

Beitrag Themenheft „Testen und Prüfen“

Pascale Schaller* und Julia Winkes*

Lernverlaufsdiagnostik (LVD) der Schreibflüssigkeit bei Kindern mit Deutsch als Zweitsprache

Progress monitoring in writing fluency for children with German as a second language

<https://doi.org/10.1515/infodaf-2022-0068>

Zusammenfassung: Der empirische Beitrag widmet sich der Lernverlaufsdiagnostik im Bereich Schreiben (LVD – Schreiben) und zeigt anhand einer Stichprobe von Kindern mit Deutsch als Erstsprache (DaE) und Deutsch als Zweitsprache (DaZ) der 3. bis 6. Klassenstufe auf, für welche Kennwerte sich die Kindergruppen in LVD – Schreiben unterscheiden. Der Fokus liegt bei der Reliabilität und der Validität des Verfahrens. Der Ansatz der Lernverlaufsdiagnostik wird vorgestellt und – basierend auf der Datenauswertung – in seinem Nutzen für die verschiedenen Kindergruppen eingeordnet.

Schlüsselwörter: Lernverlaufsdiagnostik (LVD), Deutsch als Zweitsprache, Curri-culumbasiertes Messen (CBM), Schreibflüssigkeit

Abstract: The article shows based on a sample of children with German as a first language and German as a second language in the 3rd to 6th grade level for which scoring methods in progress monitoring in writing the groups of children differ. The focus is on the reliability and validity of the procedure. The approach of learning progress monitoring is presented and – based on the data analysis – discussed in its potential for the different groups of children.

Keywords: progress monitoring in writing, German as a second language, Curriculum-Based Measurement (CBM), writing fluency

*Kontaktpersonen: Pascale Schaller, E-Mail: pascale.schaller@phbern.ch
Julia Winkes, E-Mail: julia.winkes@unifr.ch

1 Einleitung und Fragestellung

Der vorliegende Beitrag widmet sich einem Verfahren aus der Lernverlaufsdagnostik im Bereich Schreiben. Im Speziellen geht er der Frage nach, als wie reliabel und valide sich LVD – Schreiben im Deutschen erweist im Vergleich zwischen Kindern mit Deutsch als Erstsprache ($N = 175$) und Deutsch als Zweitsprache ($N = 136$).

Zudem interessiert, welche Auswertungsmethoden in welcher Klassenstufe geeignet sind und inwiefern sich hierbei die Kindergruppen unterscheiden. Für die Überprüfung der Validität wird die Einschätzung der Lehrperson herangezogen. Basierend auf den gewonnenen Resultaten wird diskutiert, inwiefern sich LVD – Schreiben auch für Kinder mit DaZ eignet und welchen Aspekten für den Einsatz des Verfahrens besonders Rechnung zu tragen ist.

2 LVD – Schreiben

2.1 Zur Funktion einer Lernverlaufsdagnostik – besonders für Kinder mit Deutsch als Zweitsprache

Mit jeder Art von Überprüfung von Sprachkompetenzen geht die Frage einher, wie das vom Kind erreichte Ergebnis zu interpretieren ist. Für viele schulische Settings, aber auch für eine Vielzahl von standardisierten Sprachtests stellt der Vergleich der Kompetenz eines Lernenden mit seiner (Lern-)Altersgruppe die Grundlage für die Interpretation dar. Im Zentrum steht dabei für die Lehrperson die Frage, ob ein Kind im Vergleich mit der Altersnorm über eine angemessene Kompetenz verfügt, sei das etwa im Lesen, in der Rechtschreibung oder im Texteschreiben. Diese Information erlaubt der Lehrperson, die Leistungsheterogenität innerhalb einer Klasse abzuschätzen und die (Leistungs-)Position der einzelnen Kinder in der Gruppe zu erfassen. Gerade für die formative Begleitung von Kindern ebenso wichtig ist die didaktisch-pädagogische Perspektive mit der Frage, ob und wie sich die Sprachkompetenzen jedes einzelnen Kindes über die Zeit hinweg entwickeln, ob sich also ein Lernfortschritt feststellen lässt – und zwar unabhängig davon, auf welchem Sprachstand das entsprechende Kind steht. Solch eine Überprüfung des Lernverlaufs weist verschiedene Vorteile auf. Sie gibt der Lehrperson Auskunft darüber, welche Kinder Lernfortschritte machen. Das wiederum gibt ihr einen Hinweis darauf, welche Kinder vom Unterricht zu profitieren vermögen. Und ebenso zeigt sich, welche Kinder in ihrem Lernverlauf stagnieren oder nur sehr langsam Fortschritte machen. Diese Fragen adressiert die sogenannte Lernverlaufs- oder Lernfortschrittsdiagnostik (im englischen Sprach-

raum unter dem Terminus Curriculum-Based Measurement bzw. CBM etabliert). Sie hat zum Ziel, anhand wiederholter kurzer Messungen über einen bestimmten Zeitraum (etwa ein Schuljahr) hinweg für verschiedene Kompetenzen den Lernverlauf der Kinder abzubilden. Dabei kommt ihr auch eine präventive Funktion zu: Sie ist eine Hilfe dafür, Kinder mit potenziellen Lernschwierigkeiten frühzeitig zu erkennen (Grosche/Casale 2021: 173; Huber/Grosche 2012: 312), indem nicht zugewartet wird, bis die Leistungsschere zwischen ihnen und ihren Peers ausreichend groß ist, um eine Entwicklungsstörung feststellen zu können. Stattdessen ermöglicht die Beobachtung des Lernverlaufs eine Prognose, ob ein Kind im aktuellen Lerntempo die Lernziele erreichen wird. Ist dies nicht der Fall, kann die Lehrperson entsprechend reagieren und gewisse Kinder enger begleiten. Dieses Vorgehen „basiert auf der empirisch gut abgesicherten Erkenntnis, dass Lernprozesse immer dann besonders effektiv verlaufen, wenn die Probleme frühzeitig erkannt werden und sowohl Lehrkräfte als auch Schulkinder möglichst kurzfristige Rückmeldungen über den Erfolg der Förderung erhalten“ (Huber/Grosche 2012: 314).

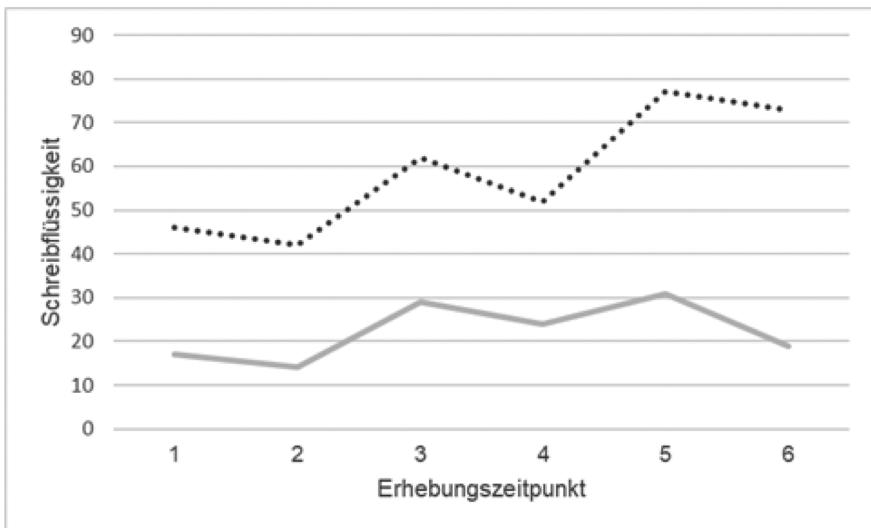


Abbildung 1: Lernverlauf der Schreibflüssigkeit von zwei Kindern der 4. Klasse

Anm.: Daten aus einem laufenden Forschungs- und Entwicklungsprojekt an einer Berner Schule; Schreibflüssigkeit gemessen in CWS (vgl. Tab. 1)

Abbildung 1 stellt exemplarisch den Lernverlauf von zwei Kindern der 4. Klasse über sechs Messzeitpunkte (Herbst bis Frühling) dar. Die Daten stammen aus einem laufenden Forschungs- und Entwicklungsprojekt an einer Schule in der Stadt Bern. Die Abbildung zeigt den Lernverlauf eines Kindes (gepunktete Linie), das

über die sechs Messzeitpunkte hinweg in der Tendenz einen Lernzuwachs verzeichnet, während das zweite Kind (durchgezogene Linie) in seinem Lernfortschritt zu stagnieren droht. Diese Beobachtung kann der Lehrperson Anlass dazu geben, das Kind enger zu begleiten beziehungsweise zu beobachten und es gezielter zu unterstützen. Dieser Vorteil der Lernverlaufsdiagnostik stellt für alle Kinder einen Gewinn dar, dürfte aber für die Begleitung der Sprachentwicklung der Kinder durch die Lehrperson gerade für Kinder mit Deutsch als Zweitsprache von besonderem Vorteil sein: Kinder mit Deutsch als Zweitsprache verfügen häufig über eingeschränktere Deutschkenntnisse als die monolingualen Peers. Das Hauptinteresse der Lehrperson dürfte daher – gerade bei Kindern mit besonders basalen Sprachkompetenzen – insbesondere darin liegen, das Kind dabei zu unterstützen, sprachliche Fortschritte zu machen. Dabei spielt der Vergleich mit der Altersnorm oder der Klasse vorerst eine zweitrangige Rolle. Verfügt die Lehrperson über ein Hilfsmittel, den Lernverlauf der Kinder zuverlässig und mit überschaubarem Aufwand regelmäßig zu überprüfen, hat sie die Möglichkeit, ihren sprachlichen Input, ihre Instruktionen oder ihre Lerngelegenheiten auf die Kinder anzupassen und vor allem jenen Kindern besondere Aufmerksamkeit zukommen zu lassen, deren Lernverlauf zu stagnieren droht. Die Lernverlaufsdiagnostik stellt ein Instrument dar, das solch eine Überprüfung des Lernverlaufs auf einfache, systematische und ökonomische Weise ermöglichen soll. Es kann mit zwei Zielsetzungen eingesetzt werden: Zum einen kann es als Screening im Sinne einer Statusdiagnostik den Lernstand eines Kindes zu einem bestimmten Zeitpunkt abbilden; zum andern vermag es aber im Sinne einer Lernverlaufsdiagnostik ebenso, den Lernverlauf eines Kindes über einen längeren Zeitrahmen hinweg zu erfassen. LVD-Verfahren stellen allgemein kurze Tests mit vergleichbarer Schwierigkeit dar, die wiederholt von der Lehrperson eingesetzt werden können und anhand eines überprüften Indikators (beispielsweise Anzahl korrekt gelesener Wörter in einer Minute) der Lehrperson Auskunft geben über eine zugrunde liegende Kompetenz (beispielsweise die Leseflüssigkeit) (Allen et al. 2019: 2). Darüber hinaus ist LVD dadurch charakterisiert, dass es die Testgütekriterien (Reliabilität, Validität, Objektivität, Testfairness etc.) erfüllt, sodass eine solide Datengrundlage für pädagogische Entscheidungen der Lehrperson besteht (Keller-Margulis/Mercer/Matta 2021; Fuchs 2017). Ihr ökonomischer Einsatz im schulischen Alltag wird gewährleistet durch eine zeitsparende Durchführung und Auswertung (in der Regel dauert die Durchführung nur wenige Minuten).

Es werden mit LVD Kompetenzbereiche fokussiert, die zentral für den Schulerfolg der Kinder sind, so etwa die Bereiche Lesen, Mathematik, Rechtschreibung und Schreiben (Romig/Therrien/Lloyd 2017: 73). Der wiederholte Einsatz von LVD hat sich entsprechend für den Schulkontext als bedeutsam und effektiv erwiesen (Voß et al. 2020: 90). Während LVD in den USA in unterschiedlichen Bereichen

breite Anwendung findet und Methode und Materialien seit Jahrzehnten erforscht werden (Huber/Grosche 2012: 319), geschieht beides im deutschsprachigen Raum nicht in vergleichbarem Umfang. Es besteht daher ein erhebliches Forschungsdesiderat im Bereich deutschsprachiger LVD-Verfahren (Voß et al. 2020: 90).

Dieses Desiderat umfasst auch verschiedene Fragen zu LVD – Schreiben, die sowohl die Eignung von LVD in der Funktion einer Statusdiagnostik als auch seine Eignung für die Lernverlaufsdiagnostik betreffen. Weitgehend ungeklärt ist beispielsweise, wie reliabel und valide LVD – Schreiben die Schreibkompetenz in der deutschen Schriftsprache abbildet. In Bezug auf die Reliabilität stellt der Umstand, dass die Leistungen der Kinder bei einer wiederholten Testung durch LVD – Schreiben innerhalb einer kurzen Zeitspanne (zum Beispiel bei wiederholten Tests innerhalb von zehn Tagen) teilweise erheblich variieren, eine besondere Herausforderung dar (vgl. Winkes/Schaller 2022a). Diese intraindividuelle Variabilität (siehe hierzu Winkes/Schaller 2022b) wirft die Frage auf, wie reliabel das Instrument zu einem Messzeitpunkt misst und inwiefern es sich für die Lernverlaufsdiagnostik über verschiedene Messzeitpunkte eignet. So wird in Abbildung 1 ersichtlich, dass auch das Kind, das in der Tendenz einen Leistungszuwachs verzeichnet, in seinen Leistungen schwankt. Ebenso ist für den deutschsprachigen Raum ungeklärt, ob sich in Bezug auf die Reliabilität und Validität für verschiedene Kindergruppen – so etwa Kinder mit DaE und Kinder mit DaZ – unterschiedliche Auswertungsverfahren (Kennwerte) eignen. Auf diesen letztgenannten Aspekt fokussiert der vorliegende Beitrag. Er widmet sich also vorerst der statusdiagnostischen Qualität von LVD – Schreiben, spezifisch der Reliabilität und der Validität des Verfahrens und den diesbezüglichen Unterschieden zwischen Kindern mit DaE und DaZ.

2.2 Zum Vorgehen bei LVD – Schreiben

LVD – Schreiben besteht aus kurzen Schreibproben, die als Antwort auf einen Anfangssatz, einen sogenannten Storystarter, verfasst werden (z.B. „Ich öffnete die Tür, dann ...“) (Hosp/Kaldenberg 2020: 78–79). Nach einer Minute Denkzeit schreiben die Kinder während drei bis sieben Minuten einen Text. Dieser Text wird daraufhin anhand verschiedener Scoringmethoden ausgewertet. Zu unterscheiden ist dabei zwischen drei verschiedenen Typen von Scoringmethoden (vgl. Tab. 1): den produktionsabhängigen Maßen, den produktionsunabhängigen Maßen und sogenannten *Accurate Production Indices* (Jewell/Malecki 2005).

Tabelle 1: verschiedene Scoringmethoden für LVD – Schreiben

Anm.: Für jeden Kennwert wird angegeben, welchen Wert das Kind in seinem Text (Abb. 2) erreichen würde.

Art des Kennwerts	Kennwert und Beispiel aus dem Kindertext (Abb. 2)
produktionsabhängige Maße	TWW (Total Words Written): Summe aller Wörter: 33 WSC (Words Spelled Correctly): alle orthografisch korrekt geschriebenen Wörter: 31 CWS (Correct Writing Sequences): korrekte Schreibsequenzen – Interpunktion, Syntax, Grammatik und Semantik werden ebenfalls berücksichtigt; die korrekte Schreibsequenz besteht aus zwei Einheiten: 27
produktionsunabhängige Maße	%CWS (Percentage Correct Writing Sequences): prozentualer Anteil korrekter Schreibsequenzen: 75
<i>Accurate Production Indices</i>	CIWS (Correct minus Incorrect Writing Sequences): $27 - 8 = 19$ CIWS + TWW (Correct minus Incorrect Writing Sequences + Total Words Written): $27 - 8 + 33 = 52$

Endlich hat es geklappt. Ich habe eine Maschine erfunden, die...

in andere Zeiten reisen kann. Ich wollte sie gerade ausprobieren, aber meine Freundin soll auch mitkommen. Ich rief ihr schnell an und einige Zeit später flogen wir schon los. Als die Maus

Abbildung 2: Beispiel eines Textes aus dem Korpus der 6. Klasse

Anm.: oval umrahmt = Beispiel korrekte Schreibsequenzen; eckig umrahmt = Beispiel inkorrekte Schreibsequenzen

Produktionsunabhängige Maße berücksichtigen die Länge des verfassten Textes nicht, während produktionsabhängige Maße diese miteinberechnen (Saddler/Asaro-Saddler 2013: 27–28). Während die Anzahl verschrifteter (TWW) oder korrekt verschrifteter (WSC) Wörter in der Schreibforschung bekannte Maße für Textauswertungen darstellen und selbsterklärend sind, soll auf die sogenannten Schreibsequenzen (CWS) genauer eingegangen werden. Schreibsequenzen berücksichtigen über die sprachliche Korrektheit des einzelnen Elements hinaus auch seine Passung zum vorangehenden sowie zum darauffolgenden Element. Eine korrekte Schreibsequenz fokussiert also die inhaltliche und sprachliche Passung zweier aufeinander folgender Elemente (McMaster/Espin 2007: 70).

Die sogenannten *Accurate Production Indices* zeigen sich als besonders nahe an holistischen Ratings von Lehrpersonen, speziell in der Mittelstufe (Saddler/Asaro-Saddler 2013: 28). Aus diesem Grund wird das Maß korrekte minus inkorrekte Schreibsequenzen (CIWS) oftmals für das Beschreiben des Lernverlaufs herangezogen. Der Kennwert zeigt sich hinsichtlich der kriterialen Validität in der Forschung als den anderen Kennwerten überlegen und wird in der Literatur insbesondere für die höheren Klassenstufen empfohlen (Payan et al. 2019).

2.3 Was misst LVD – Schreiben? Zum Konstrukt der Schreibflüssigkeit

Das Schreiben von Texten ist ein hochkomplexer Prozess, bei dem verschiedene Teilkompetenzen integriert werden müssen. Diese Teilkompetenzen betreffen zum einen sogenannte hierarchiehöhere Bereiche wie das Planen und das Generieren von Texten und zum anderen hierarchieniedrigere Bereiche wie die Motorik und die Rechtschreibung. Während im deutschsprachigen Raum die Schreibflüssigkeit im Wesentlichen auf die flüssige und geläufige Handschrift bezogen wird, umfasst die sogenannte *writing fluency* die Textmenge, „die innerhalb einer bestimmten Zeit orthographisch, grammatisch und semantisch korrekt und leserlich verfasst wird“ (Sturm 2017). Kim et al. (2017: 3) bezeichnen die Textschreibflüssigkeit als Mittler zwischen Transkription (Rechtschreibung und Schreibmotorik) und Textgenerierung auf der einen Seite und der Schreibqualität auf der anderen: „Text writing fluency facilitates writing by enabling cognitive resources to be used for higher-order cognitive processes, and it draws on transcription and text generation, then text writing fluency would play a mediating role, acting as a bridge to connect, at least partially, transcription skills and text generation with a criterion writing skill.“

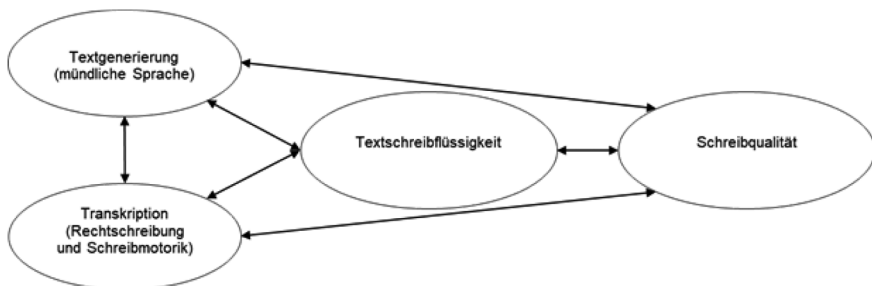


Abbildung 3: Modell der Textschreibflüssigkeit nach Kim et al. (2017: 3)

Die Textschreibflüssigkeit als Indikator für Schreibqualität steht bei LVD – Schreiben im Zentrum, da LVD – Schreiben die Menge und die Korrektheit von Texten

misst, die innerhalb einer definierten Zeitbegrenzung verschriftet werden können. Hierarchieniedrigere Schreibfähigkeiten nehmen zu Beginn des Schreiberwerbs vergleichsweise viel Raum ein, treten dann aber mit zunehmender Schreibroutine mehr Raum an hierarchiehöhere Prozesse ab. Das heißt wiederum, dass gut ausgebildete basale Schreibfähigkeiten den Schreibprozess entlasten. „In diesem Sinne ist zu erwarten, dass – analog zu anderen Domänen wie Lesen oder Mathematik – basale Fähigkeiten hierarchiehöhere Fähigkeiten zumindest teilweise vorhersagen können“ (Sturm/Nänny/Wyss 2017: 96). Zumal sich mit zunehmender Schreibroutine der Anteil der beteiligten Teilkompetenzen verändert und Modellierungen von Schreibflüssigkeit damit einen Entwicklungsaspekt enthalten, ist davon auszugehen, dass Indikatoren für die Schreibqualität für verschiedene Altersgruppen unterschiedlich adäquat sind. So gelten Flüssigkeitsmaße laut Sturm, Nänny und Wyss (2017: 96) als besonders guter Prädiktor für Schreibkompetenzen von Schreibanfängerinnen und Schreibanfängern, wobei der Zusammenhang von Schreibflüssigkeit und Textqualität auf höheren Klassenstufen zunehmend an Bedeutung zu verlieren scheint. Wenn also durch LVD-Verfahren die Schreibflüssigkeit von Kindern erhoben werden soll, ist folgenden Aspekten besonders Rechnung zu tragen:

- (1) Die Indikatoren für die Schreibqualität auf Textebene sind unter Umständen abhängig vom (Lern-)Alter der Kinder. Demzufolge ist für verschiedene Maße zu überprüfen, wie zuverlässig sie die Schreibkompetenz auf einer bestimmten Klassenstufe abbilden.
- (2) Texte zu schreiben, ist ein hochkomplexer sprachlicher Prozess, der in ein Produkt mündet, das ganz unterschiedliche sprachliche und inhaltliche Formen annehmen kann. Während etwa mit der Leseflüssigkeit die Anzahl korrekt gelesener Wörter ein und desselben Textes innerhalb einer Minute vergleichsweise präzise bestimmt werden kann, führt das Schreiben einer Geschichte zu einem Bild oder einem Anfangssatz als Stimulus für 100 Kinder zu 100 verschiedenen Texten mit unterschiedlicher inhaltlicher und sprachlicher Gestaltung. Es zeigt sich denn auch, dass die Vorhersagekraft basaler Fähigkeiten für hierarchiehöhere Fähigkeiten in der Domäne Schreiben geringer ausfällt als im Lesen oder in der Mathematik. Die hier skizzierten Schwierigkeiten sind aber auf der anderen Seite auch für standardisierte statusdiagnostische Schreibttests als Problembereiche bekannt und somit der Schreibdiagnostik allgemein inhärent (Wilson et al. 2019).

2.4 Wie gut ist LVD – Schreiben für verschiedene Gruppen von Lernenden geeignet?

Im englischen Sprachraum haben sich die kurzen Schreibproben von LVD – Schreiben generell als valide und reliable Indikatoren der Schreibkompetenz für Schüle-

rinnen und Schüler verschiedener Klassenstufen herausgestellt (Weissenburger/Espin 2005) und die Aufmerksamkeit der Forschung richtet sich zunehmend auf die Testgütekriterien des Lernverlaufs (z.B. Reliabilität und Validität des Verlaufs, Änderungssensitivität; Bestimmung von Lernkurven und durchschnittlichen Lernzuwachsraten). Während auch im deutschsprachigen Raum für einzelne Bereiche, namentlich Mathematik, Lesen und Rechtschreibung, erste Forschungsergebnisse und auch bereits Testmaterialien vorliegen (vgl. hierzu etwa die Materialien der *Lernlinie*: Voß et al. 2020), fehlt es hier bislang gänzlich an Studien zu LVD – Schreiben. Aus diesem Grund sollten sich Untersuchungen zunächst der Adäquatheit der Einzelmessung widmen. Winkes und Schaller (2022a) bearbeiten grundsätzliche Fragen zur Reliabilität, Validität, Homogenität der Testschwierigkeiten und der Änderungssensitivität von LVD – Schreiben für Schülerinnen und Schüler der Primarstufe allgemein. Darüber hinaus sollte die Eignung von LVD-Verfahren aber auch für verschiedene Gruppen von Lernenden empirisch belegt sein, um eine faire Interpretation von Testergebnissen zu gewährleisten (McMaster/Espin 2007). Studien zur Frage nach der Eignung von LVD – Schreiben für mehrsprachige Schülerinnen und Schüler existieren bislang vor allem für English Language Learners (ELL) auf der High-School-Stufe (Espin et al. 2008; Campbell/Espin/McMaster 2013). Die Resultate weisen darauf hin, dass sich ein ähnliches Muster ergibt wie bei einsprachigen Kindern; allerdings sind die Korrelationen für ELL stärker ausgeprägt. Diese Ergebnisse stehen aber in Kontrast zu den Ergebnissen von Keller-Margulis et al. (2016), die in ihrer explorativen Studie Viertklässlerinnen und Viertklässler mit unterschiedlichen Sprachhintergründen verglichen und geringere Korrelationen für ELL fanden. Übereinstimmung zeigt sich über verschiedene Studien hinweg vor allem in der Feststellung, dass weitere Forschungsbemühungen für Schülerinnen und Schüler unterschiedlicher Gruppen dringend benötigt werden.

3 Studie

3.1 Beschreibung der Studienanlage

An der Studie, deren Daten im Folgenden präsentiert werden, haben insgesamt 22 Klassen der 3., 4., 5. und 6. Klassenstufe aus den Kantonen Bern und Luzern teilgenommen. Erhoben wurden Daten zu zwei Erhebungszeitpunkten (September 2016 sowie März 2017). Zu jedem der beiden Erhebungszeitpunkte wurden innerhalb von maximal zehn aufeinanderfolgenden Tagen und mit maximal zwei Erhebungen pro Tag insgesamt zehn Texte verschriftet. Eingesetzt wurden dafür in jeder Klasse zehn unterschiedliche Storystarter. Nach der Erhebung durch die Lehrper-

son wurden die Texte der Kinder durch Studierende der Sonderpädagogik der Universität Freiburg ausgewertet.

3.2 Fragestellung

Im vorliegenden Beitrag werden folgende Fragen adressiert: (1) Für welche Kennwerte ergeben sich für die unterschiedlichen Klassenstufen Unterschiede in den erreichten Werten von Kindern mit Deutsch als Erstsprache und Kindern mit Deutsch als Zweitsprache? (2) Wie ist die Paralleltest-Reliabilität von LVD – Schreiben für Kinder mit Deutsch als Erstsprache und für Kinder mit Deutsch als Zweitsprache in den verschiedenen Klassenstufen und für die verschiedenen Auswertungsmethoden? (3) Wie ist die Validität von LVD – Schreiben für die Kinder mit Deutsch als Erstsprache und mit Deutsch als Zweitsprache?

3.3 Stichprobe

Die Gesamtstichprobe der Studie setzt sich aus 354 Schülerinnen und Schülern, verteilt auf die 3. bis 6. Klassenstufe, zusammen. Tabelle 2 veranschaulicht die Verteilung der Kinder auf die vier Klassenstufen, das Geschlecht, das Alter sowie den sprachlichen Hintergrund. Für den vorliegenden Beitrag werden die Kinder mit deutscher Erstsprache ($N = 175$) mit den Kindern, die eine andere Erstsprache haben ($N = 136$), verglichen. Nicht berücksichtigt werden diejenigen Kinder, die mit mehreren Sprachen aufwachsen ($N = 26$). Die im Folgenden berücksichtigte Teilstichprobe umfasst also jeweils 175 Kinder mit Deutsch als Erstsprache (DaE) und 136 Kinder mit Deutsch als Zweitsprache (DaZ).

Tabelle 2: Beschreibung der Stichprobe

Klasse	Anzahl Kinder N	Alter in Monaten M (SD)	Geschlecht m; w (in %)	DaE N (%)	DaZ N (%)	zweisprachig N (%)
3	72	106.4 (6.3)	55.6; 44.4	34 (47.2)	26 (36.1)	12 (16.7)
4	109	119.2 (6.7)	57.8; 42.2	57 (52.3)	35 (32.1)	11 (10.1)
5	92	134.3 (7.6)	40.2; 59.8	44 (47.8)	41 (44.6)	–
6	81	143.9 (6.1)	51.9; 48.1	40 (49.4)	34 (42.0)	3 (3.7)
Total	354	126.16 (15.14)	51.4; 48.6	175 (49.4 %)	136 (38.4 %)	26 (7.3 %)

4 Methode

4.1 LVD – Schreiben

Für die Bearbeitung der oben genannten Fragestellungen werden die Daten von t1 (Erhebung Herbst 2016) herangezogen. Die einzelnen Instrumente werden im Folgenden beschrieben. Für die Untersuchung der Paralleltest-Reliabilität werden drei Schreibproben aus t1 berücksichtigt, und zwar die 2., 3. und 4. Schreibprobe dieses Messzeitpunktes. Die erste durchgeführte Schreibprobe war die erste für die Kinder überhaupt und könnte dadurch ungünstig beeinflusst sein, dass die Kinder das Verfahren vorher nicht kannten. Aus diesem Grund wird sie nicht berücksichtigt. Die Auswahl von drei Schreibproben orientiert sich an Empfehlungen für den praktischen Einsatz von LVD – Schreiben im Schulalltag (Hosp/Kaldenberg 2020: 79). Diese drei Schreibproben sollen mit drei verschiedenen Storystartern mit vergleichbarer Schwierigkeit erhoben werden (Hosp/Kaldenberg 2020: 85). Da LVD ein Verfahren darstellen soll, das von den Lehrpersonen im Unterrichtsalltag einsetzbar ist, wäre es unzumutbar, pro Testzeitpunkt zehn Schreibproben zu erheben. Ziel sollte es daher sein, Kennwerte zu eruieren, die für ein überschaubares Maß an Text – also etwa drei Schreibproben – und mit überschaubarem Aufwand für die Lehrperson möglichst reliabel und valide eine Aussage über die Schreibkompetenz eines Kindes erlauben. Die Homogenität der Testschwierigkeiten wurde für die hier ausgewerteten Daten durch Verwendung des Vorgehens nach Klauer (2011) ermittelt: Weder für die Gruppe der Kinder mit DaE noch für die Gruppe der Kinder mit DaZ ergaben T-Tests für verbundene Stichproben (SS2 mit SS3 sowie SS3 mit SS4) signifikante Mittelwertsunterschiede mit mindestens mittlerer Effektstärke (Cohens $d \geq .50$). Dies gilt für alle untersuchten Klassenstufen. Damit kann davon ausgegangen werden, dass die Schreibproben 2, 3 und 4 aus Testzeitpunkt 1 vergleichbare Schwierigkeiten aufweisen und unterschiedlich hohe Werte der Kinder in den drei Schreibproben nicht auf unterschiedliche Schwierigkeiten zurückzuführen sind.

Während Hosp, Hosp und Howell (2007) vorschlagen, den Median aus drei Schreibproben als Indikator für die Schreibkompetenz zu einem Messzeitpunkt zu wählen, wird in der Forschung zuweilen der Mittelwert herangezogen. Analog zu Campbell, Espin und McMaster (2013) und Campbell (2010), die LVD – Schreiben für Lernende mit Englisch als Zweitsprache untersuchen, wird für den vorliegenden Beitrag der Mittelwert verwendet.

Die oben ausformulierten Fragen zu den verschiedenen Auswertungsmethoden der Kindertexte werden anhand der Analyse folgender Kennwerte vorgenommen: TWW, WSC, CWS, CIWS und CIWS + TWW (vgl. zur Beschreibung der Kennwerte Tab. 1). McMaster und Espin (2007: 78–79) merken zu Maßen, die Pro-

zentangaben beinhalten – wie %CWS (vgl. Tab. 1) –, kritisch an, dass die Daten nicht mehr gleich skaliert sind wie die übrigen Kennwerte, wodurch die Vergleichbarkeit erschwert wird. Aus diesem Grund wird dieses Maß in die vorliegenden Analysen nicht miteinbezogen.

4.2 Einschätzung der Lehrperson

Aufgrund des Fehlens von standardisierten deutschsprachigen Instrumenten zur Erhebung des Schreibens auf Textebene wird eine Einschätzung der Lehrperson als Kriteriumsmaß für die Schreibkompetenz der Kinder eingesetzt. Die Lehrperson hat für jedes Kind einzeln dessen Schreibkompetenz auf einer Skala von 1 (= sehr schlecht) bis 10 (= sehr gut) beurteilt. Diese Einschätzung der Schreibkompetenz der Kinder durch die Lehrperson bietet eine pragmatische Variante, wenngleich anzumerken ist, dass solch ein Rating verzerrt sein kann und Lehrpersonen unter Umständen gerade die Kompetenz von Kindern mit Deutsch als Zweitsprache unterschätzen (vgl. zu dieser Diskussion etwa Carigiet Reinhard 2012).

5 Ergebnisse

5.1 Leistungen der Kinder mit DaE und mit DaZ in LVD – Schreiben

Wie Abbildung 4 verdeutlicht, unterscheiden sich die von den Kindern erreichten Werte zwischen Kindern mit DaZ und Kindern mit DaE. Während sich in der 3. Klassenstufe für die produktionsunabhängigen und *Accurate Production Indices* keine bedeutsamen Unterschiede zeigen, ist das für die übrigen Klassenstufen der Fall. Auffallend ist, dass in der 5. Klassenstufe Gruppenunterschiede allein beim Kennwert CIWS festzustellen sind, was gegebenenfalls durch Besonderheiten der Stichprobe erklärt werden könnte. In den Stufen vier und sechs schneiden Schülerinnen und Schüler mit DaE mit allen Scoringmethoden signifikant besser ab als die gleichaltrigen Kinder mit DaZ, und zwar mit mindestens mittlerer Effektstärke.

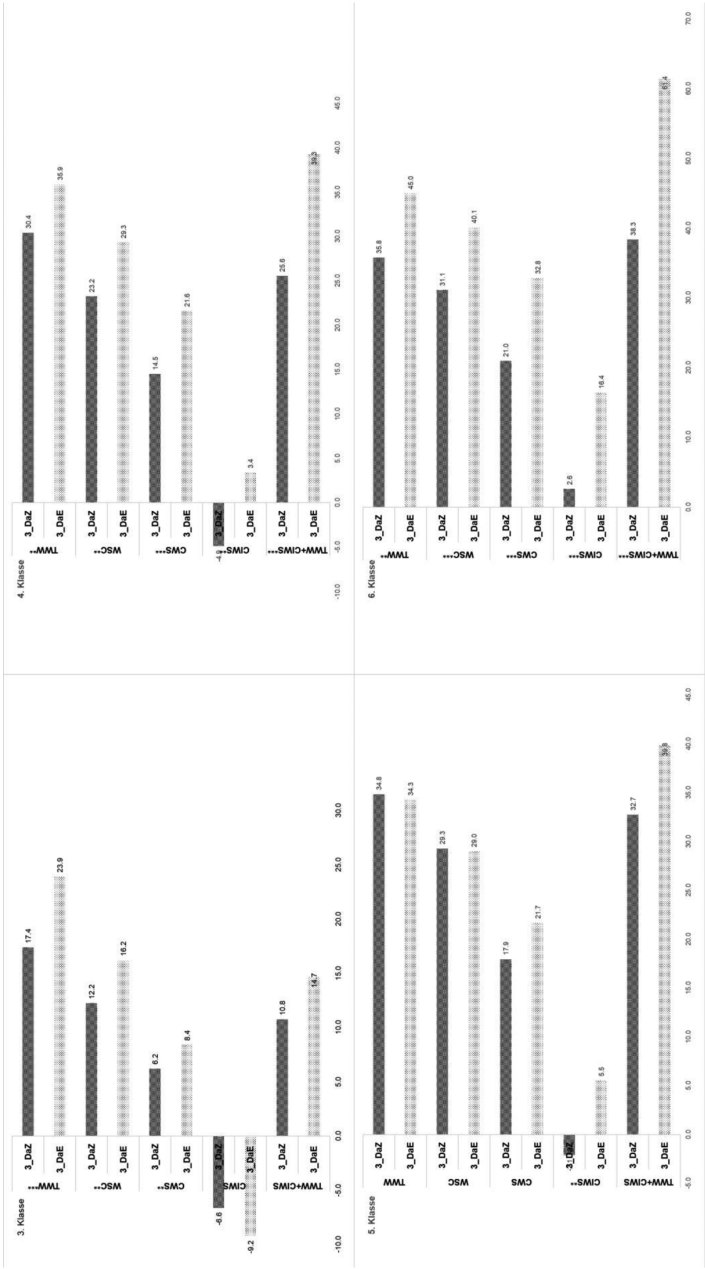


Abbildung 4: Grafik zu den erreichten Werten für verschiedene Klassenstufen und verschiedene Gruppen
Anm.: * = der T-Test für unabhängige Stichproben ergibt einen signifikanten Mittelwertsunterschied mit kleiner Effektstärke (Cohens $d \geq .20$); ** = mit mittlerer Effektstärke (Cohens $d \geq .50$); *** = mit großer Effektstärke (Cohens $d \geq .80$)

5.2 Reliabilität des Verfahrens für Kinder mit DaE und Kinder mit DaZ

Wie aus Tabelle 3 hervorgeht, sind in der dritten Klassenstufe die Paralleltest-Reliabilitäten bei Kindern mit DaE durchgehend höher als für Schülerinnen und Schüler mit DaZ. Dieses Bild zeigt sich auch in Stufe vier für die Auswertungsmethoden TWW, WSC und CIWS, nicht aber für CWS und TWW + CIWS, für welche die Reliabilitäten in vergleichbarer Höhe ausfallen. In Klassenstufe 5, welche laut den Resultaten aus Abschnitt 5.1 auch dadurch auffällt, dass dort keine Mittelwertsunterschiede zwischen DaZ und DaE festgestellt werden können (mit Ausnahme von CIWS), sind die Paralleltest-Reliabilitäten für Kinder mit DaZ höher als für die monolingual aufwachsenden Peers. Die Unterschiede zwischen den durchschnittlichen Korrelationen sind dabei aber überwiegend gering. Dies gilt für TWW und WSC auch in der sechsten Klasse. Bei den *Accurate Production Indices*, welche für die höheren Klassenstufen empfohlen werden, sind die Paralleltest-Reliabilitäten für Kinder mit DaE höher.

Tabelle 3: Paralleltest-Reliabilitäten von drei Schreibproben für die verschiedenen Kennwerte

Stufe und Gruppe	TWW	WSC	CWS	CIWS	TWW + CIWS
3_DaE	.55**	.62**	.70**	.33	.64**
	.63**	.65**	.61**	.50**	.54**
	.58**	.60**	.68**	.57**	.65**
	.58	.62	.66	.47	.61
3_DaZ	.62**	.40	.52*	.72**	.58**
	.15	.11	.16	-.01	.23
	.49*	.51*	.24	-.09;	.30
	.42	.34	.31	.209	.37
4_DaE	.79**	.78**	.72**	.66**	.74**
	.74**	.68**	.61**	.63**	.60**
	.78**	.74**	.68**	.70**	.67**
	.77	.74	.67	.66	.67
4_DaZ	.62**	.58**	.78**	.67**	.78**
	.50**	.62**	.68**	.53**	.68**
	.24	.42*	.64**	.56**	.61**
	.45	.54	.70	.59	.69
5_DaE	.50**	.67**	.72**	.64**	.67**
	.62**	.62**	.66**	.67**	.66**
	.33*	.48**	.62**	.69**	.62**
	.48	.59	.67	.67	.65

Tabelle 3: (fortgesetzt)

Stufe und Gruppe	TWW	WSC	CWS	CIWS	TWW + CIWS
5_DaZ	.66**	.67**	.72**	.78**	.72**
	.59**	.63**	.68**	.65**	.68**
	.64**	.68**	.64**	.67*	.63*
	.63	.66	.68	.70	.67
6_DaE	.55**	.57**	.56**	.66**	.67**
	.70**	.68**	.76**	.59**	.78**
	.64**	.72**	.67**	.64**	.67**
	.63	.65	.66	.63	.71
6_DaZ	.68**	.62**	.41**	.60**	.38*
	.62**	.61**	.56**	.54**	.55**
	.65**	.63**	.52**	.45**	.46**
	.65	.62	.50	.53	.46

Anm.: r = Korrelationskoeffizient nach Pearson; Korrelationen jeweils zwischen SS2 und SS3; SS2 und SS4; SS4 und SS3; Mittelwert (fett markiert). * = die Korrelation ist auf dem Niveau von 0.05 (2-seitig) signifikant; ** = die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant.

5.3 Validität des Verfahrens für Kinder mit DaE und Kinder mit DaZ

Die Validität von LVD – Schreiben sollte mindestens $r = .50$ betragen, um als ausreichend für den praktischen Einsatz als Screening oder als Lernverlaufsdiagnostik angesehen zu werden (Allen et al. 2019). Diesen Schwellenwert erreicht in der sechsten Klassenstufe keine Auswertungsmethode, in Klasse 3 nur TWW+CIWS für Kinder mit DaZ. In den Stufen 4 und 5 erweisen sich die einfachen produktionsabhängigen Werte TWW und WSC ebenfalls als nicht zufriedenstellend valide. Mit den Maßen CWS, CIWS und TWW+CIWS hingegen ergeben sich hohe Korrelationen mit der Einschätzung der Lehrperson.

Tabelle 4: Korrelationen zwischen den einzelnen Kennwerten und der Einschätzung durch die Lehrperson

Stufe und Gruppe	TWW	WSC	CWS	CIWS	TWW + CIWS
3_DaE	.04	.19	.16	.35	.20
3_DaZ	.16	.27	.48*	.14	.50*
4_DaE	.03	.31*	.58**	.72**	.60**
4_DaZ	.05	.30	.49**	.61**	.52**
5_DaE	.42**	.56**	.67**	.63**	.68**
5_DaZ	.35*	.46**	.69**	.61**	.69**
6_DaE	.03	.15	.25	.29	.26
6_DaZ	-.13	.01	.22	.40*	.26

Anm.: ** = die Korrelation ist auf dem 0,01-Niveau signifikant; * = die Korrelation ist auf dem 0,05-Niveau signifikant; Korrelationskoeffizient Spearman-Rho für ordinalskalierte Daten (Einschätzung der Lehrperson)

6 Zusammenfassende Darstellung und Diskussion

Die zusammenfassende Darstellung in Abbildung 5 verdeutlicht folgende Aspekte: Auf allen Klassenstufen zeigen sich zwischen Kindern mit Deutsch als Erstsprache und Kindern mit Deutsch als Zweitsprache bedeutsame Unterschiede in der durch LVD gemessenen Schreibflüssigkeit, wobei sich diese Unterschiede nicht zwingend für alle untersuchten Kennwerte ergeben. Dabei stellen sich CWS (korrekte Schreibsequenzen) und CIWS (korrekte minus inkorrekte Schreibsequenzen) als diejenigen Kennwerte heraus, die – mit Ausnahme der Gruppe 6_DaZ – die besten Werte für die Reliabilität erreichen. Das Gesamtbild für alle Klassenstufen und Kindergruppen weist darauf hin, dass sowohl die Reliabilität als auch die Validität von LVD – Schreiben noch weiterer Untersuchungen bedürfen. Unterschieden zwischen der Schreibflüssigkeit von Kindern mit DaZ und DaE sollte dabei besondere Aufmerksamkeit gelten. Aus den vorliegenden Ergebnissen ergeben sich für Lehrpersonen folgende praktische Schlussfolgerungen:

Die Auswertungen zu Fragestellung 1 (Leistungsunterschiede zwischen Kindern mit DaE und DaZ) dokumentieren, dass Schülerinnen und Schüler mit DaZ auch in den höheren Klassenstufen noch signifikant niedrigere Werte in der Schreibflüssigkeit erreichen als ihre Peers mit DaE. Dies ist überraschend, denn es wäre plausibel anzunehmen, dass die Kinder mit DaZ der 3. Klassenstufe zwar deutlich stärker als die Kinder mit DaE mit hierarchieniedrigeren Kompetenzen

beschäftigt sind (Campbell/Espin/McMaster 2013), sich dies aber durch den Kontakt mit der Schulsprache im Lauf der Schulzeit ändert und sich Unterschiede in der Folge tendenziell nivellieren. Die vorliegende Studie weist stattdessen darauf hin, dass bei der komplexen und anspruchsvollen Tätigkeit des Schreibens über alle Klassenstufen hinweg Differenzen zwischen Kindern mit DaE und DaZ manifest werden. Diese Beobachtung unterstreicht die Bedeutung der individuellen Bezugsnorm, denn eine ausschließliche Anwendung der sozialen Bezugsnorm benachteiligt systematisch Lernende mit DaZ. Der Blick sollte also nicht nur auf das Lernergebnis, sondern auch auf den Lernprozess (d.h. die Lernkurve) gerichtet sein, wozu sich LVD vorzüglich eignet. Dieser Einsatzbereich von LVD wurde in der vorliegenden Studie nicht angewendet, aber in Abschnitt 2.1 in Grundzügen beschrieben. Zur Anwendung in der Funktion von Lernverlaufsdiagnostik werden grafische Darstellungen verwendet, welche die Lernentwicklung über einen längeren Zeitraum visuell veranschaulichen. Die Perspektive der Lehrperson richtet sich damit speziell auch auf den Lehr-Lern-Prozess und auf die Frage, ob alle Schülerinnen und Schüler in ausreichendem Maße vom angebotenen Unterricht profitieren können, was zur Reflexion des eigenen Unterrichts anregen kann.

Die Resultate zu Fragestellung 2 (Reliabilität von LVD – Schreiben) illustrieren, dass Verfahren, welche für Kinder mit DaE in einem bestimmten Alter genaue und verlässliche Messergebnisse liefern, unter Umständen für Kinder mit DaZ nicht adäquat sind und sich dieser Effekt mit zunehmendem Alter auch umdrehen kann. Deutlich wird das beispielsweise in Bezug auf die Paralleltest-Reliabilität von LVD – Schreiben in der dritten Klasse, die für Kinder mit DaZ deutlich niedriger ausfällt als für Kinder mit DaE. In der fünften Klassenstufe hingegen kehrt sich dieses Verhältnis um. Die Verfügbarkeit verschiedener Scoringmethoden für LVD – Schreiben führt somit dazu, dass sorgfältig ausgewählt werden sollte, welche Methode für welche Gruppe von Kindern in welcher Klassenstufe sinnvoll einsetzbar ist.

Fragestellung 3 beinhaltet explorative Analysen zur Validität von LVD – Schreiben unter Einsatz des Urteils der Lehrperson als Kriteriumsmaß. Als Kernresultat kann hier bezeichnet werden, dass die einfachen und beliebten Auswertungsmaße TWW und WSC keine guten Indikatoren für die Schreibkompetenz von Schülerinnen und Schülern darstellen. Bei allen Lernenden rechtfertigen komplexere Auswertungsmethoden den leicht erhöhten Aufwand, was für Kinder mit DaE und DaZ gleichermaßen gilt.

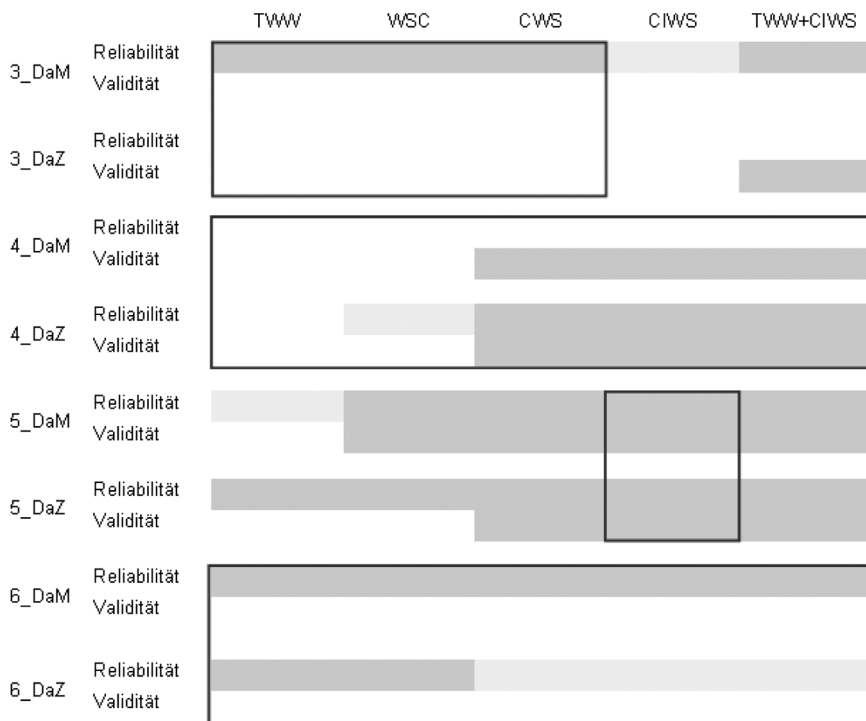


Abbildung 5: Zusammenfassende Darstellung der Ergebnisse zur Reliabilität und zur Validität

Anm.: Hervorhebungen bei Reliabilität: dunkle Balken = Reliabilität: $\alpha \geq .80$; helle Balken = Reliabilität: $\alpha \geq .70$; Hervorhebungen bei Validität: dunkle Balken = Korrelation nach Spearman's Rho mit Einschätzung der Lehrperson signifikant mit $r \geq .50$; dunkel gerahmt: Kennwerte, für die sich signifikante Mittelwertsunterschiede zwischen Kindern mit DaE und DaZ zeigen.

Die vorliegende Studie unterliegt gewissen Limitationen: Der Umfang der Gesamtstichprobe und der Teilstichproben ist eher gering, aber vergleichbar mit Studien aus dem englischsprachigen Raum, so etwa mit der Studie von Keller-Margulis et al. (2016). Hier steht die Forschung also ganz allgemein noch am Anfang und muss zwingend durch größer angelegte Studien weitergeführt werden. Die Einteilung der Kinder in Kinder mit DaE und DaZ könnte für solche Folgestudien differenzierter und zum Beispiel anhand der Kontaktzeit mit der Zweitsprache erfolgen. Ebenso könnte mitberücksichtigt werden, welche Kinder in einem bildungsnahen respektive bildungsfernen Elternhaus aufwachsen, da sich der Einfluss des sozioökonomischen Status für die schulischen Leistungen oftmals als bedeutsam erweist (Steinig 2013). Schließlich ist anzufügen, dass die hier präsentierte Auswertung Ergebnisse zu einem einzelnen Messzeitpunkt präsentiert. Aus-

geklammert bleiben die Fragen, wie sich Kinder mit DaE und DaZ über die Zeit hinweg entwickeln und – in diesem Zusammenhang – wie die Änderungssensitivität von LVD – Schreiben für verschiedene Kindergruppen aussieht.

LVD – Schreiben wird derzeit im deutschsprachigen Raum in der Praxis noch nicht eingesetzt, da diese Methode hier bislang weitgehend unbekannt ist. Angesichts der Schwierigkeiten von Lehrpersonen, das komplexe Konstrukt der Schreibkompetenz auf zuverlässige, zeitökonomische und valide Art und Weise zu erfassen, könnte LVD – Schreiben eine willkommene Perspektive im diagnostischen Repertoire darstellen. Dies gilt insbesondere deshalb, weil sie neben der Statusdiagnostik (Screening aller Schülerinnen und Schüler) auch in der Funktion als Lernverlaufsdiagnostik genutzt werden kann. Wie die Datenauswertung im Rahmen des vorliegenden Beitrags zeigt, lohnt sich bei der weiteren Erforschung und Entwicklung von LVD – Schreiben die getrennte Betrachtung verschiedener Klassenstufen und vor allem auch verschiedener Gruppen von Kindern.

Literaturverzeichnis

- Allen, Abigail A.; Jung, Pyung-Gang; Poch, Apryl L.; Brandes, Dana; Shin, Jaehyun; Lembke, Erica S.; McMaster, Kristen L. (2019): „Technical Adequacy of Curriculum-Based Measures in Writing in Grades 1–3“. In: *Reading and Writing Quarterly* 33, 1–25.
- Campbell, Heather M. (2010): „The technical adequacy of curriculum-based measurement passage copying with secondary school English language learners“. In: *Reading and Writing Quarterly* 26 (4), 289–307.
- Campbell, Heather M.; Espin, Christine A.; McMaster, Kristen (2013): „The technical adequacy of curriculum-based writing measures with English learners“. In: *Reading and Writing* 26 (3), 431–452.
- Carigiet Reinhard, Tamara (2012): *Schulleistungen und Heterogenität: Eine mehrbenenanalytische Untersuchung der Bedingungsfaktoren der Schulleistungen am Ende der dritten Primarschulklasse*. Bern: Haupt.
- Espin, Christine A.; Wallace, Teri; Campbell, Heather M.; Lembke, Erica S.; Long, Jeffrey D.; Ticha, Renata (2008): „Curriculum-Based Measurement in Writing: Predicting the Success of High-School Students on State Standards Tests“. In: *Exceptional Children* 74 (2), 174–193.
- Fuchs, Lynn S. (2017): „Curriculum-Based Measurement as the Emerging Alternative: Three Decades Later“. In: *Learning Disabilities Research & Practice* 32 (1), 5–7.
- Grosche, Michael; Casale, Gino (2021): „Response-to-Intervention (RTI) und schulische Inklusion – Grundlegender Widerspruch oder zwei Seiten einer Medaille?“. In: Mähler, Claudia; Hasselhorn, Marcus (Hrsg.): *Inklusion: Chancen und Herausforderungen*. Göttingen: Hogrefe, 169–183.
- Hosp, John L.; Kaldenberg, Erica (2020): „What is Writing Assessment for Tiered Decision Making?“. In: Dunn, Michael (Hrsg.): *Writing Instruction and Intervention for Struggling Writers: Multi-Tiered Systems of Support*. Newcastle-upon-Tyne, England: Cambridge Scholars Publisher, 70–85.

- Hosp, Michelle K.; Hosp, John L.; Howell, Kenneth W. (2007): *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York: The Guilford Press (The Guilford practical intervention in the schools series).
- Huber, Christian; Grosche, Michael (2012): „Das response-to-intervention-Modell als Grundlage für einen inklusiven Paradigmenwechsel in der Sonderpädagogik“. In: *Zeitschrift für Heilpädagogik* 8), 312–322.
- Jewell, Jennifer; Malecki, Christine Kerres (2005): „The Utility of CBM Written Language Indices: An Investigation of Production-Dependent, Production-Independent, and Accurate-Production Scores“. In: *School Psychology Review* 34 (1), 27–44.
- Keller-Margulis, Milena A.; Payan, Anita M.; Jaspers, Kathryn E.; Brewton, Christie (2016): „Validity and Diagnostic Accuracy of Written Expression Curriculum-Based Measurement for Students With Diverse Language Backgrounds“. In: *Reading & Writing Quarterly* 32 (2), 174–198.
- Keller-Margulis, Milena A.; Mercer, Sterett H.; Matta, Michael (2021): „Validity of automated text evaluation tools for written-expression curriculum-based measurement: a comparison study“. In: *Reading and Writing* 34 (10), 2461–2480.
- Kim, Young-Suk Grace; Gatlin, Brandy; Al Otaiba, Stephanie; Wanzek, Jeanne (2017): „Theorization and an Empirical Investigation of the Component-Based and Developmental Text Writing Fluency Construct“. In: *Journal of learning disabilities* 51 (4), 1–16.
- Klauer, Karl Josef (2011): „Lernverlaufsdiagnostik – Konzept, Schwierigkeiten und Möglichkeiten“. In: *Empirische Sonderpädagogik* (3), 207–224.
- McMaster, Kristen; Espin, Christine A. (2007): „Technical Features of Curriculum-Based Measurement in Writing“. In: *The Journal of Special Education* 41 (2), 68–84.
- Payan, Anita M.; Keller-Margulis A., Milena; Burrige, Andrea B.; McQuillin, Samuel D.; Hassett, Kristen S. (2019): „Assessing Teacher Usability of Written Expression Curriculum-Based Measurement“. In: *Assessment for Effective Intervention* 45 (1), 51–64.
- Romig, John Elwood; Therrien, William J.; Lloyd, John W. (2017): „Meta-Analysis of Criterion Validity for Curriculum-Based Measurement in Written Language“. In: *The Journal of Special Education* 51 (2), 72–82.
- Saddler, Bruce; Asaro-Saddler, Kristie (2013): „Response to Intervention in Writing: A Suggested Framework for Screening, Intervention, and Progress Monitoring“. In: *Reading & Writing Quarterly* 29 (1), 20–43.
- Steinig, Wolfgang (2013): „Rechtschreibung: Schicht, nicht Mehrsprachigkeit macht den Unterschied“. In: *zmi-Magazin*, 6–10.
- Sturm, Afra (2017): „Förderung hierarchieniedriger Schreibprozesse“. In: Philipp, Maik (Hrsg.): *Handbuch: Schriftspracherwerb und weiterführendes Lesen und Schreiben*. Weinheim: Beltz Juventa, 266–284.
- Sturm, Afra; Nänny, Rebekka; Wyss, Stefanie (2017): „Entwicklung hierarchieniedriger Schreibprozesse“. In: Philipp, Maik (Hrsg.): *Handbuch: Schriftspracherwerb und weiterführendes Lesen und Schreiben*. Weinheim: Beltz Juventa, 84–104.
- Voß, Stefan; Blumenthal, Yvonne; Ehrich, Kerstin; Mahlau, Kathrin (2020): „Multimodale Diagnostik als Ausgangspunkt für spezifische Förderung: Eine Darlegung am Beispiel der Rechtschreibung“. In: *Zeitschrift für Heilpädagogik* 71 (3), 88–99.
- Weissenburger, Jacalyn Wright; Espin, Christine A. (2005): „Curriculum-based measures of writing across grade levels“. In: *Journal of School Psychology* 43 (2), 153–169.
- Wilson, Joshua; Chen, Dandan; Sandbank, Micheal P.; Hebert, Michael (2019): „Generalizability of automated scores of writing quality in Grades 3–5“. In: *Journal of educational psychology* 111 (4), 619–640.

- Winkes, Julia; Schaller, Pascale (2022a): „Lernverlaufsdiagnostik Schreiben (LVD – Schreiben): Reliabilität, Validität und Sensitivität für mittelfristige Lernfortschritte im deutschsprachigen Raum“. In: *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete* 91, 1–26.
- Winkes, Julia; Schaller, Pascale (2022b): „Generalizability of Written Expression Curriculum-Based-Measurement in the German Language: What Are the Major Sources of Variability?“. *Frontiers in Education* 7: 919756. Doi: 10.3389/feduc.2022.919756

Biographische Angaben

Pascale Schaller

ist als Dozentin an der PHBern, Institut Primarstufe, und als Lehrbeauftragte an der Universität Freiburg CH im Departement für Sonderpädagogik, Abteilung Logopädie tätig. Nach dem Studium in Germanistik und Geschichte sowie dem Lehramtsstudium hat sie an der Universität Freiburg CH in Germanistischer Linguistik promoviert. Ihre Forschungsinteressen liegen in den Bereichen Spracherwerb, Sprachdiagnostik und Sprachförderung.

Julia Winkes

arbeitet als Lektorin an der Universität Freiburg CH im Departement für Sonderpädagogik, Abteilung Logopädie und Abteilung Schulische Heilpädagogik. Sie studierte an der Universität zu Köln Sprachheilpädagogik und promovierte an der Universität Freiburg CH. Ihre Forschungsthemen umfassen Fragen des normalen und beeinträchtigten Sprach- und Schriftspracherwerbs.