*. Kohlhammer.
- Juska-Bacher, B., Röthlisberger, M., Brugger, L. & Zangger, C. (2021). Lesen im 1. Schuljahr: Die Bedeutung von phonologischer Bewusstheit, Benennungsgeschwindigkeit und Wortschatz (in diesem Band).
- Kintsch, W. (1998). *Comprehension*. Cambridge University Press.
- Klauda, S. & Guthrie, J. T. (2008). Relationships of Three Components of Reading Fluency to Reading Comprehension. *Journal of Educational Psychology* 100(2), 310-321.
- Kruse, G., Rickli, U., Riss, M. & Sommer, T. (2011). „LESEN: Das Training“ – Strategien. Ernst Klett Verlag.
- LaBerge, D. & Samuels, S. J. (1974). *Toward a theory of automatic information processing in reading*. *Cognitive Psychology*, 6(2), 293–323.
- Lewis, M., Haviland-Jones, J. & Barrett, L. F. (2010). *Handbook of Emotions*. NYC: Guilford Publications.
- Mayringer, H. & Wimmer, H. (2014). *SLS 2-9. Salzburger Lese-Screening für die Klassenstufen 2-9*. Hans-Huber Verlag.
- Meyer, P.-L. (2011). *Leseförderung schwach lesender Schülerinnen und Schüler in der Grundschule – Lesen mit Hörbuch in der Jahrgangsstufe 3. Ein Forschungsprojekt in Kooperation mit dem Landesinstitut für Lehrerbildung in Hamburg*. Lüneburg (unveröffentlichte Masterarbeit).
- Möller, J. & Schiefele, U. (2004). Motivationale Grundlagen der Lesekompetenz. In: U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (S.101–124). VS Verlag.
- Muthén, L. K. & Muthén, B. O. (2012). *Mplus version 7.0 [Computer software]*. CA: Authors.
- Nix, D. (2006). *Das Lesetheater. Integrative Leseförderung durch das szenische Vorlesen literarischer Texte*. Praxis Deutsch, 199(5), 23-29.
- Nix, D. (2011). *Förderung der Leseflüssigkeit. Theoretische Fundierung und empirische Überprüfung eines kooperativen Lautlese-Verfahrens im Deutschunterricht*. Beltz Juventa.
- Pieper, I. (2019). Lese- und literarische Sozialisation. In: M. Kämper-van den Boogaart & K. Spinner (Hrsg.). *Lese und Literaturunterricht*. Band 1. (S. 87-147). Schneider Verlag Hohengehren.
- Pikulski, J. & Chard, D. (2005). Fluency: Bridge Between Decoding and Reading Comprehension. *The Reading Teacher* 58(6), 510-519.
- Plath, M. & Richter, K. (2012). Literarische Sozialisation in der mediatisierten Kindheit. Ergebnisse neuer empirischer Untersuchungen. In: G. Lange (Hrsg.), *Kinder- und Jugendliteratur der Gegenwart* (S. 485–507). Schneider Verlag Hohengehren.
- Rosebrock, C. & Nix, D. (2017). *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung*. Schneider Verlag Hohengehren.

- Rosebrock, C., Nix, D., Rieckmann, C. & Gold, A. (2011). *Leseflüssigkeit fördern. Lautleseverfahren für die Primar- und Sekundarstufe*. Klett Kallmeyer.
- Röttig, S., Schwerkolt, C. & Nottbusch, G. (2020). Die Entwicklung der Leseflüssigkeit in der Grundschule. Eine Longitudinalstudie über die interagierenden Dimension Dekodiergenauigkeit, Automatisierung, Lesegeschwindigkeit und Prosodie bei Kindern der Jahrgangsstufen 2 und 3 (in diesem Band).
- Sappok, C. (2020). Exploring Advanced Prosody – eine Best-Practice-Untersuchung in der weiterführenden Schule (in diesem Band).
- Schnotz, W. (2006). Was geschieht im Kopf des Lesers? Mentale Konstruktionsprozesse beim Textverstehen aus der Sicht der Psychologie und der kognitiven Linguistik. In: H. Blühdorn, E. Breindl & U. H. Waßner (Hrsg.), *Text - Verstehen. Grammatik und darüber hinaus* (S. 222-238). De Gruyter.
- Stephany, S., Linnemann, M., Goltsev, E. & Becker-Mrotzek, M. (2020): Prosodische Aspekte der Leseflüssigkeit als Indikator für Lesekompetenz – Analysen mithilfe prosodischer Lupenstellen (in diesem Band).
- Titz, C., Geyer, S., Ropeter, A., Wagner, H., Weber, S. & Hasselhorn, M. (Hrsg.) (2017). *Konzepte zur Sprach- und Schriftsprachförderung entwickeln*. Kohlhammer.
- Wietzke, F. (2011). *Das Lesetagebuch für Erzähltexte. 2.–4. Klasse*. Persen.
- Wietzke, F. (2013). *Das Lesetagebuch für Sachtexte. 3./4. Klasse*. Persen.
- Willenberg, H. (1999). *Lesen und Lernen – Eine Einführung in die Neuropsychologie des Textverstehens*. Spektrum.

Daria Ferencik-Lehmkuhl & Ilka Fladung

Lesediagnostik und -förderung im heterogenen Deutschunterricht

Finale Ergebnisse aus dem Projekt *Ganz In*

1 | Hintergrund: Das Projekt „Ganz In – Mit Ganztag mehr Zukunft“

Der Veröffentlichung der Ergebnisse des *Programme for International Student Assessment* (PISA; Baumert et al. 2001) aus dem Jahr 2000 folgte eine intensive bildungspolitische sowie öffentliche Debatte. Aus dieser Diskussion resultierten umfangreiche schulische Reformen, zu denen auch der Ausbau von Ganztagschulen zählte. Vor diesem Hintergrund startete im Jahr 2010 das Projekt *Ganz In. Mit Ganztag mehr Zukunft. Das neue Ganztagsgymnasium NRW*. Hierbei handelt es sich um ein kooperatives Schulentwicklungsprojekt der Universitäten der Ruhrallianz (Universität Duisburg-Essen, Ruhr-Universität Bochum, Technische Universität Dortmund), der Stiftung Mercator und des Ministeriums für Schule und Weiterbildung¹ des Landes Nordrhein-Westfalen (NRW). Hauptziel von *Ganz In* war es in der ersten Projektphase (*Ganz In I*, 2010-2015), 30 ausgewählte Gymnasien durch Fortbildungs- und Netzwerkarbeit auf ihrem Weg zu gebundenen Ganztagschulen² in der Schul- und der Unterrichtsentwicklung zu begleiten (vgl. Wendt & Bos 2015).

Angesichts verschiedener bildungspolitischer, pädagogischer und sozialpolitischer Zielsetzungen von Ganztagschule betrifft eine der zentralen Begründungslinien von *Ganz In* die Ausschöpfung von Bildungspotentialen: So gelingt die Förderung von leistungsstarken Gymnasialschülerinnen und -schülern noch nicht optimal, wenn man Deutschland mit anderen EU- und OECD-Staaten vergleicht (vgl. Wendt et al. 2013). Weiterhin besteht ein deutlicher Handlungsbedarf die Chancengerechtigkeit betreffend. So zeigen Befunde, z.B. aus der *Internationalen Grundschul-Lese-Untersuchung* (IGLU), dass nicht nur die schulische Leistung, sondern auch soziokulturelle Merkmale einen deutlichen Einfluss auf die Übergangsempfehlung der Lehrkräfte haben. Auch wenn diese nicht verbindlich ist, beeinflusst sie aber trotzdem häufig die Eltern, die ihr Kind oft nicht auf das Gymnasium, sondern auf eine andere weiterführende Schule schicken. Auch im Jahr 2016 konnte erneut der Zusammenhang zwischen der sozioökonomischen

¹ Inzwischen heißt dies Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen.

² Als gebundene Ganztagschule gilt eine Schule (im Primar- und Sekundarbereich I) dann, wenn „die Ganztagsangebote unter der Aufsicht und Verantwortung der Schulleitung organisiert und in enger Kooperation mit der Schulleitung durchgeführt werden sowie in einem konzeptionellen Zusammenhang mit dem Unterricht stehen“ (KMK 2015: 4), wenn die Schülerinnen und Schüler an mindestens drei Tagen in der Woche für je mindestens sieben Zeitstunden an diesen Ganztagsangeboten teilnehmen und ihnen an allen Tagen des Ganztagschulbetriebs ein Mittagessen zur Verfügung gestellt wird (vgl. ebd. 5).

Stellung der Familie von Schülerinnen und Schülern und der Chance, eine Gymnasialempfehlung zu erhalten, belegt werden. Bei gleichen kognitiven Fähigkeiten bzw. bei gleicher Lesekompetenz ist die Chance für ein Kind mit Eltern aus der „oberen Dienstklasse“³ fast 4-mal so hoch wie für ein „Facharbeiter“-Kind (vgl. Stubbe, Bos & Schurig 2017: 244). Damit aber gewährleistet ist, dass jede Schülerin und jeder Schüler unabhängig von Stand oder Vermögen der Eltern einen Bildungsweg einschlagen kann, der ihren oder seinen Interessen und Fähigkeiten entspricht, sind Gymnasien aufgefordert, „Schülerinnen und Schüler aus Familien mit vergleichsweise geringen bildungsförderlichen Bedingungen verstärkt aufzunehmen, Herkunft und Bildungserfolg zu entkoppeln und durch systematische und bedarfsgerechte Förderung zu höheren schulischen Leistungen und schließlich erfolgreich zum Abitur zu führen“ (Wendt & Bos 2015: 12). Im Fokus von *Ganz In* lagen aber auch Gymnasialschülerinnen und -schüler des unteren Leistungsviertels, welche aufgrund von sozialen Umständen (z.B. bildungsfernes Milieu und/oder Migrationshintergrund) ohne zusätzliche Unterstützung ihr Potential nicht ausschöpfen können. Daher wurden im Rahmen des Projektes folgende Zielsetzungen benannt:

„durch die Verzahnung der unterschiedlichen Lerngelegenheiten eine allgemeine Verbesserung der Schülerinnen- und Schülerleistungen zu erreichen;

durch eine bedarfsorientierte Entwicklung von Ganztagsangeboten der auch an Gymnasien vorzufindenden Heterogenität von Schülerschaften gerecht zu werden und durch die Ausgestaltung spezifischer Angebote verbesserte Möglichkeiten der individuellen Förderung zu schaffen, von denen insbesondere Schülerinnen und Schüler profitieren, die in ihrem häuslichen Umfeld in Bezug auf ihre individuellen Leistungspotentiale auf keine adäquate Unterstützung zurückgreifen können“ (Wendt & Bos 2015: 18).

Die Schulentwicklungsarbeit im Projekt wurde durch fachspezifische Lehrer*innen-Fortbildungen und durch die Begleitung von Arbeitstreffen auf Netzwerkebene⁴ unterstützt. Im Fach Deutsch wurde der Fokus auf die Diagnostik und Förderung in den Bereichen *Lesen*, *Schreiben* und *Orthographie* gelegt. Diejenigen Lehrkräfte der 30 Gymnasien, die an den Arbeitstreffen und den Fortbildungen teilnahmen, sollten als Multiplikatorinnen und Multiplikatoren fungieren, indem sie in ihren jeweiligen Deutschfachschaften ihre Kolleginnen und Kollegen über die Ergebnisse der Treffen informieren. Da sich dieses Konzept als überarbeitungsbedürftig erwies⁵, wurde der Zuschnitt des Projekts für die zweite Phase dahingehend geändert, dass ausschließlich mit Deutschkolleginnen und -kollegen aus fünf Gymnasien kooperiert wurde.⁶ Die

³ Die soziale Lage von Familien wurde mit Hilfe der EGP-Klassen (Erikson, Goldthorpe & Portocarero 1979) beschrieben: Obere Dienstklasse (I), Untere Dienstklasse (II), Routinedienstklasse (III), Selbstständige (IV), (Fach-)Arbeiter (V, VI), un- und angelernte Arbeiter (VII) (vgl. Stubbe, Bos & Schurig 2017: 244).

⁴ Es gab vier Schulnetzwerke und in jedem Netzwerk fanden regelmäßig Arbeitstreffen statt.

⁵ Zum einen war es den Schulen meist nicht möglich, eine Personalkontinuität über mehrere Jahre hinweg zu gewährleisten, die sich förderlich auf die Zusammenarbeit ausgewirkt hätte. Letzteres war jedoch nicht nur Terminkollisionen und gewöhnlichen Veränderungen in der Deutschfachschaft geschuldet, sondern wurde in einigen Fällen von den Gymnasien in Kauf genommen, da zeitlich befristet angestellte Lehrkräfte, z. B. Referendarinnen und Referendare oder Vertretungskräfte, entsendet wurden. Zum anderen musste festgestellt werden, dass es auch den teilnehmenden Deutschlehrkräften es in der Regel nicht gelang, ihre Kolleginnen und Kollegen gemäß dem Multiplikatorenprinzip zu informieren und vom Mehrwert der Fördermaßnahmen zu überzeugen. Daher konnten die Maßnahmen nur in wenigen Fällen etabliert werden.

⁶ In Bremerich-Vos (2019: 85–88) werden „lessons learned“ im Hinblick auf die Zusammenarbeit an der Schnittstelle Fachdidaktik – Schulpraxis formuliert.

folgenden Ausführungen beziehen sich ausschließlich auf diese zweite Phase des Projekts (*Ganz In II*, 2015 – 2018).

Der Fokus auf den Domänen *Lesen*, *Schreiben* und *Orthographie* blieb auf Wunsch der Schulen aber bestehen. Daher wurden an diesen Gymnasien jeweils in der Mitte der Schuljahre Testungen in diesen Bereichen durchgeführt. Jährlich konnten daher etwa 450 Schülerinnen und Schüler der Klassen 5 (Schuljahr 2015/2016), 6 (2016/2017) und 7 (2017/2018) getestet werden. Durchgeführt und ausgewertet wurden alle Testungen durch Mitglieder der universitären Arbeitsgruppe. Die Ergebnisse wurden den betreffenden Lehrkräften im Rahmen der Fortbildungen an den Schulen rückgemeldet und entsprechende Möglichkeiten der konkreten Förderung auf Individual- und Klassenebene diskutiert. Hierfür wurden vor allem bestehende Fördermaterialien gesichtet und ihr Nutzen gemeinsam reflektiert.

Als Vorüberlegung, bevor die Testinstrumente sowie die Fördermaßnahmen in der Domäne *Lesen* vorgestellt werden, soll *Lesekompetenz* vor dem Hintergrund großer Schulleistungstudien thematisiert werden.

2 | Lesediagnostik im Kontext großer Schulleistungstudien

Die Lesekompetenz deutscher Schülerinnen und Schüler wird regelmäßig in großen internationalen Schulleistungstudien – wie z. B. im Rahmen der *Internationalen Grundschul-Lese-Untersuchung* (IGLU) – untersucht (vgl. im Überblick Gailberger 2019). Bei IGLU bzw. PIRLS (*Progress in International Reading Literacy Study*) werden alle fünf Jahre nicht nur das Leseverständnis, sondern auch die Einstellungen und Lesegewohnheiten von Schülerinnen und Schüler am Ende der Grundschulzeit erfasst (vgl. Hußmann et al. 2017).

Das IGLU zugrundeliegende theoretische Konzept begreift *Lesen* als konstruktiven und interaktiven Prozess. Lesekompetenz (*reading literacy*) wird als die Fähigkeit verstanden, gesellschaftlich erforderliche und/oder individuell bedeutsame schriftsprachliche Formen (= Texte verschiedener Art) nicht nur zu verstehen, sondern auch – in unterschiedlichen Lebenssituationen – nutzen zu können (vgl. ebd.: 37). In IGLU 2016 wurden zur Erfassung der Lesekompetenz Verständnisaufgaben zu sechs informierenden und sechs erzählenden Texten eingesetzt. Zusätzlich zu den international vorgegebenen Tests wurden in Deutschland zwei Untertests des Instruments *Ein Leseverständnistest für Erst- bis Siebtklässler – Version II* (ELFE II) genutzt (Lenhard, Lenhard & Schneider 2018), mit welchem auch die Leistungen der leistungsschwächeren Leserinnen und Leser differenziert erfasst werden konnten.

Zur inhaltlichen Interpretation der Testwerte wurden in IGLU Kompetenzstufen gebildet. Dafür wurden die Lesefähigkeiten und die Aufgabenschwierigkeiten auf einer gemeinsamen Skala abgebildet, die in fünf gleich große Abschnitte zwischen 400 und 625 Skaleneinheiten unterteilt werden kann. Diese Kompetenzstufen sind in Tabelle 1 beschrieben (Bremerich-Vos, Wendt & Bos 2017: 92).

| Kompetenzstufe | | Skalenbereich der Fähigkeit |
|----------------|--|-----------------------------|
| V | Unter Bezug auf Textpassagen bzw. den Gesamttext Informationen ordnen und Aussagen selbstständig interpretierend und kombinierend begründen. | > 625 |
| IV | Für die Herstellung von Kohärenz auf der Ebene des Textes relevante Aspekte des Inhalts und der Darstellung erfassen und komplexe Schlüsse ziehen. | 551-625 |
| III | „Verstreute“ Informationen verständig miteinander verknüpfen | 476-550 |
| II | Explizit angegebene Informationen identifizieren und auf lokaler Ebene Kohärenz herstellen. | 400-475 |
| I | Rudimentäres Leseverständnis | < 400 |

Tab. 1: Beschreibung der Kompetenzstufen und Skalenwerte in IGLU 2016

Schülerinnen und Schüler, welche Kompetenzstufe III nicht erreichen, schaffen es weder, anspruchsvolle Interpretationsleistungen zu erbringen noch komplexe Schlussfolgerung oder einfache(re) textbasierte Verstehensleistungen zu erbringen. Es ist davon auszugehen, dass diese Leserinnen und Leser nicht nur beim Lesen selbst, sondern auch in anderen Domänen des Deutschunterrichts⁷ sowie in schriftbasierten Aufgaben des Mathematik- oder Sachunterrichts⁸ große Schwierigkeiten haben bzw. diese während der Sekundarstufe I auch in weiteren Fächern haben werden.

Betrachtet man die IGLU Ergebnisse aus dem Jahr 2016, so zeigt sich, dass im Vergleich zur ersten Erhebung im Jahr 2001 der Anteil leistungsschwacher Viertklässlerinnen und Viertklässler gestiegen ist. Betrug der Anteil der Leserinnen und Leser unter Kompetenzstufe III 2001 noch 16,9 Prozent, waren es im Jahr 2016 schon 18,9 Prozent (Bremerich-Vos, Wendt & Bos 2017: 110 ff.). Diese Schülerinnen und Schüler ‚erlesen‘ Wörter häufig noch, d.h. sie lesen lautierend oder syllabierend. Damit erfassen sie die Bedeutung dieser Wörter nicht sofort und haben aufgrund der begrenzten Kapazität des Arbeitsgedächtnisses kaum Ressourcen für komplexere Prozesse frei, wie z.B. für die Verarbeitung größerer Einheiten auf Satz- und Textebene.

Doch es sind nicht nur die schwachen Leistungen der Schülerinnen und Schüler, die von Interesse sind, sondern auch die Leistungsheterogenität, ausgedrückt durch die Streuung der Testwerte. Betrachtet man Deutschland, so fallen hier die Leistungen im Jahr 2016 deutlich heterogener aus als im Jahr 2001 (vgl. ebd.: 121). Diese (zunehmende) Heterogenität stellt natürlich nicht nur die Grundschulen, sondern auch die weiterführenden Schulen in Deutschland vor große Herausforderungen. Um schwächere Leserinnen und Leser bestmöglich zu unterstützen, ist eine gezielte (individuelle) Leseförderung, auch noch in der Sekundarstufe I, unabdingbar.

⁷ Für den Zusammenhang zwischen Lese- und Revisionskompetenz siehe auch Ferencik-Lehmkuhl (2017).

⁸ Siehe hierzu auch der Beitrag von Gailberger et al. im vorliegenden Band.

3 | Vorstellung der Test-Instrumente und Darstellung der Rückmeldung

3.1 | Überblick über die eingesetzten Instrumente

Im Folgenden werden für die drei Messzeitpunkte zunächst die Instrumente, die zur Testung der Leseleistung eingesetzt wurden, erläutert, um danach exemplarisch die Rückmeldungen, die die Lehrkräfte bekamen, vorzustellen. Zu allen drei Messzeitpunkten wurden Paper-Pencil-Tests eingesetzt. Die Durchführung der Tests erfolgte vor Ort in den Schulen jeweils durch geschulte Mitglieder der universitären Arbeitsgruppe. Zum ersten Messzeitpunkt konnte kein Lesetest eingesetzt werden, der über eine Normierung für die Jahrgangsstufe 5 bis 7 verfügt hätte.⁹ Aus folgenden, zu diesem Zeitpunkt zur Verfügung stehenden Alternativen musste daher ausgewählt werden:

ELFE 1-6: *Ein Leseverständnistest für Erst- bis Sechstklässler* (Lenhard & Schneider 2006)

FLVT 5-6: *Frankfurter Leseverständnistest für 5. und 6. Klassen* (Souvignier et al. 2008)

LESEN 6-7: *Lesetestbatterie für die Klassenstufe 6-7* (Bäuerlein, Lenhard & Schneider 2012)

LGVT 6-12: *Lesegeschwindigkeits- und Verständnistest für die Klassen 6-12* (Schneider, Schlagmüller & Ennemoser 2007)

Die zu erwartenden Deckeneffekte, die beim Einsatz der kurzen Texte beim Test ELFE I entstanden wären, schlossen diese als Testinstrument aus. Ein ähnliches Ausschlusskriterium ist für den FLVT zu formulieren. Daher wurde LESEN 6-7 ausgewählt, obwohl auch für dieses Testverfahren keine gesicherten Normen für 5. Klassen zur Verfügung standen. Zum zweiten Messzeitpunkt wurde ebenfalls LESEN 6-7 eingesetzt. Bereits zu diesem Messzeitpunkt mussten jedoch seitens der Schülerinnen und Schüler Einbußen auf motivationaler Seite festgestellt werden. Bei einem erneuten Einsatz wären daher überdies auch Wiederholungseffekte zu befürchten gewesen. Aufgrund der dargestellten Gründe wurde LESEN 6-7 im Rahmen des 3. Messzeitpunkts nicht erneut eingesetzt, sondern anstelle dessen der LGVT 6-12 genutzt.¹⁰ Tabelle 2 zeigt einen Überblick über die eingesetzten Tests und die Anzahl der Schülerinnen und Schüler, die an den Testungen jeweils teilgenommen haben. An allen drei Messzeitpunkten haben 118 Schülerinnen und Schüler teilgenommen. Bei der Interpretation der Ergebnisse ist im Folgenden daher der Umstand, dass die Testung der Leseleistung nicht als echt-längsschnittliche Untersuchung erfolgte, zu berücksichtigen.

| Messzeitpunkt | Schuljahr | getestete (Klassen) | SuS | eingesetzter Lesetest | Dauer |
|-----------------|-----------|---------------------|-----|-----------------------|--------------|
| Messzeitpunkt 1 | 2015/16 | 473 (20) | | LESEN 6-7 | 1x 3min plus |
| Messzeitpunkt 2 | 2016/17 | 479 (20) | | LESEN 6-7 | 2x 18min |
| Messzeitpunkt 3 | 2017/18 | 417 (20) | | LGVT 6-12 | 1x 4 min |

Tab. 2: Testinstrumente und Stichprobe

⁹ Mit ELFE II liegt ein solches Instrument mittlerweile vor (vgl. Lenhard, Lenhard & Schneider 2018).

¹⁰ Auch für dieses Instrument wurde in der Zwischenzeit eine erweiterte, neu normierte Auflage (LGVT 5-12+) veröffentlicht, die auch für 5. Klassen Normierungsdaten enthält (vgl. Schneider, Schlagmüller & Ennemoser 2017).

3.1.1 | LESEN 6-7

Das Verfahren dient sowohl der Überprüfung der basalen Lesekompetenz als auch des tiefergehenden Textverständnisses. Zu beiden Facetten der Lesekompetenz liegt jeweils ein Subtest vor (vgl. Bäuerlein, Lenhard & Schneider 2012). Die Testung wurde als Gruppentest im Klassenverband eingesetzt (möglich wären auch Einzeltests). Die Bearbeitungsdauer beträgt 45 Minuten mit einer reinen Bearbeitungszeit von maximal 39 Minuten.

Im Bereich der basalen Textkompetenz werden die Dekodier-Geschwindigkeit und die Dekodier-Genauigkeit auf Satzebene getestet, indem die Schülerinnen und Schüler eine Liste von 100 kurzen Sätzen wie z.B. „Tee ist ein Getränk“ lesen und diese Sätze jeweils mit einem Kreuzchen bei „richtig“ oder „falsch“ auf ihre inhaltliche Richtigkeit hin beurteilen. Binnen der Testzeit von 3 Minuten ist es nicht möglich, alle Sätze zu lesen, sodass das Ziel des Aufgabenformats darin besteht, möglichst viele Sätze zu lesen und als „richtig“ oder „falsch“ zu bewerten. Das Textverständnis wird getestet, indem je ein expositorischer und ein narrativer Text gelesen und dazu je 17 Multiple-Choice-Fragen beantwortet werden. Pro Lesetext stehen den Schülerinnen und Schülern jeweils 18 Minuten Zeit zur Verfügung. Alle MC-Items haben 5 Antwortmöglichkeiten und sind eindeutig lösbar.

3.1.2 | LGVT 6-12

Auch der LGVT 6-12 kann sowohl als Einzel- als auch als Gruppentest eingesetzt werden und wurde im Rahmen von *Ganz In* im Klassenverband eingesetzt. Die Zielsetzung des LGVT besteht darin, das Leseverständnis und die Lesegeschwindigkeit zu ermitteln (vgl. Schneider, Schlagmüller & Ennemoser 2007). Nach der Bearbeitung eines Beispiels lesen die Schülerinnen und Schüler einen narrativen Text mit dem Titel „Brot und Rosenkohl“, der aus 1727 Wörtern besteht. An 23 Stellen im Text müssen die Leserinnen und Leser aus drei Wortalternativen in Klammern diejenige auswählen, die am besten in den Textzusammenhang passt. Der erste Beispielsatz lautet:

Die Giraffe ist eines der größten Säugetiere auf der Welt, sie kann bis sechs [*Zentimeter, Meter, Kilometer*] groß werden.

Das passende Wort, in diesem Fall *Meter*, soll jeweils unterstrichen werden. Für die Bearbeitung des Lesetests haben die Schülerinnen und Schüler 4 Minuten Zeit. Die richtige Antwort erfordert gründliches, sinnverstehendes Lesen; ein ausschließliches flüchtiges Überfliegen der Texte reicht nicht aus. Um darüber hinaus sicherzustellen, dass schnelle, weniger sorgfältige Leserinnen und Leser vor Ende der Testzeit an das Ende des Textes gelangen, wurde die Länge von knapp 1800 Wörtern gewählt. Innerhalb der zur Verfügung stehenden Bearbeitungszeit ist ein sorgfältiges Lesen des gesamten Texts nicht möglich. Wenn die Zeit vorbei ist, werden die Schülerinnen und Schüler aufgefordert, kenntlich zu machen, was das zuletzt gelesene Wort war sowie einen Kreis um die Nummer, die in derselben Zeile rechts steht, zu machen. Anhand dieses Wertes kann dann die Lesegeschwindigkeit bestimmt werden.

3.2 | Auswertung und Ergebnisdarstellung

3.2.1 | Auswertung

Die Auswertung des Tests LESEN 6-7 erfolgt zunächst für jeden Subtest einzeln und kann auch von fachlich nicht einschlägigen Personen durchgeführt werden, da Schablonen genutzt werden.

Im ersten Testteil können theoretisch 100 Rohwertpunkte vergeben werden, ein Rohwertpunkt pro korrekt gelöster Aufgabe. Die Rohwerte können zunächst mit den Rohwerten der entsprechenden Normstichprobe verglichen werden oder sie können den entsprechenden Normwerten (T-Werte und Prozentränge) zugeordnet werden. Im zweiten Testteil können für die korrekte Bearbeitung aller MC-Items zu den beiden Lesetexten insgesamt 34 Rohwertpunkte erreicht werden. Da für die 5. Klassen keine Normwerte vorlagen, wurden die T-Werte auf Basis der Testergebnisse aller fünf teilnehmenden Schulen berechnet. Das heißt, die Leistungen der Schülerinnen und Schüler einer *Ganz In*-Schule wurden zu den Leistungen der anderen Schülerinnen und Schüler an *Ganz In*-Gymnasien in Bezug gesetzt statt zu denen einer Normstichprobe. Das Gesamtergebnis berechnet sich aus den T-Werten der beiden Untertests und kann ebenfalls Vergleichswerten zugeordnet werden.

Beim LGVT 6-12 erhalten die Schülerinnen und Schüler für jede korrekte Unterstreichung 2 Punkte. Wenn ein falsches Wort unterstrichen wurde oder in einer Lücke mehrere Optionen unterstrichen wurden, wird 1 Punkt abgezogen. Wird eine erforderliche Unterstreichung nicht vorgenommen, erhält die Schülerin bzw. der Schüler 0 Punkte. Theoretisch wäre daher für das Textverständnis ein maximaler Wert von 46 Punkten erreichbar. Werte über 30 kommen in der Praxis jedoch fast nie vor, da innerhalb der zu Verfügung stehenden Bearbeitungszeit ein sorgfältiges Lesen des gesamten Texts nicht möglich ist. Für die Klassenstufen 6-9 liegen für alle Schularten Prozentrangnormen für die Anzahl korrekter Unterstreichungen und für die Menge der gelesenen Wörter vor.

3.2.2 | Ergebnisrückmeldung

Nachdem die Auswertung der Testungen erfolgt war, wurden die Testergebnisse für jede Schule gesondert an die Deutschlehrkräfte zurückgemeldet. Diese Rückmeldung erfolgte in Form von Tabellen mit Rohwertpunkten bzw. T-Werten. Um einen umfassenden Blick auf die Ergebnisse zu ermöglichen, wurden den Lehrkräften nicht nur die Ergebnisse auf der Individualebene der einzelnen Schülerinnen und Schüler zurückgemeldet, sondern es erfolgte jeweils auch eine tabellarische Darstellung auf Klassen- und Schulebene.

Für die Rückmeldung auf Individualebene wurden die Ergebnisse zusätzlich farblich hervorgehoben: T-Werte unter 40 wurden als kritisch angesehen, da diese mehr als eine Standardabweichung unter dem Mittelwert der jeweiligen Vergleichsgruppe liegen. Werte unter 40 wurden daher vor einem roten Hintergrund (hier = dunkelgrau unterlegt) dargestellt; lagen die Leistungen bis zu einer Standardabweichung unter dem Mittelwert, also zwischen 40 und 50, wurden sie gelb (hier = hellgrau) markiert (vgl. Tab. 3 und Tab. 4). Insgesamt erachteten die Lehrkräfte das Format der Rückmeldungen als nachvollziehbar und gewinnbringend für die weitere Arbeit.

In den Gesprächen mit den Lehrkräften wurden die Ergebnisse zunächst vorgestellt und erläutert. Anschließend wurden die Deutschlehrkräfte jeweils dazu aufgefordert, die Ergebnisse mit ihrer subjektiven Einschätzung der Lesekompetenz einzelner Schülerinnen und Schüler zu vergleichen.

Zwar stimmten Urteile und Testergebnisse häufig überein, in der Tendenz wurden die Leseleistungen von den Lehrkräften jedoch besser eingeschätzt, als die Testergebnisse es vermuten ließen. Beispielhaft sei die Erfahrung einer Lehrperson berichtet. Diese schilderte den Fall eines Kindes, das deutlich schlechter abgeschnitten hatte als erwartet. Auch die Eltern sahen ihr Kind als eine sehr gute Leserin an. Die Testergebnisse verhalfen dazu, diese Urteile kritisch zu hinterfragen und letztendlich auch zu revidieren. Dank des anschließenden Förderprogramms konnte sich das Kind im Lesen deutlich verbessern. Ohne Testung und Reflexion der Testergebnisse wäre es trotz schlechter Leseleistungen wahrscheinlich unauffällig geblieben, da es Strategien entwickelt hatte, um die Defizite „zu vertuschen“.

4 | Rückmeldung der Testergebnisse

4.1 | Rückmeldung der Testergebnisse auf Schul- und Klassenebene

4.1.1 | Ergebnisrückmeldungen: 1. und 2. Messzeitpunkt

Die Ergebnisse der Testung auf Klassen- und Schulebene für die 5. und 6. Klassen mithilfe des Lesetests LESEN 6-7 sind in den Abbildungen 1 und 2 aufgeführt.

Die Ergebnisse im Bereich der Leseflüssigkeit zeigen zum ersten Messzeitpunkt auf beiden Auswertungsebenen recht deutlich die Leistungsheterogenität innerhalb der Projektschulen. Erreichten die Schülerinnen und Schüler der Schule 01 bei den basalen Lesefähigkeiten im Durchschnitt 41,4 von 100 Rohwertpunkten, so waren es bei den Schülerinnen und Schüler der Schule 05 hingegen 48 Rohwertpunkte. Diese Differenz entspricht „umgerechnet“ einem durchschnittlichen „Vorsprung“ von rund sieben richtig beantworteten Fragen für die getesteten Schülerinnen und Schüler in der Schule 01. Betrachtet man die Standardabweichung als Maß für die Heterogenität bzw. Streuung der Leistungen innerhalb der Klassen, so zeigen sich an Schule 02 recht homogene Leistungen. Die Standardabweichungen reichen in diesen Klassen von 8,4 bis 9,5 Rohwertpunkten, wohingegen sich bspw. innerhalb der Schule 03 die Leistungen auf Klassenebene deutlicher voneinander unterscheiden. Die Klasse 03-c erreichte zwar mit 45,2 Rohwertpunkten gemessen am Durchschnittswert der Schule eine geringe Punktzahl, jedoch beträgt die Standardabweichung lediglich 6,8 Punkte. Werte wie dieser sprechen dafür, dass für die gesamte Klasse von eher schwachen Leserinnen und Lesern auszugehen ist. In Klasse 03-e ist im Gegensatz dazu die Heterogenität sehr viel höher. Dies zeigt sich in der breiten Streuung der Ergebnisse mit einer Standardabweichung von 16 Punkten.

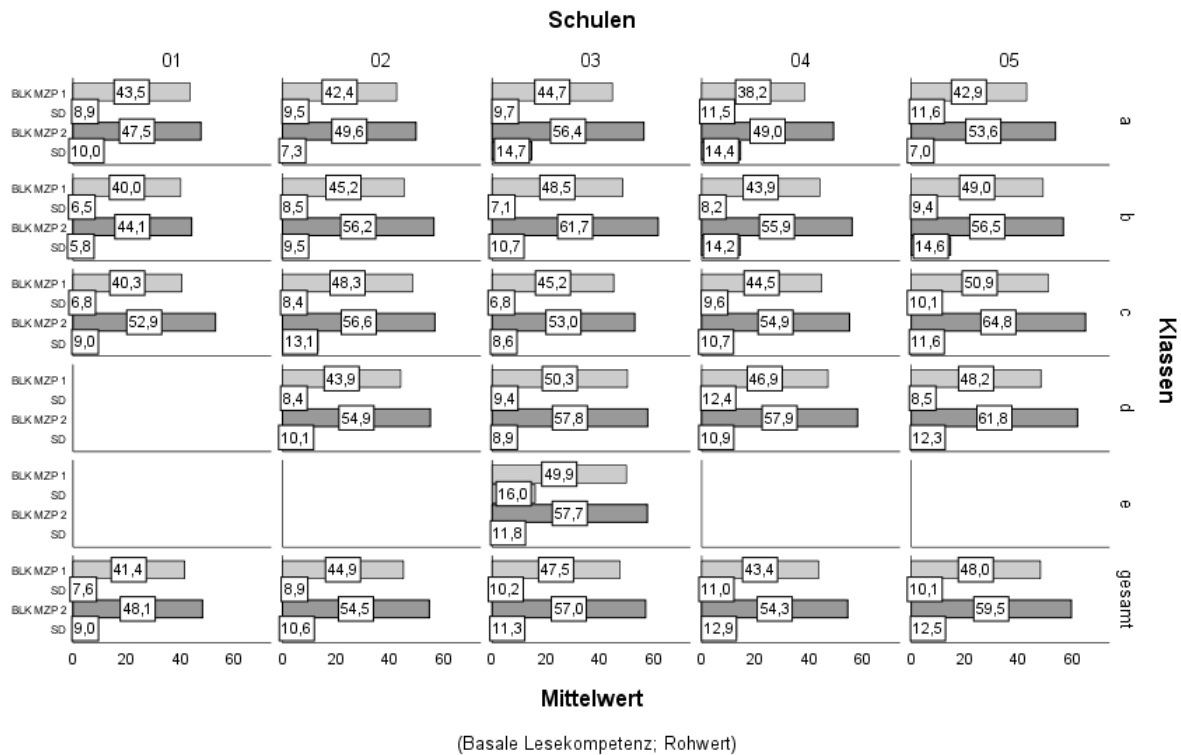


Abb. 1: Ergebnisse der Testung der basalen Lesekompetenz; MZP 1 (2015/2016) und MZP 2 (2016/2017)

Für das Textverständnis ergibt sich eine ähnliche Situation. Auch hier schneidet Schule 01 mit 11,5 Rohpunktwerten im Durchschnitt am schlechtesten und Schule 05 mit 18,7 Rohpunktwerten am besten ab. Betrachtet man auch für das Textverständnis erneut die Leistungsheterogenität, sind auch hier teils große Unterschiede innerhalb der Klassen einer Schule zu sehen. Die Schülerinnen und Schüler der Klasse 02-c erreichten durchschnittlich 10,3, die aus Klasse 02-d hingegen 14,3 Punkte. Dies entspricht einer Differenz von durchschnittlich drei richtig beantworteten Fragen. Betrachtet man erneut die Standardabweichung als Maß der Streuung, so deutet diese in den Klassen der Schule 05 auf eine geringe Leistungsheterogenität hin. Dieses Ergebnis lässt sich auch für andere Klassen anderer Schulen feststellen. Mit einer Standardabweichung von 7,7 ist in Klasse 03-e die größte Leistungsheterogenität aller Klassen des Projekts zum ersten Messzeitpunkt zu verzeichnen. Dieser Wert sticht sowohl im Kontrast zur gesamten Stichprobe als auch im Vergleich zu den anderen Klassen dieser Schule heraus.

Vergleicht man die Leistungen der Schülerinnen und Schüler von Klasse 5 zu Klasse 6 auf Schulebene, sieht man, dass überall eine Verbesserung der Leseleistung zu verzeichnen ist. Diese – teils sehr deutliche – Verbesserung auf allen Ebenen kann jedoch (aufgrund der Schwierigkeiten im Testdesign¹¹) nicht zweifelsfrei auf die jeweiligen Fördermaßnahmen zurückgeführt werden.

¹¹ Anfangs wurde ein Forschungsdesign mit Projekt- und Kontrollschulen gewählt. Allerdings führte der Versuch, den fünf Projektschulen jeweils „Zwillinge“, d.h. im Hinblick auf zentrale Merkmale vergleichbare *Ganz In-Gymnasien*, zur Seite zu stellen und dort in analoger Art und Weise zu testen, nicht zum Erfolg. Die Gründe waren vielfältig, doch der Hauptgrund liegt wohl in der fehlenden Motivation der Kontrollschulen (auf Lehrer- und Schülerseite), ausschließlich als Vergleichsgruppe zu dienen, ohne dazu eine Gegenleistung zu erhalten.

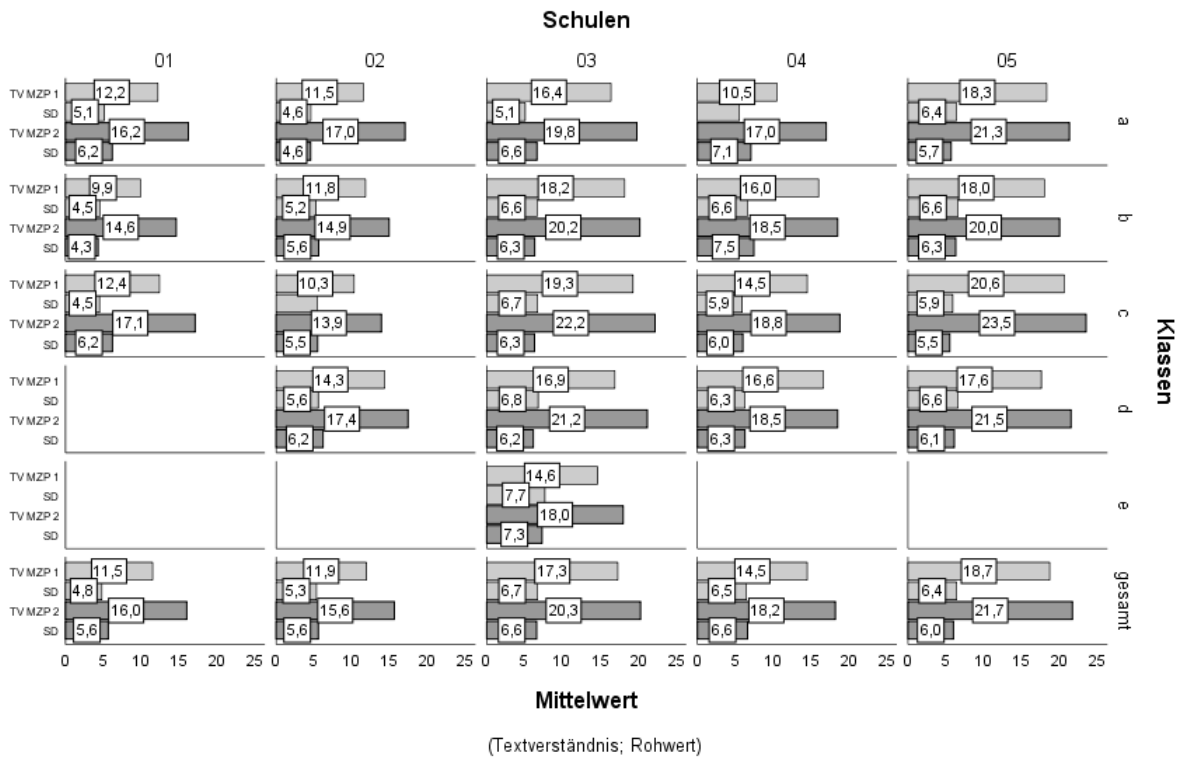


Abb. 2: Ergebnisse der Testung des Textverständnisses; MZP 1 (2015/2016) und MZP 2 (2016/2017)

Weiterhin zeigt ein Abgleich mit den Normwerten der Klassenstufe 6, dass die Leistungen der Schülerinnen und Schüler an vier von fünf Projektschulen weder im Bereich der basalen Lesekompetenz ($M = 61,2$; $SD = 13,9$) noch im Bereich des Textverständnisses ($M = 23,8$; $SD = 5,2$) an die gymnasiale Normstichprobe heranreichen.

4.1.2 | Ergebnismrückmeldungen: 3. Messzeitpunkt

Zum dritten Messzeitpunkt bestätigt sich das Bild einer insgesamt schwachen Leseleistung an den teilnehmenden Projektschulen. Auch konnte erneut eine teils erhebliche Streuung der Rohwerte auf Schulebene festgestellt werden. Die Abbildungen 3 und 4 zeigen die Ergebnisse des dritten Messzeitpunkts auf Klassen- und Schulebene für die 7. Klassen mithilfe des Lesetests LGVT 6-12.

Die leistungsstärkste Schule der ersten beiden Messzeitpunkte (Schule 05) ist auch zum dritten Messzeitpunkt erneut die einzige Schule, die im Leseverständnis mit 10,1 Punkten knapp an den gymnasialen Normwert für die 7. Klassen ($M = 11,9$) heranreicht. In der Lesegeschwindigkeit ist die Differenz mit unter zehn Wörtern auch die geringste innerhalb der Stichprobe. Im Mittel lasen die Schülerinnen und Schüler der Schule 05 642 Wörter; der Wert der gymnasialen Normstichprobe liegt bei 650 Wörtern.

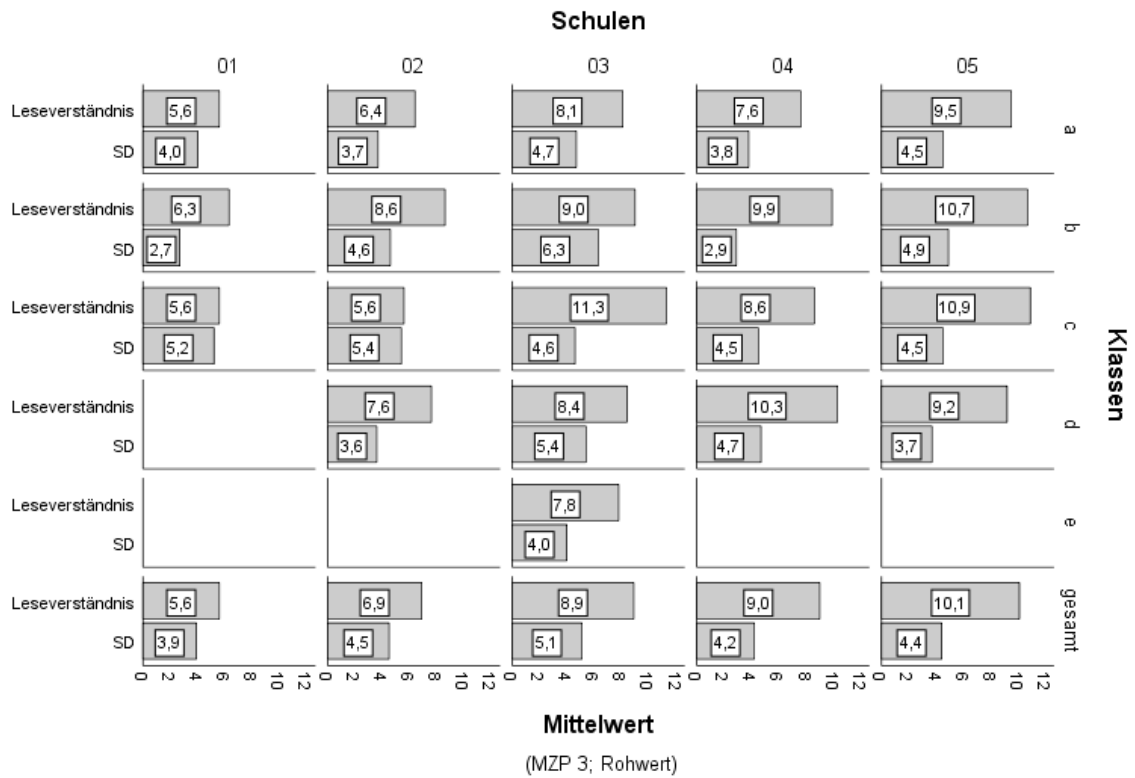


Abb. 3: Ergebnisse der Testung des Leseverständnisses; MZP 3 (2017/2018)

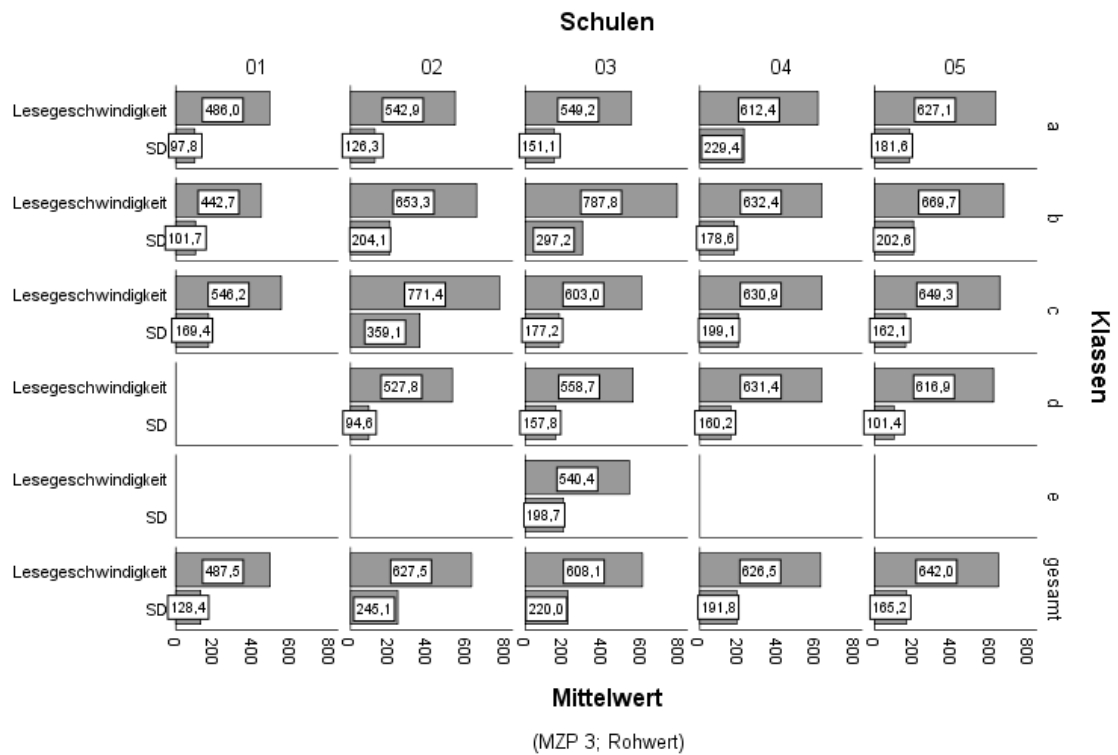


Abb. 4: Ergebnisse der Testung der Lesegeschwindigkeit; MZP 3 (2017/2018)

Bei der Lesegeschwindigkeit zeigt sich insgesamt auf Klassenebene eine etwas höhere Heterogenität als beim Leseverständnis. Diese soll im Folgenden kurz anhand der Werte aus den Schulen 03 und 02 verdeutlicht werden. Die vier Klassen der Schule 03 zeigen die größte Heterogenität. So lasen die Schülerinnen und Schüler der Klasse 03-e im Durchschnitt 247,4 Wörter weniger als die der Klasse 03-b. Nur geringfügig kleinere Unterschiede zeigen sich in Schule 02, da hier in Klasse 02-c durchschnittlich 243,6 Wörter mehr gelesen wurden als in Klasse 02-d. Des Weiteren zeigt sich in dieser Schule in Klasse 02-c eine bemerkenswerte Streuung der Werte um den Mittelwert: Mit einer Standardabweichung von 359,1 Punkten ist dieser Wert sehr hoch. Im Vergleich dazu liegt die Standardabweichung in Klasse 02-d nur bei 94,6 Punkten. Betrachtet man die Ergebnisse für Schule 02 analog zu den Ergebnissen der ersten beiden Messzeitpunkte, so zeigen sich hier erneut schwache Leistungen bei den Schülerinnen und Schülern aus Klasse 02-d, wohingegen die Klasse 02-c besonders in Bezug auf die Heterogenität ihrer Leseleistungen heraussticht.

4.2 | Rückmeldung der Testergebnisse auf Individualebene

4.2.1 | Beispiele für die Ergebnisrückmeldungen: 1. und 2. Messzeitpunkt

Nachdem bisher die Rückmeldungen auf Klassen- bzw. Schulebene vorgestellt wurden, sollen im Folgenden die Rückmeldungen auf Individualebene im Vordergrund stehen. Die folgenden Tabellen zeigen T-Werte für einzelne Schülerinnen und Schüler, wobei Werte mit einer bzw. zwei Standardabweichungen nach unten grafisch hervorgehoben sind.

Dies gilt auch für die folgende Tabelle 3, die exemplarisch die Ergebnisse der Testung der Lesekompetenz zum ersten und zweiten Messzeitpunkt aus der Klasse c aus Schule 03 zeigt. Schülerinnen und Schüler, bei denen Werte in der Tabelle fehlen, waren bei der Testung abwesend. Beim direkten Vergleich beider Klassen zeigt sich die bereits auf Klassenebene angesprochene Heterogenität der Klassen der Schule 03. Hier sind sowohl sehr starke (z.B. ID 03321) als auch eher schwache (z.B. 0332, 03315) Leserinnen und Leser zu finden (vgl. Tabelle 3).

| Schule / Klasse | Schüler-ID | Basale Lesekompe- tenz MZP 1 | Textverständnis MZP 1 | Basale Lesekompe- tenz MZP 2 | Textverständnis MZP 2 |
|--------------------|------------|---------------------------------|--------------------------|---------------------------------|--------------------------|
| | | T-Wert | T-Wert | T-Wert | T-Wert |
| 03-c | 0331 | 44 | 49 | 39 | 35 |
| | 0332 | 45 | 39 | 38 | 37 |
| | 0333 | 46 | 63 | 41 | 44 |
| | 0334 | 48 | 47 | 50 | 37 |
| | 0335 | 50 | 56 | 34 | 37 |
| | 0336 | 58 | 53 | 51 | 45 |
| | 0337 | 51 | 65 | 41 | 55 |
| | 0338 | 68 | 60 | 49 | 57 |
| | 0339 | 54 | 70 | 45 | 63 |
| | 03310 | 41 | 43 | | |
| | 03311 | 53 | 43 | 43 | 40 |
| | 03312 | 40 | 64 | 46 | 60 |
| | 03313 | 43 | 50 | 51 | 35 |
| | 03314 | 48 | 58 | 39 | 41 |
| | 03315 | 46 | 35 | 34 | 21 |
| | 03316 | 58 | 58 | 52 | 44 |
| | 03317 | 48 | 50 | 36 | 44 |
| | 03318 | - | - | 60 | 60 |
| | 03319 | 52 | 56 | 41 | 60 |
| | 03320 | 56 | 72 | 52 | 63 |
| | 03321 | 65 | 63 | 58 | 55 |
| | 03322 | 51 | 64 | - | - |
| | 03323 | 44 | 62 | 35 | 54 |
| | 03324 | 50 | 68 | 41 | 55 |
| | 03325 | 48 | 68 | 41 | 55 |
| | 03326 | 48 | 65 | 47 | 60 |

Tab. 3: Ergebnisse der Testung der Lesekompetenz auf Individualebene; Schule 03; Klasse 03-c; MZP 1 (2015/2016) und MZP 2 (2016/2017)

Bei der Interpretation der Ergebnisse muss erneut beachtet werden, dass die Leistungen der Schülerinnen und Schüler der 5. Klassen mit denen der Schülerinnen und Schüler der anderen *Ganz In*-Schulen verglichen wurden, wohingegen die Leistungen der Schülerinnen und Schüler der Klasse 6 mit den Leistungen aus der Normstichprobe für 6. Gymnasialklassen verglichen werden konnte. Da die gymnasiale Normstichprobe insgesamt leistungsstärker war als die *Ganz In*-Stichprobe, ist es verständlich, dass die Ergebnisse an *Ganz In* Schulen insgesamt etwas schlechter ausfallen. Daher ist der Schluss, dass es scheinbar zu einer Verschlechterung der Leistungen gekommen ist, nach dem ersten Blick auf die T-Werte des ersten Messzeitpunktes und zweiten Messzeitpunkt zwar naheliegend, jedoch wenig plausibel.

4.2.2 | Beispiele für die Ergebnisrückmeldungen: 3. Messzeitpunkt

Für den dritten Messzeitpunkt wird auf Individualebene – im Gegensatz zur Darstellung der ersten beiden Messzeitpunkte – eine Klasse aus Schule 02 mit eher homogenen (schwachen) Leistungen (Klasse d) präsentiert (vgl. Tabelle 4).

| Schule / Klasse | Schüler-ID | Leseverständnis MZP 3 | Lesegeschwindigkeit MZP 3 |
|-----------------|------------|-----------------------|---------------------------|
| | | <i>T-Wert</i> | <i>T-Wert</i> |
| 02/d | 0241 | 38 | 51 |
| | 0242 | - | - |
| | 0243 | - | - |
| | 0244 | 40 | 37 |
| | 0245 | 42 | 42 |
| | 0246 | - | - |
| | 0247 | 38 | 38 |
| | 0248 | 27 | 44 |
| | 0249 | - | - |
| | 02410 | 38 | 39 |
| | 02411 | 49 | 44 |
| | 02412 | 51 | 41 |
| | 02413 | 34 | 45 |
| | 02414 | 53 | 50 |
| | 02415 | - | - |
| | 02416 | - | - |
| | 02417 | 53 | 45 |
| | 02418 | 45 | 40 |
| | 02419 | 42 | 45 |
| | 02420 | - | - |
| | 02421 | - | - |
| | 02422 | 53 | 50 |
| | 02423 | 34 | 41 |
| | 02424 | 45 | 41 |
| | 02425 | 42 | 44 |
| | 02426 | 38 | 50 |
| | 02427 | 34 | 44 |
| | 02428 | 38 | 52 |

Tab.4: Ergebnisse der Testung der Lesekompetenz auf Individualebene; Schule 02; Klasse 024; MZP 3 (2017/2018)

Den dargestellten Ergebnissen zufolge ist in Klasse 02-d die Anzahl derer, die in beiden Bereichen der Lesekompetenz gute oder unauffällige Befunde zeigen, gering. Auch zum dritten

Messzeitpunkt gibt es noch drei Schülerinnen und Schüler in dieser Klasse, die nicht hinreichend flüssig lesen können; in zwei Fällen überschneiden sich diese Probleme mit erheblichen Defiziten im Leseverständnis. Weitere sieben Schülerinnen und Schüler haben ausschließlich im Bereich des Leseverständnisses defizitäre Leistungen gezeigt.

5 | Fördermaßnahmen

Im Rahmen der Ergebnismeldungen wurde mit den Lehrkräften vor Ort ausgearbeitet, wie auf die jeweilige Situation förderlich reagiert werden könne. Dazu wurden – abhängig von den individuellen Leseergebnissen sowie der Klassenleistung – Möglichkeiten der Leseförderung erläutert und ggfs. bereits bestehende Trainings vorgestellt. Gemeinsam wurde dann jeweils dasjenige Konzept, das die beste Passung auf die Gegebenheiten an die bestehenden Strukturen der Schule aufzuweisen schien, ausgewählt bzw. ausgearbeitet.

5.1 | Förderung der Leseflüssigkeit

In Deutschland wurde der Leseflüssigkeit und den verschiedenen Formen ihrer Förderung erst in den letzten Jahren verstärkt Aufmerksamkeit geschenkt (vgl. Rosebrock et al. 2011). Die dominierende Vorstellung bestand darin, dass Kinder spätestens nach Beendigung der Grundschulzeit flüssig lesen können. Inzwischen ist jedoch klar, dass die Fähigkeiten des flüssigen und sinnentnehmenden Lesens nicht nur vielen Kindern Anfang der Sekundarstufe I Probleme bereitet, sondern auch noch Jugendliche Mitte und Ende der Sekundarstufe I Förderbedarfe in diesem Bereich haben.

Populäre Verfahren wie stille Lese-Phasen oder „Trainings“ wie das in Deutschland traditionell recht frequent praktizierte laute Reihum-Vorlesen in der Klasse gelten als wenig bis nicht zielführend (vgl. zusammenfassend NICHD 2000). Gegenteiliges ist über die Befundlage zur Wirksamkeit des deutlich weniger populären lauten bzw. halblauten Lesens zu sagen. Metaanalysen englischsprachiger Studien belegen die Wirksamkeit von Formen, bei denen schwächere Leserinnen und Leser von flüenteren Schülerinnen und Schülern unterstützt werden. Die Arbeitsgruppe um Cornelia Rosebrock entwickelte vor diesem Hintergrund das kooperative Lautleseverfahren „Lautlese-Tandems“ für sechste Hauptschulklassen, das sich sehr effektiv im Hinblick auf eine Verbesserung der Leseflüssigkeit zeigte (vgl. im Folgenden Rosebrock et al. 2010; Rosebrock et al. 2011).

Bei den Lautlese-Tandems arbeiten immer in einem Zweierteam je ein stärkerer und ein schwächerer Schüler bzw. je eine stärkere und eine schwächere Schülerin zusammen. Für dieses Verfahren wurde eine Rahmenhandlung aus dem Bereich des Sports gewählt, die zusätzlich motivierend wirken soll: Die eine Partei wird „Trainer“ bzw. „Trainerin“ genannt, die andere „Spieler“/„Spielerin“ oder „Sportler“/„Sportlerin“. Der stärkere Part fungiert dabei als Lesemodell, von dem der schwächere Part des Teams durch Beobachtung bzw. Nachahmung lernt. Allerdings profitieren auch die kompetenteren Schülerinnen und Schüler, da sie lernen, nicht nur selbst intensiv auf die Lesegenauigkeit zu achten und über Wortbedeutungen nachzudenken,

sondern auch, ein lernförderliches und vor allem wertschätzendes Feedback zu geben. Die Rollenverteilung ist dabei veränderlich. So kann ein ‚Spieler‘ bei entsprechender Verbesserung später auch als ‚Trainer‘ fungieren.

Die Schülerpaare bekommen ein gemeinsames Exemplar eines kurzen Textabschnitts und lesen diesen simultan und halblaut vor. Der ‚Trainer‘ führt den Finger im Text mit und gibt so die Lesegeschwindigkeit vor. Wird ein Wort falsch gelesen und der ‚Sportler‘ korrigiert sich nicht selbst, verbessert der ‚Trainer‘ und streicht das fehlerhafte Wort an. Wird kein Fehler gemacht, erhält der ‚Spieler‘ Lob. Fühlt sie oder er sich sicher, darf auch allein gelesen werden. Der Text wird viermal gelesen. Danach überprüft die Lehrkraft den Lernerfolg. Um die besten Erfolge zu erzielen, sollten die Tandems dreimal wöchentlich für etwa 15-20 Minuten durchgeführt werden. Die Lehrkraft hält sich bei dem eigentlichen Training zurück und begleitet es hauptsächlich aus dem Hintergrund. Ihr obliegt – neben der Kontrolle des Leistungsfortschritts – aber das Erklären des Verfahrens, die Einführung der Methode und die Zusammensetzung der Tandems im Vorfeld des Trainings. Sie muss auch geeignete Texte auswählen. Für den Einstieg stellen Rosebrock et al. (2011) aber einen Reader mit Texten für 3., 4., 5. und 6. Klasse zur Verfügung. Diese Texte sind bis zu 300 Wörtern lang und auch für schwache Leserinnen und Leser gut handhabbar (vgl. Ferencik-Lehmkuhl 2019).

In den Rückmeldegesprächen wurden von den anwesenden Mitgliedern der universitären Arbeitsgruppe Leseflüssigkeitstrainings wie das hier beschriebene Lesetandem-Konzept als Fördermaßnahme für solche Schülerinnen und Schülern empfohlen, die bei dem jeweiligen Lesetest einen T-Wert von unter 40 erreichten. Dieses Verfahren war in der ersten Phase von *Ganz In* (2010–2015) bereits erfolgreich eingesetzt worden (vgl. Bremerich-Vos et al. 2015). Auf Grund seiner Eignung für den Einsatz im Klassenverbund und in Kleingruppen konnten jeweils verschiedene Einsatzmöglichkeiten diskutiert werden.

5.2 | Förderung des Leseverstehens

Zur Förderung der hierarchiehöheren Teilfähigkeiten haben sich Konzepte bewährt, die Lesestrategien vermitteln. Damit wird dem Umstand Rechnung getragen, dass verstehendes Lesen „keineswegs eine passive Rezeption von im Text kodierten Inhalten ist, sondern einen hochgradig aktiven Prozess der Rekonstruktion von Bedeutungen darstellt“ (Lenhard 2019: 27).

Lesestrategien können wie Lernstrategien im Allgemeinen als „Prozeduren bzw. Routinen, die bei der Ausführung von [...] Aufgaben eingesetzt werden“ definiert werden (Leopold 2009: 11). Um ein bestimmtes Lernziel erreichen zu können, werden durch Strategien einzelne Fähigkeiten und Fertigkeiten zu einem Plan für einen Handlungsablauf organisiert und koordiniert (vgl. ebd.).

Für die schulische Vermittlung und Einübung von Lesestrategien ist zu beachten, dass insbesondere Trainings, die kognitive Lesestrategien mit metakognitiven Selbstregulationsstrategien an inhaltlich relevanten Texten verbinden, erfolgsversprechend sind (vgl. Stöger, Sonntag & Greindl 2012: 139). Das Konzept der „Selbstregulation“ spielt sowohl in der lesedidaktischen als auch in der pädagogisch-psychologischen Forschung der jüngeren Vergangenheit eine große Rolle. Dort werden seit einigen Jahren verstärkt Lernprozesse untersucht, die der Lerner oder die Lernerin selbst steuert (Souvignier & Mokhlesgerami 2006). Die Strategieanwendung wird dabei durch wiederholte Übung und Reflexion deutlich verbessert (vgl. Karstens, Schmitz & Jost

2019: 4). Die größte Wirksamkeit kann daher Förderungsvarianten zugesprochen werden, die als regelmäßige Trainings angelegt (vgl. Souvignier & Trenk-Hinterberger 2010).

Unter *kognitiven* Lernstrategien subsumieren wir „Prozesse, die unmittelbar mit der Informationsaufnahme und Informationsverarbeitung verknüpft sind“ (Wild 2010: 483), wie etwa Wiederholungs-, Organisations- und Elaborationsstrategien. Wenn man etwa den Text oder seine Aufzeichnungen mehrmals durchliest oder auswendig lernt, übt man Strategien des Wiederholens aus. Zu den Organisationsstrategien zählt zum Beispiel das Unterstreichen. Wenn das eigene Vorwissen mit durch den Text bereitgestellten neuen Informationen verknüpft wird, werden Elaborationsstrategien angewendet. Im Gegensatz zu den *kognitiven* Strategien beziehen sich die *metakognitiven* Strategien auf das ziel- und aufgabenbezogene Planen der Leseaktivität, das Überprüfen des Textverstehens während des Leseprozesses sowie das Regulieren der Prozesse. Wird z.B. während des Lesens deutlich, dass ein Inhalt nicht verstanden worden ist, kann man z.B. eine Passage neu lesen und sein Lesetempo insgesamt verlangsamen.

Ein bekanntes Programm, das zum Ziel hat, Kompetenzen zum selbstständigen Umgang mit Texten zu vermitteln, stammt aus der Arbeitsgruppe von Gold et al. und wurde im Rahmen eines Forschungsprojekts zur Förderung der Lesekompetenz entwickelt (vgl. im Folgenden Gold et al. 2004). Das Basisprogramm „Wir werden Textdetektive“ richtet sich an Schülerinnen und Schüler der Klassen 5 und 6 im Gymnasium und in der Gesamtschule. Insgesamt nimmt die Durchführung des gesamten Programms 20-25 Unterrichtsstunden in Anspruch. Ausgewählte Lesestrategien werden vorgestellt und eingeübt, damit sie am Ende von allen Schülerinnen und Schüler selbstreguliert eingesetzt werden können. Die Fertigkeiten zur metakognitiven Überwachung, Steuerung und Regulation werden hierbei als wichtige Voraussetzungen für erfolgreiches Lernen angesehen (vgl. Glaser 2005: 40). Der Fokus der Leserinnen und Leser wird mehr in die Richtung gelenkt, dass sie selbst dazu beitragen können, dass die jeweilige Textlektüre gelingt. Dadurch haben Schülerinnen und Schüler ein deutlich wahrnehmbares Erfolgserlebnis. Indem sie Strategien anwenden, die zur jeweiligen Textsorte und Aufgabe passen, wird ein Text besser verstanden und letztendlich das Leseinteresse bzw. die Lesemotivation positiv beeinflusst.

Wurde bei den Lese-Tandems nach Rosebrock et al. (2010) eine sportliche Rahmenhandlung zur Situierung gewählt, so wird hier die Parallele zur Arbeit eines Detektivs gezogen, da dieser ebenfalls durch gezielte Planung und methodisches Vorgehen zum Ziel gelangt und „Fälle“ lösen kann. Daher werden keine „Lernstrategien“ oder „Lesestrategien“, sondern sieben sogenannte „Detektivmethoden“ vorgestellt: 1. Überschrift beachten, 2. Bildlich vorstellen, 3. Umgang mit Textschwierigkeiten, 4. Verstehen überprüfen, 5. Wichtiges unterstreichen, 6. Wichtiges zusammenfassen, 7. Behalten überprüfen. Bei der Untersuchung der Wirksamkeit des Programms wurde festgestellt, dass die Vermittlung von Lesestrategien im Unterricht im Vergleich zu einer herkömmlich unterrichteten Kontrollgruppe besser gelingt und dass sich das Leseverständnis verbessern ließ (vgl. Mokhlesgerami 2004). Der Umfang der Verbesserung beträgt – wenn man die Leistungen der „trainierten“ Schülerinnen und Schüler mit denen ihrer Mitschülerinnen und Mitschüler aus der herkömmlich unterrichteten Kontrollgruppe vergleicht – im Mittel etwa eine halbe Standardabweichung ($d = .46$), was als Effekt von mittlerer Höhe gedeutet werden kann.

In den Rückmeldegesprächen wurde das Programm für solche Schülerinnen und Schüler vorgeschlagen, die zwar schon über basale Lesekompetenz verfügten, d.h. flüssig lesen konnten, aber noch Schwächen beim Leseverständnis zeigten. Auch hier wurde der flexible Umgang mit Blick auf unterschiedliche Unterrichtssettings diskutiert. Eine Anregung seitens der universitären Forschungsgruppe bestand darin, die Durchführung des Programms über einen längeren Zeitraum „zu strecken“, anstatt es ohne Unterbrechung im Unterricht durchzuführen. Der Befürchtung einer „Verwässerung“ des Programms stand gegenüber, dass Zeit dafür gewonnen wurde, die Anwendung der Detektivmethoden in weiteren Unterrichtsstunden im Deutschunterricht sowie im Unterricht anderer Fächer zu üben. Empirische Untersuchungen zur Wirkung schulischer Lesestrategietrainings kommen übereinstimmend immer wieder zu Schlüssen wie dem folgenden: „Hierbei ist jedoch wichtig, dass die vermittelten Strategien von Lehrkräften immer wieder aufgegriffen und auf neue Textarten und Lerninhalte angewendet werden. Dies kann innerhalb eines Faches umgesetzt werden, sollte möglichst jedoch auch fachübergreifend geschehen“ (Stöger, Sonntag & Greindl 2012: 155; vgl. auch Souvignier & Trenk-Hinterberger 2010).

Über Fragen der konkreten Umsetzung hinaus äußerten die Lehrpersonen im Rahmen der Rückmeldegespräche Rückfragen, die drei Bereichen zuzuordnen sind. Zum einen betrafen viele Fragen die Kombinationsmöglichkeiten des vorgeschlagenen Trainingsprogramms mit bereits an der Schule etablierten Verfahren wie z.B. der 5-Schritt-Lesemethode. Zweitens wurde oft die Frage aufgeworfen, welches Textmaterial geeignet sei, um die Strategien einzuüben. Dabei konnte auf eine von den Autoren des Programms getroffene Auswahl an Sach- und narrativen Texten verwiesen werden. Darüber hinaus wurde angeregt, an eigenen, schulinternen Zusammenstellungen spannender und informierender Texte zu arbeiten. Eine solche Liste könnte zum Beispiel in der Fachkonferenz Deutsch diskutiert werden und allen Lehrpersonen des Fachs zugänglich gemacht werden. Drittens wurden auf Wunsch einiger Lehrkräfte zusätzlich Möglichkeiten der Kombination eines Lesestrategietrainings mit kooperativen (peer-gestützten) Arbeitsformen vorgestellt (vgl. Ferencik-Lehmkuhl, Schwinning & Bremerich-Vos 2015).

6 | Feedback der Lehrkräfte

Im Bereich *Lesen* war ich doch manchmal erstaunt. Im Grunde besagten die Testergebnisse: Unsere Schüler lesen mit Hauptschulkompetenz. Und mein Eindruck war, dass zumindest, was die Texte angeht, es nicht so war, dass die Schüler nichts kapieren. (L_01_01)

Das vorangestellte Zitat stammt von einer Lehrperson, die an Schule 01 an der Durchführung des Projekts *Ganz In* beteiligt war.¹² Die Lehrperson führt im weiteren Verlauf des Interviews weiterhin aus, dass es insbesondere die niedrigen Werte im Bereich der Lesegeschwindigkeit zum dritten MZP waren, die für Erstaunen sorgten:

Wir haben uns zwar in den Klassen 5 und 6 daran gewöhnt, dass wir den Schülern unglaublich viel Zeit geben. Aber in der 7? Das war dann für mich so ein Aha-Erlebnis. Da war ich in dem Bereich doch erstaunt, wie langsam sie lesen. (L_01_01)

¹² Die Zitate der Lehrkräfte, die in diesem Teilkapitel vorgestellt werden, stammen aus Interviews, die im Rahmen der Projektevaluation durchgeführt wurden. Die gesamten Interviews werden in Bremerich-Vos (2019) vorgestellt und diskutiert.

Dieses Zitat belegt beispielhaft, dass es für Lehrkräfte in höheren Jahrgangsstufen oft nicht mehr so präsent ist, dass auch die hierarchieniedrigen Teilprozesse des Lesens bei schwachen Schülerinnen und Schülern noch viel Zeit in Anspruch nehmen. Den Jugendlichen ausreichend Zeit für die Lektüre einzuräumen, wird in Klasse 7 nicht mehr als so notwendig erachtet wie in den Klassen 5 und 6.

Vom Einsatz der Lautlese-Tandems im Regelunterricht zur Förderung der Leseflüssigkeit berichteten die am Projekt beteiligten Lehrpersonen durchweg positiv:

Und die Methode des Tandemlesens scheint tatsächlich gut zu funktionieren. Bei den Kindern, die da für ein halbes Jahr etwa bei der Stange bleiben, lässt sich sehen, dass das etwas bringt. (L_01_05)

Ähnlich wie diese Lehrkraft aus Schule 05 berichteten mehrere Lehrerinnen und Lehrer davon, dass bei kontinuierlichem Einsatz im Regelunterricht schnell Fortschritte sichtbar gewesen seien. Betrachtet man das Feedback zur Implementation und zu den Durchführungsmodalitäten der Methode, so fallen die Rückmeldungen allerdings je nach Schule sehr unterschiedlich aus. Einige Lehrpersonen meldeten dem universitären Projektteam zurück, dass die Kinder in den von ihnen unterrichteten Klassen Lautlese-Verfahren nur sehr widerwillig durchgeführt hätten, weswegen sie das Training nach einiger Zeit aussetzten.

Diejenigen Lehrpersonen, die kontinuierlich mit Lautlese-Tandems arbeiteten, bewerteten insbesondere die vorgesehene Sportmetaphorik als hilfreich, insofern man als Lehrerin oder Lehrer stets betone, wie gut die Schülerinnen und Schüler bereits fortgeschritten seien und dass trotzdem noch Verbesserungsmöglichkeiten bestünden. Auch in Schule 04 wird vom Einsatz der Lesetandems berichtet, wobei eine Beschränkung auf die Klassen 5 und 6 formuliert wird.

In den Interviews äußerten mehrere Lehrpersonen überdies, dass die Förderung der Leseflüssigkeit in heterogenen Klassen meist in Fördergruppen ausgelagert wurde. In diesem Kontext ist auch die folgende Äußerung einer Lehrkraft aus Schule 05 zu verstehen, wenn diese vom Lesepatentprogramm berichtet:

Im Moment versuchen wir, ein Lesepatentprogramm so richtig ans Laufen zu bringen. Wir haben das so angelegt, dass die Kinder der fünften Klasse, die in diesen Lesetests besonders schlecht abschnitten, einen Lesepatent bekommen sollen. Das haben wir im letzten Jahr gemacht mit Schülern von der Klasse 9 bis zur Q2, die sich zweimal in der Woche mit den Kindern treffen sollten, um dann 15 bis 20 Minuten zu lesen. [...] Wir haben relativ viele Kinder in der 5, die nicht gut lesen können. (L_02_05)

Bei der Durchführung zeigten sich jedoch Probleme, die eine weitere Lehrperson aus dem Kollegium derselben Schule wie folgt beschreibt:

Das Problem bei der Förderung mit Lesepatent ist, dass die Bereitschaft der förderbedürftigen Schülerinnen und Schüler, an freiwilligen Maßnahmen teilzunehmen, oft leider nicht so groß ist, wie wir uns das wünschen würden. (L_01_05)

Da sich das Konzept der Lesepatent an Schule 05 noch in der Erprobung befindet, können an dieser Stelle leider keine weiteren Aspekte zu dessen Gelingensbedingungen vor Ort ausgeführt werden.

Betrachtet man die Aussagen der Lehrpersonen zur Förderung der hierarchiehoher Teilfertigkeiten, so fallen auch hier zunächst die grundsätzlich positive Bewertung, allerdings auch Schwierigkeiten mit der Umsetzung, auf. So berichtete eine Lehrperson aus Schule 04 z.B., dass

sie zwar versuche, bei neuen Texten immer wieder auf die Lesestrategien hinzuweisen, jedoch Probleme mit dem kontinuierlichen Einsatz feststellte.

7 | Zusammenfassung

Die Ergebnisse der Testungen der Lesefähigkeiten aus dem *Ganz In II*-Kontext bestätigen überwiegend die schwachen Leistungen der Leserinnen und Leser im Rahmen von großen Schulleistungsstudien wie IGLU. Es konnte weiterhin gezeigt werden, dass Leseförderung auf Individual- aber auch auf Klassenebene auch noch ein Thema für die gymnasiale Erprobungsstufe und sogar für die siebten Klassen sein sollte. Dies betrifft nicht nur die hierarchiehöheren Teilfähigkeiten, sondern zum Teil auch noch die Leseflüssigkeit.

Der Einsatz der standardisierten und normierten Lesetests wurde von allen Lehrkräften grundsätzlich als hilfreich für die Förderung ihrer diagnostischen Kompetenz angesehen. Trotzdem wurde eine umfassende Diagnostik in Klasse 5, die auch einen Lesetest beinhaltete, nach Ende der Testungen zunächst nur in einem Gymnasium durch das *Ganz In*-Projektteam implementiert:

In Schule 05 konnte ein co-konstruktiv erarbeitetes Konzept eingesetzt werden, bei dem neben der Diagnose in der Domäne *Lesen* auch eine qualitative Fehleranalyse in der *Orthographie* durchgeführt wurde. Aus der Diagnostik resultierten individuelle Fördermaßnahmen.¹³ Positiv ist auch anzumerken, dass dieses Konzept einer Art „Schuleingangsdiagnostik“ auch an Schulen, die nicht am Projekt beteiligt waren, transferiert werden konnte¹⁴ – obwohl die Diagnostik an der Schule selbst nach einigen Durchläufen auf die Testungen im *Lesen* beschränkt wurde. Nach Aussage der an *Ganz In* beteiligten Lehrkräfte gestaltete sich auch an diesem Gymnasium der Transfer des Erarbeiteten innerhalb der eigenen Fachschaft zunehmend problematisch. Die Akzeptanz für ein umfassendes Konzept der Diagnose und Förderung war nur vorübergehend innerhalb der gesamten Deutschfachschaft vorhanden, bevor zunehmend Stimmen laut wurden, die einen zu hohen Zeitaufwand bei einem für sie fraglichen Nutzen wähten und zu wenig Unterstützung seitens der Schulleitung befürchteten. Daher wurde die Diagnostik auch an Schule 05 – trotz der vorhandenen umfangreichen Expertise und positiven Erfahrungen – nach einigen Durchläufen auf die Testungen im Bereich des Lesens beschränkt, „weil sie gut handhabbar ist.“ (L_01_05).

8 | Literatur

Baumert, J. (Eds.) (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Leske + Budrich.

Bäuerlein, K., Lenhard, W. & Schneider, W. (2012). *LESEN 6-7. Lesetestbatterie für die Klassenstufen 6–7. Verfahren zur Erfassung der basalen Lesekompetenz und des Textverständnisses*. Hogrefe.

¹³ Des Weiteren konnten Bestrebungen verzeichnet werden, dass an die Diagnostik ein Konzept zur LRS-Diagnose und Förderung gekoppelt werden sollte.

¹⁴ An einem weiteren Gymnasium in Nordrhein-Westfalen wurden mit Beginn des Schuljahres 2018/19 zu Beginn von Klasse 5 Testungen im *Lesen* und in der *Orthographie* durchgeführt.

- Bremerich-Vos, A. (2019). Tücken des Transfers. Interviews mit an Ganz In II beteiligten Deutsch-Lehrkräften. In: A. Bremerich-Vos, D. Ferencik-Lehmkuhl, S. Schwinning, & I. Fladung (Hrsg.), *Diagnostik und Förderung im gymnasialen Deutschunterricht. Ganz In II. Materialien für die Praxis*. Waxmann.
- Bremerich-Vos, A., Wendt, H. & Bos, W. (2017). Lesekompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In: A. Hußmann, A. Wendt, A. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 79-142). Waxmann.
- Bremerich-Vos, A., Ferencik-Lehmkuhl, D., Schwanenberg, J. & Schwinning, S. (2015). Zwischenbilanz der fachdidaktischen Arbeit im Fach Deutsch. In: H. Wendt & W. Bos (Hrsg.), *Auf dem Weg zum Ganztagsgymnasium. Erste Ergebnisse der wissenschaftlichen Begleitforschung zum Projekt Ganz In* (S. 237-263). Waxmann.
- Erikson, R., Goldthorpe, H. J. & Portocarero, L. (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *British Journal of Sociology*, 30(4), 415-441.
- Ferencik-Lehmkuhl, D. (2019). Lesen. Diagnostik und Vorschläge zur Förderung der basalen Lesefähigkeit und des strategischen Lesens. In: A. Bremerich-Vos, D. Ferencik-Lehmkuhl, S. Schwinning & I. Fladung (Hrsg.), *Diagnostik und Förderung im gymnasialen Deutschunterricht. Ganz In II. Materialien für die Praxis*. Waxmann.
- Ferencik-Lehmkuhl, D. (2017). *Revisionskompetenz fördern in der Sekundarstufe I. Effekte eines Trainings der Revisionsfertigkeiten bei Schülern der 5. und 6. Klasse*. Schneider Verlag Hohengehren.
- Ferencik-Lehmkuhl, D., Schwinning, S. & Bremerich-Vos, A. (Hrsg.) (2015). Schreiben und Lesen fördern. Vorschläge zur Praxis des Deutschunterrichts. *Ganz In -Materialien für die Praxis*. Waxmann.
- Gailberger, S. (2019). Modellierung von Lesekompetenz. In M. Kämper-van den Boogaart & K. Spinner (Hrsg.), *Deutschunterricht in Theorie und Praxis. Lese- und Literaturunterricht. Konzeptionelle und empirische Grundlagen*, 11(1). Schneider, 273-346.
- Glaser, C. (2005). *Förderung der Schreibkompetenz bei Grundschulern. Effekte einer integrierten Vermittlung kognitiver Schreibstrategien und selbstregulatorischer Fähigkeiten*. Dissertation.
- Gold, A., Mokhlesgerami, J., Rühl, K., Schreblowski, S. & Souvignier, E. (2004). *Wir werden Textdetektive [Lehrermanual und Arbeitsheft]*. Vandenhoeck & Ruprecht.
- Hußmann, A., Wendt, H., Kasper, D., Bos, W. & Goy, M. (2017). Ziele, Anlage und Durchführung der Internationalen Grundschul-Lese- Untersuchung (IGLU 2016). In: A. Hußmann, H. Wendt, W. Bos, D. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. Mcelvany, T. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 29-77). Waxmann.
- Karstens, F.; Schmitz, A. & Jost, J. (2019). Gutes Material zur Förderung von selbstreguliertem Lesen von Sachtexten im Deutschunterricht? Analyse exemplarischen Fördermaterials für die Sekundarstufe I unter Einbezug pädagogisch- psychologischer und deutschdidaktischer Perspektiven. *Leseforum. Online-Plattform für Literalität* (1).
- KMK – Ständige Konferenz der Kultusministerien der Länder in der Bundesrepublik Deutschland (2015). *Allgemein bildende Schulen in Ganztagsform in den Ländern in der Bundesrepublik Deutschland - Statistik 2009 bis 2013*. http://www.kmk.org/fileadmin/pdf/Statistik/GTS_2013_Bericht.pdf [16.09.2021].
- Lenhard, W. & Schneider, W. (2006). *ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler*. Hogrefe.

- Lenhard, W., Lenhard, A. & Schneider, W. (2018). *ELFE II. Ein Leseverständnistest für Erst- bis Siebtklässler - Version II*. Hogrefe.
- Lenhard, W. (2019). *Leseverständnis und Lesekompetenz: Grundlagen - Diagnostik - Förderung*. Kohlhammer.
- Leopold, C. (2009). *Lernstrategien und Textverstehen. Spontaner Einsatz und Förderung von Lernstrategien*. Waxmann.
- Mokhlesgerami, J. (2004). *Förderung der Lesekompetenz. Implementation und Evaluation eines Unterrichtsprogramms in der Sekundarstufe I*. Verlag Dr. Kovac.
- NICHD – National Institute of Child Health and Human Development (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. U.S. Government Printing Office.
- Rosebrock, C., Rieckmann, C., Nix, D. & Gold, A. (2010). Förderung der Leseflüssigkeit bei leseschwachen Zwölfjährigen. *Didaktik Deutsch*, 28, 33-58.
- Rosebrock, C., Nix, D., Rieckmann, C. & Gold, A. (2011). *Leseflüssigkeit fördern. Lautleseverfahren für die Primar- und Sekundarstufe*. Kallmeyer; Klett.
- Schneider, W., Schlagmüller, M. & Ennemoser, M. (2007). *LGVT 6-12. Lesegeschwindigkeits- und verständnistest für die Klassen 6-12*. Hogrefe.
- Schneider, W., Schlagmüller, M. & Ennemoser, M. (2017). *LGVT 5-12+. Lesegeschwindigkeits- und Verständnistest für die Klassen 5-12+*. Hogrefe.
- Souvignier, E. & Mokhlesgerami, J. (2006). Using self-regulation as a framework for implementing strategy-instruction to foster reading comprehension. *Learning and Instruction* 16(1), 57-71.
- Souvignier, E. & Trenk-Hinterberger, I. (2010). Implementation eines Programms zur Förderung selbst-regulierten Lesens. Verbesserung der Nachhaltigkeit durch Auffrischungssitzungen. *Zeitschrift für Pädagogische Psychologie*, 24(3), 207-220.
- Souvignier, E., Trenk-Hinterberger, I., Adam-Schwebe, S. & Gold, A. (2008). *FLVT 5-6. Frankfurter Leseverständnistest für 5. und 6. Klassen*. Hogrefe.
- Stöger, H., Sonntag, C., & Greindl, T. (2012). Selbstreguliertes Lesen mit Sachtexten - ein Trainingsprogramm. In: M. Philipp & A. Schilcher (Hrsg.), *Selbstreguliertes Lesen: Ein Überblick über wirksame Leseförderansätze* (S. 39-154). Klett Kallmeyer.
- Stubbe, T., Bos, W. & Schurig, M. (2017). Der Übergang von der Primar- in die Sekundarstufe. In: A. Hußmann, H. Wendt, W. Bos, D. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. Mcelvany, T. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 235-250). Waxmann.
- Wendt, H. & Bos, W. (Hrsg.) (2015). *Auf dem Weg zum Ganztagsgymnasium. Erste Ergebnisse der wissenschaftlichen Begleitforschung zum Projekt Ganz In*. Waxmann.
- Wendt, H., Willems, A., Tarelli, I., Euen, B. & Bos, W. (2013). Ausreichend geförderte Talente? Zu den deutschen Ergebnissen von leistungsstarken Viertklässlerinnen und Viertklässlern in IGLU 2011 und TIMSS 2011. In: C. Fischer (Hrsg.), *Schule und Unterricht adaptiv gestalten. Fördermöglichkeiten für benachteiligte Kinder und Jugendliche* (S. 23-34). Waxmann.
- Wild, K.-P. (2010). Lernstrategien und Lernstile. In: D. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (S. 479-485). Beltz.

Maik Philipp

Liest du noch oder sourcst du schon?

Sourcing als neuer Bestandteil kompetenten Lesens (multipler, teils digitaler Dokumente)

1 | Einleitung: Was Sourcing ist

Es scheint, als erlebe das Lesen derzeit eine Transformation durch die Digitalisierung und durch das Internet, zu dem wir über mobile Endgeräte Zugänge in eine Bibliothek mit Myriaden von Dokumenten haben. Die Notwendigkeit, schriftliche Texte bzw. Dokumente mit verschiedenen Modalitäten rezeptiv zu verstehen, hat nichts von ihrer Wichtigkeit eingebüßt – auch und gerade um an einer Gesellschaft mit digitaler Kommunikation teilzuhaben. Dennoch ist eine Transformation beim Lesen derzeit dahingehend zu beobachten, dass neben die statischen Print-Texte die digitalen (Hyper-)Texte treten, welche wiederum andere und zusätzliche Verstehensleistungen erfordern – nicht zuletzt aufgrund der schier unendlichen Zahl, die einen selektiven und kritischen Umgang verlangen. Quantitäten und Qualitäten des Lesens verändern sich demnach, und das ist auch lesedidaktisch folgenreich.

Die Digitalisierung und die Möglichkeit, auf Dokumente zugreifen zu können, wirft nämlich aus lesedidaktischer Sicht diverse Fragen auf (Bräten, Braasch & Salmerón 2020; Magliano et al. 2017; Salmerón et al. 2018). Drei der Fragen bzw. der sich daraus ergebenden Probleme hängen damit zusammen, dass Leseleistungen bei digitalen Texten offenbar schlechter ausfallen, dass damit das Lesen multipler Dokumente erforderlich ist und, sich daraus ergebend, auch erhöhte Anforderungen wegen der Nutzung nicht nur inhaltlicher Informationen, sondern auch von Metadaten an die Leser*innen stellen.

- Die erste Frage ist offensichtlich und betrifft die *Medialität* von Dokumenten. Diese Frage ist deshalb relevant, weil beispielsweise inzwischen metaanalytisch gut belegt ist, dass das Leseverstehen bei digitalen Texten im Vergleich zu analog präsentierten Texten schlechter ist. Dies tritt gerade bei Sachtexten und bei aktuelleren Studien deutlicher zutage (Delgado et al. 2018).
- Zweitens besteht insbesondere als Begleiterscheinung des digitalen Lesens die Notwendigkeit, *multiple Dokumente* zu verstehen und intertextuelle Kohärenz zu stiften, um dokumentenübergreifendes Verstehen einerseits als Ressource zu nutzen und andererseits als eine kognitive Repräsentation erhöhten Anspruchsgrades überhaupt erst aufzubauen (Cho, Afflerbach & Han 2018; Salmerón et al. 2018).

- Drittens und für die Zwecke dieses Beitrags besonders wichtig ist, dass die vielen Internet-Dokumente, auf die man stößt, aufgrund der mangelnden editorialen Qualitätschranken zu einem Nebeneinander einschätzungsbedürftiger Dokumente führen. Dafür braucht es *Metadaten*, die beispielsweise in Zeiten von Fake News als Ressource nicht nur nutzbar, sondern geradezu nutzungsbedürftig erscheinen (Bråten, Stadtler & Salmerón 2017).

An diesem letztgenannten Punkt setzt der vorliegende Beitrag an. Er fokussiert auf die Einschätzungsbedürftigkeit von Informationen und Dokumenten. Sie wird im gegenwärtigen Diskurs der Leseforschung als *Sourcing* gefasst und gewinnt zunehmend an Bedeutung (Scharrer & Salmerón 2016; Bråten et al. 2017; Rouet & Potocki 2018). Das Sourcing bildet eine Familie strategischer kognitiver Prozesse, mittels derer Metadaten gewonnen und verarbeitet werden. Solche Metadaten als Bezugspunkt des Sourcings sind Informationen über andere Informationen, seien es Inhalte, seien es Dokumente, seien es Sammlungen von Dokumenten.

1.1 | Sourcing – eine Definition und ihre Bestandteile

Doch was wird unter Sourcing verstanden? Exemplarisch angeführt werden kann hierfür eine Begriffsbestimmung aus einer norwegischen Überblicksarbeit, in welcher die Förderung des Sourcing als narrativer Review behandelt wurde. Das Sourcing definieren Brante und Strømsø (2018: 777) als *das kognitive Identifizieren und Repräsentieren von Metadaten, um aufgabenbezogene Inhalte und Relevanz eines Dokuments vorherzusagen, zu interpretieren und zu beurteilen*, also gesamthaft zu evaluieren. Diese knappe Definition enthält diverse kompakt kombinierte, prozessbezogene Komponenten, die einer vertiefenden Betrachtung wert sind:

- *Metadaten*: Metadaten sind im weitesten Sinne Daten über Dokumente, die entweder explizit vorliegen oder aber durch Recherche und/oder vorwissensbasierte Inferenzen (re-)konstruierbar sind. Es handelt sich hierbei um eine Melange von Informationen wie die Daten zu Personen, welche ein Dokument hergestellt haben, der Erscheinungskontext im weitesten Sinne (Medium, Zeitpunkt, Publikationsart, Dokumentenart), aber auch der Stil bzw. die Rhetorik von Dokumenten.
- *Identifizieren und Repräsentieren*: Damit Metadaten für verschiedene Zwecke genutzt werden können, müssen sie als solche überhaupt erst einmal wahrgenommen und als informationell relevant erachtet werden. Das heißt: Eine lesende Person darf sie nicht ignorieren, sondern muss vielmehr die eigene Dignität von Metadaten als verarbeitungswürdige und -bedürftige Einheit erkennen.
- *Inhalt*: Gemäß der Definition geht es beim Sourcing um die Kopplung von Inhalt und Metadaten, die aktiv miteinander kognitiv verknüpft werden müssen. Das heißt, es wäre verfehlt, Inhalt und Metadaten separiert zu repräsentieren, sondern man muss die wechselseitige Bedingtheit rekonstruieren, eben genau damit das semantische Potenzial als (In-)Kongruenz beider Informationsarten erkennbar wird. Die Verknüpfung von Inhalt und Metadaten ist auf vorwissensbasierte Integrationsprozesse angewiesen und eher bottom-up gesteuert.
- *Relevanz*: Mit der Relevanz ist eine andere, externale und eher top-down ablaufende Form der Informationsverarbeitung von Metadaten gemeint. Diese hat einen starken Bezug zu dem mit dem Lesen verbundenen Zweck und damit auch zur (selbst oder von Dritten) gestellten Aufgabe. Dadurch wird die Aufmerksamkeit auf möglicherweise andere Metadaten

gelenkt, etwa um sich zu entscheiden, ob ein Dokument als solches überhaupt für die Nutzung infrage kommt, etwa bei Suchmaschinenergebnissen.

- *Evaluieren*: Das Evaluieren, welches gemäß der obigen Definition das prospektive Prognostizieren sowie das synchrone bzw. retrospektive Interpretieren und Beurteilen umfasst, stellt eine stark auf Vorwissen basierende Form des Inferierens dar. Hier geht es um die Einschätzung von Wahrheitsgehalt/Plausibilität bzw. Vertrauenswürdigkeit, das Rekonstruieren von kommunikativen Absichten, aber auch den Vergleich verschiedener Perspektiven im Falle multipler Dokumente bzw. eingebetteter Quellen.

In der kurzen Betrachtung jener o.g. Komponenten der Definition des Sourcing wurde schon deutlich, dass es diverse kognitive Prozesse sind, die für das Sourcing nötig sind: Eine lesende Person muss Metadaten finden, vorwissensbasiert interpretieren, mit (erwarteten) Inhalten aus Dokumenten inferenzbasiert verknüpfen und aufgrund (möglicherweise) mehrdimensionaler Evaluationsprozesse begründete Einschätzungen inhaltlicher Informationen vornehmen.

1.2 | Aufbau und Ziele des Kapitels

Das Sourcing avanciert derzeit zu einem neuen Bestandteil kompetenten Lesens, und diese Entwicklung zeichnet sich prototypisch beim Lesen multipler, inhaltlich widersprüchlicher Dokumente ab, wie es in digitalen Kontexten häufig der Fall ist. In diesem Ausschnitt des Lesens – einem Bereich, dem mutmaßlich zukünftig eine steigende Bedeutung zukommen wird (Magliano et al. 2017) – lassen sich wie unter dem Brennglas gegenwärtige Potenziale und Problem-lagen gleichermaßen betrachten und analysieren. Hier setzt das Kapitel an, indem es mithilfe vor allem theoretischer Zugänge konturiert, welche Bedeutung das Sourcing hat, und diese multiplen theoretischen Perspektiven werden mit empirischen Befunden unterfüttert. Wegen der primären Fokussierung auf die Theorie haben die aggregierten empirischen Anteile einen ausdrücklich illustrierenden Charakter (vgl. Philipp 2020 für eine ausführlichere und vollständige Präsentation des Forschungsstands).

Im Folgenden werden also das Sourcing und seine Bedeutung für das Lesen – bzw. genauer: für das Leseverstehen – vertiefend behandelt. Dabei wird im- wie explizit das Verstehen kontroverser, teils auch argumentativer Dokumente im Zentrum stehen. Das Kapitel entfaltet die Thematik in mehreren Schritten. Der Abschnitt 2 bildet den Kern des Kapitels. Er entfaltet mithilfe von vier komplementären Theorien resp. Theorieentwürfen, wie das Sourcing als Verbund verschiedener Prozessierungen mithilfe von Inferenzen angereicherter Metadaten gegenwärtig modelliert wird. Dabei geht es nicht nur um die theoretisch begründete Wichtigkeit des Sourcing beim Lesen multipler Dokumente, sondern auch um die generelle Nutzbarkeit von Metadaten als eigene Informationsquelle mit genuin wichtigen Informationen. Hinzu kommen zwei Vertiefungen eines zentralen kognitionspsychologischen Modells, welche das Verhältnis des Sourcing im Verbund mit anderen (meta-)kognitiven Prozessen spezifizieren. Das Fazit in Teilkapitel 3 geht schließlich darauf ein, welche prioritären lesedidaktischen Implikationen sich aus alldem ergeben.

2 | Wozu Sourcing dient: Funktionen einer Prozessgruppe für das Lesen

Das eingangs definierte Sourcing hat verschiedene Funktionen, was sich in der Einschätzung niederschlägt, dass es sich um ein ‚Regenschirm-Konzept‘ (Scharrer & Salmerón 2016: 1540) handelt. Konkret kann die strategische kognitive Verarbeitung von Metadaten unterschiedlichen Zwecken dienen, verschiedene Grade an Vernetzung, Vollständigkeit und Verarbeitungstiefe annehmen und insbesondere beim Lesen aus unterschiedlichen Gründen zu einem tieferen Verstehen führen, wenn es glückt, vorwissensbasierte Inferenzen zu bilden. Welche Funktionen das Sourcing erfüllt, wird in diesem Teilkapitel behandelt, und zwar folgendermaßen: Den Einstieg bildet ein Tableau über drei dominante Funktionen des Sourcings (s. Teilkap. 2.1), von denen dann zwei miteinander eng verknüpfte Funktionen weiterbehandelt werden, und zwar sowohl und primär theoriebezogen als auch schlaglichtartig empirisch (2.2). Zwei Perspektiven sind leitend: Zum einen geht es im Teilkapitel 2.2.1 um allgemeine kognitions- und sozialpsychologische Modellierungen, wie Metadaten in (schriftlichen) Kommunikationssituationen genutzt werden (können). Zum anderen erfolgt eine Vertiefung, wie das Sourcing dazu verwendet wird, im Falle vorliegender semantischer intertextueller Konflikte und Inkohärenzen vorzugehen und diese aufzulösen (2.2.2).

2.1 | Welche Funktionen dem Sourcing im Kontext des Lesens im Allgemeinen zukommen

Das Sourcing ist als Familie verschiedener kognitiver Prozesse ein Regenschirmkonzept; der Terminus vereint also unter sich unterschiedliche Prozesse, die teils unterscheidbaren, wenngleich verwandten Zwecken dienen (Scharrer & Salmerón 2016). Diese Zwecke des Sourcings demonstriert die Übersicht in Tabelle 1. Sie enthält drei prototypische Funktionen des Sourcings, welche sich nicht gegenseitig ausschließen, sondern eine Schnittmenge aufweisen und teils sogar aufeinander aufbauen. Je nach Funktion wird ein anderer Schwerpunkt gesetzt: Sourcing kann für eine begründete Auswahl genutzt werden, und zwar im doppelten Sinne. Einerseits können Metadaten dazu dienen, sich insbesondere bei (1) digitalen Dokumenten aus dem Internet für oder gegen das Lesen zu entscheiden, wobei dies natürlich auch auf analoge, also Print-Dokumente zutrifft. Andererseits dient das Sourcing dazu, (2) in bereits gelesenen Dokumenten Informationen auszuwählen, die für die lesende Person eine angestrebte Glaubwürdigkeit haben, um für das Leseverstehen verwendet zu werden. Daraus ergibt sich die dritte Funktion, nämlich das (3) Nutzen von Metadaten auch für ein umfassendes Leseverstehen, etwa und insbesondere beim Lesen von multiplen Dokumenten. Aus diesem Feld der Leseforschung speist sich momentan ein ergiebiger Strom an theoretischen Modellierungen und empirischen Ergebnissen. Um die beiden letztgenannten, miteinander korrespondierenden Funktionen des Sourcings – die Unterstützung für eine begründete Informationsauswahl sowie für das Leseverstehen – wird es im nächsten Teilkapitel gehen.

| | | |
|---|---|--|
| 1) Entscheidungen im digitalen Navigationsverhalten | 2) Hilfe für begründete Informationsauswahl und -interpretation | 3) Unterstützung im Leseverstehen |
| Filterung von Websites, die man besucht bzw. nicht besucht aufgrund von Metadaten und anderen Informationen | zielbezogene Glaubwürdigkeitseinschätzung bei geringem eigenem Vorwissen, um die (vorläufige) Belastbarkeit von inhaltlichen Aussagen abzuschätzen und sich für oder gegen die Nutzung von Informationen zu entscheiden | Nutzung von Informationen, um das Gesamtverständnis singulärer bzw. multipler Dokumente zu unterstützen, indem «Dokumentenknotten» systematisch mit verstehensrelevanten Informationen aufgefüllt werden |

Tab. 1: Drei prototypische Funktionen des Sourcings in Bezug auf das Lesen (Quelle: Darstellung basierend auf Scharrer & Salmerón 2016: 1540–1542)

2.2 | Funktionen des Sourcings im Besonderen: multiple Funktionen bei der Interpretation von Informationen und dem verstehenden Lesen

Die Nutzung von Metadaten qua Sourcing stellt einen momentan stark an Konjunktur gewinnenden Zweig der Forschung dar. Besondere Aufmerksamkeit hat das Sourcing nicht zuletzt deshalb erfahren, weil die Metadaten als Informationskategorie für das Verstehen von Texten bzw. ganzen Diskursen in bisherigen dominierenden Modellierungen zu kurz gekommen sind, was inzwischen als korrekturbedürftig gilt (Braasch & Bråten 2017). Hier haben die Modelle zum Leseverstehen multipler Dokumente die Bedeutung des Sourcings als Verwendung nutzungsbedürftiger Metadaten betont (Philipp 2018, 2020), auch wenn der Ertrag des Sourcings keineswegs nur auf multiple Dokumente als Lesegrundlage limitiert ist. Vielmehr eint gegenwärtige Perspektiven auf das Sourcing, dass es ebenfalls für singuläre Texte mit eingebetteten Quellen nötig wirkt (Stadtler 2017; Strømsø 2017). Anders ausgedrückt: Die Genese theoretischer Modellierungen war ein Motor für eine überfällig wirkende Erweiterung bestehender Modelle, der Wirkungsbereich dieser Erweiterungen reicht aber weit über den Spezialfall des Lesens multipler Dokumente hinaus. Wenn im Folgenden die Entfaltung des Sourcings am Beispiel des Lesens multipler Dokumente erfolgt, dann geschieht dies im Wissen um die Exemplarität der Darstellung.

2.2.1 | Perspektive 1: Sourcing als integraler Bestandteil des Verstehens singulärer und vor allem multipler Dokumente

2.2.1.1 | Theoretische Perspektiven

Das nach gegenwärtigem Stand wichtigste Modell zum Leseverstehen multipler Dokumente hat eine Arbeitsgruppe rund um Anne Britt und Jean-François Rouet (Perfetti, Rouet & Britt 1999; Britt & Rouet 2012; Britt, Rouet & Braasch 2013) zunächst als deskriptive Modellierung vorgelegt und dann weiterentwickelt. Das ‚Dokumentenmodell‘ (s. Abbildung 1) ist eine dezidierte Erweiterung des ‚Construction-Integration Model‘ von Kintsch, indem es dessen drei Repräsentationsebenen (wortwörtliche Repräsentation, propositionale Repräsentation und Situationsmodell) um zwei weitere ergänzt. Dies ist zum einen die Ebene des ‚Intertextmodells‘, welche

für die Sourcingprozesse einschlägig ist, und zum anderen gibt es noch die Ebene des ‚integrierten mentalen Modells‘, welche dem Umstand der multiplen Dokumente als Grundlage des Aufbaus kognitiv verschmolzener Repräsentationen Rechnung trägt. Das Modell ist produktorientiert, da es die idealtypischen Komponenten beschreibt und systematisiert, ohne jedoch zu spezifizieren, welche Prozesse dazu führen. Nach gegenwärtigem Stand sind zwei Hauptprozesse dafür notwendig: das Sourcing und das intra- und intertextuelle Integrieren, also das kohärente, regelgeleitete Verknüpfen von Propositionen auf der Basis von Inferenzen (Philipp 2020).

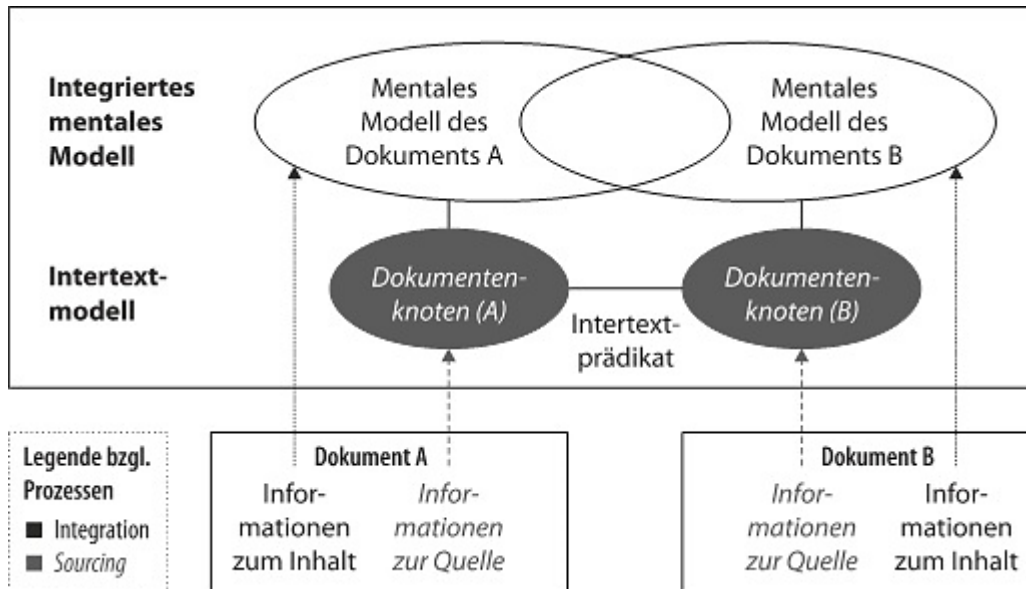


Abb. 1: Überblick über die Komponenten des Dokumentenmodells mit Zuordnung zu den beiden Hauptprozessen des Lesens multipler Dokumente (Quelle: Darstellung basierend auf Britt & Rouet 2012: 285)

Die Ebene des ‚integrierten mentalen Modells‘ berücksichtigt explizit, dass Personen mehr als ein Dokument – also einen Text mit zugehörigen Metadaten – lesen und dabei vor allem spezifische *Integrationsprozesse* zu bewältigen haben, bei denen Inhalte aus Dokumenten („Propositionen“) beteiligt sind. Dies wird an drei Stellen im grafisch dargestellten Modell aus Abbildung 1 evident:

- Erstens betrifft dies die überlappenden mentalen Modelle einzelner Dokumente, also die Kombination von Bestandteilen intratextueller mentaler Modelle. Hierbei geht es darum, intertextuelle Kohärenz zu stiften.
- Zweitens gibt es zwischen den Ebenen des ‚Intertextmodells‘ und ‚integrierten mentalen Modells‘ vertikale Verquickungen, die als ‚Quelle-Inhalt-Verbindungen‘ bezeichnet werden. Auch hierbei handelt es sich um zunächst einzeldokumentspezifische, inferenzbasierte Verknüpfungen. Sie verbinden die (interpretierten) Metadaten aus dem ‚Dokumenten-knoten‘ auf der Ebene des ‚Intertextmodells‘ mit den Inhalten aus den Dokumenten auf der Ebene des ‚integrierten mentalen Modells‘.
- Drittens sind intertextuelle Integrationsleistungen auch auf der Ebene des ‚Intertextmodells‘ zu leisten. Sie betreffen die ‚Intertextprädikate‘ zwischen den Dokumenten-knoten. Die ‚Intertextprädikate‘ bezeichnen das Verhältnis zwischen Dokumenten, welche starke

Abstraktionsleistungen von der lesenden Person erfordern, weil jemand auf der Aggregat-ebene ganzer Dokumente bestimmen muss, in welcher intertextuellen Relation einzelne Dokumente zueinanderstehen.

Das *Sourcing* ist im Vergleich zum Integrieren eindeutiger im ‚Dokumentenmodell‘ verortbar, nämlich auf der Ebene des ‚Intertextmodells‘ und dort im ‚Dokumentenknotten‘. Dieser ‚Dokumentenknotten‘ ist das Resultat des Sourcings, und er enthält eine Sammlung von Metadaten, welche ihrerseits untereinander per Inferenzbildung miteinander verknüpft werden müssen und die Basis für Evaluationen und weitere Inferenzen bilden. Der ‚Dokumentenknotten‘ fungiert damit als eine Art Steckbrief für Metadaten, die im Sinne des ‚Dokumentenmodells‘ ihren Bezugspunkt in gesamten Dokumenten haben. Bei eingebetteten Quellen, also beispielsweise Zitaten von Personen in Dokumenten, kann sich dieser Bezugspunkt verändern, weil dann Einzelaussagen mit Metadaten zur Quelle dieser Aussagen verknüpft werden. Hierin liegt auch die genuine Schnittmenge des Sourcings sensu ‚Dokumentenmodell‘ zu Modellen des Leseverstehens einzelner Texte/Dokumente.

Das ‚Dokumentenmodell‘ enthält implizit die Prämisse, dass Metadaten und mit ihm das Sourcing als Prozess des Nutzens dieser Metadaten bedeutsam für das Verstehen von multiplen Dokumenten sind. Doch warum genau ist das Sourcing ein integraler Bestandteil des Leseverstehens multipler, aber auch einzelner Dokumente? Hierfür lieferten Britt et al. (2013) eine Erklärung, indem sie betonen, dass *Dokumente als ‚soziale Artefakte‘* oder auch *‚Diskursteilnehmer‘* (‚Discourse Participants‘) zu kontextualisieren sind. In diesem Sinne bilden Inhalt und Herkunft der Aussagen nebst weiteren Metadaten eine bedeutungstragende, kommunikative Gesamteinheit, die ohne die Berücksichtigung von Metadaten als Spuren eines sozialen Kontexts nachgerade unvollständig ist. Demnach ist der Inhalt eines Dokuments nicht sinnvoll verstehbar ohne das quellenbezogene Wissen, welche Person mit welchem Motiv und mit welcher Expertise Aussagen in welchem kommunikativen Kanal mit welcher angestrebten Adressatenschaft in Umlauf gebracht hat. Es sind demnach zuvorderst vorwissensbasierte evaluative Fähigkeiten nötig (Cho et al. 2018), die beispielsweise in Lesekompetenzstudien wie PISA in den Kompetenzstruktur- und -stufenmodellen zuoberst rangieren (OECD 2019).

Die eben wiedergegebene Auffassung, dass Metadaten und Inhalte eine Einheit bilden und daher nicht isoliert voneinander betrachtet werden sollten, impliziert, dass die Metadaten eine bedeutungstragende bzw. -modifizierende Eigenschaft haben. Auch das ist explikationsbedürftig, denn die Folgefrage lautet: Warum und wie transportieren Metadaten semantisch relevante Informationen? Weil das Sourcing die Vertrauenswürdigkeit von sozialen Artefakten betrifft, sind aus der Perspektive der Leseforschung hierfür sozialpsychologische Erklärungsansätze von Interesse. Diesen Gedanken greifen neuere Publikationen dezidiert auf (Bråten et al. 2017; Wegener, Patton & Haugtvedt 2018) und führen als eine Theorie das beginnend in den 1970er Jahren entwickelte, wirkmächtige *‚Elaboration Likelihood Model‘* (ELM, ‚Modell der Elaborationswahrscheinlichkeit‘; Petty & Wegener 1999) an. Das ELM fokussiert auf die persuasive Kommunikation und modelliert, wie es basierend auf dieser Kommunikation zu mehr oder minder stabilen Einstellungsänderungen kommt, die auf der namensgebenden Elaboration – also auf Prüfung und konstruktive Anreicherung von Aussagen durch die lesende Person – basieren.

Zentral ist im komplexen ELM, dass sowohl die Inhalte von (schriftlichen) Äußerungen als auch Metadaten verarbeitet werden, und zwar entweder peripher oder zentral, d. h. oberflächlich-

affirmativ oder tiefenorientiert-kritisch prozessiert werden. Diese beiden Arten von (regelgeleiteten, in der ELM als ‚Postulate‘ bezeichneten) Verarbeitungen verschiedener Informationen führen dann zu ephemeren oder stabilen Einstellungen. Gemäß ELM können Metadaten aus verschiedenen Gründen wie wenig Ablenkung, Zeit, Vorwissen, Motivation und Relevanz eine entscheidende, bisweilen sogar die entscheidendere Rolle darin spielen, kommunikative Aussagen zu verarbeiten. Das ELM spezifiziert anhand verschiedener Postulate, dass Personen anfällig dafür sind, Metadaten teils zu bevorzugen und sie eben genau nicht zu ignorieren. Dafür sind komplexe Wirkgefüge zuständig, die das ELM als Zusammenspiel verschiedener Mechanismen umreißt.

Für die Zwecke dieses Kapitels lässt sich festhalten, dass der Rückgriff auf Metadaten bei Laien dann unwahrscheinlicher ist, wenn ihnen wenige zeitliche Ressourcen bleiben, sie auf ein geringes Vorwissen zurückgreifen können, sie für eine tiefergehende, strategische Informationsverarbeitung zu unmotiviert sind und/oder sie inhaltlichen Aussagen eine geringe Relevanz zugestehen. Dies ist, um es ausdrücklich zu betonen, kein defizientes Vorgehen. Vielmehr unterstreicht es, dass Metadaten generell zugestanden wird, bei persuasiver Kommunikation als Informationsquelle eigener Dignität zu fungieren.

Synoptisch lässt sich damit festhalten: Dem ‚Dokumentenmodell‘ kommt das Verdienst zu, als erstes Modell des Leseverstehens multipler Dokumente die Komponenten kartiert und systematisiert zu haben, welche nach wie vor die Vorstellung davon prägen, welche Informationen auf welche Art kognitiv repräsentiert werden. Insbesondere die Trennung von inhaltlich-propositionalen Informationen auf der einen Seite, die eine lesende Person sowohl intra- und vor allem intertextuell prozessieren, transformieren und über Dokumentengrenzen hinweg repräsentieren muss, und den Metadaten auf der anderen Seite hat sich als analytisch fruchtbar erwiesen und viele weitere theoretische und empirische Arbeiten stimuliert. In dieser Trennung sind die beiden Hauptprozesse, das Integrieren und das Sourcing, implizit angelegt. Die Notwendigkeit des Sourcings wurde über die Modellierung von Dokumenten als soziale Artefakte nochmals betont, und mithilfe des ‚Elaboration Likelihood Model‘ ist das Sourcing bzw. die Verarbeitung der Metadaten – zumindest im Kontext der persuasiven Kommunikation – als Prozess durchaus aufgewertet worden.

2.2.1.2 | Empirische Perspektiven

Die theoretischen Modellierungen und Überlegungen sind bereits Gegenstand der empirischen Überprüfung geworden, und gegenwärtig ist ein reges und zunehmendes Forschungsinteresse zu attestieren (Bråten et al. 2017). Aus Platzgründen ist es nicht möglich, die Forschungsbefunde hier detailliert wiederzugeben (s. dazu Philipp 2020), daher nimmt der Beitrag einige schlaglichtartige Perspektiven auf die wichtigsten, Dokumentenmodell-nahen Forschungsbefunde ein. Die Belege für die empirische Evidenz fallen aus dem genannten darstellungsökonomischen Grund ebenfalls sparsam aus.

Die aktuelle Forschung lässt sich folgendermaßen bündeln, und es gibt Befunde sowohl in Bezug auf Prozesse als auch auf Produkte des Verstehens. Bei vielen Befunden ist zu konzedieren, dass sie aus Studien mit Studierenden stammen und damit ein schmales Segment von Leser*innen derzeit in der Forschung überrepräsentiert ist:

- *Die Funktionen des Sourcings sind empirisch beobachtbar in Prozessen und Produkten:* Insbesondere bei guten Leser*innen, speziell jenen mit hoher Expertise in einer Domäne, ist das strategische Sourcing beobachtbar, und zwar durchaus in Bezug auf die diversen Funktionen des Sourcings (Cho et al. 2018; Salmerón et al. 2018; Hahnel et al. 2019a; Wineburg 1991).
- *Das Konsultieren und das Nutzen von Metadaten korrespondieren mit anderen Prozessen des verstehenden Umgangs mit multiplen Dokumenten:* Das Sourcing ist als Prozessgruppe nur ein Ausschnitt des allgemeinen kognitiven Vorgehens. Nur wenige Studien haben dies bislang mittels Korrelationsanalysen verifiziert, doch dort, wo es erfolgte, trat das Sourcing im Verbund mit anderen zielführenden und damit strategischen Prozessen auf (Hahnel et al. 2019b; Anmarkrud, Bråten & Strømsø 2014; Cho et al. 2017). Die Zusammenhänge waren je nach Art der Erfassung der Prozesse unterschiedlich stark ausgeprägt.
- *Das Sourcing korrespondiert mit besseren Verstehensleistungen, also Produktmaßen in expressiven und rezeptiven Aufgaben:* Studien, in denen mit lautem Denken oder computerbasiertem Tracking Sourcingprozesse erfasst wurden, konnten diverse Zusammenhänge zwischen Prozessen und Produkten in expressiven Aufgaben nachweisen. Wer quantitativ extensiver Sourcing betrieb, verknüpfte Inhalte und Metadaten stärker (Anmarkrud et al. 2014; List, Alexander & Stephens 2017), übernahm mehr Inhalte (Barzilai, Tzadok & Eshet-Alkalai 2015; List et al. 2017), strukturierte eigene Texte besser (Anmarkrud et al. 2014; Barzilai et al. 2015) und beschrieb nicht nur Relationen von Dokumenten korrekter, sondern konnte auch die Konflikte zwischen Dokumenten korrekter inferenzbasiert auflösen (Merkt, Werner & Wagner 2017). Das intertextuelle Leseverstehen war bei vollständigerem Sourcing ebenfalls ausgeprägter (Hahnel et al. 2019b).

Der Forschungsstand ist momentan noch deutlich ausbaufähig, dennoch gibt es einige empirische Hinweise, welche für die Annahmen aus dem ‚Dokumentenmodell‘ sprechen. Strategische, zielbezogene Sourcingprozesse treten im Verbund mit anderen strategischen (meta-)kognitiven Prozessen auf. Das Sourcing hat empirisch differenzielle Funktionen, und es ist verstehensförderlich. Die Ergebnisse unterstreichen in ihrer Gesamtheit mithin die theoretisch postulierte Bedeutung des Sourcings für das Verstehen multipler Dokumente.

2.2.2 | Perspektive 2: Sourcing als Basis für das Verstehen konfligierender Quellen und Dokumente

2.2.2.1 | Theoretische Perspektiven

Nachdem im Teilkapitel 2.2.1.1 theoretische Modellierungen in Form des ‚Dokumentenmodells‘ und des ‚Elaboration Likelihood Model‘ als Rahmen dienten, um Sourcing und Sourcingprozesse als verständnisnotwendige Komponenten beim Lesen vor allem multipler Dokumente zu beschreiben, findet an dieser Stelle eine mikrotheoretische Vertiefung statt. Das ‚Dokumentenmodell‘ ist mit seiner Architektur als produktorientiertes Komponentenmodell nämlich nur sehr bedingt dazu geeignet, die Bedingungen und Abläufe bei den Prozessen zu modellieren,

was beim ‚Elaboration Likelihood Model‘ als Flussmodell besser möglich ist, aber für das Verstehen multipler Dokumente noch nicht spezifiziert wurde.

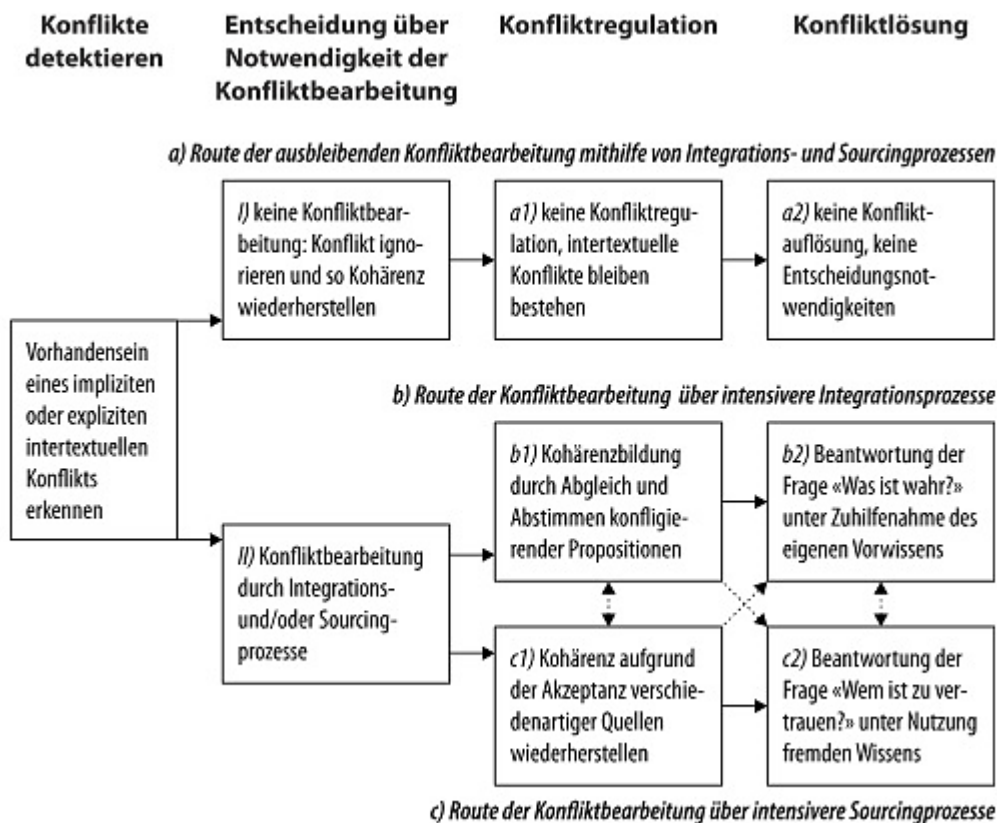


Abb. 2: Heuristik zu den Routen des Umgangs mit intertextuellen Konflikten (teils eigene, teils fremdbasierte Darstellung der Routen basierend auf Braasch & Bråten 2017 sowie Stadtler & Bromme 2014; die Sequenzialität und Trennung in einzelne Routen der Konfliktbearbeitung ist eine starke Vereinfachung – realiter sind die Routen keine deterministischen Wege, was sich auch in den möglichen Kreuzpfaden von der Konfliktregulation zur Lösung (b1–c2, c1–b2) und den gestrichelten Doppelpfeilen zwischen den Stationen b1 und c1 sowie b2 und c2 manifestiert – Quelle der Darstellung: Philipp 2021: 77)

Hier nehmen zwei Modelle gleichsam eine Zwischenstellung ein, wobei sie sich dichter am ‚Dokumentenmodell‘ lokalisieren lassen: zum einen das ‚Discrepancy-Induced Source Comprehension Model‘ (DISCM, ‚Modell des diskrepanzinduzierten Quellenverständnisses‘; Braasch & Bråten 2017) und zum anderen das ‚Content-Source Integration Model‘ (CSIM, ‚Modell der Inhalt-Quelle-Integration‘; Stadtler & Bromme 2014). Beide Modelle eint, dass sie das ‚Dokumentenmodell‘ insofern erweitern, als sie genauer bestimmen, wann Sourcingprozesse stattfinden und in welchem Verhältnis sie zu den ebenfalls nötigen Integrationsprozessen stehen. Dies erfolgt in beiden Modellen durch eine Fokussierung auf ein in der Forschung besonders dominant untersuchtes intertextuelles Verhältnis von Dokumenten, nämlich inhaltlich konfliktierende Dokumente. Hier ist die objektiv nicht vorhandene intertextuelle Kohärenz Auslöser des Sourcings. In diesem Sinne fungiert das Sourcing als Reparaturmechanismus für inhaltlich fehlgeschlagene Prozesse der Herstellung intertextueller Kohärenz.

Während CSIM und DISCM darin übereinstimmen, dass Sourcingprozesse die Folge notwendigerweise missglückender Integrationsprozesse sind, unterscheiden sich beide Modelle darin,

wie genau sie dieses Verhältnis beschreiben. Das *DISCM* ist im Vergleich das sparsamere Modell: Es postuliert im Sinne einer Wenn-dann-Formulierung die wissenschaftliche Hypothese, dass beim Auftreten von widersprüchlichen Aussagen in Form von intratextuellen, intertextuellen bzw. zwischen Text und lesender Person zu situierenden Konflikten, ein (verstärktes) Sourcing stattfindet. Das Sourcing soll in dem Fall dazu dienen, erkannte Inkohärenzen zu erklären und dadurch Kohärenz wiederherzustellen. Das *DISCM* benennt explizit auch singuläre Texte als Geltungsbereich, in denen es zu inhaltlichen Widersprüchen zwischen eingebetteten Quellen kommt, etwa bei zwei Aussagen von politischen Gegnern zu einem Thema (Braasch & Bräten 2017).

Das *CSIM* ist komplexer, wie es die im Wesentlichen auf dem *CSIM* basierende Abbildung 2 in ihrem Aufbau demonstriert. Das *CSIM* ist ein mehrphasiges Modell, welches seinen Ursprung darin hat, wie Laien mit widersprüchlichen Informationen umgehen, welche sie in der Wissenschaftskommunikation antreffen. Das Modell stammt also aus dem Paradigma der Experten-Laien-Kommunikation. Das ist deshalb so erwähnenswert, weil die Sourcingprozesse (s. Schritte c1 und c2 unten in der Darstellung) sich aus den Überlegungen zur ‚kognitiven Arbeitsteilung‘ innerhalb der Experten-Laien-Kommunikation speisen. Wie mit den erwähnten Schritten c1 und c2 schon implizit angesprochen wurde, ist das Sourcing – anders als im *DISCM* – nur eine Option unter mehreren. Das *CSIM* unterscheidet grundsätzlich drei Routen, die eine Person beim Erkennen von im- oder explizit vorhandenen intertextuellen Konflikten einschlagen kann. Neben dem kognitiv sparsamen Ignorieren des Konflikts (Route a) besteht eine kognitive aufwändigere Variante des Umgangs mit intertextuellen Konflikten darin, einen Konflikt zu bearbeiten, indem intensivere Integrationsbemühungen stattfinden (Route b) und/oder das Sourcing forciert wird (Route c). Das *CSIM* sieht vor, dass Personen inhaltliche Konflikte regulieren (jeweils in den Schritten a1, b1 und c1), was hier meint, dass sie vor allem die Art, das Ausmaß und die Auslöser des Konflikts erkennen und präzisieren, und darüber hinausgehend die Konflikte auch lösen. Die Konfliktlösung (dargestellt in den Schritten a2, b2 und c2) bildet damit den Schlusspunkt in den kognitiven Prozessen und Entscheidungen der lesenden Personen (Stadtler & Bromme 2014).

Im Falle des Sourcings ist die Route c) aus dem Modell in Abbildung 2 von Belang, denn sie markiert den sourcingbasierten Weg, mit intertextuellen Konflikten umzugehen. Wie auch beim *DISCM* muss eine Leserin bzw. ein Leser den Konflikt nicht nur erkannt, sondern auch als auflösungsbedürftig erachtet haben. Zwei Schritte sind gemäß dem *CSIM* denk- und gangbar:

- Im Falle der *Konfliktregulation* (c1) fungiert das Sourcing dafür, Gründe für den Konflikt aus zu erkennenden und mit dem Inhalt zu verknüpfenden Metadaten zu rekonstruieren. Hierfür sind Inferenzen und Verbindungen zwischen den einzelnen Komponenten des gesamten Dokumentenmodells erforderlich. Jemand muss also erklären können, aufgrund welcher Metadaten (etwa unterschiedliche Expertise der Autor*innen bzw. unterschiedliche Intentionen und Motive) der Konflikt erklärbar ist. Dies mündet in eine dem Dokumentenmodell nahe kognitive Repräsentation.
- Der Schritt der *Konfliktlösung* (c2) geht darüber hinaus, weil die Erklärung des Konflikts nicht dafür ausreicht, den Konflikt im Sinne einer Entscheidung über die Plausibilität bzw. die Glaubwürdigkeit zu entscheiden. Da das Vorwissen von Laien hierfür nicht ausreicht, beurteilen sie anhand der Metadaten weniger die angenommene Korrektheit von Aussagen, die in der Route b) den entscheidungsrelevanten Ausschlag gibt, sondern

beantworten mithilfe der Metadaten vorläufig und mit Restunsicherheit die Frage, welcher Quelle in puncto Expertise und Benevolenz zu trauen ist. Dies wird als ‚Evaluationen aus zweiter Hand‘ bezeichnet (Stadtler & Bromme 2014). Damit ist die Konfliktlösung regelbasiert, weil sie sich auf das Sourcing stützt, und bildet somit eine heuristische Annäherung.

Zusammenfassend lässt sich konstatieren: Das DISCM und das CSIM modellieren Sourcing als Folge eines aufgrund der gegebenen Widersprüche erschwerten intertextuellen Integrierens. Voraussetzung für das Sourcing als Reparatur intertextueller Inkohärenz ist demnach das Erkennen von Diskrepanzen. Das DISCM postuliert als Folge lediglich, dass Sourcing nach detektierten Inkohärenzen erfolgt. Das CSIM beschreibt, dass Sourcing nur eine von mehreren Optionen ist, nämlich für den Fall, dass die Inkohärenz beseitigt werden muss und dass dazu das eigene Vorwissen nicht ausreicht. Zudem kann das Sourcing in diesem Falle dazu dienen, das Zustandekommen der Inkohärenz/der Widersprüche zu erklären (in der Terminologie des CSIM: Konfliktregulation) oder sogar darüber hinausgehend zu lösen, indem die Positionen mithilfe der Metadaten in ihrer Glaubwürdigkeit abgewogen werden. Obwohl das CSIM hier spezifischer ist als das DISCM, erlaubt es in seiner gegenwärtigen Form und Komplexität noch wenig genaue Aussagen, wie Leser*innen verfahren, da das Modell implizit von komplexen Bedingungsgefügen ausgeht, welche den Kontext und Merkmale der Person inkludieren, doch diese Faktoren sind unterspezifiziert.

2.2.2.2 | Empirische Perspektiven

Wie bereits in im Teilkapitel 2.2.1.2 werden auch an dieser Stelle einige empirische Befunde zusammengestellt, mit denen bereits einige der theoretischen Grundannahmen sowohl des DISCM als auch des CSIM unterfüttert werden konnten. Hier dominieren ebenfalls Untersuchungen mit Studierenden als Testpersonen, und der Forschungsstand ist bei aller erfreulichen Tendenz der Zunahme von empirischen Studien dezidiert fragmentarisch. Dennoch finden sich bereits Befunde bzw. Befundmuster bei Produkt- und Prozessmaßen, die im Folgenden den beiden Modellen DISCM und CSIM direkt zugeordnet werden können:

- *Intertextuelle inhaltliche Konflikte gehen mit einer erhöhten strategischen Sourcingaktivität einher (DISCM):* Je mehr Konflikte ein Set an Dokumenten aufweist, desto quantitativ umfassender fallen wiederholte Sourcingaktivitäten aus (Hahnel et al. 2019a). Insbesondere stark intertextuell konfligierende Stellen evozieren vermehrtes Sourcing (Strømsø & Bråten 2014), und Studien mit Eye-Tracking zeigten, dass Leser*innen gezielt solche Stellen in Dokumenten konsultierten, welche über Metadaten Hinweise auf Glaubwürdigkeitsurteile zu versprechen schienen (Kammerer & Gerjets 2014; Gottschling, Kammerer & Gerjets 2019).
- *Leser*innen nutzen mit Inferenzen angereicherte Metadaten dazu, Konflikte zu erklären und zu lösen (CSIM):* Unter diesen Aspekt fallen zwei Schritte aus der Abbildung 2, nämlich die Kohärenzwiederherstellung durch die Akzeptanz divergierender Quellen und Perspektiven (Schritt c1) und die regelgeleitete, metadatenbasierte Beantwortung der Frage der Glaubwürdigkeit zur Auflösung eines inhaltlichen Konflikts (Schritt c2).

- *Ad c1:* Personen scheinen bei Konflikten eine Sensitivität für mögliche Erklärungen in Abhängigkeit von Metadaten zu haben. So lasten sie bei Konflikten Laien mögliche Fehler an, während sie bei Expert*innen andere Divergenzursachen wie die Komplexität des Themas bemühen (Stadtler, Scharrer & Bromme 2013). Außerdem neigen Leser*innen bei Konflikten in unterschiedlich benevolenten Dokumenten dazu, die Motive der Personen stärker als Konfliktursache zu sehen (Gottschling et al. 2019).
- *Ad c2:* Auch bei der Konfliktauflösung gibt es empirische Hinweise. So stimmen Personen im Falle unterschiedlich glaubwürdiger Dokumente den Dokumenten mit zweifelhafter Glaubwürdigkeit weniger zu (Gottschling et al. 2019) und nutzten in eigenen Texten eher die Positionen des Dokuments mit höherer Glaubwürdigkeit (Kobayashi 2014).

Auch dieser zweite selektive Forschungsüberblick illustriert, dass dem Sourcing eine wichtige, in diesem Fall kohärenzstiftende Funktion zukommt. Diese Funktion erfüllt das Sourcing, indem über vorwissensbasierte Inferenzen Metadaten mit Inhalten verknüpft werden, sodass erklärbar wird, wie inhaltliche Differenzen zwischen Dokumenten zustande kommen und wie sich intertextuelle Widersprüche über die Entscheidungen für oder gegen die Glaubwürdigkeit einzelner Quellen bzw. ganzer Dokumente aus Sicht einer lesenden Person auflösen lassen. Hierfür bedarf es der Konzertierung lesestrategischer Prozesse seitens der lesenden Person, die dadurch eine erhöhte kognitive Belastung einerseits bei der Aufgabenbearbeitung generieren, andererseits aber nahezu unvermeidlich dadurch überhaupt erst zu einer kognitiven Problemlösung gelangen (Hahnel et al. 2019b; Sweller 2011).

3 | Fazit: Was sich aus alledem für die Lesedidaktik prioritär ergibt

Dieses Kapitel hat das Thema behandelt, dass das sozialpsychologisch verortbare Sourcing einen neuen Bestandteil des kompetenten Lesens darstellt. Dieser neue Bestandteil wird möglicherweise den Status einer neuen Kompetenzdimension der Lesekompetenz erlangen (Alexander & the Disciplined Reading and Learning Research Laboratory 2012), wie es sich in Kompetenzstrukturmodellen bspw. bei PISA mit dem Reflektieren und Bewerten bereits deutlich ankündigt (OECD 2019). Nicht allein deshalb verdient er Aufmerksamkeit. Auch aus Sicht der neueren Theoriearbeit und gemäß der zunehmenden empirischen Forschung zum Lesen multipler Dokumente wirkt das Sourcing als hochbedeutsam, zumal es verschiedene Funktionen erfüllt. Diese Funktionen korrespondieren mit der Vertrauenswürdigkeit von Informationen und Dokumenten, welche eine lesende Person aufgrund der metadatenbasierten Interpretation und Evaluation zuweist. Allerdings schälen sich in der Forschung zum Sourcing bereits deutliche Hinweise darauf heraus, dass das Sourcing auch ausbleiben, fehlschlagen oder isoliert erfolgen kann (Barzilai et al. 2015; Philipp 2020; Strømsø & Bråten 2014).

Was sind die sich daraus ergebenden lesedidaktischen Implikationen? Ehe es hierauf eine priorisierte und provisorische Antwort gibt, muss noch der Hinweis erfolgen, dass lesedidaktische Maßnahmen in Hinblick auf das Sourcing nicht isoliert erfolgen sollten, sondern im Verbund mit Maßnahmen, die auf das Verstehen von singulären Texten mit eingebetteten Quellen und multiplen Dokumenten abzielen und in denen das Sourcing seine Funktionalität erfahrbar entfaltet (Brante & Strømsø 2018; Philipp 2019, 2020). Das Sourcing hat schließlich in seinen vielfältigen Zwecken immer einen extrinsischen Anlass im Sinne einer Funktionalität, die auch für die Lernenden klar erkennbar sein sollte. Hieraus speisen sich zwei lesedidaktische Implikationen, die von hoher Priorität sind:

- 1) Die *erste Implikation* betrifft eine *grundsätzliche Notwendigkeit, Sourcing im Allgemeinen zu fördern*. Hierbei hat die Lesedidaktik den Vorteil, dass sich mit dem Sourcing eine Schnittmenge zwischen Lese- und Mediendidaktik ergibt, etwa wenn es darum geht, Fake News auf die Schliche zu kommen und dabei auch Strategien des Sourcings anzuwenden. Hierfür zeichnen sich vielversprechende Wege ab, das Sourcing auf verschiedene Weisen zu vermitteln (Brante & Strømsø 2018; Philipp 2019) – und zwar durchaus in Bezug auf seine verschiedenen Funktionen (s. Teilkap. 2.1).
- 2) Die *zweite Implikation* besteht darin, die *Kombination von Metadaten und Inhalten aus Dokumenten* gezielter zum Gegenstand der Vermittlung zu machen. Das Sourcing als Reparaturmechanismus missglückender Kohärenzbildung setzt voraus, dass Personen mit Informationen verschiedener Qualitäten Inferenzen bilden, die sie dann zielbezogen nutzen. Wie Brante und Strømsø (2018) allerdings monieren, ist selbst bei effektiven Maßnahmen häufig noch zu implizit, auf welche Metadaten Lernende fokussieren sollen, wodurch die strategische Nutzung dieser Metadaten erschwert wird. Doch existiert ein sich in der Interventionsforschung abzeichnender Trend, nach dem die Kombination von Sourcing- und Integrationsprozessen bereits in einer Vielzahl von lesestrategischen Förderansätzen erfolgt ist. Dadurch ist es geglückt, diverse Verstehensleistungen zu erhöhen, darunter spezifische im Sourcing, aber auch solche, die auf intertextuelle Kohärenzbildung fokussierten (Philipp 2020).

Die Lesedidaktik steht damit gegenwärtig einerseits vor der Herausforderung, andererseits und zugleich vor der Chance, das Sourcing zum Gegenstand der systematischen Leseförderung zu machen und den bisherigen Förderansätzen hinzuzufügen (Philipp 2020). Hierin deutet sich ein nötiger Wandel an, der sich etwa in dem Ausdruck ‚staatsbürgerliche Internet-Denkfähigkeit‘ (‚Civic Online Reasoning‘; McGrew et al. 2018: 166) als Zielvorstellung wiederfindet, nämlich als die durch die Digitalisierung erforderliche Fähigkeit, soziale und politische Informationen online effektiv zu recherchieren, zu evaluieren und zu verifizieren. Das Sourcing ist hierbei als Teil des kompetenten Lesens dezidiert mitgemeint und fungiert damit als Bedingung der gesellschaftlichen Teilhabe.

Wenn im Titel des Beitrags die Frage aufgeworfen wurde ‚Liest du noch oder sourcst du schon?‘, dann impliziert die Formulierung, dass mit dem Sourcing eine neue und anspruchsvolle Facette oder auch Kompetenzdimension zum Konstrukt der Lesekompetenz hinzutritt (OECD 2019). Freilich ist das Sourcing keineswegs neu, weil es das Sourcing in Domänen wie der Geschichte schon seit Langem gibt (Wineburg 1991). Doch wird es infolge des aktuell beobachtbaren Wandels des Lesens aufgewertet, der vor allem von der Digitalisierung und ihren Folgen vorangetrieben wird. Ob das Sourcing dabei langfristig in den Kernbereich des Lesens vordringt oder im Peripheriebereich verbleibt, ist noch keine ausgemachte Sache. Die Chancen stehen aber gut, dass mit dem Sourcing ein langfristiger Begleiter der Lese- (und Schreib-)Didaktik auf den Plan getreten ist.

4 | Literatur

- Alexander, P. A. & The Disciplined Reading and Learning Research Laboratory (2012). Reading into the Future. Competence for the 21st Century. *Educational Psychologist* 4, 259–280.
- Anmarkrud, Ø., Bråten, I. & Strømsø, H. I. (2014). Multiple-Documents Literacy. Strategic Processing, Source Awareness, and Argumentation when Reading Multiple Conflicting Documents. *Learning and Individual Difference*, 64–76.
- Barzilai, S., Tzadok, E. & Eshet-Alkalai, Y. (2015). Sourcing while Reading Divergent Expert Accounts. Pathways from Views of Knowing to Written Argumentation. *Instructional Science* 6, 737–766.
- Braasch, J. L. G. & Bråten, I. (2017). The Discrepancy-Induced Source Comprehension (D-ISC) Model. Basic Assumptions and Preliminary Evidence. *Educational Psychologist* 3, 167–181.
- Brante, E. W. & Strømsø, H. I. (2018). Sourcing in Text Comprehension. A Review of Interventions Targeting Sourcing Skills. *Educational Psychology Review* 3, 773–799.
- Bråten, I., Braasch, J. L. G. & Salmerón, L. (2020). Reading Multiple and Non-Traditional Texts. New Opportunities and New Challenges. In: E. Moje, P. Afflerbach, P. Enciso & N. K. Lesaux (Eds.), *Handbook of Reading Research* (S. 79-98). Routledge.
- Bråten, I., Stadtler, M. & Salmerón, L. (2017). The Role of Sourcing in Discourse Comprehension. In: M. F. Schober, D. N. Rapp & M. A. Britt (Eds.), *The Routledge Handbook of Discourse Processes*, 2. Aufl. (S. 141-168). Routledge.
- Britt, M. A. & Rouet, J.-F. (2012). Learning with Multiple Documents. Component Skills and Their Acquisition. In: J.R. Kirby & M. J. Lawson (Eds.), *Enhancing the Quality of Learning. Dispositions, Instruction, and Learning Processes* (S. 276-314). Cambridge University Press.
- Britt, M. A., Rouet, J.-F. & Braasch, J. L. G. (2013). Documents as Entities. Extending the Situation Model Theory of Comprehension. In: M. A. Britt, S. R. Goldman & J.-F. Rouet (Eds.), *Reading. From Words to Multiple Texts* (S. 160-179). Routledge.
- Cho, B.-Y., Afflerbach, P. & Han, H. (2018). Strategic Processing in Accessing, Comprehending, and Using Multiple Sources Online. In: J. L. G. Braasch, I. Bråten & M.T. McCrudden (Eds.), *Handbook of Multiple Source Use* (S. 133-150). Routledge.
- Cho, B.-Y., Woodward, L., Li, D. & Barlow, W. (2017). Examining Adolescents' Strategic Processing during Online Reading with a Question-Generating Task. *American Educational Research Journal* 4, 691–724.
- Delgado, P., Vargas, C., Ackerman, R. & Salmerón, L. (2018). Don't Throw Away Your Printed Books. A Meta-Analysis on the Effects of Reading Media on Reading Comprehension. *Educational Research Review*, 23–38.
- Gottschling, S., Kammerer, Y. & Gerjets, P. (2019). Readers' Processing and Use of Source Information as a Function of Its Usefulness to Explain Conflicting Scientific Claims. *Discourse Processes* 5–6, 429–446.
- Hahnel, C., Kröhne, U., Goldhammer, F., Schoor, C., Mahlow, N. & Artelt, C. (2019a). Validating Process Variables of Sourcing in an Assessment of Multiple Document Comprehension. *British Journal of Educational Psychology* 3, 524–537.
- Hahnel, C., Schoor, C., Kröhne, U., Goldhammer, F., Mahlow, N. & Artelt, C. (2019b). The Role of Cognitive Load in University Students' Comprehension of Multiple Documents. *Zeitschrift für Pädagogische Psychologie* 2, 105–118.

- Kammerer, Y. & Gerjets, P. (2014). Quellenbewertungen und Quellenverweise bei Lesen und Zusammenfassen wissensbezogener Informationen aus multiplen Webseiten. *Unterrichtswissenschaft 1*, 7–23.
- Kobayashi, K. (2014). Students' Consideration of Source Information during the Reading of Multiple Texts and Its Effect on Intertextual Conflict Resolution. *Instructional Science 2*, 183–205.
- List, A., Alexander, P. A. & Stephens, L. A. (2017). Trust but Verify. Examining the Association between Students' Sourcing Behaviors and Ratings of Text Trustworthiness. *Discourse Processes 2*, 83–104.
- Magliano, J. P., McCrudden, M. T., Rouet, J.-F. & Sabatini, J. P. (2017). The Modern Reader. Should Changes to How We Read Affect Research and Theory? In: M.F. Schober, D. N. Rapp & M.A. Britt (Eds.), *The Routledge Handbook of Discourse Processes*, 2. Aufl. (S. 343-361). Routledge.
- McGrew, S., Breakstone, J., Ortega, T., Smith, M. & Wineburg, S. S. (2018). Can Students Evaluate Online Sources? Learning from Assessments of Civic Online Reasoning. *Theory & Research in Social Education 2*, 165–193.
- Merkt, M., Werner, M. & Wagner, W. (2017). Historical Thinking Skills and Mastery of Multiple Document Tasks. *Learning and Individual Differences*, 135–148.
- OECD (2019). *PISA 2018 Assessment and Analytical Framework*. OECD.
- Perfetti, C. A., Rouet, J.-F. & Britt, M. A. (1999). Toward a Theory of Documents Representation. In: H. van Oostendorp & S. R. Goldman (Eds.), *The Construction of Mental Representations during Reading* (S. 99-122). Lawrence Erlbaum.
- Petty, R. E. & Wegener, D. T. (1999). The Elaboration Likelihood Model. Current Status and Controversies. In: S. Chaiken & Y. Trope (Eds.), *Dual-Process Theories in Social Psychology* (S. 37–72). Guilford Press.
- Philipp, M. (2018). *Multiple Modelle des Leseverstehens multipler Texte. Eine Synopse aktueller kognitiver Modellierungen aus lesedidaktischer Perspektive*. www.leseforum.ch/sysModules/obxLeseforum/Artikel/646/2018_3_de_philiS.pdf [16.09.2021].
- Philipp, M. (2019). *Multiple Wege führen nach Rom. Ergebnisse einer quantitativen Sekundäranalyse effektiver Fördermaßnahmen zur Verbesserung der Sourcing- und Integrationsprozesse in der Nutzung multipler Texte*. www.leseforum.ch/sysModules/obxLeseforum/Artikel/665/2019_1_de_philiS.pdf [16.09.2021].
- Philipp, M. (2020). *Multiple Dokumente verstehen. Theoretische und empirische Perspektiven auf Prozesse und Produkte des Lesens mehrerer Dokumente*. Beltz Juventa.
- Philipp, M. (2021). Fakt oder Fake? Zu Produkten, Prozessen und Anforderungen des verstehenden Lesens multipler, konfligierender (Online-)Dokumente. In: A. Wegner, J. Frisch, E. Vetter & M. Busch (Hrsg.), *Bedingungen und Bezüge politischer und sprachlicher Bildung* (S. 70-84). Wochenschau Verlag.
- Rouet, J.-F. & Potocki, A. (2018). From Reading Comprehension to Document Literacy. Learning to Search for, Evaluate and Integrate Information across Texts. *Infancia y Aprendizaje 3*, 415–446.
- Salmerón, L., Strømsø, H. I., Kammerer, Y., Stadler, M. & van den Broek, P. (2018). Comprehension Processes in Digital Reading. In: M. Barzilla, J. Thomson, S. Schroeder & P. van den Broek (Eds.), *Learning to Read in a Digital World* (S. 91-120). John Benjamins.
- Scharrer, L. & Salmerón, L. (2016). Sourcing in the Reading Process. Introduction to the Special Issue. *Reading and Writing 8*, 1539–1548.
- Stadler, M. (2017). The Art of Reading in a Knowledge Society. Commentary on the Special Issue on Models of Multiple Text Comprehension. *Educational Psychologist 3*, 225–231.

- Stadtler, M. & Bromme, R. (2014). The Content-Source Integration Model. A Taxonomic Description of How Readers Comprehend Conflicting Scientific Information. In: D. N. Rapp & J. L. G. Braasch (Eds.): *Processing Inaccurate Information. Theoretical and Applied Perspectives from Cognitive Science and the Educational Sciences* (S. 379-402). MIT Press.
- Stadtler, M., Scharrer, L. & Bromme, R. (2013). *How Do Readers Explain the Occurrence of Conflicts in Science Texts? Effects of Presentation Format and Source Expertise. Proceedings of the Annual Meeting of the Cognitive Science Society*, 3448–3453.
- Strømsø, H. I. (2017). Multiple Models of Multiple-Text Comprehension. A Commentary. *Educational Psychologist* 3, 216–224.
- Strømsø, H. I. & Bråten, I. (2014). Students' Sourcing while Reading and Writing from Multiple Web Documents. *Nordic Journal of Digital Literacy* 2, 92–111.
- Sweller, J. (2011). Cognitive Load Theory. In: J. P. Mestre & B. H. Ross (Eds.), *Psychology of Learning and Motivation* (S. 37-76). Academic Press.
- Wineburg, S. S. (1991). Historical Problem Solving. A Study of the Cognitive Processes Used in the Evaluation of Documentary and Pictorial Evidence. *Journal of Educational Psychology* 1, 73–87.
- Wegener, D. T., Patton, K. M. & Haugtvedt, C. P. (2018). A Social Psychological Perspective on Multiple Source Use. Elaboration and Persuasion. In: J. L. G. Braasch, I. Bråten & M.T. McCrudden (Eds.), *Handbook of Multiple Source Use* (S. 79-95). Routledge.

Carolin Meier

Leseförderung mit Antolin?

1 | Forschungsgegenstand

Lesen gilt, in der heutigen Mediengesellschaft vielleicht mehr denn je, als Schlüsselqualifikation für gesellschaftliche Teilhabe. Aufgrund dessen ist es zunächst einmal die Aufgabe der Grundschule, bei allen Schülern*innen gleichermaßen sowohl Lesekompetenz auszubilden und zu festigen, zu der bekanntlich auch die Lesemotivation gehört (vgl. Schaffner et al. 2004: 93; Artelt et al. 2001: 70; Bos 2017: 11; Bock 2010: 30; Pfaff-Rüdiger 2011: 42). Große Hoffnung wird dabei auf die Integration digitaler Medien in den Leseunterricht gesetzt. Ein Beispiel für die Verbindung von Leseförderung und den Einsatz von digitalen Medien im Kontext des Grundschulunterrichts ist das Online-Portal Antolin der Westermann Gruppe. Es ist kostenpflichtig im Internet nutzbar und richtet sich an Schüler*innen der ersten bis zehnten Klasse. Bei Antolin handelt es sich um eine Datenbank mit Quizfragen zu Büchern aus dem Kinder- und Jugendbereich, welche im Multiple-Choice-Format (immer drei Choices mit zwei Distraktoren) gestellt sind. Die Schüler*innen können zwischen drei Schwierigkeitsgraden der Fragen wählen, die durch eine blaue, rote oder grüne Kappe auf dem Kopf des lesenden Rabens (das Antolin-Symbol) gekennzeichnet sind. Die blaue Kappe bedeutet, dass ausschließlich Fragen zum Inhalt des Buches gestellt werden. Die rote Kappe symbolisiert, dass neben Fragen zum Inhalt auch Fragen zur Struktur des Buches, zur Charakterisierung der Protagonisten, „zum Nachdenken“ (Homepage Antolin) und zur eigenen Meinung gestellt werden (Homepage Antolin). Die grüne Kappe symbolisiert, dass es sich um leichte Fragen für Leseanfänger und schwache Leser handelt (Homepage Antolin). Einen Einblick bieten die folgenden Beispielfragen aus einem Quiz zum Buch *Großvater und die Wölfe* von Per Olov Enquist. Es handelt sich um ein Quiz mit der roten Kappe und das Buch wird ab der vierten Klasse empfohlen.

Großvater und die Wölfe
von: Enquist, Per Olov

Verbleibende Zeit: 19:20

Überspringen

?

!

Wovon hat Mina geträumt, als sie in der Nacht aufwachte und fürchterlich weinen musste?

- Von Einbrechern, die ihre Kuschtierchen klauten
- Von Wölfen, die sie auf dem Weg zum Kindergarten angefallen haben
- Von einem Krokodil, das sie in den Po gebissen hat

1/15 0 Punkte Weiter

Abb. 1: Beispieltitem aus Antolin

© Westermann Verlag

Großvater und die Wölfe
von: Enquist, Per Olov

Verbleibende Zeit: 18:07

?

!

"Hör auf zu phantasieren", sagt der Vater, "sonst wirst du noch wie ..."

- dein Großvater." ✓
- Pippi Langstrumpf."
- Pinocchio."

 RICHTIG: Du bekommst 4 Punkte.

0 Punkte Weiter

Abb. 2: Beispieltitem aus Antolin

© Westermann Verlag

Großvater und die Wölfe
von: Enquist, Per Olov


Verbleibende Zeit: 17:20

?

! Man muss sich zwingen, nicht mehr an das Fürchterliche zu denken.

Man muss etwas Großes tun, damit einen nichts mehr erschrecken kann, z. B. eine Expedition (eine große Wanderung) unternehmen.

Man muss sich schlau machen über das, was einem Angst macht. Dann vergeht die Angst. x



FALSCH: Dir werden 4 Punkte abgezogen.

-4 Punkte [Weiter](#)

Abb. 3: Exemplarische Rückmeldung durch Antolin

© Westermann Verlag

Großvater und die Wölfe
von: Enquist, Per Olov

Du hast insgesamt -4 von 60 Punkten erreicht.

| | |
|-------------------------|-------------|
| Richtige Antworten | 7 |
| Falsche Antworten | 8 |
| Übersprungene Antworten | 0 |
| Leistung | 46 % |



[Zur Meinungsfrage \(ohne Bewertung\)](#) [Meinungsfrage überspringen](#)

15/15  -4 Punkte

Abb. 4: Exemplarische Auswertung in Antolin

© Westermann Verlag

Bei der Bearbeitung der Quizfragen können für richtig beantwortete Fragen Punkte erzielt werden, falsche Antworten führen zu Punktabzug. Die Anzahl der zu erreichenden Punkte orientiert sich an der Schwierigkeitsstufe des zugrunde liegenden Textes, der mit der Klassenstufe korreliert. Letztlich wird aus der Punktbilanz die Leseleistung errechnet.⁶⁸ Es erfolgt eine statistische Darstellung der Daten, die sowohl für die Schüler*innen als auch für die Lehrer*innen einsehbar ist (vgl. Kepser 2013: 580; Abraham & Kepser 2009: 201; Homepage Antolin).

Auf der Internetseite des Programms wird deutlich, welche Zielsetzung verfolgt werden soll: Die Nutzung von Antolin fördere Schüler*innen „auf ihrem Weg zum eigenständigen Lesen und in der Entwicklung der eigenen Leseidentität. [...] Es wird intrinsisches Interesse für Literatur geweckt“ (Homepage Antolin). Außerdem mache es gezielte Leseförderung möglich. Nach der Überarbeitung der Internetseite im Jahr 2017 werden die Ziele etwas anders formuliert. Seitdem heißt es, Antolin schaffe Anreize zum Lesen, mache motivieren leicht, Kinder üben flüssiges und sinnerfassendes Lesen und bekämen Anerkennung (Homepage Antolin). Ein Leitspruch des Programms ist: „Mit Antolin zum Lesen motivieren“ (Homepage Antolin). Zur Motivation beitragen sollen hauptsächlich das Sammeln von Punkten, die Aussicht auf eine Urkunde am Ende des Projekts sowie – so banal es klingt – die Arbeit im Internet. Aber auch das Erfassen von Inhalten, und damit die Lesefähigkeiten auf Satz- und Textebene, würden durch die Nutzung von Antolin verstärkt geschult. Und schließlich sei der Einsatz von Antolin eine unkomplizierte und sinnvolle Möglichkeit, Leseförderung und Einsatz neuer, digitaler Medien zu verbinden (Homepage Antolin).

Wie ist es um Antolins Beitrag zur institutionellen schulischen Leseförderung aber tatsächlich bestellt? Im Jahr 2017 untersuchten Viertel et al., inwieweit Antolin an Grundschulen in Niedersachsen angewendet wird und welchen Stellenwert Lehrkräfte einer solchen Form der onlinegestützten Leseförderung beimessen. Die Ergebnisse der Studie zeigen eine starke Präsenz von Antolin an niedersächsischen Grundschulen. Nutzen bereits 89 Prozent der befragten Grundschullehrkräfte internetbasierte Angebote zur Leseförderung im Allgemeinen, beziehen sich 84 Prozent ausschließlich auf Antolin. 64 Prozent der befragten Lehrer*innen verwenden das Portal im regulären Unterricht, teilweise regelmäßig (sprich wöchentlich). 61 Prozent der befragten Lehrkräfte geben an, dass die Nutzung von Antolin im schulinternen Curriculum vorgeschrieben sei (vgl. Viertel et al. 2017: 154 ff.). Die Informationen und Empfehlungen auf dem deutschen Bildungsserver und den Bildungsservern der einzelnen Bundesländer lassen darüber hinaus auf eine recht intensive Nutzung von Antolin in ganz Deutschland schließen, wird doch auf allen bekannte Bundes- und Landesservern wie auch auf der Internetseite der Stiftung Lesen aktiv auf die Internetseite des Antolin-Portals verlinkt; einige Länder empfehlen die Arbeit

⁶⁸ Für Bücher der Klasse 1 ist 1 Punkt pro Frage (10 Fragen) vorgesehen = 10 Punkte

Ab der Klasse 2 werden immer 15 Fragen gestellt, daraus ergibt sich folgendes Punkteschema:

Klasse 2: 2 Punkte = 30 Punkte

Klasse 3: 3 Punkte = 45 Punkte

Klasse 4: 4 Punkte = 60 Punkte

Klasse 5: 5 Punkte = 75 Punkte

Klasse 6: 6 Punkte = 90 Punkte

Jugendbuch: 7 Punkte = 105 Punkte

Jugendbuch ab Klasse 9: 9 Punkte = 135 Punkte

Beispielrechnung: Ein*e Schüler*in hat ein Quiz mit 15 Fragen bearbeitet und davon 10 Fragen richtig beantwortet: $10 : 15 = 0,67 \times 100 = 66,67 \%$ Leseleistung (vgl. Homepage Antolin).

mit Antolin sogar ganz konkret, sei es primär zur Leseförderung oder um digitale Medien einzusetzen (vgl. Bildungsserver).

So kann also festgestellt werden, dass in den zurückliegenden Jahren einige neue Lernsoftwares entwickelt und auf den Markt gebracht wurden, allerdings fehlen uns empirische Untersuchungen zu Einsatz und Erfolg eben dieser Programme (vgl. Möbius 2014: 350; Pieper 2015: 131). Überraschenderweise fand bisher auch noch keine empirische Überprüfung über die Wirksamkeit von Antolin statt, was doch aber aufgrund der weiten Verbreitung des Programms dringend notwendig erscheint – und zwar auf zwei Ebenen: Zum einen müssen die Herausforderung der Leseförderung und ihrer Umsetzung mithilfe von Antolin betrachtet werden. Und zum anderen gilt es zu überprüfen, ob der Einsatz digitaler Medien in dieser Form für den Leseunterricht tatsächlich bereichernd ist.

Die Frage, die meinem Forschungsprojekt also konkret zugrunde liegt, ist, ob Antolin als bundesweit bekannte und augenscheinlich einschlägige Form der Leseförderung unter Einbeziehung digitaler Medien tatsächlich ‚hält, was es verspricht‘. Folgende drei erkenntnisleitende Forschungsfragen sollen dabei beantwortet werden:

- Führt Antolin zu einer Erhöhung der intrinsischen Lesemotivation der Schüler*innen?
- Unterstützt Antolin die Schüler*innen dabei, eine positive Einstellung zum Lesen zu entwickeln?
- Und kann Antolin als ein positiver Faktor bei der Anbahnung einer gelingenden Lesesozialisation von Schüler*innen angesehen werden?

Das Ziel der vorliegenden Arbeit ist also die empirisch validierte Generierung von Hypothesen und Theoriebildung zur Wirkung des Einsatzes von Antolin auf die Lesesozialisation von Schüler*innen. Es handelt sich um eine qualitative Querschnittstudie, in der die Datenerhebung mittels leitfadengestützter Gruppendiskussionen mit Viertklässler*innen erfolgte. Ergänzt wurde diese durch Einzelinterviews mit Siebtklässler*innen, da auch der längerfristige Einfluss der intensiven Nutzung des Antolin-Programms in der Grundschule ermittelt werden sollte. Aufgrund des zeitlichen Rahmens meines Forschungsprojektes handelt es sich um einen künstlichen Längsschnitt.

2 | Untersuchungsdesign

Die Forschung mit Kindern stellt besondere Anforderungen an das Forschungsdesign und ist daher „immer auch eine Methodenforschung“ (Fuhs 2012: 93). Um Einstellungen von Kindern zu ermitteln, ist es wichtig, dass die Datenerhebung offen gestaltet und interpretatives Vorgehen bei der Datenauswertung möglich ist. Aus diesem Grund sind qualitative Methoden für diese Studie am besten geeignet (vgl. Heinzl 2012b: 22).

Grundlegend für dieses Forschungsprojekt ist zunächst einmal das Verständnis des Begriffs Einstellung. Da Einstellungen nicht direkt beobachtbar sind, handelt es sich um hypothetische Konstrukte. Die Erforschung der Einstellung dient folglich dazu, diese latente Dimension erfassbar zu machen (vgl. Bierbrauer 2005: 139; Meinefeld 1977: 9). Hierfür wurden Einstellungen als Subjektive Theorien konzeptualisiert. Grundlegend ist hierfür die Konzeption von Groeben et al. (1988), die mit ihrem Forschungsprogramm Subjektive Theorien die Erforschung der subjek-

tiven Sicht methodisch und theoretisch am detailliertesten ausgearbeitet haben, was sich innerhalb der qualitativen Forschung fest etabliert hat. Im Allgemeinen geht es zunächst einmal darum, die Perspektive der Beforschten zu verstehen und auf Grundlage dieser Theorien zu bilden. Im Zentrum der Erforschung Subjektiver Theorien stehen das Subjekt und seine Sichtweisen, Erfahrungen und Handlungen (vgl. Steinke 1999: 53). Es handelt sich um ein „[...] theoretisches Konstrukt [...], dessen Beschreibung und Qualität auf der Innensicht eines Erkenntnis-Objekts basiert und somit dessen Sinnbezüge und Intentionen einschließt [...]“ (Groeben et al. 1988: 27). Groeben et al. (1988) definieren Subjektive Theorien als „Kognitionen der Selbst- und Weltsicht, [...] als komplexes Aggregat mit (zumindest impliziter) Argumentationsstruktur, [...] das auch die zu objektiven (wissenschaftlichen) Theorien parallelen Funktionen [...] der Erklärung, Prognose, Technologie erfüllt“ (Groeben et al. 1988: 19). Diese Definition bezeichnen sie als die „weite Variante“ (Groeben et al. 1988: 19) des Konzepts. Die Definition Subjektiver Theorien der engen Variante ist dadurch ergänzt, dass Subjektive Theorien „im Dialog-Konsens aktualisier- und rekonstruierbar sind [...] [und] deren Akzeptierbarkeit als ‚objektive‘ Erkenntnis zu prüfen ist [...]“ (Groeben et al. 1988: 22). Aufgrund dieser unterschiedlichen Varianten, die mit dem Begriff Subjektive Theorie verbunden sind, und da ich in diesem Forschungsprojekt weder eine kommunikative noch explanative Validierung vornehme, sondern ausschließlich der weiten Definition des Konstrukts folge, wird der Begriff der Einstellung für den definierten Untersuchungsgegenstand verwendet.

2.1 | Methoden der Datenerhebung

Da subjektive Interpretationen in der Regel sozial verankert sind und soziale Erfahrungen eine wichtige Rolle bei der Entstehung von Einstellungen spielen (vgl. Bierbrauer 2005: 138; Aronson et al. 2014: 218 ff.; Steinke 1999: 53), habe ich mit der Methode der Gruppendiskussion als Datenerhebungsverfahren gearbeitet und folgte dabei der Konzeption Werner Mangolds (1960), da mein Erkenntnisinteresse die Erhebung der Gruppenmeinung war und nicht darin bestand, die Äußerungen der einzelnen Kinder individuell zu betrachten.⁶⁹ Für die Forschung mit Kindern wurde die Gruppendiskussionsmethode allerdings bisher nur selten angewandt und musste daher an die besonderen Anforderungen des vorliegenden Projektes angepasst werden. Die Gruppendiskussion ermöglicht in besonderem Maße eine Datenerhebungssituation, die die Verbalisierung der Subjektiven Theorien der Kinder begünstigt. Dies gilt vor allem für Kinder zwischen etwa neun und elf Jahren (vgl. Vogl 2005: 29; Heinzl 2012a: 104 ff.). Für die Ermittlung von Subjektiven Theorien älterer Kinder wie der Siebtklässler*innen ist wiederum der Einsatz von Leitfadenterviews zielführender (vgl. Kunze 2004: 206; Groeben & Scheele 2000). Eine Teilstrukturierung der Datenerhebungssituation ist in beiden Fällen für die Aktualisierung und Erhebung Subjektiver Theorien sinnvoll, da den Kindern auf diese Weise ausreichend Freiraum für eigene Schwerpunktsetzungen und Themen eingeräumt und trotz-

⁶⁹ Mangold ist der Ansicht, dass die gegenseitige Beeinflussung der Teilnehmer während der Gruppendiskussion verhindert, dass sich tatsächlich Einzelmeinungen herauskristallisieren lassen (vgl. Mangold 1960: 28) und dass es sich stattdessen eher um eine kollektive Meinung handelt, die am Ende einer Gruppendiskussion steht (vgl. Lamnek 2005: 57). Als Erkenntnisinteresse der Gruppendiskussion benennt Mangold die Ermittlung der „informellen Gruppenmeinung“ (Mangold 1960: 6). Diese wird während der Diskussion „arbeitsteilig vorgetragen. Die Sprecher bestätigen, ergänzen, berichtigen einander, ihre Äußerungen bauen aufeinander auf [...]. Die Gruppenmeinung ist keine ‚Summe‘ von Einzelmeinungen, sondern das Produkt kollektiver Interaktionen.“ (Mangold 1960: 49).

dem sichergestellt wird, dass die im Hinblick auf das Erkenntnisinteresse wichtigen Aspekte angesprochen werden. Die von Groeben & Scheele (2000) vorgeschlagene Abfolge von zunächst offenem Erzählimpuls und anschließend spezifischeren Fragen macht beide Ziele mittels Leitfadeninterview erreichbar.

2.2 | Auswertungsmethode

Um inhaltlich-thematische Aspekte herauszuarbeiten, wie es für die Erhebung von Einstellungen notwendig ist, eignet sich zur Auswertung die qualitative Inhaltsanalyse nach Mayring (2015). Diese ist sowohl für das Datenmaterial, das aus den Gruppendiskussionen als auch aus den Leitfadeninterviews hervorgegangen ist, nutzbar. Die Methode ermöglicht es, neben manifesten auch latente Inhalte und Deutungen interpretativ zu erschließen. Um subjektive Sichtweisen aus auf diese Art und Weise generiertem Material herauszuarbeiten, hat sich die Strukturierung als gewinnbringendste Form der qualitativen Inhaltsanalyse erwiesen, und da es sich bei der Ermittlung von Einstellungen um ein thematisch-inhaltliches Forschungsinteresse handelt, liegt es nah, nach inhaltlichen Gesichtspunkten zu strukturieren.

Der induktiven Orientierung und dem Ziel der Theoriegenerierung wird die Methode durch eine kleine Modifikation gerecht. Das deduktive Analyseverfahren anhand des auf dem Leitfaden für Diskussionen und Interviews basierenden Kodierleitfadens wird durch einen zusätzlichen Arbeitsschritt ergänzt, der eine induktive Ausdifferenzierung der Kategorien vorsieht. Auf diese Weise bietet die Auswertung anhand der qualitativen Inhaltsanalyse ausreichend Offenheit für die Erfassung kindlicher subjektiver Theorien. Dieses wird im Abschnitt Auswertung expliziert.

2.3 | Untersuchungsaufbau

Im Folgenden soll nun das Untersuchungsdesign des Forschungsprojekts zusammengefasst werden. Der Fokus der Datenerhebung in diesem Forschungsprojekt lag auf den Gruppendiskussionen mit Viertklässler*innen. Nach dem Urteil der teilnehmenden Lehrkräfte sollten die Kinder in der vierten Klasse bereits ausreichend Lesefähigkeiten erworben haben, um selbstständig und auch lustorientiert lesen zu können (vgl. dazu auch Bamberger 2000: 158). Insgesamt wurden 14 Gruppendiskussionen mit jeweils sechs bis zehn Kindern komparativ ausgewertet. Zusätzlich wurde auch der längerfristige Einfluss der intensiven Nutzung des Antolin-Programms in der Grundschule ermittelt. Hierfür wurden 31 Siebtklässler*innen mittels Leitfadeninterviews befragt.

Bei den Viertklässlergruppen handelte sich um natürliche Gruppen, die jeweils hinsichtlich der schulischen Antolin-Nutzung sowie der sekundären Lesesozialisation und meist auch bezüglich ihrer primären Lesesozialisation homogen sind. Die Einordnung bezüglich der primären Lesesozialisation erfolgt anhand der Angaben der Lehrer*innen und Schulleiter*innen und wurde mithilfe eines Elternfragebogens⁷⁰ kontrolliert. Die Auswahl der Stichproben in dieser Untersuchung erfolgte anhand des Kriteriums der Nutzung des Antolin-Programms im Unterricht, die

⁷⁰ Bei der Konzeption der Elternbefragung wurde der Fragebogen von Richter und Plath aus ihrer Forschung zur Lesemotivation in der Grundschule (vgl. Richter & Plath 2007) als Vorbild genommen, entsprechend der Fragestellung dieser Arbeit gekürzt und mithilfe der theoretischen Erkenntnisse zur primären Lesesozialisation ergänzt sowie einige Variablen, die zur Erfassung der primären Lesesozialisation in der IGLU-Studie verwendet wurden (vgl. Stubbe et al. 2007: 305), berücksichtigt.

Informationen zur Antolin-Nutzung der Viertklässler*innen basieren auf den Angaben der Lehrkräfte. Die folgende Tabelle bietet einen Überblick über die Diskussionsgruppen.

| Name der Schule ⁷¹ | Antolin-Nutzung (ja/nein/gemischt) | Ausprägung der primäre Lesesozialisation (eher stark/eher schwach) | Code |
|-------------------------------|---------------------------------------|---|------|
| Michael Ende-GS Gruppe 1 | ja | eher schwach | A_sw |
| Michael Ende-GS Gruppe 2 | ja | eher schwach | A_sw |
| Michael Ende-GS Gruppe 4 | ja | eher schwach | A_sw |
| Michael Ende-GS Gruppe 3 | ja | eher stark | A_st |
| Erich Kästner-GS Gruppe 2 | ja | eher stark | A_st |
| Erich Kästner-GS Gruppe 3 | ja | eher stark | A_st |
| Paul Maar-GS | nein | eher schwach | B_sw |
| Kerstin Gier-GS | nein | eher schwach | B_sw |
| Joanne K. Rowling-GS | nein | eher schwach | B_sw |
| Cornelia Funke-GS | nein | eher stark | B_st |
| Otfried Preußler-GS | nein | eher stark | B_st |
| Janosch-GS | nein | eher stark | B_st |
| Christine Nöstlinger-GS | gemischt | eher schwach | C_sw |
| Erich Kästner-GS Gruppe 1 | gemischt | eher stark | C_st |

Tab. 1: Diskussionsgruppen (Meier 2019: 252)

Ist bei den Viertklässler*innen von Antolin-Nutzung (vgl. Tab. 1 Code A) die Rede, bedeutet dies, dass Antolin sowohl innerhalb des Unterrichts als auch zu Hause von den Schüler*innen genutzt wird und dass die Arbeit mit dem Online-Portal einen festen Platz im Stundenplan hat.

Keine Antolin-Nutzung (vgl. Tab. 1 Code B) bedeutet, dass diese Integration in den Unterricht nicht gegeben ist. In vielen Fällen haben die Schüler*innen aber dennoch einen Zugang zu dem Portal und können es freiwillig zu Hause nutzen.

Außerdem existieren Mischformen der Nutzung (vgl. Tab. 1 Code C). In diesem Fall wird Antolin in der Regel nicht in der Schule eingesetzt, die Nutzung zu Hause ist aber verpflichtend für die Schüler*innen und wird von der Lehrkraft kontrolliert.

Da es sich um eine komparative Analyse handelt, wurden ebenso viele Schülergruppen befragt, die das Antolin-Programm intensiv in der Schule nutzen (vgl. Tab. 1 Code A), wie Gruppen, die in der Schule nicht damit arbeiten (vgl. Tab. 1 Code B).

⁷¹ Es handelt sich um Pseudonyme, um die Anonymität zu gewährleisten. Alle Schulen befinden sich eher im ländlichen Raum von Nordrhein-Westfalen und Niedersachsen. Pro Gruppe nahmen sechs bis zehn Kinder teil. Die Schwankung hat forschungspraktische Gründe. Alle Untersuchungsgruppen waren gemischtgeschlechtlich, da eine geschlechtsspezifische Betrachtung im Rahmen meiner Untersuchung nicht beabsichtigt war.

Ein weiteres Kriterium war die primäre Lesesozialisation der Kinder: Die Hälfte der befragten Schülergruppen besteht aus Kindern, die weniger explizite oder implizite Förderung im Rahmen ihrer primären Lesesozialisation erfahren haben (vgl. Tab. 1 Code sw). Außerdem wurden sieben Schülergruppen mit mehr Leseerfahrung aus der Familie befragt (vgl. Tab. 1 Code st). Die primäre Lesesozialisation wurde von der Lehrkraft eingeschätzt und mittels eines Elternfragebogens kontrolliert.

Das Ziel war es, diese Gruppen jeweils untereinander zu vergleichen und damit in Kontrast setzen zu können. Die Auswahl der einzelnen Kinder pro Gruppe übernahmen die Lehrkräfte. Sie wurden gebeten, Schüler*innen auszuwählen, die einen Querschnitt durch das Leseleistungsspektrum der Klasse repräsentierten.

2.4 | Datenerfassung

Die Gruppendiskussionen fanden in einem Raum an der jeweiligen Schule statt, der so vorbereitet war, dass die Schüler*innen in Kreisform beziehungsweise kreisähnlicher Form platznahmen. Das Ziel hierbei war die Bezugnahme der Schüler*innen aufeinander, die möglichst selbstständige Organisation der Diskussion und die Abwendung des Fokus von mir als Diskussionsleiterin. Zum Einstieg in die Diskussion wurde ein Bild als Diskussionsstimulus genutzt, der eine Phase des freien Assoziierens und Erzählens anregen sollte. Zu sehen war ein Kind, das auf einem Bücherstapel sitzt und liest. In den Gesprächsbeiträgen dieses Teils der Gruppendiskussionen zeigten sich die Einstellungen der Kinder besonders deutlich und konnten vor dem Hintergrund der restlichen Äußerungen sehr gewinnbringend interpretiert werden, denn in der Regel erfolgten hier die Herstellung einer Verbindung zu ihrer eigenen Situation und Äußerungen bezüglich Identifikation oder Animosität. Im Anschluss an die Einstiegsphase wurden offene Fragen gestellt, die allen Diskussionen in Form eines Leitfadens zugrunde lagen. Die längsten Diskussionen dauerten etwa 45 Minuten und hier war deutlich zu spüren, dass die Schüler*innen zum Ende ermüdeten. Folglich scheint es empfehlenswert, die Diskussionen eher etwas kürzer anzulegen.

In einer zweiten Phase der Datenerhebung wurden 31 Siebtklässler*innen mittels leitfadengestützter Einzelinterviews⁷² befragt, wobei sich die Einordnung der Antolin-Nutzung auf die Grundschulzeit bezieht und den Angaben, die die Kinder in einem Schülerfragebogen im Anschluss an das Interview machten sowie den Erzählungen während der Interviews entnommen ist. Interviewt wurden sowohl Schüler*innen, die das Gymnasium besuchen, als auch Schüler*innen, die auf die Real- und die Gesamtschule gehen. Auf diese Weise ist auch hier die Abbildung eines Querschnitts durch das Leseleistungsspektrum gewährleistet. Auch diese Datenerhebung erfolgte in einem Raum an der jeweiligen Schule. Der offene Einstiegsimpuls wurde an das Alter der Schüler*innen angepasst und erfolgte nicht mithilfe eines Bildes, sondern in Form eines Erzählimpulses, der eine freie Darstellung der Assoziationen der Schüler*innen zum Thema Lesen veranlassen sollte.

Sowohl die Gruppendiskussionen als auch die Einzelinterviews wurden ausschließlich akustisch mithilfe eines Tonaufnahmegeräts erfasst und vollständig in ein schriftliches Transkript umgewandelt, das als Basis des Auswertungsprozesses diente.

⁷² Der Leitfaden ist an dem für die Gruppendiskussionen orientiert.

2.5 | Auswertung

Die Gruppendiskussionen und Leitfadeninterviews wurden jeweils separat und wie bereits erläutert mittels qualitativer Inhaltsanalyse nach Mayring (2015) analysiert. Die detaillierte qualitative Analyse des Materials erfolgte zunächst separat für jede Gruppe und Siebtklässler*in. Der Ablauf der inhaltsanalytischen Auswertung folgte dem Ablaufmodell der inhaltlich strukturierenden qualitativen Inhaltsanalyse nach Mayring und wurde durch induktive Kategorienbildung ergänzt (vgl. Mayring 2015: 97ff.). Die Analyse des Materials begann damit, dass deduktiv, anhand des Leitfadens, die thematischen Hauptkategorien bestimmt und ein Kodierleitfaden erstellt wurde. Der erste Kodierungsprozess bestand darin, das Material Zeile für Zeile zu untersuchen und die einzelnen Textabschnitte den Hauptkategorien zuzuordnen. Nachdem die Textteile den Hauptkategorien zugeordnet wurden, wurden induktiv Subkategorien bestimmt, das bedeutet, die noch sehr allgemeinen Hauptkategorien wurden weiter ausdifferenziert. Es folgte ein zweiter Kodierungsprozess, bei dem nun das Material wieder von Anfang an genau betrachtet und Textstellen den Subkategorien zugeordnet wurden. Das Extrahieren der kodierten Textteile erfolgte automatisch, da die Auswertung mithilfe des Programms MAXQDA durchgeführt wurde. Im Anschluss an die Kodierung des Datenmaterials wurde dieses je Sub- und Hauptkategorie paraphrasiert und zusammengefasst. Im Zentrum des Interesses standen die Haupt- und Subthemen, die aus dem Material extrahiert wurden (vgl. Kuckartz 2014: 79 ff.), denn besonders die inhaltlichen Ergebnisse sind von zentraler Wichtigkeit. Außerdem waren Zusammenhänge zwischen den Subkategorien in einer Hauptkategorie oder Zusammenhänge der einzelnen Hauptkategorien untereinander von Interesse (vgl. Kuckartz 2014: 94 ff.). Auf Basis dieser Ergebnisse wurden verschiedene Einstellungen der Viertklässlergruppen und Siebtklässler*innen zum Lesen ermittelt. Um die Ergebnisse der einzelnen Gruppendiskussionen und Leitfadeninterviews anschließend in Beziehung zueinander zu setzen, wurden sie mit dem Ziel der Hypothesengenerierung und Theoriebildung miteinander verglichen. Dies erfolgte mithilfe der Gruppierungen, die in der Tabelle 1 dargestellt sind. Die im Folgenden zusammengefassten Ergebnisse sind als Hypothesen zu charakterisieren und auf die Gruppendiskussionen fokussiert, denn diese bilden den Schwerpunkt meines Forschungsprojekts.

3 | Zusammenfassung der Ergebnisse

Insgesamt konnten anhand des Datenmaterials der Gruppendiskussionen vier Einstellungen zum Lesen extrahiert werden. Inwiefern die Antolin-Nutzung mit diesen korreliert, ist in der folgenden Tabelle 2 dargestellt. Da Einstellungen sehr vielschichtige Konstrukte sind, ist diese Einteilung stark generalisiert und nicht immer trennscharf abgrenzbar.

(1) Eskapistisches Lesen

Für Schüler*innen, die eskapistisch lesen, ist das Lesen eine Freizeitbeschäftigung. Sie lesen vorrangig zur Unterhaltung sowie zur Entspannung und die Leseaktivität beruht auf Freiwilligkeit. Das Lesen ermöglicht es den Schüler*innen, aus ihrem Alltag zu fliehen, hinein in die Fantasiewelten der Bücher. Dies führt gelegentlich zu Flow-Erlebnissen. Oftmals ist das eskapistische Lesen allerdings gegenstandsspezifisch motiviert. Schulisch motiviertes Lesen spielt für eskapistisch motivierte Schüler*innen nur eine sehr rudimentäre Rolle.

(2) Intrinsisch motiviertes kompetitives Lesen zu Lernzwecken

Für Schüler*innen, die diese Einstellung aufweisen, dient das Lesen vorrangig schulischen Zielen, allerdings sind sie intrinsisch motiviert, diese zu erreichen. Das Lesen ist für diese Schüler*innen eine Schlüsselqualifikation für Erfolg und da sie erfolgreich sein möchten, lesen sie. Sie empfinden das Lesen nicht als Pflicht, sondern messen ihm besondere Wichtigkeit bei und bewerten es positiv.

(3) Dualität von eskapistischem und kompetitivem Lesen

Die Schüler*innen, bei denen diese Einstellung zu erkennen ist, befinden sich in einer Art Konflikt: Lesen ist für sie zum einen eine Freizeitbeschäftigung, zum anderen aber auch Arbeit für die Schule. Diese beiden Funktionen des Lesens konkurrieren, meist zulasten des genussorientierten Lesens, woraus oftmals eine Frustration der Schüler*innen resultiert.

(4) Extrinsisch motiviertes kompetitives Lesen als Pflichterfüllung

Schüler*innen, die primär kompetitiv lesen, haben kaum oder keine positiven Erfahrungen mit dem Lesen und mit Büchern gemacht. Sie lesen ausschließlich funktional und extrinsisch motiviert, um ihre schulische Pflicht zu erfüllen. Spaß haben sie beim Lesen nicht (Meier 2019: 439).

Die Schüler*innen, die regelmäßig mit dem Programm arbeiten, lesen hauptsächlich kompetitiv und extrinsisch motiviert oder befinden sich in einem Konflikt zwischen ursprünglich vorhandener intrinsischer Lesemotivation und damit verbundenem Lesen zu eskapistischen Zwecken und der schulischen Verpflichtung zum Lesen, die letztlich ihre Einstellung zum Lesen dominiert. Tatsächlich als Lesepersönlichkeiten, die primär intrinsisch motiviert lesen, sind nur Schüler*innen zu charakterisieren, die Antolin maximal freiwillig zu Hause nutzen (Code B). Die primäre Lesesozialisation der Schüler*innen spielt dabei den Ergebnissen zufolge keine entscheidende Rolle (vgl. Tab. 2).

Anstatt intrinsischer Lesemotivation scheint die Nutzung von Antolin eher extrinsische Lesemotivation zu fördern und dadurch dazu beizutragen, dass die Kinder das Lesen mehr mit Arbeit verknüpfen als mit Vergnügen, was dem Anschein nach die Entwicklung einer positiven Einstellung zum Lesen eher behindert. Die Gründe dafür werden im folgenden Absatz zusammengefasst und sind im Detail nachzulesen in Meier (2019).

| Einstellung 1: eskapistisches Lesen | Einstellung 2: intrinsisch motiviertes kompetitives Lesen | Einstellung 3: Dualität von eskapistischem und kompetitivem Lesen | Einstellung 4: extrinsisch motiviertes kompetitives Lesen |
|--|---|---|---|
| | | Michael Ende-GS Gr. 1 (A_sw) | |
| | | | Michael Ende-GS Gr. 2 (A_sw) |
| | | Michael Ende-GS Gr. 4 (A_sw) | |
| | | Michael Ende-GS Gr. 3 (A_st) | |
| | | | Erich-Kästner-GS Gr. 2 (A_st) |
| | | Erich-Kästner-GS Gr. 3 (A_st) | |
| | Paul Maar-GS (B_sw) | | |
| Kerstin Gier-GS (B_sw) | | | |
| Joanne K. Rowling- GS (B_sw) | | | |
| Cornelia Funke- GS (B_st) | | | |
| Otfried Preußler- GS (B_st) | | | |
| Janosch-GS (B_st) | | | |
| | Christine Nöstlinger- GS (C_sw) | | |
| | | Erich-Kästner-GS Gr. 1 (C_st) | |

Tab. 2: Übersicht LeseEinstellung der Viertklässler*innen (Meier 2019: 441)

Die Viertklässler*innen, die im Zentrum dieser Studie stehen, befinden sich dem angenommenen Verlauf der Lesesozialisation zufolge in der Phase des Viellesens. Demnach ist davon auszugehen, dass vor allem die Kinder mit eher stärkerer primärer Lesesozialisation auch ohne schulische Interventionen intrinsisch zum Lesen motiviert sind und das Lesen für sie eskapistische Zwecke erfüllt. Daher wäre es in dieser Phase ausreichend, wenn die Schule Lesestoffe bereitstellte und die Kinder bei der Lektüreauswahl unterstützte. Darin, dass die Leseförderung mittels Antolin die Kinder jedoch zusätzlich extrinsisch motiviert, ist ein Erklärungsansatz dafür zu finden, dass sich der Großteil der Schüler*innen, die Antolin im Leseunterricht nutzen, in einem Konflikt zwischen dem Wunsch nach eskapistischem, genussorientierten Lesen und dem durch die extrinsische Motivation fokussierten kompetitiven, pflichterfüllenden und funktionalen Lesen befinden.

Um gezielte und individuelle Leseförderung betreiben zu können, ist zunächst eine genaue Beobachtung des Leseverhaltens, der Lesekompetenz und der Leseentwicklung der Schüler*innen notwendig, um daraus geeignete Lesefördermethoden abzuleiten. Diese zu erleichtern, ist

ein weiteres Ziel, das der Einsatz von Antolin verspricht. Die Übersicht darüber, wie viele Fragen richtig oder falsch beantwortet wurden, letztlich also die erreichte Punktzahl, ist jedoch als Leistungsmessung zu beurteilen, welche ausschließlich Objektivität fingiert und keine Grundlage für eine angemessene Bewertung ist, insbesondere, weil Kinder, die zwar viel lesen, aber Antolin nicht so intensiv nutzen, auf diese Weise falsch eingeschätzt werden. Es konnte außerdem bereits festgestellt werden, dass ausschließlich die Beobachtung der Lesemenge und des Leseverständnisses für eine professionelle Diagnose des Leseförderbedarfs nicht ausreichen. Notwendig sind offene Gespräche über Leseerfahrungen, anhand derer Lehrer*innen feststellen können, wo die Probleme der Schüler*innen liegen. Solche Gespräche über Bücher werden jedoch durch den Einsatz von Antolin oftmals komplett aus dem Unterricht verdrängt.

Dass die Aktivitäten im Antolin-Programm von der Lehrkraft so genau überwacht werden können, führt vor allem bei leseschwachen Schüler*innen dazu, dass sie verstärkt Angst haben, Fehler zu machen. Gesteigert wird dies durch die Art und Weise der Fehlerkommunikation. Aus den Diskussionen und Interviews ist zu schließen, dass viele Lehrkräfte zudem im Zuge der Vergabe von Urkunden Rankings veröffentlichen, in denen die Schüler*innen ihre Leistung im Vergleich zu der ihrer Mitschüler*innen sehen. Sowohl die daraus folgende starke extrinsische Motivation sowie das Gefühl des Leistungsdrucks konnten bei den Schüler*innen, die Antolin intensiv nutzen, herausgearbeitet werden. Sie lesen, wenn sie von der Lehrkraft zur Nutzung von Antolin verpflichtet werden, um Demütigungen in der Klasse und das Missfallen der Lehrkraft zu vermeiden, beziehungsweise um deren Anerkennung zu gewinnen. Dies steht im Kontrast zum eigentlichen Ziel, intrinsische Lesemotivation und eine positive Einstellung zum Lesen auszubilden. Diese Art der Rückmeldung führt vor allem bei leseschwachen Kindern, die im Fokus von Leseförderung stehen sollten, vielfach zur Schwächung ihres Selbstkonzeptes und senkt ihre Lesemotivation. Daher profitieren den Ergebnissen zufolge ausschließlich leistungsorientierte, misserfolgsvermeidend motivierte und/oder ohnehin lesestarke Schüler*innen von den Rückmeldungen bei Antolin. Doch der damit zusammenhängende Leistungsdruck bewirkt auch bei diesen Schüler*innen in den meisten Fällen, dass nicht die intrinsische Motivation verstärkt wird, sondern ausschließlich die extrinsische und der funktionale Charakter des Lesens.

Dies zeigt sich besonders deutlich darin, dass Schüler*innen, die Antolin intensiv nutzen, teilweise keinen Sinn darin sehen, Bücher zu lesen, zu denen sie keine Fragen bei Antolin beantworten können.

Das ist allerdings bei Schüler*innen, die bereits sehr stabil intrinsisch leseemotiviert sind, nicht in dem Maße festzustellen. Sie versuchen, ihre genussorientierte LeseEinstellung beizubehalten, indem sie sich gegen die Leistungsorientierung auflehnen. Im schulischen Kontext scheint dies jedoch in der Regel aufgrund der Selektions- und Allokationsfunktion aussichtslos zu sein. Die intensive Nutzung von Antolin führt in vielen Fällen dazu, dass den Schüler*innen vermittelt wird, dass nur das funktionale, kompetitive Lesen, nicht aber das eskapistische wertgeschätzt wird. Je geringer intrinsische Lesemotivation und eine genussorientierte LeseEinstellung ausgeprägt sind, desto mehr scheint dies von den Schüler*innen übernommen zu werden.

Besonderer Wert wird hinsichtlich der Leseförderung der Auswahl der Lesestoffe beigemessen: Entsprechen diese den Interessen der Kinder, verknüpfen sie das Lesen mit positiven Gefühlen wie Freude und Vergnügen, was dazu führt, dass sie intrinsisch motiviert lesen. Ist dies nicht der Fall, erleben sie das Lesen in der und für die Schule als Zwang. In diesem Zusammenhang

ist es wichtig, dass die Kinder lernen, Lesestoffe selbst bedürfnisgerecht und interessengemäß auszuwählen.

Anhand der Gruppendiskussionen konnte jedoch festgestellt werden, dass die Nutzung von Antolin die Buchauswahl der Schüler*innen stark beeinflusst, sodass sie in den freien Lesestunden nicht, wie es die Idee des Ansatzes des Vielleseverfahrens (vgl. Rieckmann 2010) ist, Bücher lesen, die ihren Interessen entsprechen, sondern Bücher, mit denen sie viele Punkte erreichen können, die sie aber oftmals nicht interessieren. Folglich wird zwar eine Individualisierung des Unterrichts ermöglicht, allerdings bewirkt der Einsatz von Antolin genau das Gegenteil von dem, was eigentlich das Ziel solcher Ansätze ist: Aufgrund des hohen Grades an Fremdbestimmtheit und der starken Fokussierung des Leistungsaspekts führt dieser dazu, dass es eher zu einer Einschränkung der Freiheiten kommt. Die für Vielleseverfahren charakteristische Motivation zum vermehrten Lesen mittels extrinsischer Anreize ist auch in der Konzeption von Antolin zu finden und führt, den Ergebnissen zufolge, tatsächlich dazu, dass die Schüler*innen mehr lesen – allerdings nicht lieber.

4 | Schluss

Anhand der Ergebnisse des hier vorgestellten Forschungsprojekts konnte also festgestellt werden, dass sich die Nutzung digitaler Medien in Form von Antolin kurzfristig stark auf das Leseverhalten, die Lesemotivation, die Buchauswahl und auch auf die Einstellung der Schüler*innen zum Lesen im Allgemeinen auswirkt, und zwar unabhängig von ihrer primären Lesesozialisation. Die Auswertung der Gruppendiskussionen zeigt, dass die auf der Internetseite von Antolin genannten und eingangs zitierten Ziele, speziell zur Förderung der intrinsischen Lesemotivation und zur Unterstützung der Kinder bei der Entwicklung einer eigenen Leseidentität durch die intensive Nutzung des Programms, nicht erfüllt werden. Ebenso können auf Basis der Auswertung der für dieses Projekt erhobenen Daten auch die erwarteten Vorteile des computerunterstützten Unterrichts im Hinblick auf den intensiven Einsatz von Antolin nicht belegt werden. Es zeigt sich vielmehr (und ganz im Gegenteil zu der mit Antolin verbundenen Hoffnung nach ‚Innovation‘ i.w.S.), dass mit der Nutzung des Antolin-Programms ein didaktisch rückständiger Ansatz wieder zum Einsatz kommt, welcher als drill-and-test-Konzept⁷³ (vgl. Nattland & Kerres 2009: 319 f.) zu charakterisieren ist.

Hinsichtlich der langfristigen Auswirkungen des intensiven Einsatzes von Antolin zur Leseförderung, der mittels der Leitfadeninterviews mit Siebtklässler*innen erhoben wurde, kann Gleiches jedoch nicht festgestellt werden – im Gegenteil: In etwa 60 Prozent der Fälle beeinflusste die Nutzung von Antolin in der Grundschule die Einstellung der Kinder zum Lesen den Ergebnissen zufolge nicht nachhaltig. Eine positive sowie eine negative Beeinflussung sind jeweils in 20 Prozent der Fälle feststellbar. Bezüglich des Ziels, die Lesekompetenz zu erhöhen, können auf Basis der im Zuge dieser Studie erhobenen Daten keine Aussagen getroffen werden. Diese Frage müsste in einer separaten Studie untersucht werden.

⁷³ Geläufiger ist die Bezeichnung drill-and-practice-Programme, bei denen nach dem Informationsinput vom Computer gestellte Fragen, meist im Multiple-Choice-Format, beantwortet werden müssen. Dieser gibt auch die Rückmeldung, ob die Aufgabe richtig oder falsch gelöst wurde. Hierbei werden dann Fehlerprotokolle angelegt und gespeichert, was das Kind schon kann und was noch nicht (vgl. Kepser 2013: 580). Es gibt jedoch auch Kritik an der Bezeichnung drill-and-practice. Da es sich weniger um Üben, sondern mehr um Abfragen von Wissen handelt, wird der Begriff drill-and-test-Software teilweise bevorzugt (vgl. Nattland & Kerres 2009: 319 f.).

5 | Literatur

- Abraham, U. und Kepser, M. (2009). *Literaturdidaktik Deutsch. Eine Einführung*. Erich Schmidt Verlag.
- Aronson, E., Wilson, T. & Akert, R. (2014). *Sozialpsychologie*. Pearson.
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In: J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69-137). Leske + Budrich.
- Bierbrauer, G. (2005). *Sozialpsychologie*. Verlag W. Kohlhammer.
- Bock, K. (2010). *Kinderalltag – Kinderwelten. Rekonstruktive Analysen von Gruppendiskussionen mit Kindern*. Verlag Barbara Budrich.
- Bos, W. (2017). Vorwort. In: A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E-M. Lankes, N. McElvany, T.C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 11-12). Waxmann Verlag.
- Fuhs, B. (2012). Kinder im qualitativen Interview – Zur Erforschung subjektiver kindlicher Lebenswelten. In: F. Heinzel (Hrsg.), *Methoden der Kindheitsforschung. Ein Überblick über Forschungszugänge zur kindlichen Perspektive* (S. 80-103). Beltz Juventa Verlag.
- Groeben, N. & Scheele, B. (2000). Dialog-Konsens-Methodik im Forschungsprogramm Subjektive Theorien. *Forum Qualitative Sozialforschung*, 1, Nr.2, Art. 10 <http://nbn-resolving.de/urn:nbn:de:0114-fqs0002105> [16.09.2021].
- Groeben, N., Wahl, D., Schlee, J. und Scheele, B. (1988). *Das Forschungsprogramm Subjektive Theorien*. A. Francke Verlag.
- Heinzel, F. (2012a). Gruppendiskussion und Kreisgespräch. In: F. Heinzel (Hrsg.), *Methoden der Kindheitsforschung. Ein Überblick über Forschungszugänge zur kindlichen Perspektive* (S. 104-115). Beltz Juventa Verlag.
- Heinzel, F. (2012b). Qualitative Methoden in der Kindheitsforschung. Ein Überblick. In: F. Heinzel (Hrsg.), *Methoden der Kindheitsforschung. Ein Überblick über Forschungszugänge zur kindlichen Perspektive* (S. 22-35). Beltz Juventa Verlag.
- Kepser, M. (2013) Computer im Literaturunterricht. In: V. Frederking, H.-W. Huneke, A. Krommer & C. Meier (Hrsg.), *Taschenbuch des Deutschunterrichts. Band 2 Literatur- und Mediendidaktik* (S. 568-592). Schneider Verlag Hohengehren.
- Kuckartz, U. (2014). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*, 2. Auflage. Beltz Juventa Verlag.
- Kunze, I. (2004). *Konzepte von Deutschunterricht. Eine Studie zu individuellen didaktischen Theorien von Lehrerinnen und Lehrern*. VS Verlag für Sozialwissenschaften.
- Lamnek, S. (2005). *Gruppendiskussion. Theorie und Praxis*, 2. Auflage. Beltz Verlag.
- Mangold, W. (1960). *Gegenstand und Methode des Gruppendiskussionsverfahrens. Aus der Arbeit des Instituts für Sozialforschung*. Europäische Verlagsanstalt.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagentexte und Techniken*. Beltz Verlag.
- Meier, C. (2019). *Webbasierte Leseförderung in der Grundschule am Beispiel von Antolin. Eine empirische Studie zur Lesesozialisationsforschung*. Peter Lang.
- Meinefeld, W. (1977). *Einstellungen und soziales Handeln*. Rowohlt Taschenbuch Verlag.

- Möbius, T. (2014). Empirische Forschung zum Einsatz digitaler Medien im Deutschunterricht. Ein Überblick. In: V. Frederking, A. Krommer & T. Möbius (Hrsg.), *Digitale Medien im Deutschunterricht* (S. 337-358). Schneider Verlag Hohengehren.
- Nattland, A. & Kerres, M. (2009). Computerbasierte Medien im Unterricht. In: K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht*, 2. Auflage (S. 317-324). Julius Klinkhardt.
- Pfaff-Rüdiger, S. (2011). *Lesemotivation und Lesestrategien. Der subjektive Sinn des Bücherlesens für 10- bis 14-Jährige*. Lit Verlag.
- Pieper, I. (2015). Lese- und Literarische Sozialisation. In: M. Kämper-van den Boogaart & K. H. Spinner, *Lesen und Literaturunterricht. Teil 1 Geschichte und Entwicklung. Konzeptionelle und empirische Grundlagen* (S. 87-147). Schneider Verlag Hohengehren.
- Richter, K. & Plath, M. (2007). *Lesemotivation in der Grundschule. Empirische Befunde und Modelle für den Unterricht*, 2. Auflage. Juventa Verlag.
- Rieckmann, C. (2010). *Leseförderung in sechsten Hauptschulklassen. Zur Wirksamkeit eines Vielleseverfahrens*. Schneider Verlag.
- Schaffner, E., Schiefele, U., Drechsler, B. & Artelt, C. (2004). Lesekompetenz. In: M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 93-110). Waxmann Verlag.
- Steinke, I. (1999). *Kriterien qualitativer Forschung. Ansätze zur Bewertung qualitativ-empirischer Sozialforschung*. Juventa Verlag.
- Stubbe, T. C., Buddeberg, I., Hornberg, S. & McElvany, N. (2007). Lesesozialisation im Elternhaus im internationalen Vergleich. In: W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 299-327). Waxmann Verlag.
- Viertel, M., Ehrenspeck-Kolasa, Y. & Spies, A. (2017). Digitale Leseförderung an Grundschulen zwischen Anspruch und Wirklichkeit. Eine Untersuchung zur Nutzung und Bewertung der web-basierten Leseförderung ‚Antolin‘ durch Grundschullehrkräfte in Niedersachsen (NuBeAn). In: K. Mayrberger, J. Fromme, P. Grell & T. Hug (Hrsg.), *Jahrbuch Medienpädagogik 13. Vernetzt und entgrenzt – Gestaltung von Lernumgebungen mit digitalen Medien* (S. 151-164). Springer VS.
- Vogl, S. (2005). Gruppendiskussionen mit Kindern: methodische und methodologische Besonderheiten. *ZA- Information/ Zentralarchiv für Empirische Sozialforschung*, 57, 28-60. www.nbn-resolving.de/urn:nbn:de:0168-ssoar-198469 [16.09.2021].

Linkliste Homepage Antolin

| Titel der Homepage | online verfügbar unter: | zuletzt abgerufen: |
|--|---|--------------------|
| Antolin in der Grundschule: Anregungen und Tipps aus der Praxis zur Leseförderung | www.antolin.de/all/lesefoerderung_antolin.jsp | 16.09.2021 |
| Antolin-Prospekt | www.antolin.de/all/downloads/prospekt.pdf | 16.09.2021 |
| Das leistet Antolin | https://www.antolin.de/all/info/das_leistet_antolin.jsp | 16.09.2021 |
| Das Online-Programm zur Leseförderung von Klasse 1 bis 10. Präsentation zur Einführung | https://www.antolin.de/all/downloads/antolin_ppt.pdf | 16.09.2021 |
| Hinweise für Eltern: Häufig gestellte Fragen. | https://antolin.westermann.de/all/info/haeufig_gestellte_fragen.jsp | 16.09.2021 |
| Mit Antolin arbeiten | https://antolin.westermann.de/all/info/mit-antolin-arbeiten.jsp | 16.09.2021 |
| Nutzerhandbuch | www.antolin.de/all/downloads/nutzerhandbuch.pdf | 16.09.2021 |
| Quizfragen – der Kern von Antolin | https://antolin.westermann.de/all/info/quizfragen_der_kern.jsp | 16.09.2021 |
| Was ist Antolin? | www.antolin.de/all/howto.jsp | 16.09.2021 |

Linkliste Homepages Bildungsserver

| Titel der Homepage | Online verfügbar unter: | zuletzt abgerufen: |
|---|---|--------------------|
| Deutscher Bildungsserver: Antolin – Ermunterung zum Lesen. | www.bildungsserver.de/db/mlesen.html?Id=16809 | 16.09.2021 |
| Bildungsportal des Landes Nordrhein-Westfalen: Leseförderung. Links. | https://www.bildungspartner.schulministerium.nrw.de/Bildungspartner/Lesefoerderung/Lesefoerderung-Links/ | 16.09.2021 |
| Bildungsserver Berlin-Brandenburg (a): Buch in der Vitrine. | https://bildungsserver.berlin-brandenburg.de/buch-in-vitrine | 16.09.2021 |
| Bildungsserver Berlin-Brandenburg (b): Überregionale Partner. | https://bildungsserver.berlin-brandenburg.de/lesecurriculum-ueberregional/ | 16.09.2021 |
| Bildungsserver Mecklenburg-Vorpommern: Grundschule. | www.bildung-mv.de/eltern/schule-und-unterricht/schularten/grundschule/index.html | 16.09.2021 |
| Bildungsserver Rheinland-Pfalz (a): Lesen mit digitalen Medien fördern | https://lesen.bildung-rp.de/lesekompetenz-foerdern/lesen-im-unterricht/lesen-mit-digitalen-medien-foerdern.html | 16.09.2021 |
| Bildungsserver Sachsen-Anhalt: Arbeit mit Antolin - Zusammenarbeit mit der Stadt- und Schulbibliothek Landsberg Heike Gstöttner, Gymnasium Landsberg. | www.bildung-lsa.de/schule/schul__und_modellversuche/prolesen/projektbeispiele.html?INH_ID=9025 | 30.03.2017 |
| Institut für Qualitätsentwicklung an Schulen (2008): Eingangsphase | www.schleswig-holstein.de/DE/Landesregierung/IQSH/Publikationen/PDFDownloads/Unterrichts- | 16.09.2021 |

| | | |
|---|--|------------|
| an Grundschulen. Individuelle Förderung im Unterricht ausbauen. Deutsch. | entwicklung/SchulUnterrichtsgestaltung/Downloads/EingangsphaseOrdner/deutsch.pdf?blob=publicationFile&v=3 | |
| n-21 (a): Leseförderung mit Antolin. | www.n-21.de/staticsite/staticsite.php?menuid=176&topmenu=25&noscript=1&noscript=1&noscript=1 | 30.03.2017 |
| n-21 (b): Lesen. Kinderbücher und Lesen. | www.n21.de/staticsite/staticsite.php?menuid=186&topmenu=4&submenu=166&noscript=1&noscript=1&noscript=1&noscript=1&noscript=1 | 16.09.2021 |
| n-21 (c): Unterrichtsmaterialien. Online-Materialien für den Unterricht in der Grundschule. | www.n-21.de/staticsite/staticsite.php?menuid=129&topmenu=4&submenu=6 | 30.03.2017 |
| Hamburger Bildungsserver: Lesekompetenz anbahnen und fördern. | https://bildungsserver.hamburg.de/grundschule-lesen/ | 16.09.2021 |
| Hessischer Bildungsserver (a): Antolin - Landeslizenz zum 31.01.2012 beendet. | www.medien.bildung.hessen.de/liz/antolin/fin.html | 16.09.2021 |
| Hessischer Bildungsserver (b): Antolinschulen nach Medienzentren 2011 | www.medien.bildung.hessen.de/liz/antolin/Antolinschulen.html | 16.09.2021 |
| Institut für Qualitätsentwicklung an Schulen (2008): Eingangsphase an Grundschulen. Individuelle Förderung im Unterricht ausbauen. Deutsch. | www.schleswig-holstein.de/DE/Landesregierung/IQSH/Publikationen/PDFDownloads/Unterrichtsentwicklung/SchulUnterrichtsgestaltung/Downloads/EingangsphaseOrdner/deutsch.pdf?blob=publicationFile&v=3 | 16.09.2021 |
| Landesbildungsserver Baden-Württemberg (a): Die lange Lesenacht. | www.schulebwde/unterricht/paedagogik/ideen-poolleseorderung/grundschule/leseprojekt/lesenacht | 16.09.2021 |
| Landesbildungsserver Baden-Württemberg (b): Adressen und Links. | https://www.schule-bw.de/themen-und-impulse/ideen-pool-lesen/sekundarstufe/sekundarstufe1/adressen-und-links | 16.09.2021 |
| Saarland: Bibliotheken und Leseförderung. | www.saarland.de/22142.htm | 16.09.2021 |
| Niedersächsischer Bildungsserver Deutsch: Deutsch in der Grundschule. | www.nibis.de/nibis.php?menid=9578 | 16.09.2021 |
| Thüringer Schulportal: Schule und Bibliothek | www.schulportal-thueringen.de/schule_bibliothek?csth=Antolin | 16.09.2021 |
| Stiftung Lesen: Dossier: Internetseiten für Kinder. | www.stiftunglesen.de/initiativen-und-aktionen/digitales/1234 | 16.09.2021 |

Miriam Dittmar

Biologielehrmitteltexte sinnvoll lesen und verstehen

Eine Untersuchung zu Text-Bild-Integrationsprozessen bei Siebtklässler*innen

1 | Einleitung

Sprache und Lernen hängen sehr eng zusammen. Fehlen die sprachlichen Kompetenzen, so wirkt sich das auch immer auf das fachliche Lernen aus. Diese Erkenntnis gilt für alle Fächer, und nicht nur für den Deutschunterricht (Baumert & Schümer 2001; Bolte & Pastille 2010). Der große Einfluss sprachlicher Kompetenz von Schüler*innen auch außerhalb des Deutschunterrichts macht sich beim fachlichen Lernen mit Texten bemerkbar. Denn das Lesen von schulischen Fachtexten gewinnt im Verlauf der schulischen Sozialisation immer mehr an Bedeutung (Rincke 2010), so dass die Lesekompetenz als überfachliche Kompetenz eine immer wichtigere Rolle spielt. Doch das fachliche Lernen mit diesen Texten kann nur gelingen, wenn der Text auch von den Lesenden verstanden wird, also die Passung zwischen Lesenden und Text gegeben ist (Leisen 2013).

Gerade in den naturwissenschaftlichen Fächern zeigt sich aber immer wieder, dass ein Großteil der im Unterricht eingesetzten Lehrmitteltexte von den Schüler*innen nicht ausreichend gut verstanden wird, dass auch von einem fachlichen Lernen ausgegangen werden könnte (für Physik: Merzyn 1994; Starauschek 2003; für Biologie: Dittmar et al. 2017; Schmellentin et al. 2017). Das liegt zum einen an der sprachlichen Komplexität der Lernmaterialien, die als stark bildungssprachliche Texte grundsätzlich hohe, mitunter auch zu hohe Ansprüche an die Lernenden stellen (Schleppegrell 2008; von Borries 2012; für eine ausführliche Diskussion zu den Merkmalen schulischer Fachtexte siehe auch Schneider et al. 2019). Zum anderen müssen Schüler*innen beim Lesen von naturwissenschaftlichen Lehrmitteltexten in der Lage sein, mit ungewohnten Textformaten und verschiedenen Textsorten umzugehen. Denn der Text enthält nicht nur den verbalen Fließtext, sondern auch Bilder und Diagramme sowie die in den Lehrmitteln typischen Paratexte wie Anleitungen, Fragen und Aufgaben (Lindauer & Schneider 2007). Für ein verstehendes Lesen müssen diese verschiedenen Textteile miteinander in Bezug gesetzt werden, was aber oft nicht gelingt, da die Bezüge zwischen diesen Textteilen nicht explizit im Text dargestellt sind. Dittmar et al. (2017) diskutieren diese Komplexitätsmerkmale auf der Textstrukturebene bei Biologielehrmitteltexten im Detail: So sind zum Beispiel die Text-Bild- und Bild-Bild-Bezüge

im Text häufig nicht sichtbar gemacht (z.B. durch Abbildungsverweise). Zudem stehen Abbildungen weit entfernt von den dazugehörigen Textstellen oder sie unterbrechen Textabschnitte und müssen daher übersprungen werden. Aber auch der Fließtext selbst ist meist nicht in einer gewohnten Form dargestellt, sondern in einer Spaltengliederung wiedergegeben, so dass die Lesewege auch durch den Fließtext nicht immer eindeutig ersichtlich sind. Ein lineares Lesen durch den Lehrmitteltext führt bei diesen diskontinuierlichen Texten nicht zum Erfolg.

Damit Verstehen gelingen kann, braucht es eine gute Passung der Lehrbuchtexte an die literarischen Kompetenzen der Lernenden, denn Textschwierigkeit und Textverstehen werden sowohl von text- als auch von personenbezogenen Merkmalen beeinflusst (Mesmer, Cunningham & Hiebert 2012). Um diese Passung zu erreichen, macht es Sinn, sowohl den Text als auch die Lernenden im Blick zu halten und Maßnahmen für beide Seiten zu entwickeln sowie deren Wirkung zu überprüfen (Leisen 2013; Schmellentin et al. 2017). Die Wirkung textseitiger Maßnahmen wurde im Forschungsprojekt „Textverstehen in den naturwissenschaftlichen Schulfächern“ (Dittmar et al. 2017; Schmellentin et al. 2017; Schneider et al. 2018; Schneider et al. 2019) genauer untersucht. Ein grundsätzliches Resultat dieser Forschung ist, dass bei Lernenden in der Sekundarschule auf Niveaustufe Sek-A (erweiterte Ansprüche: leistungsstärkere Lernende) die textseitigen Maßnahmen in Form von Textanpassungen mit einem signifikant höheren Wissenszuwachs nach der Textlektüre einhergehen, als wenn diese Textanpassungen nicht vorgenommen wurden. Gleichzeitig zeigt sich aber, dass die Textanpassungen besonders bei den Lernenden auf der Niveaustufe Sek-B (Grundansprüche: leistungsschwächere Lernende) alleine nicht ausreichen. Einen signifikant höheren Wissenszuwachs beim angepassten Text konnten letztere nur dann erzielen, wenn Textanpassungen mit didaktischen Maßnahmen verknüpft wurden und das Verstehen mit gezielten Fragen an den Text angeleitet und diskursiv unterstützt wurde (Schneider et al. 2018).

Schüler*innen müssen also in ihrem Leseprozess unterstützt werden. Um zu erkennen, wie dieser am effektivsten zu unterstützen ist, macht es Sinn, die kognitiven Prozesse, die beim Lesen ablaufen, zu betrachten. Nicht ganz einig ist man sich dabei, ob sich das Textverstehen additiv hierarchisch aufbaut (Kintsch & van Dijk 1978; Schnotz 1994; van Dijk & Kintsch 1983), oder eine mentale Kohärenzbildung auf mehreren mentalen Repräsentationsebenen parallel stattfindet (Kintsch 1988, 1998). Es wird aber angenommen, dass unterschiedliche Textverstehens-ebenen am Kohärenzaufbau beteiligt sind: Zunächst einmal wird die Textoberfläche dekodiert und eine Textoberflächenrepräsentation gebildet. Aus dem Gelesenen werden dann semantische Propositionen erschlossen. Diese bilden die Textbasis. Die Bedeutungen von Sätzen und Textteilen werden als mentale Repräsentationen im Arbeitsgedächtnis mit vorhandenem Wissen verknüpft, um schließlich als Situationsmodell im Langzeitgedächtnis abgespeichert zu werden. Damit solche mentalen Repräsentationen der Textinhalte überhaupt aufgebaut werden können, müssen Beziehungen zwischen Sätzen und Textteilen hergestellt werden, d.h. lokale und globale Kohärenz muss aufgebaut werden (Kintsch 1998; Rickheit & Schade 2000). Hierbei hilft textseitig eine nachvollziehbare inhaltliche Strukturierung (van Dijk & Kintsch 1983).

Trotzdem bleiben textseitig Verbindungen zwischen Textabschnitten immer auch implizit, da nicht alle Verknüpfungen explizit dargestellt werden (können). Wissensgeleitete Inferenzen bilden und logisches Schlussfolgern sind daher beim Leseprozess notwendig (Christmann 2010; Kintsch & van Dijk 1978). Um auf Basis eines Textes solch ein Situationsmodell konstruieren zu

können, sind beim Lesenden zwei Ebenen von Lesekompetenzen wichtig, die aufeinander aufbauen. Wenn die hierarchieniedrigen Lesefertigkeiten wie das Wortverstehen oder auch das Gewinnen lokaler Informationen nicht gelingen, können auch die hierarchiehöheren Textverstehensprozesse wie das Gewichten und Verknüpfen von Informationen nicht funktionieren (Lenhard 2013).

Bei schulischen Fachtexten kommt oft erschwerend hinzu, dass neben den textlichen Verknüpfungen auch noch Verknüpfungen mit Abbildungen gemacht werden müssen, damit ein globaler Kohärenzaufbau gelingen kann. Textabschnitte müssen also nicht nur miteinander, sondern auch mit den passenden Bildern inhaltlich verknüpft werden, vor allem dann, wenn sehr komplexe Wissenskonzepte aus dem Text heraus aufgebaut werden sollen. Diese wichtigen Text-Bild-Bezüge sind dabei sowohl auf der Ebene der basalen Lesekompetenz zu verorten, wenn auf der Textoberfläche erkannt werden muss, welches Bild zu welcher Textstelle gehört, als auch in den hierarchiehohen Prozessen, wenn für den globalen Kohärenzaufbau Verknüpfungen von Informationen aus Text und zugehöriger Abbildung notwendig werden.

Der Vorgang des globalen Kohärenzaufbaus aus Text-Bild-Gefügen fordert das Arbeitsgedächtnis stark, weil die sprachlichen und die visuell-bildhaften Informationen in zwei unterschiedlichen Kanälen verarbeitet werden, aber für die Bildung des Situationsmodells wieder zusammengeführt werden müssen (Mayer 2009; Paivio 1986; Schnotz & Bannert 2003). Schnotz und Bannert (2003) sprechen hier von *deskriptionalen* und *depiktionalen* Repräsentationszweigen („descriptive and depictive representations“, ebd.: 143). Damit also schließlich der globale Kohärenzaufbau gelingen kann, müssen die kognitiv anspruchsvollen Teilprozesse *Selektion*, *Organisation*, *Transformation* und *Integration* (Mayer 2009; Schlag & Ploetzner 2011), parallel in beiden Repräsentationszweigen ausgeführt und abgeglichen werden: Zunächst werden die relevanten sprachlichen und visuellen Einheiten selektiert. Die selektierten Informationen werden mental organisiert und geordnet. So entsteht ein Wissensbestand mit sprachlich repräsentierten und ein Wissensbestand mit bildhaft repräsentierten Informationen. Werden Text- und Bildinformationen in das jeweils andere Format überführt, so spricht man von Transformation. Über Transformationsvorgänge und Integration in bestehendes Wissen werden die Informationen in das Situationsmodell eingebettet. Für das Entstehen des Situationsmodells braucht es folglich, dass erkannt wird, welche bildlichen und welche sprachlichen Informationen zusammengehören.

Betrachtet man die Vorgänge des Textverstehens unter dem Aspekt des Wissenserwerbs und des Lernens – eine vorrangige Funktion des Lesens, vor allem in den naturwissenschaftlichen Fächern – dann sind diese kognitiven Teilprozesse Selektion, Organisation, Transformation und Integration von sowohl sprachlichen als auch visuellen Informationen notwendig, um von dem alten Wissenszustand „Vorwissen“ zum neuen Wissenszustand „Nachwissen“ zu gelangen. Das Situationsmodell, welches über das Verstehen des Text-Bild-Gefüges aufgebaut wird, ist dabei maßgeblich an der Veränderung des Wissenszustands beteiligt.

Aus lesedidaktischer Sicht wird dabei immer klarer, dass diese hochkomplexen Verstehensprozesse, gerade auch für sprachlich schwächere Lernende, systematisch zu entwickelnde wie zu implementierende Unterstützungsmaßnahmen erfordern, da der Textverstehensprozess (wie

in der zitierten Literatur beschrieben) nur dann ungestört ablaufen kann, wenn es währenddessen nicht zu einer Überbelastung des Arbeitsgedächtnisses kommt (Chandler & Sweller 1991). Das kann gerade bei Text-Bild-Kombinationen schnell passieren, weil es zu einem *Split Attention Effekt* kommen kann, also zu einer kognitiven Überlastung aufgrund der Aufteilung der Aufmerksamkeit zwischen den zwei verschiedenen Zeichensystemen (Kürschner & Schnotz 2007; Mayer & Moreno 1998). Es ist also ein Ziel der Maßnahmen, das Arbeitsgedächtnis möglichst zu entlasten. Dies kann dadurch erreicht werden, dass die Lernenden schon durch die Art der Informationsgestaltung unterstützt werden, und damit die sachfremde kognitive Belastung (*extraneous cognitive load*) reduziert wird. Zudem kann die kognitive Belastung gesteuert werden, indem instruktionale Maßnahmen ergriffen werden, die verstehensfördernde Aktivitäten induzieren und damit die lernrelevante kognitive Belastung (*germane cognitive load*) gezielt erhöht wird (Brünken, Seufert & Zander 2005).

Wie nun Schüler*innen in ihrem Leseprozess gewinnbringend unterstützt werden können, damit beschäftigt sich der sprachensible bzw. sprachbewusste Fachunterricht (Leisen 2013; Lindauer, Schmellentin & Beerenwinkel 2016). Grundpfeiler dieses didaktischen Ansatzes sind neben sprachlichen Anpassungen bei den Lesetexten die vier Leseschritte, die den Leseprozess von Schüler*innen steuern und dabei gleichzeitig Lesestrategien vermitteln und einüben (Lindauer et al. 2013). An Kintschs (1998) Textverstehensebenen orientiert,

- dient der erste Leseschritt dazu, das Vorwissen zu aktivieren,
- wird im zweiten Leseschritt das Verstehen der Textbasis gesichert,
- werden im dritten Leseschritt lokale und globale Kohärenzen des Textinhaltes erschlossen, um das Verstehen größerer Einheiten, wie Absätze oder Unterkapitel zu erlangen,
- so dass, zeitgleich dazu und kumulativ, quasi im vierten Leseschritt schließlich das Situationsmodell des Textes konstruiert wird, das durch Reflexion des Gelesenen und durch Abgleich mit vorhandenem Wissen abgesichert wird.

Wie das Lesen insbesondere informationsdichter Texte, wie sie in den naturwissenschaftlichen Fächern zum Einsatz kommen, angeleitet und strukturiert werden kann, stellen Lindauer et al. (2013) ausführlich und auch an unterrichtspraktischen Beispielen dar. Doch auf die Rolle, die Bilder und Text-Bild-Bezüge beim Verstehen naturwissenschaftlicher Schulbuchtexte spielen, wird nur ansatzweise eingegangen. So wird vor allem darauf hingewiesen, dass das Bildmaterial einer Schulbuchdoppelseite zum Überblicken des Textes und der Vorwissensaktivierung genutzt werden kann. Wie aber genau das Integrieren von Abbildungen in den Lese- und Verstehensprozess gesteuert und angeleitet werden kann, bleibt vage.

Um eine gezielte Leseprozesssteuerung für die in den naturwissenschaftlichen Schulfächern typischen Text-Bild-Gefüge zu entwickeln, ist es von Vorteil, erst einmal den unangeleiteten Leseprozess von Schüler*innen mit solchen Lehrmitteltexten zu untersuchen und die dabei entstehenden Schwierigkeiten mit dem Leseweg durch den Text zu analysieren. Daher geht die vorliegende Studie der Frage nach, wie Schüler*innen der 7. Jahrgangsstufe einen komplexen Biologielehrmitteltext lesen und inwieweit ihre Lesewege durch den Text Auswirkungen auf den Wissenserwerb zeigen.

2 | Methoden

Die Daten der Studie stammen aus Leseprozessbeobachtungen, die im Rahmen des Projektes „Textverstehen in den naturwissenschaftlichen Schulfächern“ (Dittmar et al. 2017; Schmellentin et al. 2017; Schneider et al. 2018, 2019) durchgeführt wurden⁷⁴.

Dem Projekt liegt ein Mixed-Methods-Design (vgl. Müller 2018) zugrunde: Insgesamt wurden drei empirische Studien durchgeführt. Zwei sind qualitativ angelegte Leseprozessbeobachtungen (Dittmar et al. 2017; Schmellentin et al. 2017) und eine ist eine größer angelegte Interventionsstudie quantitativer Art (Schneider et al. 2018). Die Daten der hier vorgestellten Studie entstammen den Leseprozessbeobachtungen mit zwei Textvarianten eines Lehrbuchtextes zum Thema Atmung, einem Originallehrmitteltext (Beuck et al. 2012) und einem überarbeiteten Text⁷⁵. Bei dem für die vorliegende Studie ausgewählten Originallehrmitteltext handelt es sich um einen von drei Lehrmitteltexten aus gängigen Schweizer Biologie- und Naturwissenschaftslehrmitteln, die als Grundlage für das Projekt dienten, weil sie thematisch dem Lehrplan der siebten Klasse entsprachen und als in ein Thema einführende Texte der Vermittlung von Lerninhalten dienten. Alle drei Texte wurden im Projekt einer umfassenden linguistischen Textanalyse unterzogen, um typische Textkomplexitätsmerkmale für das Fach Biologie zu bestimmen (vgl. Schmellentin et al. 2017). Der ausgewählte Text zum Thema Atmung lieferte diesbezüglich umfangreiche Komplexitätsmerkmale, so dass er am geeignetsten für eine textseitige Überarbeitung erschien. Der überarbeitete Text zum Thema Atmung wurde aufgrund von empirisch ermittelten Prinzipien zur besseren Verständlichkeit auf der Basis des Originallehrmitteltextes erstellt. Diese Prinzipien ergaben sich aus den tatsächlichen Textschwierigkeiten (Mesmer et al. 2012), also aus Verstehensschwierigkeiten, die bei der Leserschaft im Leseprozess aufgetreten sind. Aus diesen Verstehensschwierigkeiten in Kombination mit den textlinguistisch ermittelten Textkomplexitätsmerkmalen wurden sechs Kategorien von insgesamt 34 Überarbeitungsprinzipien ermittelt, die im Folgenden kurz aufgezählt und mit Einzelbeispielen erläutert werden:

- *Kategorie 1 – Prinzipien zum Layout*: Typographische Mittel (z.B. Fettdruck) wurden bewusst und kohärent eingesetzt und thematische Sinneinheiten auch grafisch als Einheit präsentiert.
- *Kategorie 2 – Prinzipien zur Inhaltsorganisation und Gliederung (Textkohärenz)*: Es wurde auf eine schrittweise Themenentfaltung und eine explizite Themeneinführung geachtet (z.B. advanced organizers), der Text wurde auf die wesentlichen Informationen fokussiert, Nebenkonzpte wurden weggelassen und Redundanzen geschaffen.
- *Kategorie 3 – Prinzipien zur Kohäsion*: Bezüge im Text wurden eindeutig und explizit dargestellt.

⁷⁴ Das Projekt „Textverstehen in den naturwissenschaftlichen Schulfächern“ wurde von 2013-2016 an der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz unter der Leitung von Hansjakob Schneider und Claudia Schmellentin sowie der Mitarbeit von Eliane Gilg und Miriam Dittmar durchgeführt. Finanziert wurde das Projekt vom Schweizerischen Nationalfonds (SNF).

⁷⁵ Beide Texte sind in den Abbildungen 3, 4 & 5 einzusehen. Der überarbeitete Text ist zudem in Schneider et al. (2019) zugänglich.

- *Kategorie 4 – Prinzipien zur Bildgestaltung*: Die Abbildungen wurden auf die wesentlichen Informationen fokussiert und möglichst alle wichtigen visualisierbaren Sachverhalte auch durch Abbildungen unterstützt.
- *Kategorie 5 – Prinzipien zur syntaktischen Komplexität*: Komplexe syntaktische Einschübe und komplexe Nominalphrasen wurden vermieden.
- *Kategorie 6 – Prinzipien zu (Fach-)Wortschatz und Morphologie (Lexik)*: Fachbegriffe wurden explizit erläutert, es wurde auf Fachbegriffskonstanz geachtet und morphologisch komplexe Wörter wurden im Text vermieden oder aufgeschlüsselt.

Die Überarbeitungsprinzipien und ihre Herleitung sind in Schneider et al. (2019) detailliert beschrieben.

Da die vorliegende Studie Daten zu Lesewegen durch den Text und im Speziellen auch zu Text-Bild-Bezügen präsentiert, werden hier nun ein bisschen ausführlicher die Textüberarbeitungen beschrieben, die die Inhaltorganisation, Textstruktur und die Abbildungen betreffen: Die einzelnen Themen wurden in Unterkapitel gegliedert (Prinzipienkategorie 1) und durch neugestaltete, zum Teil zusätzliche Bilder unterstützt (Prinzipienkategorie 4). Dabei wurde auf eine leseprozesslogische Text-Bild-Platzierung geachtet (Prinzipienkategorie 2), d.h. die Abbildungen stehen immer vor der entsprechenden Textstelle, so dass im linearen Lesefluss zuerst die Abbildung betrachtet wird und danach erst die dazugehörige Textpassage. Die Text-Bild-Bezüge sind mit Abbildungsverweisen im Text explizit dargestellt (Prinzipienkategorie 3). Bei der Neugestaltung der Abbildungen wurde auf eine Fokussierung der wesentlichen Informationen geachtet und Nebenkonzpte weggelassen (Prinzipienkategorie 4). Die Themenentfaltung der einzelnen Konzepte geschieht schrittweise, wobei alle Informationen zu einem Thema auch in einem Abschnitt stehen (Prinzipienkategorie 2). Trotz Textanpassung enthielten beide Texte zur Atmung die gleichen Wissenskonzepte.

An den Leseprozessbeobachtungen zu den Lesewegen durch den Text nahmen sechzehn Lernende der siebten Klassenstufe (Durchschnittsalter 13 Jahre, 11 Monate; Min = 13 Jahre, 2 Monate; Max = 14 Jahre, 10 Monate, acht Mädchen, acht Jungen) teil. Je acht der Lernenden besuchten Schulen des Niveaus „erweiterte Ansprüche“ (Sek-A) und je acht Schulen des Niveaus „Grundansprüche“ (Sek-B). Zwischen den Schüler*innen aus den beiden Schulniveaus konnte kein Unterschied in der Lesekompetenz festgestellt werden. Bei beiden Gruppen handelt es sich um mittelstarke Leser*innen (T-Werte für Lesegeschwindigkeit: Min = 45, Max = 54; T-Werte für Leseverstehen: Min = 37, Max = 58; LGVT 6-12 mit der Normtabelle für die Gesamtschule 7. Klasse (Schneider, Schlagmüller & Ennemoser 2007)). Die Erfahrungen der Schüler*innen mit Biologielehrmitteltexten muss als unterschiedlich eingestuft werden. Lediglich kann angenommen werden, dass alle Schüler*innen seit ca. einem halben Jahr (mit Eintritt in die Sekundarschule zum siebten Schuljahr) Unterricht im Fach Natur und Technik bekamen und persönliche Erfahrungen aus dem Sachunterricht der Grundschulzeit mitbrachten. Im Gespräch mit den Lehrpersonen wurde sichergestellt, dass das Thema Atmung sowie Texte zum Thema Atmung noch nicht durchgenommen worden sind. Das Vorwissen der Proband*innen wurde diesbezüglich erhoben und konnte als eher niedrig eingestuft werden.

Jeder Schüler*in wurde eine der beiden Textversionen (Originaltext oder überarbeiteter Text) zugeteilt und sie wurden beim Lesen videografiert (N = 8 pro Textversion). Während der Leseprozessbeobachtung befanden sie sich zusammen mit der Versuchsleitung in einem ruhigen Raum der Schule. Die Testung erfolgte als Einzeltestung.

Die Leseprozessbeobachtung gliederte sich in zwei Sequenzen, eine Lesesequenz und eine Fragesequenz. Während der Lesesequenz haben die Versuchspersonen den Text still für sich gelesen, wobei sie mit dem Finger nachfahren sollten, so dass zu jedem Zeitpunkt ersichtlich ist, an welcher Textstelle oder Abbildung sich die Schüler*in aufhält. Aus diesen Daten lassen sich die Lesewege durch den Text ableiten. Wie die Versuchspersonen während des Lesens mit dem Finger den Text nachfahren sollten, wurde vorher anhand eines Beispielfideos instruiert und anhand eines Beispieltextes geübt.

Nach der Lesesequenz folgte die Fragesequenz. Während dieser beantworteten die Versuchspersonen Fragen zum Text mithilfe des Textes in einer Interviewsituation mit der Versuchsleitung, in welcher die Schüler*innen auch aufgefordert wurden, ihre Antworten zu den Fragen mit entsprechenden Textstellen zu begründen. Daraus ergaben sich problematische schülerseitige Ereignisse, die als Textverstehensschwierigkeiten interpretiert werden können. Das waren z.B. Falschbeantwortungen einer Frage, obwohl vorher die richtige Textstelle gelesen wurde, oder nicht korrekte Äußerungen von Schüler*innen in ihren Erklärungen und Begründungen, obwohl der gelesene Text diesen Sachverhalt anders darstellte. Diese problematischen schülerseitigen Ereignisse geben Hinweise auf das Konzeptverstehen.

3 | Auswertung

Die qualitative Auswertung der Lesewege der Schüler*innen durch den Text während der Lesesequenz wurde mit der Videokodiersoftware ELAN (ELAN 2019) vorgenommen. Dazu wurden zwei Videos verschiedener Ansichten des/der Lesenden (aufgenommen mit zwei Videokameras) synchronisiert und in einem ersten Schritt dahingehend zeitlich segmentiert, wann ein Bezug zwischen Textteilen hergestellt wurde. Als Textteil galt ein Absatz oder ein Bild. Absatzüberschriften wurden zu den jeweiligen Absätzen gezählt, Gesamttitel galten als eigener Textteil. Die Segmente erstrecken sich folglich über den zeitlichen Abschnitt, während dem die Lesenden in einem Textteil verweilen. In einem zweiten Schritt wurde auf die gebildeten Segmente kodiert, ob die Schüler*innen den Fließtext lesen oder eine Abbildung betrachten. Zudem wurde bei jedem Wechsel zwischen Textteilen bestimmt, welche Art Bezug dabei gemacht wurde, also ob ein Text-Text-, Text-Bild-, Bild-Bild- oder Bild-Text-Bezug durchgeführt wurde. Da die Analyse nur ganze Absätze bzw. ganze Abbildungen einbezieht, ist die Analyse der Lesewege eine sehr grobe Analyse. So gibt sie zum Beispiel wiederholendes Lesen innerhalb eines Absatzes nicht wieder. Da der Fokus der Studie aber auf den gemachten Text-Bild-Bezügen liegt, wird dieser Feinheitsgrad der Analyse als ausreichend angesehen.

Darüber hinaus wurde noch die Richtung der Lesewege kodiert, d.h., ob im Text vorwärts gelesen wurde oder zurück zu früheren Passagen gegangen wurde. Außerdem wurde untersucht, ob der jeweilige Bezug linear erfolgte, also der Anordnung der Textteile auf der Lehrmittelseite folgte oder nichtlinear, also Textteile übersprungen worden sind. Aus dieser Analyse ergeben

sich vier mögliche Leserichtungen: „linear vor“, „linear rück“, „nicht-linear vor“ und „nicht-linear rück“.

Abschließend wurde dann für jeden Leseweg kodiert, ob der Leseweg als sinnvoll oder nicht sinnvoll im Hinblick auf das Textverstehen gelten kann. Sinnvolle Lesewege wurden im Rahmen der textlinguistischen Analyse im Projektteam für die beiden Lehrmitteltex-te im Konsensverfahren bestimmt, da sie eine Grundlage für die Textüberarbeitung darstellten (z.B. leseprozesslogische Text-Bild-Platzierung in Prinzipienkategorie 2). Diese Bestimmung der sinnvollen Lesewege orientierte sich vornehmlich an inhaltlichen Aspekten des Textes. So wurde beispielsweise ein Sprung zu einer Abbildung, die nicht zum Inhalt des gerade gelesenen Textabschnittes passt, als „nicht-sinnvoll“ kodiert, hingegen das Betrachten einer Abbildung mit passendem Inhalt zu vorher gelesenen Textabschnitt als sinnvoller Leseweg. Als weitere nicht-sinnvolle Lesewege galten, wenn z.B. so in den Text eingesetzt wurde, dass wesentliche Textabschnitte zum Textverständnis ausgelassen wurden oder wenn aufgrund der Spaltenstruktur der korrekte Spaltenwechsel nicht gelang (und dadurch für das Verständnis wichtige Textabschnitte oder Bilder ausgelassen wurden). Diese nicht-sinnvollen Lesewege betrafen ausschließlich die Ebene der lokalen Kohärenzbildung. Auf der Ebene der globalen Kohärenzbildung wurden auch Lesewege als sinnvoll kodiert, wenn Textteile nacheinander gelesen oder betrachtet wurden, bei denen sich die Inhalte unterschieden (z.B. Sprünge zwischen Abschnitten, Bildern, Unterkapiteln, Titeln). Jegliche Zweifelsfälle wurden durchgängig als sinnvolle Lesewege kodiert, damit individuellen Lesegewohnheiten Raum gegeben wird.

Die Kodierungen auf den Segmenten wurden ausgezählt und deskriptiv statistisch dargestellt, um über diesen Weg Hinweise auf Textunterschiede zu bekommen. Für die statistischen Berechnungen zu den Textvergleichen und die tabellarischen Gegenüberstellungen wurden Textlänge und Anzahl der Bildflächen über beide Texte so gewichtet, dass der Originaltext in Bezug auf den überarbeiteten Text normiert wurde. Dazu wurden verschiedene Faktoren berechnet, die in die statistischen Berechnungen miteingehen, um die textversionsspezifischen Unterschiede in Bezug auf Textlänge und Anzahl der Abbildungen und Absätze auszugleichen (vgl. Tabelle 1).

| | Originaltext | überarbeiteter Text | Faktor zur Angleichung |
|---|--------------|---------------------|------------------------|
| Anzahl Bildflächen | 5 | 4 | 0,8 |
| Anzahl Einzelbilder | 6 | 6 | 1,0 |
| Anzahl Zeichen | 2926 | 4416 | 1,51 |
| Anzahl Wörter | 490 | 809 | 1,65 |
| Anzahl Sätze | 44 | 84 | 1,91 |
| Anzahl Absätze | 11 | 16 | 1,45 |
| Anzahl Textteile insgesamt (Absätze und Bildflächen) | 16 | 20 | 1,25 |

Tab.1: Textlängenangleichung des Originaltextes an den überarbeiteten Text

4 | Ergebnisse

Lesewege und Bezüge zwischen Textteilen

Insgesamt konnten beim Originaltext $n(\text{gesamt}) = 132$ Bezüge zwischen Textteilen und beim überarbeiteten Text $n(\text{gesamt}) = 297$ Bezüge zwischen Textteilen beobachtet werden. Da der Originaltext weniger Textteile aufweist als der überarbeitete Text, wurde die Anzahl der Bezüge an die Textlänge des überarbeiteten Textes mit dem Faktor 1,25 angepasst (vgl. Tabelle 2), so dass man $n(\text{gesamt_angepasst}) = 165$ Bezüge beim Originaltext erhält. Vergleicht man die Leserichtungen der Versuchspersonen, so lässt sich kein großer Unterschied zwischen den beiden Texten erkennen. Die Leser*innen beider Texte lesen bevorzugt linear vorwärts, wobei alle möglichen Leserichtungen, wenn zum Teil auch selten, ebenfalls eingeschlagen werden (vgl. Fehler! Verweisquelle konnte nicht gefunden werden.).

| Leserichtung | Originaltext | überarbeiteter Text |
|-------------------|--------------|---------------------|
| | Anzahl (%) | Anzahl (%) |
| linear vor | 101 (61%) | 194 (65%) |
| linear rück | 5 (3%) | 32 (11%) |
| nicht-linear vor | 39 (23%) | 40 (13%) |
| nicht-linear rück | 20 (12%) | 31 (10%) |
| gesamt | 165 (100%) | 297 (100%) |

Tab. 2: Verteilung der Leserichtungen

Ein anderes Bild zeigt sich jedoch, wenn man sich die Art der Herstellung von Bezügen anschaut, die zwischen Textteilen gemacht werden. Beim Originaltext werden insgesamt $n(\text{Text-Text}) = 70$ Bezüge zwischen Textabschnitten, $n(\text{Text-Bild}) = 27$ Bezüge zwischen Text und Bild und $n(\text{Bild-Bild}) = 35$ Bezüge zwischen Bildern hergestellt, beim überarbeiteten Text sind das $n(\text{Text-Text}) = 87$ Bezüge zwischen Textabschnitten, $n(\text{Text-Bild}) = 161$ Bezüge zwischen Text und Bild und $n(\text{Bild-Bild}) = 49$ Bezüge zwischen Bildern. Da der überarbeitete Text mehr Textabschnitte, aber weniger Bildflächen enthält, und damit die Wahrscheinlichkeit beim Lesen des Originaltextes einen Text-Text-Bezug zu machen kleiner, für einen Text-Bild-Bezug aber grösser ist, wird die Gesamtzahl der Text-Text-Bezüge beim Originaltext mit dem Faktor 1,45 erweitert, die Text-Bild-Bezüge jedoch um den Faktor 0,8 vermindert (vgl. Abbildung 1). Ein Vergleich der beiden Textversionen zeigt deutlich, dass beim Originaltext vor allem Bezüge zwischen Fließtextabschnitten hergestellt werden und nicht so sehr zu den Bildern. Beim überarbeiteten Text hingegen geschehen mehr als die Hälfte aller Bezüge zwischen Text und Bildern (vgl. Abbildung 1). Diese Unterschiede sind signifikant ($\text{Chi}^2(2, n = 501) = 88,945, p < .001$).

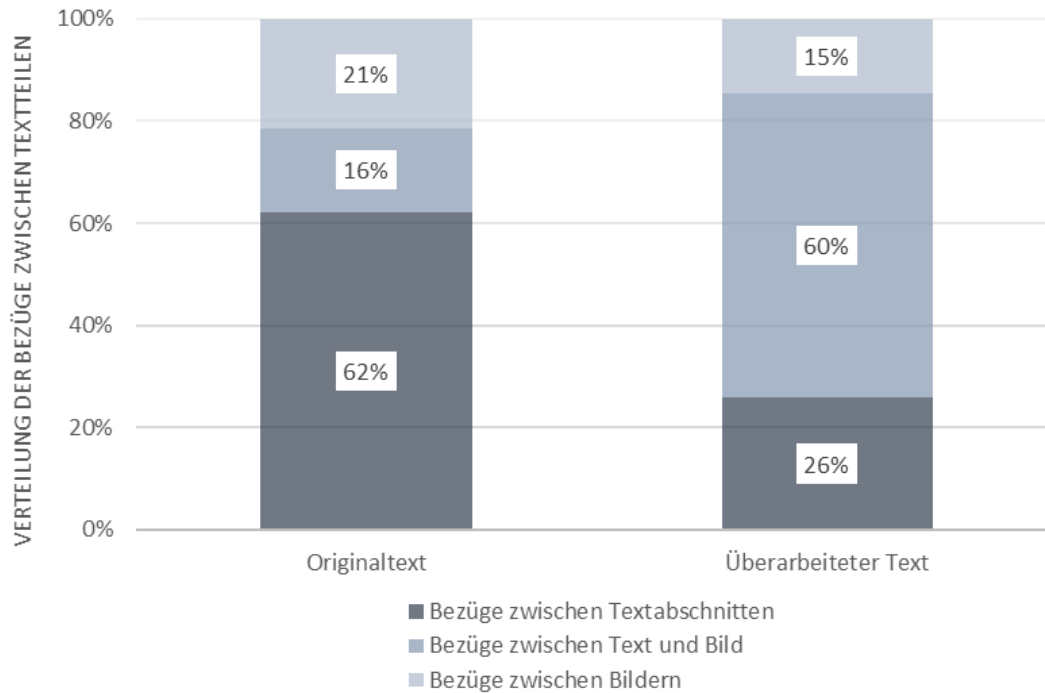


Abb.1: Prozentuale Verteilung der Bezüge zwischen Textteilen und Abbildungen

Dabei muss zusätzlich im Blick behalten werden, dass auch die Gesamtanzahl aller Bezüge, die beim Lesen des Originaltextes gemacht werden, viel kleiner ist als beim Lesen des überarbeiteten Textes (vgl. Abbildung 1).

Dieser Unterschied im Leseverhalten der Schüler*innen zwischen den beiden Textversionen wird auch deutlich, wenn man die durchschnittliche kumulative Verweildauer betrachtet, also wie lange sich jede Versuchsperson mit dem Fließtext bzw. mit den Bildern beschäftigt (vgl. Abbildung 2). Setzt man die Verweildauer beim Text und bei den Bildern miteinander ins Verhältnis, dann zeigt sich, dass sich die Versuchspersonen beim Originaltext im Durchschnitt 14 Prozent ihrer Lesezeit mit den Bildern beschäftigen und 86 Prozent mit dem Fließtext. Damit beschäftigen sie sich ca. siebenmal länger mit dem Fließtext als mit den Bildern. Beim überarbeiteten Text verweilen die Probanden 28 Prozent ihrer Lesezeit bei den Bildern und 72 Prozent beim Fließtext; das ist nur in etwa dreieinhalbmal länger. Der Unterschied vom Verhältnis der durchschnittlichen kumulativen Verweildauer bei Fließtext und Bild ist zwischen den Textversionen signifikant mit großer Effektstärke ($t = -2,813$; $df = 14$; $p = .016$; $d = 1,44$).

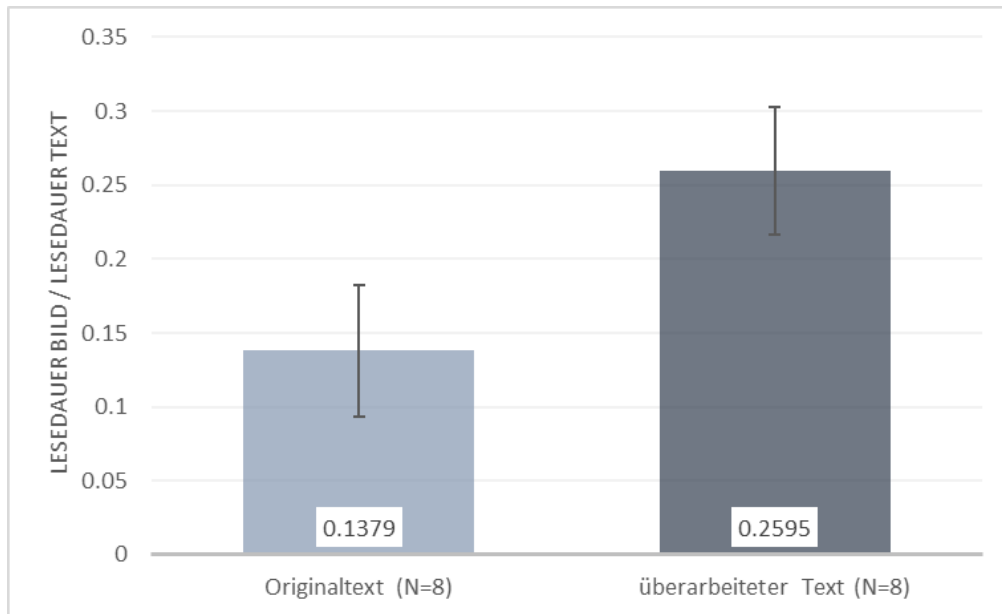


Abb.2: Verhältnis der durchschnittlichen kumulativen Verweildauer bei Bild und Text

Dass die Bilder beim überarbeiteten Text häufiger zum Leseprozess hinzugezogen werden, zeigt sich auch dann, wenn man sich die Anzahl der Bildbetrachtungen anschaut. Beim Originaltext werden während des Leseprozesses die Bilder meist nur einmal bis gar nicht hinzugezogen, beim überarbeiteten Text deutlich öfter (vgl. Abbildung 3).

Es stellt sich nun zusätzlich noch die Frage, wann sich die Schüler*innen während des Lesens entscheiden, zu einem Bild zu gehen. Abbildung 4 zeigt die Bezüge aus dem Text heraus zu einem Bild mit gelben Pfeilen markiert, graue Pfeile stellen Bild-Bild-Bezüge dar. Man erkennt, dass die Versuchspersonen, die den Originaltext gelesen haben, hauptsächlich Bild-Bild-Bezüge machen, wobei die Bilder unabhängig vom Text in der Reihe gelesen werden. Beim überarbeiteten Text werden wesentlich häufiger Text-Bild-Bezüge gemacht.

Stoffwechsel des Menschen

1 Atemwege. A (Übersicht); B (Bronchien und Lungenflügel); C (Endbronchien mit Lungenbläschen); D Schleimhaut mit Flimmerepithelien

3 Atmung

3.1 Wie wir atmen

Beim Atmen strömt die Luft durch die beiden Nasenhöhlen in ein verzweigtes System von Nasenmuscheln und Nebenhöhlen, die in unseren hohlen Oberkiefer- und Stirnhöhlen liegen. Die Wände dieser Höhlen sind mit einer leuchtigen Schleimhaut ausgekleidet. In der Schleimhaut der Nasenmuscheln liegen die Kiemenzotten. Die Schleimhaut wärmt die Einatemluft auf Körpertemperatur vor und befeuchtet sie. Außerdem bleiben an ihrem wässrigen Schleim Staub, Bakterien und Viren haften. Größere Verunreinigungen werden von Haaren im Nasenraum abgefangen.

Über den Rachen, wo sich Nasen- und Mundraum vereinigen, gelangt die Luft zum Kehlkopf. Er trennt Speiseröhre und Luftröhre. Die Luftröhre teilt sich in die beiden Hauptbronchien. Jede versorgt einen der

Flimmerepithelien zum Kehlkopf transportiert, wo sie hinuntergeschluckt werden. Im Magen tötet die Magensäure alle Krankheitserreger ab. Die Atemkanälchen enden in Trauben aus winzigen Lungenbläschen. Sie besitzen Wände, die so dünn sind wie die von Seifenblasen. Durch sie werden Sauerstoff und Kohlenstoffdioxid ausgetauscht. Außen ist die Lunge vom Lungenfell umhüllt. Zwischen ihm und dem Rippenfell, das den Brustkorb auskleidet, befindet sich ein flüssigkeitsgefüllter Spalt. Dadurch kann sich die Lunge beim Atmen im Brustraum verschieben.

Bei der Atmung wird Luft abwechselnd eingeatmet und ausgeatmet. Dies bewirkt vor allem das **Zwerchfell**, eine dünne Muskelmembran, die unterhalb der Lunge

Stoffwechsel des Menschen

quer durch den Bauchraum gespannt ist. Es trennt den Brustkorb mit Lunge und Herz vom Bauchraum mit Magen, Darm und anderen Organen. Ist das Zwerchfell entspannt, bildet es eine Kuppel, die vom unteren Ende der Rippen bis hoch in die Herzgegend reicht. Der Innenraum des Brustkorbs, den die Lunge ausfüllt, ist dann klein. Dies ist der Zustand beim Ausatmen. Spannt sich das Zwerchfell, wird es flach und das Volumen des Brustkorbs vergrößert sich. Dabei entsteht ein Unterdruck und wir atmen ein. Gleichzeitig werden die Bauchorgane nach unten gedrückt und die Bauchwand wölbt sich vor. Deshalb nennt man diese Form der Atmung auch **Bauchatmung**. Erschlafft das Zwerchfell, verkleinern sich Brustraum und Lungenvolumen. Dabei erhöht sich der Druck und wir atmen wieder aus.

Solche Atembewegungen entstehen auch durch Ausdehnung des Brustkorbs. Bei dieser **Brustatmung** heben Zwischenrippenmuskeln das Brustbein und die Rippen an. Dadurch dehnt sich der Brustkorb nach vorne und zur Seite. Er vergrößert sein Volumen und Luft strömt ein. Erschlaffen die Zwischenrippenmuskeln wieder, verringert sich das Volumen des Brustkorbs. Wir atmen aus.

Bei großen Anstrengungen verstärken alle Muskeln des Oberkörpers die Atembewegungen. Das trifft auch für das Husten und Niesen zu, mit dem wir Fremdkörper aus den Atemwegen entfernen.

2 Atmung. A **Bauchatmung**; B **Brustatmung**

Die Atemluft gelangt durch Nase, Rachen, Kehlkopf, Luftröhre und Bronchien in die Lunge. Zwerchfell und Zwischenrippenmuskeln bewirken durch Vergrößerung und Verkleinerung des Lungenvolumens die Atembewegungen.

- 1 Beschreibe den Weg der Atemluft.
- 2 Begründe den Ratschlag, durch die Nase statt durch den Mund einzatmen.
- 3 Sitzende Arbeitshaltung kann dazu führen, dass Teile der Lunge nicht ausreichend belüftet werden. Begründe, warum man im Sitzen nur flach atmet.

3

ast der Lungenvene

Lungenbläschen

3

Knoorpel

Drüsenzelle

Flimmerepithelien

Flimmerepithelien

Die Atmung

Wie die in diesem Text lernt

Bei der Atmung wird Luft in die Lunge transportiert und wieder aus der Lunge ausgesaugt. Sind die Atmung bei uns, den Körper mit dem lebenswichtigen Sauerstoff zu versorgen. Die Lunge ist das wichtigste Atemorgan. In diesem Text lernst du, welchen Weg die Luft zur Lunge nimmt. Du erfährst auch, wie die Luft in den Atemwegen gereinigt wird. Zudem lernst du zwei Atmungsbewegungen kennen, die die Atmung ermöglichen: die Bauchatmung und die Brustatmung.

Reinigung der Luft in den Atemwegen

Die Luft wird auf dem Weg von der Nase zur Lunge gereinigt. Die erste Reinigungsstufe ist in der Nase. Hier liegen die Nasenhöhle, größere Verästelungen wie kleine Verästelungen wie Staub, Viren oder Bakterien werden mithilfe von Schleimzellen abgefangen. Alle Atemwege von der Nase bis zu den Bronchien haben Schleimhäute. Diese sind feucht und klebrig. Die klebrigen Verunreinigungen werden mit dem Schleim abgefangen. Die Schleimhäute haben winzige Härchen. Diese Härchen wippen aus Flimmerepithelien. Die Flimmerepithelien bewegen sich ständig und transportieren so die Verunreinigungen, Bakterien und Viren zum Kehlkopf. Dort werden sie hinuntergeschluckt und gelangen in den Magen. Im Magen werden die Bakterien und Viren von der Magensäure abgetötet.

Die Schleimhäute in der Nase fangen Verunreinigungen ab und streifen und belutschen die Luft. Es ist deshalb ratsam, durch die Nase zu atmen und nicht durch den Mund. Wenn man durch die Nase atmet, kommt gereinigte und gewärmte Luft in die Lunge.

Zwei Atmungsarten

Es gibt zwei Arten von Atembewegungen, die Bauchatmung (Weite C1 und C2) und die Brustatmung (Weite D1 und D2). Beide Atembewegungen werden durch Muskeln bewirkt.

Die Bauchatmung:
Das Zwerchfell ist die Muskel, die die Brustatmung bewirkt. Muskeln können angespannt oder entspannt sein. Im angespannten Zustand ist das Zwerchfell flach (Weite C1), im entspannten Zustand ist es nach oben gewölbt (Weite C2).

Die Brustatmung:
Das Zwerchfell ist die Muskel, die die Brustatmung bewirkt. Muskeln können angespannt oder entspannt sein. Im angespannten Zustand ist das Zwerchfell flach (Weite D1), im entspannten Zustand ist es nach oben gewölbt (Weite D2).

Die Atemwege

Über die Atemwege gelangt die Luft in die Lunge, die aus zwei Lungenflügeln besteht.

1. Nasenhöhle, 2. Nasenmuschel, 3. Rachenraum, 4. Kehlkopf, 5. Luftröhre, 6. Bronchien, 7. Lungenflügel

Der Mensch atmet die Luft durch die Nase (1) oder den Mund (2) ein. Die Luft gelangt über den Rachenraum (3) zum Kehlkopf (4). Beim Schlucken trennt sich die Speiseröhre und die Luftröhre (5). Die Luft gelangt in die Luftröhre. Die Luftröhre teilt sich in zwei Äste (6). Diese Äste nennt man Bronchien (6). Durch die Luftröhre und die Bronchien beim Atmen nicht verschleimen, haben sie Ringe aus Knorpel. Diese Ringe nennt man Knorpelringe (7) und Knorpel ist hart aber doch flexibel. Die Knorpelringe haben die Luftröhre und die Bronchien offen. Die Luft gelangt durch die Luftröhre und die Bronchien in die Lunge. In der Lunge teilen sich die Bronchien in immer kleinere Bronchien. Am Ende der Bronchien befinden sich winzige Lungenbläschen (8). Durch die dünne Haut der Lungenbläschen gelangt Sauerstoff in das Blut. Wie dies funktioniert, erfährst du in einem anderen Text.

Die Atembewegungen

Die Atembewegungen ermöglichen, dass die Luft in die Lunge transportiert und wieder ausgesaugt wird. In diesem Abschnitt wird zuerst beschrieben, wie die Organe aufgebaut sind, die für die Atembewegungen wichtig sind. Danach werden die Atembewegungen erklärt.

1. Luftröhre, 2. Rippen, 3. Zwischenrippenmuskeln, 4. Zwerchfell, 5. Lungenflügel, 6. Rippenfell

In dem Brustraum (Weite B) befinden sich das Herz und die Lunge (9). Die Brustraum ist von schmalen Knochen umschlossen, dem Rippen (10). Zwischen dem Rippen hat ein Muskel. Diese nennt man Zwischenrippenmuskeln (10). Rippen und Zwischenrippenmuskeln bilden eine Art Korb um die Lunge. Man nennt ihn deshalb Brustkorb. Unter dem Brustkorb befindet sich das Zwerchfell (11). Das Zwerchfell trennt den Brustraum vom Bauchraum. In ihm steckt Muskel, der wie eine Haut aussieht.

Die Bauchatmung

Zwischen dem Brustkorb und der Lunge hat es zwei flache Hautschichten: das Lungenfell (1) und das Rippenfell (2). Das Lungenfell umhüllt die Lunge. Das Rippenfell liegt an der Lungenfell (3) und ist mit einem flüssigen Stoff bedeckt. Zwischen dem Lungenfell und dem Rippenfell befindet sich Flüssigkeit. Diese Flüssigkeit bewirkt, dass die Lunge flach und sich trotzdem verschieben können. So kann die Lunge den Bewegungen der Rippen folgen. Wie die Bewegungen beim Atmen funktionieren, wird es nächsten Abschnitt beschrieben.

Zwei Atmungsarten

Es gibt zwei Arten von Atembewegungen, die Bauchatmung (Weite C1 und C2) und die Brustatmung (Weite D1 und D2). Beide Atembewegungen werden durch Muskeln bewirkt.

Die Bauchatmung:
Das Zwerchfell ist die Muskel, die die Brustatmung bewirkt. Muskeln können angespannt oder entspannt sein. Im angespannten Zustand ist das Zwerchfell flach (Weite C1), im entspannten Zustand ist es nach oben gewölbt (Weite C2).

Die Brustatmung:
Das Zwerchfell ist die Muskel, die die Brustatmung bewirkt. Muskeln können angespannt oder entspannt sein. Im angespannten Zustand ist das Zwerchfell flach (Weite D1), im entspannten Zustand ist es nach oben gewölbt (Weite D2).

Fragebogen

Lies noch einmal den Absatz 'Reinigung der Luft in den Atemwegen' und bearbeite die folgenden Fragen auf dem Blatt.

1. Nenne alle richtigen Antworten.
 - a) Die Schleimhäute sind feucht und klebrig.
 - b) Die Schleimhäute transportieren die Luft.
 - c) Die Schleimhäute reinigen die Luft.
 - d) Die Schleimhäute bilden die Luft.
2. Was wird von dem Schleimhaut abgefangen?
 - a) Staub
 - b) Viren
 - c) Bakterien
 - d) Alle oben genannten
3. Wo gibt es in der Atemwege Schleimhaut?
 - a) Nur in der Lunge
 - b) Nur in der Nase
 - c) In der Nase und in der Lunge
 - d) In der Nase und im Rachenraum
4. Fasse in 3-4 Sätzen zusammen, welche Funktionen die Schleimhäute bei der Atmung haben.

Abb.3: Anzahl der Bildbetrachtungen pro Versuchsperson (N=16) beim Originaltext (oben) und überarbeiteten Text (unten); jeder Kreis stellt eine Versuchsperson dar, die Zahl im Kreis zeigt, wie oft diese Versuchsperson die Abbildung während des Leseprozesses hinzugezogen hat.

Stoffwechsel des Menschen

3 Atmung

3.1 Wie wir atmen

Beim Atmen strömt die Luft durch die beiden Nasenlöcher in ein verzweigtes System von Nasenschleimhäuten und Nebenhöhlen, die in unseren hohlen Oberkiefer- und Stirnknochen liegen. Die Wände dieser Höhlen sind mit einer feuchten Schleimhaut ausgekleidet. In der Schleimhaut der Nasenschleimhäute liegen die Kiemenzotten. Die Schleimhaut wärmt die Einatemluft auf Körpertemperatur vor und befeuchtet sie. Außerdem bleiben an ihrem wässrigen Schleim Staub, Bakterien und Viren haften. Größere Verunreinigungen werden von Haaren im Nasenraum abgefangen.

Über dem Rachen, wo sich Nasen- und Mundraum vereinigen, gelangt die Luft zum Kehlkopf. Er trennt die Speiseröhre und Luftröhre. Die Luftröhre teilt sich in die beiden Hauptbronchien. Jede versorgt einen der Lungenflügel.

Flimmerhärchen zum Kehlkopf transportiert, wo sie hinuntergeschluckt werden. Im Magen tötet die Magensäure alle Krankheitserreger ab. Die Atemkanäle enden in Trauben aus winzigen Lungenbläschen. Sie besitzen Wände, die so dünn sind wie die von Seifenbläschen. Durch sie werden Sauerstoff und Kohlendioxid ausgetauscht. Außen ist die Lunge vom Lungenfell umhüllt. Zwischen ihm und dem Rippenfell, das den Brustkorb auskleidet, befindet sich ein flüssigkeitsgefüllter Spalt. Dadurch kann sich die Lunge beim Atmen im Brustraum verschieben.

Bei der Atmung wird Luft abwechselnd eingeatmet und ausgeatmet. Dies bewirkt vor allem das Zwerchfell, eine dünne Muskelmembran, die unterhalb der Lunge quer durch den Bauchraum gespannt ist. Es trennt den Brustkorb mit Lunge und Herz vom Bauchraum mit Magen, Darm und anderen Organen. Ist das Zwerchfell entspannt, bildet es eine Kugel, die vom unteren Ende der Rippen bis hoch in die Herzgegend reicht. Der Innenraum des Brustkorbs, den die Lunge ausfüllt, ist dann klein. Dies ist der Zustand beim Ausatmen. Spannt sich das Zwerchfell, wird es flach und das Volumen des Brustkorbs vergrößert sich. Dabei entsteht ein Unterdruck und wir atmen ein. Gleichzeitig werden die Bauchorgane nach unten gedrängt und die Bauchwand wölbt sich vor. Deshalb nennt man diese Form der Atmung auch Bauchatmung. Durch die Ausdehnung des Brustkorbs strömt die Luft durch die Bronchien in das Brustbein und die Rippen. Durch die Dehnung dehnt sich der Brustkorb nach vorne und nach hinten. Er vergrößert sein Volumen und saugt Luft ein. Erschlaffen die Zwischenrippenmuskeln wieder, verringert sich das Volumen des Brustkorbs: Wir atmen aus. Bei großen Anstrengungen verstärken die Muskeln des Oberkörpers die Atembewegungen. Dadurch auch für das Husten und Niesen zum Abwurf von Fremdkörpern aus den Atemwegen.

194

Stoffwechsel des Menschen

3.1 Wie wir atmen

Zwischen dem Brustkorb und der Lunge hat es zwei dünne Hautschichten: das Lungenfell (1) und das Rippenfell (2). Das Lungenfell umhüllt die Lunge. Das Rippenfell liegt an der Lungenfellwand und an den Rippen. Zwischen dem Lungenfell und dem Rippenfell befindet sich Flüssigkeit. Diese Flüssigkeit bewirkt, dass die Lunge und das Rippenfell aneinander kleben und sich nicht voneinander verschieben können. So kann die Lunge den Bewegungen der Rippen folgen. Wie die Bewegungen beim Atmen funktionieren, wird im nächsten Abschnitt beschrieben.

Zwei Atemarten

Es gibt zwei Arten der Atmung: die Brustatmung (3) und die Bauchatmung (4). Die Brustatmung erfolgt durch die Brustmuskeln, die Bauchatmung durch die Bauchmuskeln.

Die Brustatmung: Die Brustmuskeln bewirken die Brustatmung. Sie heben und senken die Rippen.

Die Bauchatmung: Die Bauchmuskeln bewirken die Bauchatmung. Sie ziehen den Bauch nach unten.

Einatmen: Beim Einatmen werden die Zwischenrippenmuskeln mit der Brust zusammengezogen. Die Rippen nach vorne. Der Brustkorb wölbt sich dadurch nach oben. Die Bewegung des Brustkorbs und die Vergrößerung des Brustkorbs drückt die Luft in die vergrößerte Lunge. Wir atmen ein.

Ausatmen: Wenn sich die Zwischenrippenmuskeln entspannen, dann senken sich die Rippen wieder (5). Der Brustkorb wird kleiner und damit auch die Lunge. Dadurch wird die Luft aus der Lunge gedrückt und wir atmen wieder aus.

Fragen

Les auch einmal den Absatz 'Reinigung der Luft in den Atemwegen' und bearbeite die folgenden Fragen auf dem Blatt.

1. Nennen alle richtigen Antworten.
 - a) Die Schleimhäute wärmen und befeuchten die Luft.
 - b) Die Schleimhäute transportieren die Luft.
 - c) Die Schleimhäute reinigen die Luft.
 - d) Die Schleimhäute bilden die Luft.
2. Was wird von den Schleimhäuten abgehalten? Zähle drei Dinge auf.
3. Wo gibt es in den Atemwegen Schleimhäute? Zähle drei Stellen auf.
4. Fasse in 2-3 Sätzen zusammen, welche Funktion die Schleimhäute bei der Atmung haben.

195

Die Atmung

Was ist in diesem Text neu?

Bei der Atmung wird Luft in die Lunge transportiert und wieder aus der Lunge ausgeatmet. Was die Atmung ist, was die Rippen mit dem lebenswichtigen Sauerstoff zu tun haben, die Lunge ist das wichtigste Atemorgan. In diesem Text erfährst du, welchen Weg die Luft zur Lunge nimmt. Zudem lernst du zwei Muskelbewegungen kennen, die die Atmung ermöglichen: die Brustatmung und die Bauchatmung.

Die Atemwege

Über die Atemwege gelangt die Luft in die Lunge, die aus zwei Lungenflügel besteht.

Die Atembewegungen

Die Atembewegungen ermöglichen, dass die Luft in die Lunge strömt und wieder ausgeatmet wird. In diesem Abschnitt lernst du, wie die Organe aufgebaut sind, die die Atembewegungen wichtig sind. Danach werden die Atembewegungen beschrieben.

Reinigung der Luft in den Atemwegen

Die Luft wird auf dem Weg von der Nase zur Lunge gereinigt. Die erste Reinigung findet in der Nase statt. Hier fangen die Nasenschleimhäute grobere Verunreinigungen ab. Kleineren Verunreinigungen wie Staub, Viren oder Bakterien werden mithilfe von körpereigenen Abwehrstoffen abgefangen. Alle Anstriche von der Nase bis zu den Bronchien haben Schleimhäute. Diese sind feucht und klebrig. Die klebrigen Verunreinigungen bleiben an den Schleimhäuten kleben. Die Schleimhäute haben winzige Härchen. Diese Härchen reiben nun alle Verunreinigungen ab. Die Härchen bewegen sich ständig und transportieren so die Verunreinigungen, Bakterien und Viren zum Kehlkopf. Dort werden sie hinuntergeschluckt und gehen in den Magen. Im Magen werden die Bakterien und Viren durch die Magensäure abgetötet.

Die Schleimhäute in der Nase transportieren die ab- und wärmere und befeuchtete Luft in die Lunge. Dies geschieht durch die Nase zu den Bronchien. Die Luft ist dadurch gereinigt und gewärmt.

Zwischen dem Brustkorb und der Lunge hat es zwei dünne Hautschichten: das Lungenfell (1) und das Rippenfell (2). Das Lungenfell umhüllt die Lunge. Das Rippenfell liegt an der Lungenfellwand und an den Rippen. Zwischen dem Lungenfell und dem Rippenfell befindet sich Flüssigkeit. Diese Flüssigkeit bewirkt, dass die Lunge und das Rippenfell aneinander kleben und sich nicht voneinander verschieben können. So kann die Lunge den Bewegungen der Rippen folgen. Wie die Bewegungen beim Atmen funktionieren, wird im nächsten Abschnitt beschrieben.

Zwei Atemarten

Es gibt zwei Arten der Atmung: die Brustatmung (3) und die Bauchatmung (4). Die Brustatmung erfolgt durch die Brustmuskeln, die Bauchatmung durch die Bauchmuskeln.

Die Brustatmung: Die Brustmuskeln bewirken die Brustatmung. Sie heben und senken die Rippen.

Die Bauchatmung: Die Bauchmuskeln bewirken die Bauchatmung. Sie ziehen den Bauch nach unten.

Einatmen: Beim Einatmen werden die Zwischenrippenmuskeln mit der Brust zusammengezogen. Die Rippen nach vorne. Der Brustkorb wölbt sich dadurch nach oben. Die Bewegung des Brustkorbs und die Vergrößerung des Brustkorbs drückt die Luft in die vergrößerte Lunge. Wir atmen ein.

Ausatmen: Wenn sich die Zwischenrippenmuskeln entspannen, dann senken sich die Rippen wieder (5). Der Brustkorb wird kleiner und damit auch die Lunge. Dadurch wird die Luft aus der Lunge gedrückt und wir atmen wieder aus.

Fragen

Les auch einmal den Absatz 'Reinigung der Luft in den Atemwegen' und bearbeite die folgenden Fragen auf dem Blatt.

1. Nennen alle richtigen Antworten.
 - a) Die Schleimhäute wärmen und befeuchten die Luft.
 - b) Die Schleimhäute transportieren die Luft.
 - c) Die Schleimhäute reinigen die Luft.
 - d) Die Schleimhäute bilden die Luft.
2. Was wird von den Schleimhäuten abgehalten? Zähle drei Dinge auf.
3. Wo gibt es in den Atemwegen Schleimhäute? Zähle drei Stellen auf.
4. Fasse in 2-3 Sätzen zusammen, welche Funktion die Schleimhäute bei der Atmung haben.

Abb.4: Lesewege zu den Abbildungen (Originaltext oben, überarbeiteter Text unten) mit Anzahl der Versuchspersonen (Nummerierung und Dicke des Pfeils), die diesen Leseweg eingeschlagen haben. Graue Pfeile zeigen Bild-Bild-Bezüge, gelbe Pfeile Text-Bild-Bezüge.

Jedoch sind höchstwahrscheinlich nicht alle Bezüge zwischen den Textteilen, die von den Schüler*innen gemacht werden, auch als sinnvoll für den Leseprozess anzusehen. So ist es beispielsweise für den Leseprozess und das Textverstehen nicht förderlich, wenn Lesende beim Originaltext vom Ende des Fließtextes her zum Bild 1C gehen, da in den letzten Absätzen des Textes die Atembewegungen Bauch- und Brustatmung beschrieben werden, das Bild jedoch den Aufbau der Lungenbläschen mit Endbronchien und Blutversorgung darstellt. Trotzdem schlugen 4 Versuchspersonen (50%) genau diesen Leseweg aufgrund der Anordnung von Text und Bild ein. Sinnvoller wäre es hier, sich die Bilder zur Bauch- und Brustatmung oben rechts anzuschauen. Diese Verknüpfung gelingt aber nur einer Sek-A Schülerin (vgl. dazu Abbildung 4).

Auch beim überarbeiteten Text passieren nicht alle Bezüge, die zwischen Textteilen gemacht werden, an sinnvoller Stelle für das Textverstehen, jedoch die meisten. Abbildung 4 zeigt in Bezug auf sinnvolle Lesewege deutliche Unterschiede zwischen den Textversionen. Diese Unterschiede sind immer dann signifikant, wenn Bilder in den Leseprozess mit einbezogen werden.

| | Originaltext | | | überarbeiteter Text | | | Unterschied Chi-Quadrat nach Pearson |
|----------------------|-----------------------|----------------------------------|--------------|-----------------------|--------------------------------|--------------|--|
| | sinnvolle Lesewege | nicht sinn- volle Lesewege | gesamt | sinnvolle Lesewege | nicht sinnvolle Lesewege | gesamt | |
| Text-Bild- Bezug | 5 (36%) | 9 (64%) | 14 (100%) | 76 (94%) | 5 (6%) | 81 (100%) | Chi ² (1, n=95) = 32.083 p <.001 |
| Bild-Text- Bezug | 8 (62%) | 5 (38%) | 13 (100%) | 76 (95%) | 4 (5%) | 80 (100%) | Chi ² (1, n=93) = 14.325 p <.001 |
| Bild-Bild-Be- zug | 29 (83%) | 6 (17%) | 35 (100%) | 47 (96%) | 2 (4%) | 49 (100%) | Chi ² (1, n=84) = 4.042 p <.044 |
| Text-Text- Bezug | 66 (94%) | 4 (6%) | 70 (100%) | 80 (92%) | 7 (8%) | 87 (100%) | Chi ² (1, n=157) = 0.324 n.s. |

Tab. 3: Anzahl sinnvoller und nicht sinnvoller Textteilbezüge pro Textversion

Zusammenfassend sieht man, dass es Schüler*innen einer siebten Klasse mit dem überarbeiteten Text besser gelingt, die Bilder während des Textlesens zu integrieren: Sie verweilen länger bei den Bildern, ziehen sie häufiger in ihren Leseprozess mit ein und machen mehr sinnvolle Bezüge zwischen dem Text und den Bildern. Das zeigt, dass schon rein textseitige Maßnahmen, wie eine Anpassung der Lehrmitteltexte an eine strukturierte, auf der Textoberfläche offensichtliche Inhaltsorganisation (z.B. durch Absatzüberschriften), eine leseprozesslogische Text-Bild-Platzierung und die Bereitstellung von Bildverweisen im Fließtext Sekundarschüler*innen der siebten Klasse einen hilfreichen Rahmen bieten, um ihren Leseprozess erfolgreich zu steuern. Inwieweit dieses auch einen Einfluss auf das Textverstehen hat, soll im Folgenden anhand von Daten zum Konzeptverständnis aufgezeigt werden.

Konzeptverstehen

Die Daten zum Konzeptverstehen werden hier anhand zweier Beispielkonzepte „Aufbau der Atemwege“ und „Aufbau des Brustraumes“ aufgezeigt und interpretiert. Beide hier vorgestellten Konzepte sind in beiden Textversionen sowohl sprachlich als auch visuell repräsentiert (vgl. Abbildung 5). Als Unterschied zwischen den Textversionen fällt jedoch auf, dass beim überarbeiteten Text Fließtext und Bilder zum jeweiligen Konzept räumliche Nähe und eine inhaltlich kohärente Struktur aufweisen, während das beim Originaltext nicht der Fall ist (für eine detaillierte Analyse der Texte in Bezug auf Kohärenz in der Inhaltsorganisation siehe Dittmar et al. 2017).

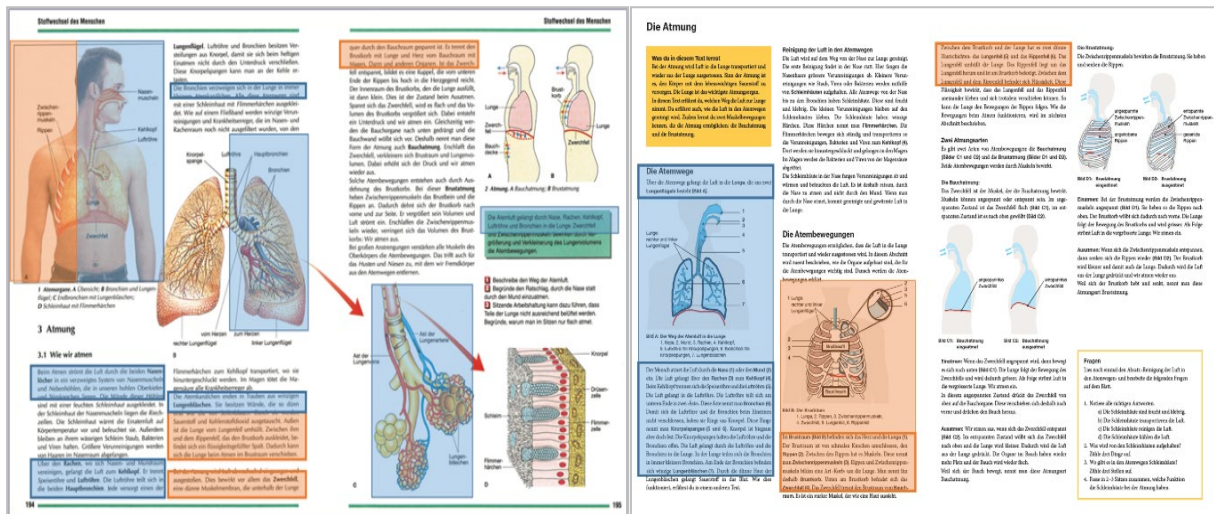


Abb. 5: Verteilung der Informationen zu den Konzepten Atemwege (blau) und Brustraum (orange). Links der Originaltext, rechts der überarbeitete Text.

Das Konzeptverstehen der Schüler*innen wurde anhand ihrer Performanz in der Fragesequenz bestimmt (siehe Abschnitt Methoden). Inwieweit Fachkonzepte verstanden oder nicht verstanden wurden, lässt sich von den Antworten der Schüler*innen während der Fragesequenz ableiten, indem den Versuchspersonen aufgrund ihrer Antworten problematische und unproblematische Ereignisse zugeteilt wurden. Tabelle 4 zeigt einen Vergleich der problematischen Ereignisse mit den unproblematischen Ereignissen bei Fragen zu den Wissenskonzepten „Atemwege“ und „Aufbau Brustraum“ (für eine Analyse der anderen im Text enthaltenen Wissenskonzepte siehe Dittmar et al. 2017).

Man erkennt, dass der Anteil der problematischen Ereignisse beim Originaltext doch sehr groß ist, d.h., ein ausreichendes Verständnis der Konzepte „Atemwege“ und „Brustraum“ gelingt mit diesem Lehrmitteltext zur Atmung nicht. Da auch die sinnvollen Lesewege mit diesem Text nicht gelingen, könnte es hier einen Zusammenhang geben. Diese Interpretation muss aber unter Vorbehalt gemacht werden, weil die Studie nicht so angelegt war, dass die Textanpassungen kontrolliert nur für die Organisation von Text und Bild durchgeführt worden sind. Es wurden ebenso andere Textanpassungsprinzipien, wie z.B. eine vereinfachte Syntax, explizitere Begriffsdefinitionen etc. vorgenommen, die ebenfalls das Konzeptverstehen beeinflussen können. Daher sind diese Daten nur als Indizien zu werten.

| Konzept | | problematische | unproblematische | gesamt |
|-----------|---------------------|----------------|------------------|------------|
| | | Ereignisse | Ereignisse | |
| | | Anzahl (%) | Anzahl (%) | Anzahl (%) |
| Atemwege | Originaltext | 63 (64%) | 36 (36%) | 99 (100%) |
| | überarbeiteter Text | 38 (28%) | 98 (72%) | 136 (100%) |
| Brustraum | Originaltext | 29 (78%) | 8 (22%) | 37 (100%) |
| | überarbeiteter Text | 16 (13%) | 107 (87%) | 123 (100%) |

Tab. 4: Anzahl der problematischen und unproblematischen Ereignisse während der Fragesequenz bei Fragen zu den Wissenskonzepten Atemwege und Brustraum

5 | Diskussion

Die Daten geben Hinweise darauf, dass Siebtklässler*innen Mühe damit haben, Bilder ausreichend in den Leseprozess mit einzubeziehen. Zum Teil ignorieren sie Abbildungen als Informationsquelle, machen problematische Text-Bild-Bezüge und/oder lesen die Bilder unabhängig vom Text. Je weniger ihr Leseprozess durch die Textstrukturierung gesteuert wird, desto schwerer fällt ihnen die Integration von Text und Bild sowie ein sinnvoller Leseweg durch den Text. Die Folge ist, dass sich das Konzeptverstehen aus dem Leseprozess nicht genügend aufbaut.

Natürlich stellt sich in Bezug auf die vorliegenden Daten die methodische Frage, ob der Auftrag, beim Lesen des Textes mit dem Finger nachzufahren, nicht den Leseprozess an sich dahingehend stört, dass ein lineares Lesen bevorzugt wird und ein Springen zu weit entfernten Abbildungen eher unterbunden wird, da es die kognitiven Ressourcen aufgrund der motorischen Extraleistung (Augen- evtl. Kopf- plus Finger-/Armbewegungen) zusätzlich belastet. Das würde einen Vorteil für den überarbeiteten Text bedeuten, da hier die Abbildungen näher an ihre korrespondierenden Textstellen gerückt wurden.

Mit dem Einsatz eines Eyetrackers hätte diese zusätzliche kognitive Belastung reduziert werden können. Ein Blick auf in der Beobachtung von Leseprozessen verorteten Eyetracking-Studien unterstützt jedoch die Ergebnisse der vorliegenden Studie. So zeigt zum Beispiel Jian (2016), dass die Augenbewegungen taiwanesischer Viertklässler*innen beim Lesen eines Biologielehrmitteltextes im Gegensatz zu Erwachsenen vor allem zwischen Textabsätzen und zwischen Bildern hin und her gingen, nicht jedoch zwischen Text und Bild. Ähnliches finden Hannus & Hyönä (1999), die zeigen konnten, dass das Leseverhalten finnisch sprechender Viertklässler*innen stark auf den Fließtext fokussiert und nur sehr wenig auf die Abbildungen. Auch in dieser Studie gab es Gruppenunterschiede dahingehend, dass Kinder mit höheren intellektuellen Fähigkeiten auch besser in der Lage waren, ihre visuelle Aufmerksamkeit zwischen Fließtext und Bildern aufzuteilen.

Hinweise darauf, dass die Integration von Abbildungen während des Leseprozesses vom Leseziel abhängt, gibt die Studie von Zhao et al. (2014). Sie konnten zeigen, dass Jugendliche Abbildungen vor allem dann nutzen, wenn Fragen zum Text beantwortet werden sollen und nicht so sehr beim unspezifischen Auftrag, den Text als Ganzes zu lesen.

Auch dass das Textverstehen beim Lesen naturwissenschaftlicher Texte davon abhängt, wie gut Text und Bilder während des Leseprozesses integriert werden, wird durch eine Eyetracking-

Studie unterstützt: So konnten Mason, Tornatora & Pluchino (2013) bei Italienisch sprechenden Viertklässler*innen zeigen, dass die Kinder, welche häufiger zwischen Fließtext und Abbildungen hin und her schauten sowie die Abbildungen länger betrachteten, auch besser beim Lernen von sowohl Fakten- als auch Transferwissen abschnitten.

Zusammenfassend kann man also sagen, dass ein unangeleitetes sinnvolles Lesen von Text-Bild-Kombinationen Kindern und jüngeren Jugendlichen (noch) nicht gelingt. Dagegen bewirken jedoch minimale Lesesteuerungselemente (wie etwa Abbildungsverweise im Text, räumlich nahe Text-Bild-Platzierung sowie gezielte Instruktionen in Form von Fragen), dass beim Leseprozess eine vermehrte Text-Bild-Integration abläuft.

Es macht also Sinn, den Leseprozess von Schüler*innen auch in den höheren Klassenstufen anzuleiten. Textanpassungen in Bezug auf die Inhaltsorganisation und Abbildungsverweise sind da ein erster Schritt, doch ihre Umsetzung in Form von Lehrmittelerneuerungen sind sehr langwierig. Explizite Leseanleitungen hingegen können unabhängig von einer Lehrmittelneuaufgabe zu bestehenden Lehrmitteltexten entwickelt werden und sie organisieren zudem den Leseprozess noch stärker als die reinen Textanpassungen. Außerdem unterstützt die Nutzung von Leseanleitungen den Erwerb von Lesestrategien. Von daher wäre es sinnvoll, fachspezifische und stufenspezifische Leseprozesssteuerungen für diese komplexen Lehrmitteltexte zu entwickeln.

Wie schon in der Einleitung erwähnt, widmet sich der sprachbewusste Fachunterricht der Leseprozesssteuerung von schulischen Fachtexten (Lindauer et al. 2013), hat aber die spezifischen Schwierigkeiten, die Text-Bild-Kombinationen aufwerfen, noch nicht konsequent integriert. Eine Unterstützung der drei kognitiven Teilprozesse beim Lesen komplexer Sachtexte – Selektion, Organisation und Transformation – ist zwar in Bezug auf Organisationsprozesse angedeutet (Concept Map, Ablauf-, Struktur-, Prozessschema erstellen, Ober-/Unterbegriffe sammeln etc.), jedoch wird nicht beschrieben, wie diese kognitiven Prozesse systematisch und auch im Zusammenhang mit informationstragenden Bildern aktiviert werden können. Wie eine Leseanleitung zur Aktivierung der kognitiven Teilprozesse aussehen könnte, schlagen Schlag & Ploetzner (2011) vor. Sie zeigen in ihrer Studie, welche Wirksamkeit Instruktionen auf das Konzeptlernen mit naturwissenschaftlichen Texten haben, wenn diese Selektions- und Organisationsstrategien zusammen mit der Verknüpfung von Text und Bild anregen. Dazu ließen sie Mittelschüler*innen einer 6. Klasse einen Biologietext mit der folgenden Anleitung lesen:

- „Lies den ersten Satz vom Text und unterstreiche die Begriffe, die dir am wichtigsten erscheinen“ (Selektion Text).
- „Markiere dann die dazugehörigen Elemente in der Abbildung“ (Selektion Abbildung).
- „Nun beschrifte die markierten Elemente in der Abbildung mit den unterstrichenen Begriffen“ (Organisation Text-Bild).

Zusätzlich sollten die Schüler*innen den Text noch mit eigenen Worten zusammenfassen und eine Skizze über die wichtigsten Informationen aus dem Text zeichnen (Transformation). Im Gegensatz zu einer Kontrollgruppe, die nur den Auftrag erhielt, das Gelesene mit eigenen Worten zusammenzufassen, erreichte die Interventionsgruppe wesentlich höhere Nachwissenswerte und zwar in allen drei Wissensarten, namentlich im Faktenwissen, Konzeptwissen und Transferwissen (siehe auch Schlag 2011).

Um also noch einmal auf die vier Leseschritte des sprachbewussten Fachunterrichts zurückzukommen, sollten folgende Ergänzungen beim Lesen informationsdichter Texte in Betracht gezogen werden:

- Im Leseschritt *eins*, wenn es darum geht, dem Text zu begegnen, einen Überblick über den Text zu generieren und das Vorwissen zu aktivieren, wäre es sicherlich von Bedeutung, den Fokus der Lesenden direkt auf Abbildungen mit Situierungsfunktion zu lenken, auf Abbildungen also, die Szenarien bieten, um den Text in einen inhaltlichen Rahmen zu stellen. Zusätzlich könnten auch schon in diesem ersten Leseschritt die Funktionen weiterer Abbildungen im Text geklärt werden, bevor das genaue Lesen gestartet wird. Zu schauen wäre hier also z.B., ob dekorative Bilder vorkommen und welche Abbildungen informierende Bilder mit darstellender (Zeigefunktion), organisierender (Konstruktionsfunktion) oder transformierender Funktion sind (Ballstaedt 2012; Weidenmann 2002). Dieses Vorgehen bereitet den kognitiven Prozess der Selektion vor, indem etwa rein dekorative Bilder aus dem Aufmerksamkeitsbereich genommen werden können (siehe auch *seductive details* von Garner, Brown, Sanders & Menke 1992; Harp & Mayer 1998).
- Im *zweiten* Leseschritt, wenn der Text auf lokaler Kohärenzebene bearbeitet wird, ist es wichtig, auch die Abbildungen als Informationsquellen im Blick zu behalten. Hier stellt sich für die Lesenden die Schwierigkeit, dass zur Informationsgenerierung das visuelle Argument der Abbildung erschlossen werden muss (Weidenmann 1994). Da Bilder immer mehrdeutig sind, ist dafür die Fähigkeit erforderlich, Darstellungs- und Steuerungscode lesen zu können. Darstellungscode in Bildern fokussieren auf kritische Merkmale der Wahrnehmung von Objekten und Szenen, indem sie z.B. Kontraste, Begrenzungen und Umrisse übertreiben. Steuerungscode lenken die Rezeption des visuellen Arguments. Sie steuern den Blickverlauf, heben Details hervor und regen kognitive Operationen wie Vergleichen und Schlussfolgern an. Typische Steuerungscode sind Pfeile, Vergrößerungen oder Hervorhebungen, sie erfolgen aber auch oft sprachlich in Form von Bildlegenden, Bildüberschriften und Bildbeschriftungen (Weidenmann 1994). Das Lesen und Nutzen der Darstellungs- und Steuerungscode braucht Anleitung und Übung. Zudem könnte es von Vorteil sein, die einzelnen Abbildungen zunächst einmal ohne den Text zu „lesen“ und damit einen möglichen Split Attention Effekt zu verhindern, um dann schließlich im Leseschritt drei (siehe unten) die Informationen aus Abbildungen und Fließtext zusammenbringen zu können.
- Aber nicht nur innerhalb der Bilder sollte der Wahrnehmungsprozess gesteuert werden, sondern auch zwischen Bildern bzw. zwischen Fließtext und Abbildung. Hier gilt es, lokale Kohärenzen im Text für die Lesenden sichtbar zu machen, indem die Bedeutung der unterschiedlichen „Kohäsionsmittel“ zwischen Fließtext und Abbildungen, zu denen zum Beispiel explizite Mittel wie Abbildungsverweise, aber auch implizite wie Nähe, Pfeile, Lupen etc. zählen, zu klären. Mit der Steuerung der Wahrnehmungsprozesse hin zu diesen verbindenden Elementen werden die verschiedenen Bezüge im Text für die Lesenden sichtbar. Im Leseschritt *zwei* sollten also neben dem Gewinnen lokaler Informationen aus dem Fließtext zum einen die Wahrnehmungsprozesse zum Lesen der Abbildungen angeleitet und gesteuert werden, zum zweiten die Informationen, die Abbildungen generieren, erschlossen werden und zum dritten die Verknüpfungen von Fließtext und Bildern auf der Textoberfläche sichtbar gemacht werden.

- Der nachfolgende Leseschritt *drei* zielt auf die hierarchiehöheren Textverstehenskompetenzen ab. Der Text soll verarbeitet und lokale und globale Kohärenz innerhalb von Abschnitten bzw. über den Gesamttext gebildet werden. Damit dies gelingen kann, müssen die einzelnen Informationen aus dem Fließtext und den Abbildungen in einen Zusammenhang gebracht werden, d.h. die Abbildungen müssen als Bestandteile der einzelnen Konzepte, die der Text vermitteln soll, im Blick behalten werden. Hierbei kommen, wie schon in der Einleitung erwähnt, die drei kognitiven Teilprozesse Selektion, Organisation und Transformation ins Spiel. Selektions- und Organisationsprozesse werden dabei am besten mithilfe der Kombination aus Fließtext und Abbildungen angeregt. So kann es zum Beispiel die Entscheidung unterstützen, welche Informationen im Fließtext als wichtig erachtet werden sollten, wenn erkannt wird, dass bestimmte Informationen auch in den Abbildungen wiederaufgenommen worden sind. Durch die Redundanz werden wichtige Inhalte fokussiert (Selektion) und können dann in Organisationsprozessen weiterverarbeitet werden. Und auch für die Organisation von Inhalten sind dann wieder die Abbildungen entscheidend, da sie z.B. räumliche Strukturen oder zeitliche Prozesse darstellen und damit eine Organisationsmöglichkeit vorgeben. Der stetige Abgleich während des Leseprozesses zwischen Abbildungs- und Fließtextinhalten führt dann zu Transformationsprozessen, also zu einer Versprachlichung der bildlichen Informationen und einer Verbildlichung der sprachlichen Informationen, so dass daraus ein umfangreiches mentales Modell gebildet werden kann, welches schließlich in das bestehende Wissen zu integrieren ist. Durch Aufträge zur Versprachlichung (z.B. Beschreibungen) und bildlichen Darstellung (z.B. Diagramme) werden diese Transformationsprozesse zusätzlich angeregt.
- Leseschritt *vier* dient schließlich dazu, das Textverständnis zu sichern, also die Integrationsprozesse in vorhandenes Wissen anzuregen. Hier verortet sich zudem die Reflexion über den Text, aber auch über das eigene Textverstehen, so dass in diesem letzten Leseschritt auch noch einmal über die verschiedenen Funktionen der Bilder im Text und die verwendeten Strategien zur Text-Bild-Integration nachgedacht werden kann. Das gebildete mentale Modell wird an dieser Stelle also nochmal kritisch ‚befragt‘ und in diesem Sinne abgesichert.

Bilder unterstützen also den Wissensaufbau, wenn sie sinnvoll im verbalen Textkörper platziert sowie thematisch passend und damit kohärent verwendet sind. Das aber wird in vielen Lehrmitteltexten, die Schüler*innen als Lernangebot gegeben werden, nicht erfüllt. Außerdem stellt das Lesen von Text-Bild-Kombinationen an sich aufgrund spezieller Anforderungen Lernende vor verschiedene Herausforderungen, so dass das Lesen und das Verstehen von schulischen Fachtexten unbedingt didaktisch zu begleiten ist. Dabei sollte das Ziel instruktorischer Unterstützung darin bestehen, solche Hilfen anzubieten, die den Wissenserwerb unterstützen, ohne das Ausmaß an zusätzlicher kognitiver Belastung zu stark zu erhöhen.

6 | Literatur

- Ballstaedt, S.-P. (2012). *Visualisieren: Bilder in wissenschaftlichen Texten*. UVK Verlagsgesellschaft.
- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In: J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider & M. Weiss (Hrsg.), *PISA 2000 Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323–407). Leske + Budrich.
- Beuck, H.-G., Dobers, J., Rabisch, G. & Zeeb, A. (2012). *Erlebnis Biologie 2*. Schroedel.
- Bolte, C. & Pastille, R. (2010). Naturwissenschaften zur Sprache bringen. Strategien und Umsetzung eines sprachaktivierenden naturwissenschaftlichen Unterrichts. In: G. Fenkart, A. Lembens & E. Erlacher-Zeitlinger (Hrsg.), *Sprache, Mathematik und Naturwissenschaften* (S. 26–46). Studienverlag.
- Brünken, R., Seufert, T. & Zander, S. (2005). Förderung der Kohärenzbildung beim Lernen mit multiplen Repräsentationen. *Zeitschrift für Pädagogische Psychologie*, 19(1/2), 61–75. <https://doi.org/10.1024/1010-0652.19.12.61>
- Chandler, P. & Sweller, J. (1991). Cognitive Load Theory and the format of instruction. *Cognition and Instruction*, 8(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2
- Christmann, U. (2010). Lesepsychologie. In: M. Kämper-van den Boogaart & K. Spinner (Hrsg.), *Lese- und Literaturunterricht, Teil 1* (S. 148–200). Schneider.
- Dittmar, M., Schmellentin, C., Gilg, E. & Schneider, H. (2017). Kohärenzaufbau aus Text-Bild-Gefügen: Wissenserwerb mit schulischen Fachtexten. *Leseforum Schweiz. Literalität in Forschung und Praxis*, 1(Multimodalität), 1–19.
- ELAN (Version 5.8) [Computer Software]. (2019). Max Planck Institute for Psycholinguistics, The Language Archive. <https://archive.mpi.nl/tla/elan>.
- Garner, R., Brown, R., Sanders, S. & Menke, D. J. (1992). “Seductive details” and learning from text. In: K. A. Renninger, S. Hidi & A. Krapp (Hrsg.), *The role of interest in learning and development* (S. 239–254). Erlbaum.
- Hannus, M. & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low- and high-ability children. *Contemporary Educational Psychology*, 24(2), 95–123. <https://doi.org/10.1006/ceps.1998.0987>
- Harp, S. F. & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90(3), 414.
- Jian, Y.-C. (2016). Fourth graders’ cognitive processes and learning strategies for reading illustrated biology texts: Eye movement measurements. *Reading Research Quarterly*, 51(1), 93–109. <https://doi.org/10.1002/rrq.125>
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163–182. <https://doi.org/10.1037/0033-295X.95.2.163>
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Kintsch, W. & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394. <https://doi.org/10.1037/0033-295X.85.5.363>
- Kürschner, C. & Schnotz, W. (2007). Konstruktion mentaler Repräsentationen bei der Verarbeitung von Text und Bildern. *Unterrichtswissenschaft*, 35(1), 48–67.
- Leisen, J. (2013). *Handbuch Sprachförderung im Fach: Sprachsensibler Fachunterricht in der Praxis*. Klett.

- Lenhard, W. (2013). *Leseverständnis und Lesekompetenz: Grundlagen - Diagnostik - Förderung*. Kohlhammer.
- Lindauer, T., Schmellentin, C. & Beerenwinkel, A. (2016). Sprachbewusster Naturwissenschaftsunterricht: Werkstattbericht zu einem transdisziplinären Entwicklungsprojekt. In: I. Winkler & F. Schmidt (Hrsg.), *Interdisziplinäre Forschung in der Deutschdidaktik*. «Fremde Schwestern' im Dialog (S. 225–246). Peter Lang.
- Lindauer, T., Schmellentin, C., Beerenwinkel, A., Hefti, C. & Furger, J. (2013). *Sprachbewusst unterrichten. Eine Unterrichtshilfe für den Fachunterricht im Auftrag des Bildungsraums Nordwestschweiz*. Abgerufen von <https://irf.fhnw.ch/handle/11654/16585> [16.09.2021].
- Lindauer, T. & Schneider, H. (2007). Lesekompetenz ermitteln: Aufgaben im Unterricht. In: A. Bertschi-Kaufmann (Hrsg.), *Lesekompetenz, Leseleistung, Leseförderung: Grundlagen, Modelle und Materialien* (S. 109–125). Kallmeyer.
- Mason, L., Tornatora, M. C. & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, 60(1), 95–109. <https://doi.org/10.1016/j.compedu.2012.07.011>
- Mayer, R. E. (2009). *Multimedia Learning*. Cambridge University Press.
- Mayer, R. E. & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90(2), 312–320. <https://doi.org/10.1037/0022-0663.90.2.312>
- Merzyn, G. (1994). *Physikschulbücher, Physiklehrer und Physikunterricht. Beiträge auf der Grundlage einer Befragung westdeutscher Physiklehrer*. IPN.
- Mesmer, H. A., Cunningham, J. W. & Hiebert, E. H. (2012). Toward a theoretical model of text complexity for the early grades: Learning from the past, anticipating the future. *Reading Research Quarterly*, 47(3), 235–258. <https://doi.org/10.1002/rrq.019>.
- Müller, C. (2018). Mixed Methods: Theorie und Praxis methodenpluraler Forschung. In: J. Boelmann (Hrsg.), *Empirische Forschung in der Deutschdidaktik* (S. 161–172). Schneider Hohengehren.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press, Clarendon Press.
- Rickheit, G. & Schade, U. (2000). Kohärenz und Kohäsion. In: K. Brinker, G. Antos, W. Heinemann & S. F. Sager (Hrsg.), *Text- und Gesprächslinguistik: Ein internationales Handbuch zeitgenössischer Forschung*. De Gruyter.
- Rincke, K. (2010). Von der Alltagssprache zur Fachsprache. Bruch oder schrittweiser Übergang? In: G. Fenkart, A. Lembens & E. Erlacher-Zeitlinger (Hrsg.), *Sprache, Mathematik und Naturwissenschaften* (S. 47–62). Studienverlag.
- Schlag, S. (2011). *Kognitive Strategien zur Förderung des Text- und Bildverstehens beim Lernen mit illustrierten Sachtexten: Theoretische Konzeptualisierung und empirische Prüfung*. Logos.
- Schlag, S. & Ploetzner, R. (2011). Supporting learning from illustrated texts: Conceptualizing and evaluating a learning strategy. *Instructional Science*, 39(6), 921–937. <https://doi.org/10.1007/s11251-010-9160-3>
- Schmellentin, C., Dittmar, M., Gilg, E. & Schneider, H. (2017). Sprachliche Anforderungen in Biologielehrmitteln. In: B. Ahrenholz, B. Hövelbrinks & C. Schmellentin (Hrsg.), *Fachunterricht und Sprache in schulischen Lehr-/Lernprozessen* (S. 23–91). Narr.

- Schneider, H., Dittmar, M., Gilg, E. & Schmellentin, C. (2018). Textseitige Maßnahmen zur Unterstützung des Leseverstehens im Biologieunterricht. *Didaktik Deutsch*, 45, 94–116.
- Schneider, H., Gilg, E., Dittmar, M. & Schmellentin, C. (2019). Prinzipien der Verständlichkeit in Schulbüchern der Biologie auf der Sekundarstufe 1. In: B. Ahrenholz (Hrsg.), *Sprache im Fach*. De Gruyter.
- Schneider, W., Schlagmüller, M. & Ennemoser, M. (2007). *Lesegeschwindigkeits- und -verständnistest für die Klassenstufen 6-12 (LGVT)*. Hogrefe.
- Schnotz, W. (1994). *Aufbau von Wissensstrukturen: Untersuchungen zur Kohärenzbildung beim Wissenserwerb mit Texten*. Beltz.
- Schnotz, W. & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. [https://doi.org/10.1016/S0959-4752\(02\)00017-8](https://doi.org/10.1016/S0959-4752(02)00017-8)
- Staraschek, E. (2003). Ergebnisse einer Schülerbefragung über Physikschulbücher. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 135–146.
- van Dijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.
- Weidenmann, B. (1994). Informierende Bilder. In: B. Weidenmann (Hrsg.), *Wissenserwerb mit Bildern: Instruktionale Bilder in Printmedien, Film/Video und Computerprogrammen* (S. 9–58). Huber.
- Weidenmann, B. (2002). Abbilder in Multimediaanwendungen. In: L. J. Issing & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia und Internet: Lehrbuch für Studium und Praxis* (S. 2–11). Beltz.
- Zhao, F., Schnotz, W., Wagner, I. & Gaschler, R. (2014). Eye tracking indicators of reading approaches in text-picture comprehension. *Frontline Learning Research*, 2(4), 46–66.