



Das Nachsprechen von Nichtwörtern zur Testung bilingualer Kinder*

Zum Einsatz der einzelsprachunabhängigen Items des LITMUS-CL-NWRT zur Differenzierung von Kindern mit und ohne Sprachentwicklungsstörungen

Testing bilingual children with a Nonword Repetition Task

Using the language-independent items of LITMUS-CL-NWRT to distinguish between children with and without Developmental Language Disorders

Juliane Hinnerichs

Zusammenfassung

Hintergrund: Die Differenzierung zwischen bilingualen Kindern, bei denen eine Sprachentwicklungsstörung (SES) vorliegt und denen, deren geringe Deutschkenntnisse lediglich auf unzureichenden Input zurückzuführen sind, gilt als Herausforderung. Die Überprüfung beider/aller erworbener Sprachen ist oft nicht möglich, deshalb scheint das Nachsprechen von Nichtwörtern (NWRT) als einzelsprachunabhängige Methode vielversprechend. Der LITMUS-CL-NWRT wurde gezielt zur Testung bilingualer Kindern mit diversen sprachlichen Hintergründen konstruiert.

Ziele: Nach einer ersten Studie zur Auswirkung verschiedener Auswertungsvorgehensweisen auf die Ergebnisse des LITMUS-CL-NWRT soll dessen Eignung zur Differenzierung von bilingualen Kindern mit und ohne SES überprüft werden. Anschließend soll untersucht werden, ob der LITMUS-CL-NWRT dabei eine bessere diagnostische Validität aufweist als ein einzelsprachspezifischer NWRT (LS-NWRT).

Methode: 103 monolinguale und bilinguale Kinder zwischen 3;7 und 13;9 Jahren wurden im Rahmen des BiliSAT-Projekts getestet und daraufhin jeweils in eine Gruppe mit und ohne SES aufgeteilt. Die NWRT-Ergebnisse wurden mit einer Kovarianzanalyse verglichen, anschließend wurde die Sensitivität und Spezifität der Testungen berechnet.

Ergebnisse: Beide NWRTs differenzieren Kinder mit und ohne Sprachentwicklungsstörungen voneinander unabhängig davon, ob diese bilingual oder monolingual aufwachsen. Der LITMUS-CL-NWRT zeigt bessere Ergebnisse zur diagnostischen Validität als der einzelsprachspezifische NWRT.

Schlussfolgerungen: Sowohl die Itemkonstruktion als auch die Auswertungsvorgehensweise beeinflussen die Ergebnisse von NWRTs. Einzelsprachunabhängig konstruierte Nichtwörter eignen sich besser als LS-NWRTs zur Testung bilingualer Kinder.

Schlüsselwörter

Bilingualismus, Sprachentwicklungsstörungen, Nachsprechen von Nichtwörtern

* Dieser Beitrag hat das double-blind Peer-Review-Verfahren durchlaufen.

Abstract

Background: The distinction between children with Developmental Language Disorders (DLD) and those with low German language skills simply due to insufficient input is regarded as a challenge. The assessment of both/all acquired languages is often impossible, thus making approaches based on repetition of nonwords a promising language-independent tool. The LITMUS-CL-NWRT was specifically designed to test bilingual children with various language backgrounds.

Aims: After a first study on the impact of different scoring procedures on the results of the LITMUS-CL-NWRT, its suitability to distinguish between bilingual children with or without DLD is to be assessed. Subsequently, it has to be checked if the diagnostic accuracy of the LITMUS-CL-NWRT is higher than that of a language-specific NWRT (LS-NWRT).

Methods: 103 monolingual and bilingual children, aged 3;7 - 13;9, were tested within the BiliSAT-project and divided up into two groups with and without DLD. The results of the NWRTs were compared with an analysis of covariance, afterwards test sensitivity and specificity were determined.

Results: Independent of bilingualism as an influencing factor, with both NWRTs, children with and without DLD can be distinguished. The LITMUS-CL-NWRT shows a higher diagnostic accuracy than the LS-NWRT.

Conclusion: The item construction as well as the scoring procedure affect the results of NWRTs. Language-independent constructed nonwords are more suitable than LS-NWRTs to test bilingual children.

Keywords

Bilingualism, Developmental Language Disorders, Nonword Repetition Test

1 Einleitung – Zur Testung von bilingualen Kindern

Der Umgang mit bilingualen¹ Kindern im sprachtherapeutischen Kontext ist eine Herausforderung. Ein Drittel der befragten Sprachtherapeut:innen geben an, bei der Diagnostik mehrsprachiger Kinder überfordert zu sein und mehr als die Hälfte fühlt sich bei der Auswahl geeigneter Testverfahren unsicher (Marxen, Lüke & Cholin 2022, S.103). Lange Zeit wurden Testungen und Therapieansätze in der Annahme entwickelt, dass Kinder lediglich eine Sprache erwerben. In den letzten Jahren entstand zunehmend ein Bewusstsein für die steigende Anzahl bilingualer Kinder in Deutschland, wobei „praxistaugliche sprachwissenschaftliche fundierte Verfahren für mehrsprachig aufwachsende Kinder in Deutschland“ noch in der Entwicklung sind (Chilla 2022, S.125). Die korrekte Einschätzung der sprachlichen Fähigkeiten bilingualer Kinder ist schwierig. Ein Grund hierfür ist die Überschneidung der Symptome einer SES mit den Phänomenen des unauffälligen Zweitspracherwerbs (Chilla 2022; Scherger 2022). Ist der Wortschatz im Deutschen aufgrund des begrenzten Inputs so klein oder liegt eine Störung im semantisch-lexikalischen Bereich vor? Wurde die korrekte Verbstellung aufgrund der kurzen Sprachkontaktdauer noch nicht erworben oder ist der Erwerb grammatischer Strukturen generell beeinträchtigt? Die Berichte von Fehldiagnosen zeigen, dass es in der Praxis schwierig ist, die sprachliche Entwicklung bilingualer Kinder korrekt einzuschätzen (Chilla 2022; Grimm & Schulz 2014; Rothweiler 2007).

Die Prävalenzrate für eine SES², die „nicht auf sensorische, organische, mentale oder gravierende sozial-emotionale Defizite zurückgeführt“ werden kann, beträgt circa 5-8% (De Langen-Müller, Kauschke, Kiese-Himmel, Neumann & Noterdaeme 2011, S.32), unabhängig davon, ob eine oder mehrere Sprachen erworben werden. Hierbei wird immer wieder betont, dass bei mehrsprachigen Kindern alle Sprachen von der Störung betroffen sein müssen (Chilla 2022). Sobald sich in einer Sprache eine unauffällige Entwicklung abzeichnet, können keine angeborenen Spracherwerbsdefizite vorliegen. Die Auffälligkeiten in der bzw. den anderen Sprachen müssen dann auf die Inputsituation zurückzuführen sein.

Idealerweise müssten also die Fähigkeiten in allen erworbenen Sprachen überprüft werden, dies ist jedoch oft nicht möglich. So mangelt es einerseits an normierten Testverfahren in der Erstsprache sowie andererseits auch an Sprachtherapeut:innen, die diese beherrschen. Des Weiteren gilt es als problematisch, die Normwerte monolingualer Kinder für den bilingualen Erwerbskontext zu verwenden, unabhängig davon, ob es sich um die Erst- oder Zweitsprache handelt (Chilla 2022). Für die meisten Sprachtestungen im Deutschen liegen keine gesonderten

1 Der Begriff bilingual wird hier für Kinder verwendet, die mindestens zwei, gegebenenfalls aber auch drei oder mehr Sprachen erwerben.

2 Andere Bezeichnungen: Umschriebene SES (USES), Spezifische SES (SSES); im Englischen nach aktuellem Konsens Developmental Language Disorder (DLD; Bishop, Snowling, Thompson & Greenhalgh 2017)

Normwerte für bilinguale Kinder vor. Wenn bilinguale Normen erhoben wurden, so scheint es fast unmöglich mit diesen die vielfältigen bilingualen Spracherwerbsbiografien zufriedenstellend abzudecken, wie beispielsweise den simultanen, frühen oder späten sukzessiven Erwerb und die darauf folgenden Variationen der Sprachkontaktdauer. So enthält beispielsweise die Linguistische Sprachstandserhebung – Deutsch als Zweitsprache (LiSe-DaZ; Schulz & Tracy 2011) für bilinguale Kinder mehrere Normwerttabellen pro Altersgruppe, die je nach Kontaktdauer zum Deutschen ausgewählt werden können. Dennoch gibt es natürlich immer wieder Kinder, für deren Alter und Kontaktdauer keine entsprechende Normwerttabelle vorliegt.

Somit wird offensichtlich, dass bei bilingualen Kindern die Einordnung der sprachlichen Fähigkeiten in der Einzelsprache als „auffällig“ und „unauffällig“ oft schwierig ist. Eine sinnvolle Unterstützung bei der Differentialdiagnose von bilingualen Kindern wäre demnach ein Verfahren, das zugrundeliegende Sprachverarbeitungsfähigkeiten unabhängig von den Einzelsprachen erfasst.

1.1 Nachsprechen von Nichtwörtern

Das Nachsprechen von Nichtwörtern wird in diesem Kontext als vielversprechende Methode angesehen, da hiermit einzelsprachunabhängige Fähigkeiten überprüft werden sollen. Mit diesem Verfahren sollen also nicht bereits erworbene sprachliche Fähigkeiten (zum Beispiel im Bereich Lexikon oder Grammatik) untersucht werden, sondern Prozesse, von denen angenommen wird, dass sie der Sprachverarbeitung zugrunde liegen. Das Ziel ist hierbei, die Funktionsweise zugrundeliegender Spracherwerbsmechanismen unabhängig von einzelsprachspezifischem Wissen zu testen und hiermit den diagnostischen Prozess bei bilingualen Kindern mit Verdacht auf SES zu erleichtern.

Die Annahme, dass beim Nachsprechen von Nichtwörtern einzelsprachunabhängige Fähigkeiten überprüft werden, wird damit begründet, dass die Items neu konstruierte Lautfolgen sind, die keine realen Wörter repräsentieren und somit für alle Kinder gleichermaßen unbekannt sind. Mit dem Nachsprechen von Nichtwörtern soll die Leistungsfähigkeit des phonologischen Kurzzeitgedächtnisses³ gemessen werden, da die kurzfristige Speicherung von neuem verbalen Input als essentielle Voraussetzung für den Spracherwerb gilt (Gathercole & Baddeley 1990; Hasselhorn & Körner 1997; Pierce, Genesee, Delcenserie & Morgan 2017). Der NWRT ist ein bereits seit langem im sprachtherapeutischen Alltag eingesetztes Testinstrument, einer aktuellen Umfrage nach setzen ihn zwei Drittel der Sprachtherapeut:innen ein (Heeg et al. 2022). Die Idee, unbekannte Lautfolgen nachzusprechen, geht auf Grete Mottier (1951) zurück und wurde von ihr zur Testung von Kindern mit Lese-Rechtschreib-Störungen (LRS) entwickelt (Gamper, Keller, Messerli, Moser & Wüst, 2012). Heutzutage werden NWRTs in diversen Testungen und Sprachscreenings zur Erkennung von LRS und auch SES eingesetzt, sowie auch in einigen Sprachstandsverfahren⁴.

Jedoch wurde in Bezug auf die bislang eingesetzten Nichtwörter mehrfach betont, dass diese bei weitem nicht so einzelsprachunabhängig sind wie teilweise angenommen (Ennemoser, Poupoua & Hartung 2012 u.a.). So wird an dem Beispiel des Items „Glösterkeit“ aus dem SETK 3-5 (Grimm 2015) deutlich, dass hier sowohl das deutsche Morphem „keit“ verwendet wird als auch die Konsonantenverbindung /gl/, welche nicht unbedingt in der entsprechenden Erstsprache eines bilingualen Kindes vorkommt. Ebenso kommt der Laut /œ/ zwar nicht nur im Deutschen vor, jedoch gibt es viele Sprachen dieser Welt, die diesen Vokal nicht in ihrem Inventar haben. Somit wird offensichtlich, dass dieses Item für ein monolinguales Kind in Deutschland wesentlich vertrauter klingt – selbst wenn es dieses noch nie zuvor gehört hat – als für ein Kind mit wenig Kontakt zum Deutschen.

Die Nachsprecheleistungen werden durch Vertrautheit mit der Sprache, nach der die Nichtwörter konzipiert wurden, beeinflusst: So zeigen zahlreiche Studien, dass bilinguale Kinder durch NWRTs benachteiligt werden, wenn die Items nach den Wortbildungsregeln der weniger dominanten Sprache konstruiert wurden (u. a. Messer et al. 2010; Parra, Hoff & Core 2011; Sorenson Duncan & Paradis 2016).

3 Dieser zugrundeliegende Mechanismus wird unter anderem auch als verbales Kurzzeitgedächtnis, verbales oder phonologisches Arbeitsgedächtnis oder phonologische Schleife bezeichnet (Rescorla 2014; Schöler & Brunner 2008; Schwob et al. 2021). Die meisten Autor:innen beziehen sich dabei auf das Arbeitsgedächtnismodell nach Baddeley & Hitch (1994).

4 Siehe unter anderem BISC von Jansen, Mannhaupt, Marx & Skowronek (2002); Delfin 4 von Fried, Briedigkeit, Isele & Schunder (2009); HASE von Schöler & Brunner (2008); KISS von Holler-Zittlau, Euler & Neumann (2011); PhoMo-Kids von Stadie & Schöppe (2013); SETK 3-5 von Grimm (2015) und SSV von Grimm, Aktas & Kießig (2003)

Grimm (2015, S.66/75) sieht in dieser Benachteiligung kein Problem, indem sie darauf verweist, dass bei bilingualen Kinder mit den Nichtwörtern des SETK 3-5 die schon erworbenen Sprachfähigkeiten des Deutschen untersucht werden und nicht – wie bei den monolingualen Kindern – das Vorliegen einer SES. Des Weiteren führt sie an, dass „man gar nicht erst ernsthaft darüber nachdenken“ möchte, wie Nichtwörter konstruiert werden müssten, um möglichst viele unterschiedliche Sprachen zu repräsentieren (Grimm 2015, S.75).

1.2 LITMUS-CL-NWRT

Im Rahmen des internationalen Forschungsnetzwerkes COST Action IS0804 stellte sich ein Team von Wissenschaftler:innen genau dieser Herausforderung. Es wurde ein NWRT entwickelt, dessen Items in ihrem Aufbau mit möglichst vielen Sprachen kompatibel sind, der LITMUS-CL-NWRT (Language Impairment Testing in Multilingual Settings – Crosslinguistic Nonword-Repetition-Test; Chiat 2015). Zugleich betont Chiat (2015, S.133), dass selbst diese Items nicht komplett neutral, also uneingeschränkt einzelsprachunabhängig sind: “However we construct non-words, they will be more or less like real word phonology in different languages with respect to length, prosody, segmental constituents and their frequency and phonetic realisation of segments”. So kann eine bestimmte Wortlänge oder Silbenstruktur in einer Sprache als typisch gelten, während sie in einer anderen Sprache wiederum eher selten oder gar nicht vorkommt (Chiat 2015, S.133).

Leonard (2014, S.3) erklärt beispielsweise die Tatsache, dass italienische Vorschulkinder drei- und viersilbige Nichtwörter besser nachsprechen können als englische Vorschulkinder, damit, dass italienische Wörter im Schnitt länger sind und die italienischen Kinder somit besser mit der Aussprache langer Wörter vertraut sind. Des Weiteren finden Szewczyk, Marecka, Chiat & Wodniecka (2018, S.33) in ihren Daten, dass polnische Kinder im Gegensatz zu englischsprachigen Kindern keine zusätzlichen Schwierigkeiten durch Konsonantenverbindungen innerhalb der ihnen präsentierten Nichtwörter haben und erklären dies damit, dass im Polnischen vergleichsweise viele Konsonantenverbindungen vorkommen.

Obwohl Nichtwörter also niemals gleichermaßen neutral in Bezug auf alle Sprachen der Welt konstruiert werden können, so zeigen bisherige Studien zum LITMUS-CL-NWRT vielversprechende Ergebnisse in Bezug auf dessen möglichst einzelsprachunabhängige Itemkonstruktion. So wurde die Testung bereits in über 15 Ländern erprobt und die hierbei erhaltenen Daten von mehr als 1800 monolingualen und bilingualen Proband:innen zeigen, dass die bilingualen Kinder, unabhängig von ihrer Kontaktdauer, beim LITMUS-CL-NWRT nicht schlechter als monolinguale Kinder abschneiden (Polišenská et al. 2022). Gleichzeitig finden sich in diesen Daten bedeutsame Unterschiede zwischen Kindern mit sprachlichen Auffälligkeiten und sprachlich unauffälligen Kindern (im Folgenden TD-Kinder für engl. Typically Developing). Des Weiteren haben Untersuchungen aus den Niederlanden und Irland gezeigt, dass sich der LITMUS-CL-NWRT besser zur Testung bilingualer Kinder eignet als ein einzelsprachspezifisch konstruierter NWRT (im Folgenden LS-NWRT für engl. Language Specific Nonword Repetition Test) (Antonijevic-Elliott et al. 2020; Boerma et al. 2015; Verhagen & Andringa 2021). Die Autor:innen zweier Metaanalysen zum Einsatz von NWRTs bei der Identifikation von Kindern mit SES kommen ebenfalls zu dem Schluss, dass sich einzelsprachunabhängige Nichtwörter besser zur Testung bilingualer Kinder eignen als einzelsprachspezifische Items (Ortiz 2021, S. 2275; Schwob et al. 2021, S.3589).

Der LITMUS-CL-NWRT wurde auch für Testungen im deutschsprachigen Raum adaptiert (Hinnerichs, Polišenská & Chiat 2016; siehe auch Hinnerichs 2016) und dort zunächst anhand einer kleinen Stichprobe (n=24) erprobt. Um die Testung mit einem LS-NWRT zu vergleichen wurde der Untertest Nachsprechen von Kunstwörtern des Heidelberger auditiven Screenings in der Einschulungsuntersuchung (HASE-NK; Schöler & Brunner 2008; siehe auch 3.2) herangezogen. Es zeigte sich hierbei, dass der LITMUS-CL-NWRT die monolingualen und bilingualen Kinder mit und ohne Sprachentwicklungsstörungen gut voneinander unterscheiden kann, jedoch nicht besser als der HASE-NK. Die Ergebnisse von Hinnerichs (2016) sollten jedoch aufgrund der kleinen Stichprobe mit Vorsicht betrachtet werden. In einer groß angelegten Untersuchung von Zaretsky, Lange & Hey (2022; n=1441) zeigte sich hingegen, dass die bilingualen Kinder in Deutschland durch die zwei eingesetzten LS-NWRTs benachteiligt werden, während sie im LITMUS-CL-NWRT sogar besser abschneiden als die monolingualen Kinder. Zaretsky et al. (2022) setzten jedoch nicht die Items des HASE-NK sondern die des KiSS (Holler-Zittlau, Euler & Neumann 2011) und des SSV (Grimm, Aktas & Kießig 2003) zum Vergleich mit dem LIT-

MUS-CL-NWRT ein. Da die Items dieser beiden LS-NWRTs im Vergleich zum HASE-NK insofern komplexer und einzelsprachspezifischer sind, dass deren Items Konsonantenverbindungen (Beispiel aus KiSS: Verklasenaft) und Morpheme des Deutschen (Beispiel aus SSV: Pristobierlichkeit) enthalten, können die Ergebnisse von Zaretsky et al. (2022) nur bedingt mit denen von Hinnerichs (2016) verglichen werden. Beim Vergleich von NWRTs sollte also nicht nur berücksichtigt werden, ob diese einzelsprachspezifisch oder einzelsprachunabhängig konstruiert wurden, die Itemkonstruktion muss in jedem Fall im Detail betrachtet werden. In Bezug auf die Itemkonstruktion von Nichtwörtern gibt es zahlreiche Faktoren – hier seien beispielhaft die Länge, die phonologische Komplexität und die Nähe zu realen Wörtern genannt – die einen Einfluss auf die Nachsprecheleistungen der Proband:innen ausüben können (für einen Überblick siehe Szewczyk et al. 2018).

1.3 Zum Auswertungsvorgehen bei NWRTs

Ein weiterer Punkt, der Unterschiede zwischen NWRT-Studien hervorrufen kann, ist das Auswertungsvorgehen. Während das Vorgehen bei Elisionen einheitlich ist – die Tilgung von Lauten oder Silben wird immer als Fehler gewertet –, so variiert die Vorgehensweise bei Additionen und Substitutionen zwischen den Studien. Laut- und Silbenadditionen werden zwar in den meisten Studien als Fehler gewertet, jedoch schließen Boerma et al. (2015, S.1753) diese mit der Begründung, dass dabei lediglich etwas hinzugefügt wird, aber kein Informationsverlust verzeichnet werden kann, von der Fehlerwertung aus.

In Bezug auf Substitutionen wird meist zwischen Vokalen und Konsonanten unterschieden. Während minimale Vokalabweichungen oft nicht als Fehler gewertet werden (beispielsweise [u/o] oder [i/ə]; Chiat, Polišenská, Szewczyk & Kapalková 2016 und Tuller et al. 2018), so schließen Abed Ibrahim & Fekete (2019) jegliche Vokalabweichungen bei der Fehlerwertung aus. Bei Konsonanten werden oft Abweichungen bei der Stimmgebung toleriert, da bilinguale Kinder hierdurch benachteiligt werden können. So wird Stimmhaftigkeit je nach Sprache unterschiedlich realisiert (Dos Santos & Ferré 2016, S.5) und zudem liegt nicht in allen Sprachen die stimmhafte und stimmlose Variante eines Lautes vor. Im Standardarabischen gibt es beispielsweise nur den stimmhaften Labial /b/, nicht aber das stimmlose /p/ (Watson 2002, S.14). Des Weiteren werden oft konstante Substitutionen von Konsonanten bei der Fehlerwertung ausgeschlossen, da diese als „unreife“ Sprache oder Artikulationsschwierigkeit eingestuft werden und nicht als Symptom einer Sprachentwicklungsstörung (Abed Ibrahim & Fekete 2019, S.11; Boerma et al. 2015, S.1753; Chiat et al. 2016). Hierbei wird nicht berücksichtigt, dass konstante Substitutionen nicht nur ein Symptom phonetischer, also artikulatorischer Defizite sein können, sondern auch auf phonologische Schwierigkeiten hinweisen können, welche im deutschsprachigen Raum als Symptom einer Sprachentwicklungsstörung eingestuft werden (Kauschke & Vogt 2019, S. 179). Ebenso sprechen die Beobachtungen, dass Aussprachestörungen oft mit einer Sprachentwicklungsstörung einhergehen können (Vuolo & Goffmann 2020, S.2) und phonetische und phonologische Defizite ohnehin oft nicht klar voneinander abgrenzbar sind (Weinrich & Zehner 2017, S.37/38), dafür, konstante Substitutionen im Rahmen eines NWRTs als Fehler zu werten (Hinnerichs in Vorb.).

2 Zielsetzung und Forschungsfragen

Das Ziel der vorliegenden Untersuchung ist, zunächst anhand der Manipulation verschiedener Auswertungskriterien eine Auswertungsvorgehensweise festzulegen, die bestmöglich zwischen Kindern mit und ohne SES differenziert ohne dabei bilinguale Kinder zu benachteiligen (Studie 1). Daraufhin soll unter Anwendung dieser Auswertungsvorgehensweise untersucht werden, ob sich der LITMUS-CL-NWRT als Testinstrument im bilingualen Kontext eignet und dabei bessere Ergebnisse als ein bereits etablierter LS-NWRT aufweist (Studie 2).

Hierfür sollen folgende Forschungsfragen beantwortet werden:

- Inwiefern wird die Differenzierungsfähigkeit des LITMUS-CL-NWRT beeinflusst, wenn Additionen, Vokalsubstitutionen und konstante Substitutionen als Fehler gewertet werden? (Studie 1)
- Mit welcher Auswertungsvorgehensweise differenziert der LITMUS-CL-NWRT Kinder mit und ohne Sprachentwicklungsstörungen am besten voneinander? (Studie 1)

- Unterscheidet der LITMUS-CL-NWRT monolinguale und bilinguale Kinder mit und ohne Sprachentwicklungsstörungen? (Studie 2)
- Gibt es zwischen dem LITMUS-CL-NWRT und dem HASE-NK Unterschiede bezüglich der diagnostischen Validität? (Studie 2)

3 Methoden

Die hier beschriebenen Daten wurden im Rahmen des BiliSAT-Projekts (BILInguale Entwicklung von Schüler:innen mit Arabisch und Türkisch als Erstsprache; DFG-CH-1112/4-1; Prof. Dr. Chilla, Universität Flensburg/Prof. Dr. Hamann, Universität Oldenburg) erhoben. Abgesehen von den beiden NWRTs wurden mit den Proband:innen des Projekts zahlreiche weitere Testungen in der Erst- und Zweitsprache durchgeführt, um ihren Sprachstand adäquat einschätzen zu können. Für die Erhebungen waren 3-4 Testsitzungen notwendig, die meisten Kinder wurden zuhause getestet, einige auch in der Schule, im Kindergarten oder in einer logopädischen Praxis.

Die Proband:innen wurden entweder von der Autorin selbst oder von wissenschaftlichen Mitarbeiter:innen und Hilfskräften getestet. Die Auswertungen des LITMUS-CL-NWRT und HASE-NK wurden von der Autorin selbst vorgenommen und mit den Auswertungen einer anderen Wissenschaftlerin (Interraterin) abgeglichen.

3.1 Proband:innen

Insgesamt wurde der LITMUS-CL-NWRT mit 103 Kindern durchgeführt, davon wurden bis auf drei Kinder auch alle mit dem HASE-NK getestet. Die Kinder wurden in vier Gruppen aufgeteilt, jeweils zwei Gruppen mit sprachlich unauffälligen Kindern (monolinguale Gruppe: MoTD; bilinguale Gruppe: BiTD) und zwei Gruppen mit Kindern, bei denen eine Sprachentwicklungsstörung (engl. DLD) diagnostiziert wurde (MoDLD und BiDLD).

3.1.1 Gruppenzuordnung

Um festzustellen, bei welchen Kindern eine Sprachentwicklungsstörung vorliegt, wurden deren sprachliche Fähigkeiten im Deutschen und Türkischen bzw. Arabischen getestet. Für das Türkische wurde dabei der TELD-3-T (Turkish Test of Early Language Development; Form A; Topbaş & Güven 2011) durchgeführt, im Arabischen der ELO-L (Batterie d'Evaluation du Langage Oral chez l'enfant libanais; Zebib et al. 2017). Das Deutsche wurde mit der PLAKSS-II (Psycholinguistische Analyse kindlicher Aussprachestörungen; Fox-Boyer 2014), dem WWT (Wortschatz- und Wortfindungstest für 6 bis 10-Jährige; Glück 2011), dem TROG-D (Test zur Überprüfung des Grammatikverständnisses; Fox 2009) und der LiSe-DaZ (Linguistische Sprachstandserhebung – Deutsch als Zweitsprache; Schulz & Tracy 2011) getestet. Ein monolinguales Kind wurde der Gruppe mit Sprachentwicklungsstörungen zugeordnet, wenn es auf mindestens zwei sprachlichen Ebenen (Phonologie, Lexikon rezeptiv, Grammatik) im auffälligen Bereich lag, das heißt unter $-1,25$ SD (Kriterium nach Tomblin, Records & Zhang 1996; siehe auch Hamann & Abed Ibrahim 2017). Zum Vorgehen bei Kindern, die zum Testzeitpunkt jünger bzw. älter als die Normgruppen einzelner Testungen waren, sei auf die ausführlichen Beschreibungen bei Hinnerichs (in Vorb.) verwiesen. Bei den bilingualen Kindern wurden die Normabweichungen zur Definition einer Auffälligkeit je nach Sprachdominanz angepasst (mit Ausnahme der LiSe-DaZ, die bereits Normwerte für bilinguale Kinder enthält).

Die Sprachdominanz wurde auf der Basis eines ausführlichen Elternfragebogens (PaBiQ, Tuller 2015) berechnet (Dominante Sprache: $-1,5$ SD; nicht-dominante Sprache: $-2,25$ SD; ausgeglichener „balanced“ Sprachstatus: $-1,75$ SD; Thordardottir 2015 und Tuller et al. 2018). Es wurden hierbei Informationen zum Sprachkontakt in den ersten vier Lebensjahren, sowie auch zum aktuellen Sprachgebrauch erfragt (für eine ausführliche Beschreibung hierzu siehe Abed Ibrahim & Fekete 2019, S.8). Durch die Elternbefragung wurde ebenfalls das Vorliegen von Hörstörungen und anderen Beeinträchtigungen ausgeschlossen, mit der Raven's Coloured Progressive Matrices (CPM; Bulheller & Häcker 2002) wurde überprüft, ob der nonverbale IQ im Normbereich liegt.

3.1.2 Gruppenbeschreibung

Eine Übersicht der Proband:innen findet sich in Tabelle 1. Hier wird deutlich, dass die Kinder eine große Altersspanne umfassen (3;7 - 13;9) und auch die Kontaktdauer zum Deutschen, sowie das Alter zu Erwerbsbeginn (im Folgenden: AoO, nach dem engl. Age of Onset) in den bilin-

gualen Gruppen stark variieren. Die BiTD-Gruppe hat mehr Kontaktdauer ($M = 5;0$ Jahre) als die BiDLD-Gruppe ($M = 4;6$ Jahre), der Unterschied ist jedoch nicht signifikant ($p > .05$; $U = 347.5$). Die Unterschiede bezüglich des Alters zu Erwerbsbeginn des Deutschen sind hingegen bedeutsam ($p < .01$; $U = 181.5$). Die BiTD-Kinder hatten mit durchschnittlich 4;10 Jahren deutlich später Kontakt zum Deutschen als die Kinder der BiDLD-Gruppe ($M = 2;2$ Jahre). Dieser Unterschied scheint widersprüchlich zu den Angaben der Kontaktdauer und lässt sich dadurch erklären, dass die Kinder der BiTD-Gruppe auch deutlich älter sind ($M = 9;8$ Jahre) als die der BiDLD-Gruppe ($M = 6;8$), wodurch sie trotz eines höheren AoO eine etwas längere Kontaktdauer aufweisen. Die MoDLD- und MoTD-Gruppe liegen mit ihren Altersdurchschnittswerten zwischen den Gruppen der bilingualen Kinder (MoDLD: $M = 8;0$ Jahre; MoTD: $M = 8;7$), die Altersunterschiede sind nur zwischen der BiDLD- und BiTD-Gruppe signifikant ($p < .01$). Dies wird bei den statistischen Berechnungen berücksichtigt.

Tab. 1: Übersicht zu Alter und Sprachkontaktdaten aller Teilnehmer:innen (nach Hinnerichs in Vorb.)

Gruppe	n	Alter (in Monaten)		Kontaktdauer zum Deutschen (in Monaten)		AoO des Deutschen (in Monaten)	
		M (SD)	Spanne	M (SD)	Spanne	M (SD)	Spanne
MoTD	18	102.7 (26.0)	51-145	-			
MoDLD	14	96.0 (34.3)	43-153	-			
BiTD	58	115.7 (28.1)	49-165	59.9 (37.0)	13-135	57.6 (37.6)	0-126
BiDLD	13	80.2 (12.4)	67-113	54.3 (20.4)	17-83	25.9 (19.0)	0-59
Gesamt	103	106.3 (29.6)	43-165	58.7 (34.6)	13-135	51.5 (36.9)	0-126

Anhand der Angaben zum AoO wird deutlich, dass manche Kinder schon von Geburt an Kontakt zum Deutschen hatten, also simultan bilingual aufwachsen, während andere Kinder sogar erst mit über 10 Jahren Kontakt zum Deutschen hatten. Insgesamt wachsen zwei Drittel der bilingualen Kinder dieser Untersuchung sukzessiv bilingual auf, während ein Drittel dem simultanen Erwerb (AoO < 36 Monate) zugeordnet werden kann. Der sozio-ökonomische Status wurde durch den Bildungsgrad der Mutter festgelegt (Hoff 2006), es fanden sich keine signifikanten Unterschiede zwischen den Gruppen (Hinnerichs in Vorb.).

Für die Berechnungen von Studie 2 wurden die Ergebnisse aller in Tabelle 1 aufgeführten Kinder miteinbezogen. Für Studie 1 zum Auswertungsvorgehen wurde nur ein Teil der Proband:innen berücksichtigt, da zum Zeitpunkt der Analysen noch nicht alle Daten erhoben und ausgewertet waren. Hierfür wurden nur Kinder der BiTD- und MoDLD-Gruppe ausgewählt und verglichen, da die Unterscheidung dieser Gruppen meist als größte Herausforderung gilt: Wenn die Nachteile, die eventuell durch den geringeren Sprachkontakt bilingualer Kinder entstehen zu denselben, vergleichsweise niedrigen Testergebnissen führen wie die Symptome einer Sprachentwicklungsstörung, dann werden die Kinder der BiTD-Gruppe fälschlicherweise als sprachlich auffällig diagnostiziert. Ein Auswertungsvorgehen, durch welches die BiTD- und MoDLD-Gruppe bestmöglich voneinander differenziert werden, kann also als sensitiv für die Defizite einer Sprachentwicklungsstörung gelten ohne dabei bilingualer Kinder zu benachteiligen. Für Studie 2 wurden 22 Kinder der BiTD-Gruppe (Alter Ø 10;2 Jahre) mit 10 Kindern der MoDLD-Gruppe (Alter Ø 9;0 Jahre) verglichen.

3.2 Testungen

Der LITMUS-CL-NWRT enthält 16 Items, wobei zu jedem Item vier bis sechs Alternativen konstruiert wurden, aus denen das passendste für die jeweilige Sprache ausgesucht werden kann. Für das Deutsche wurde die Testung nach Hinnerichs (2016) adaptiert und enthält nur Items, die weder im Deutschen, noch in den Erstsprachen der hier getesteten Kinder als reale Wörter vorkommen (Hinnerichs, Polišenská & Chiat 2016). Bei der Lautauswahl der Items wurde darauf geachtet, dass diese mit möglichst vielen Sprachen kompatibel sind. Die Nichtwörter enthalten die drei Vokale /a, i, u/, und die elf Konsonanten /p, b, t, d, k, g, s, z, l, m, n/ (siehe Tabelle 2). Die Testung besteht aus jeweils vier 2-, 3-, 4- und 5-Silbern, alle Silben sind offen und haben nur einen Konsonanten im Anlaut (CV-Struktur). Die Items werden mit neutraler Prosodie präsentiert, das heißt, dass alle Silben gleichermaßen betont werden.

Die Items des HASE-NK entstammen dem Untertest Nachsprechen von Kunstwörtern aus dem HASE (Schöler & Brunner 2008; siehe auch 1.2). Sie bestehen ebenfalls nur aus CV-Silben,

die Lautauswahl ist jedoch etwas größer als beim LITMUS-CL-NWRT. So enthält der HASE-NK sechs Vokale (/a, e, ə, i, o, u/) und elf Konsonanten (/p, b, d, k, g, f, v, ʋ, l, m, n/), also drei Vokale mehr als der LITMUS-CL-NWRT (siehe Tabelle 2). Das Betonungsmuster entspricht dem Deutschen, Zweisilber werden beispielsweise auf der ersten Silbe betont, Dreisilber auf der zweiten Silbe.

Tab. 2: Nichtwörter des LITMUS-CL-NWRT und HASE-NK

Testung	2-Silber	3-Silber	4-Silber	5-Silber
LITMUS-CL-NWRT	/naki/ - 1	/nalidu/ - 11	/zipalita/ - 13	/litapimudi/ - 9
	/zipu/ - 6	/zipula/ - 2	/gazulumi/ - 8	/sipunakila/ - 10
	/tula/ - 7	/panuti/ - 3	/nugitala/ - 5	/tulikazimu/ - 15
	/lumi/ - 16	/luniga/ - 14	/litisagu/ - 12	/maluzikupa/ - 4
HASE-NK	/mali/	/lakemo/	/pukavoʋə/	
	/lufa/	/vunoʋə/	/fodəkina/	
	/bida/	/godepi/	/ʋibanelu/	

Die Nichtwörter des HASE-NK wurden, wie im Manual vorgegeben, in der hier aufgeführten Reihenfolge präsentiert, also von oben nach unten, beginnend in der linken Spalte. Die Nichtwörter des LITMUS-CL-NWRTs wurden im Gegensatz zu denen des HASE-NK nicht mit steigender Silbenanzahl präsentiert, sondern in der durch die Zahlen in Tabelle 2 angegebenen Reihenfolge.

Die beiden Testungen wurden meistens direkt hintereinander durchgeführt. Die Items des LITMUS-CL-NWRT wurden dabei im Rahmen einer Power Point-Präsentation abgespielt, die Nichtwörter werden im Kontext einer Geschichte „Zauberwörter“ genannt, mit denen eine kaputte Perlenkette wieder „ganz gezaubert“ werden kann (Polišenská & Kapalková 2014; siehe Abbildung 1). Die Items des HASE-NK wurden ebenfalls als Zauberwörter eingeführt, jedoch nur als Audioaufnahme, ohne Power-Point-Präsentation, abgespielt, um möglichst nahe an den Vorgaben des Manuals zu bleiben (Schöler & Brunner 2008).

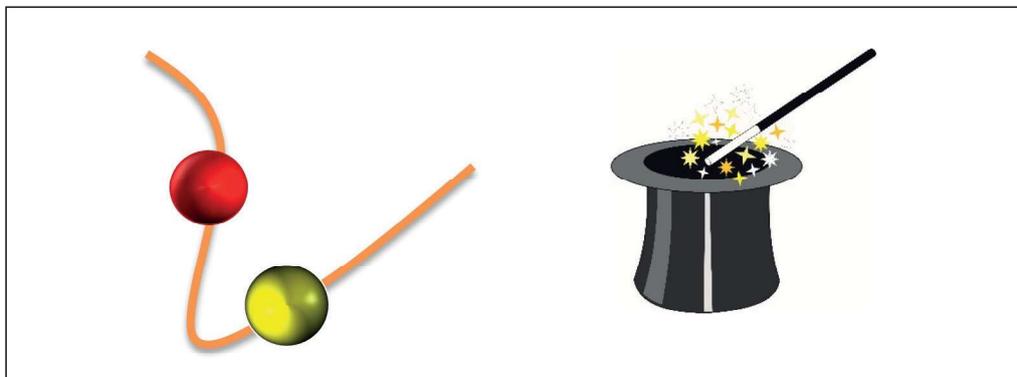


Abb. 1: Power-Point-Präsentation zur Durchführung des LITMUS-CL-NWRT: „Perlen auf eine Kette zaubern“ (Polišenská & Kapalková 2014)

3.3 Auswertungsvorgehensweisen

Um den Einfluss verschiedener Auswertungsvorgehensweisen zu vergleichen, wurde zunächst festgelegt, welche Abweichungen grundsätzlich als Fehler gelten und welche Abweichungen untersucht werden sollten, da sie in bisherigen Studien nicht einheitlich als Fehler bzw. zu tolerierende Abweichungen eingestuft wurden. So werden Laut- und Silbenelisionen und Laut- und Silbenumstellungen als Fehler gewertet, ebenso auch Nullreaktionen oder Itemwiederholungen, die keinerlei Ähnlichkeit mehr mit dem vorgegebenen Nichtwort haben. Konsonantensubstitutionen werden ebenfalls als Fehler gewertet, mit Ausnahme von Abweichungen bei der Stimmgebung, da hier die bilinguale Erwerbssituation unabhängig vom Vorliegen einer Sprachentwicklungsstörung zu Abweichungen führen kann (siehe 1.3).

Des Weiteren wurde in Studie 1 untersucht, inwiefern sich die Wertung von Laut- und Silbenadditionen, konstanten Substitutionen und Vokalsubstitutionen auf die Differenzierungsfähigkeit des LITMUS-CL-NWRT auswirkt. In Bezug auf Vokalsubstitutionen werden diese bei minimalen Abweichungen grundsätzlich nicht als Fehler gewertet, um ebenso wie in Bezug auf

die Stimmhaftigkeit von Konsonanten die bilinguale Erwerbssituation zu berücksichtigen. Ob es jedoch empfehlenswert ist, jegliche Vokalsubstitution von der Fehlerwertung auszuschließen (siehe 1.3), soll mit den vorliegenden Daten überprüft werden.

4 Ergebnisse

Das Auswertungsvorgehen, das durch die Ergebnisse von Studie 1 festgelegt wurde, diente als Basis für die Auswertungen zu Studie 2. Es wurde mit einem Interraterabgleich überprüft, die Interrater-Reliabilität ergab eine fast vollkommene Übereinstimmung ($k = .997$; Hinnerichs in Vorb.).

Alle statistischen Berechnungen wurden mit SPSS 27 durchgeführt. Die Ergebnisse sind Auszüge des Dissertationsprojekts der Autorin (Hinnerichs in Vorb.).

4.1 Ergebnisse von Studie 1

Um der Frage nachzugehen, welchen Einfluss verschiedene Auswertungsvorgehensweisen auf die Unterscheidung der BiTD- und MoDLD-Gruppe haben, wurde zunächst verglichen, wie oft Additionen, Vokalsubstitutionen und konstante Substitutionen in den beiden Gruppen vorkommen. Aufgrund der kleinen Stichprobe wurde der Mann-Whitney-U-Test herangezogen. Dieser ergab, dass sowohl Additionen als auch Vokalsubstitutionen signifikant häufiger in der MoDLD- als in der BiTD-Gruppe vorkommen (Additionen: $p < .05$; $U = 58.5$; Vokalsubstitutionen: $p < .01$; $U = 44.5$). Konstante Substitutionen kamen in keiner Gruppe vor und werden deshalb für die weiteren Berechnungen nicht berücksichtigt.

Die deskriptive Auswertung zeigt, dass Additionen zwar bei sieben der zehn Kinder der MoDLD-Gruppe vorkommen, jedoch nicht alle diese Kinder durch den Ausschluss von Additionen einen höheren Gesamtpunktwert erreichen (siehe Tabelle 3). Der Grund hierfür liegt darin, dass viele Items, bei deren Nachsprechen Additionen verzeichnet werden konnten, noch weitere Abweichungen enthalten - beispielsweise Elisionen - und somit ohnehin als falsch gewertet würden (Hinnerichs in Vorb.). Ebenso führen die 22 Vokalsubstitutionen, die in der MoDLD-Gruppe beobachtet und laut Kriterium 2 nicht als Fehler gewertet wurden, nur bei drei Kindern zu einem Anstieg von jeweils einem Punkt (siehe Tabelle 3).

Tab. 3: Übersicht zum Auftreten der Fehlerarten (Studie 1)

Fehlerarten	MoDLD	BiTD
Kriterium 1: Additionen		
Anzahl der Kinder, bei denen diese Abweichung vorkommt:	7/10 Kinder (70%)	5/22 Kinder (23%)
Anzahl der Additionen:	10	6
Auswirkung auf das Ergebnis:	3 mal + 1 Punkt	2 mal + 1 Punkt
Kriterium 2: Vokalsubstitutionen		
Anzahl der Kinder, bei denen diese Abweichung vorkommt:	8/10 Kinder (80%)	6/22 Kinder (27%)
Anzahl der Vokalsubstitutionen:	22	10
Auswirkung auf das Ergebnis:	3 mal + 1 Punkt	3 mal + 1 Punkt

Deshalb wurde überprüft, ob das häufigere Auftreten von Additionen und Vokalsubstitutionen in der MoDLD-Gruppe letztendlich dazu führt, dass diese hierdurch tatsächlich schlechter als die BiTD-Gruppe im LITMUS-CL-NWRT abschneidet.

Zu diesem Zweck wurde verglichen, wie hoch die Differenz der Mittelwerte bei Anlegung verschiedener Auswertungsvorgehensweisen ist. In Tabelle 4 befindet sich eine Beschreibung der verschiedenen Auswertungsvorgehensweisen.

Tab. 4: Übersicht und Beispiele zu den Auswertungsvorgehensweisen in Studie 1 (nach Hinnerichs in Vorb.)

Auswertungsvorgehensweisen	Beispiele
Strenge Auswertung Kriterien 1 und 2 werden als Fehler gewertet (minimale Vokalabweichungen werden toleriert)	/nugikatala/ statt /nugitala/ > falsch /zipulita/ statt /zipalita/ > falsch /panoti/ statt /panuti/ > korrekt
Kriterium 1: Laut- und Silbenadditionen zählen NICHT als Fehler	/nugikatala/ statt /nugitala/ > korrekt
Kriterium 2: Vokalsubstitutionen zählen NICHT als Fehler (alle Vokalabweichungen werden toleriert)	/zipulita/ statt /zipalita/ > korrekt
Milde Auswertung Kriterien 1 und 2 werden NICHT als Fehler gewertet	/nugikatala/ statt /nugitala/ > korrekt /zipulita/ statt /zipalita/ > korrekt /panoti/ statt /panuti/ > korrekt

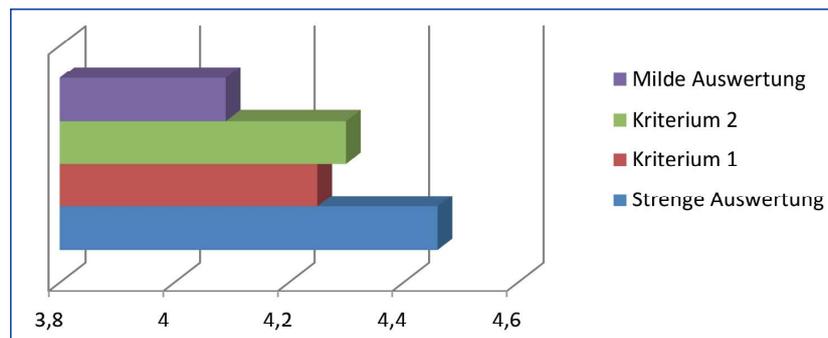
So werden, wie unter 3.3 beschrieben, einige Abweichungen, wie Elisionen und Umstellungen, immer als Fehler gewertet, während andere Abweichungen, wie Veränderungen der Stimmhaftigkeit oder minimale Vokalabweichungen, grundsätzlich toleriert werden. In Bezug auf Additionen (Kriterium 1) und Vokalsubstitutionen (Kriterium 2) wird untersucht, wie sich deren Wertung auf die Endergebnisse des LITMUS-CL-NWRTs auswirkt. Vergleicht man zunächst die Ergebnisse des Mann-Whitney-U-Tests, der nach jedem dieser Auswertungsvorgehensweisen durchgeführt wurde, so zeigt sich, dass alle Auswertungsvorgehensweisen die BiTD- und MoDLD-Gruppe signifikant voneinander unterscheiden (strenge Auswertung: $p < .01$; $U = 13.5$; Kriterium 1: $p < .01$; $U = 15.00$; Kriterium 2: $p < .01$; $U = 11.5$; milde Auswertung: $p < .01$; $U = 11.5$). Die BiTD-Gruppe schneidet also unabhängig von der Auswertungsvorgehensweise immer besser beim LITMUS-CL-NWRT ab.

In einem nächsten Schritt wurden die Ergebnisse für jede Gruppe im Mittel berechnet, woraufhin die Differenz der beiden Mittelwerte gebildet wurde. Tabelle 5 fasst diese Werte für die vier Auswertungsvorgehensweisen zusammen.

Tab. 5: Statistische Werte zu den einzelnen Auswertungsvorgehen im Gruppenvergleich

Statistische Werte	Auswertungsvorgehensweisen			
	Strenge Auswertung	Kriterium 1	Kriterium 2	Milde Auswertung
Mittelwert (SD)	BiTD: 14.36 (1.29) MDLD: 9.9 (2.88)	BiTD: 14.45 (1.3) MDLD: 10.2 (2.9)	BiTD: 14.5 (1.19) MDLD: 10.2 (2.9)	BiTD: 14.64 (1.09) MDLD: 10.5 (2.92)
Minimum-Maximum (Spannweite)	BiTD: 11-16 (5) MDLD: 6-14 (8)	BiTD: 11-16 (5) MDLD: 6-14 (8)	BiTD: 12-16 (4) MDLD: 6-14 (8)	BiTD: 12-16 (4) MDLD: 6-14 (8)
Mittelwertsdifferenz	4.46	4.25	4.3	4.14

Mit Abbildung 2 wird verdeutlicht, dass durch die strenge Auswertung mit 4.46 Punkten die größte Mittelwertdifferenz zwischen der MoDLD- und BiTD-Gruppe zustande kommt. Die geringste Differenz erfolgt mit 4.14 Punkten durch die milde Auswertung, die Anwendungen von Kriterium 1 und 2 liegen dazwischen. Insgesamt liegen die Mittelwertdifferenzen nicht weit auseinander, der maximale Unterschied beträgt 0.32 Punkte.



Werte von oben nach unten: 4.14; 4.3; 4.25; 4.46

Abb. 2: Mittelwertdifferenzen zwischen der BiTD- und MoDLD-Gruppe bei den verschiedenen Auswertungsvorgehensweisen (aus Hinnerichs in Vorb.)

4.2 Ergebnisse von Studie 2

Mit Studie 2 soll zunächst die Frage beantwortet werden, ob der LITMUS-CL-NWRT monolinguale und bilinguale Kinder mit und ohne Sprachentwicklungsstörungen voneinander unterscheiden kann. Vergleicht man die Mittelwerte der vier Gruppen, so wird deutlich, dass die beiden TD-Gruppen beim LITMUS-CL-NWRT besser abschneiden als die DLD-Gruppen. So erreicht die MoTD-Gruppe durchschnittlich 13.61 Punkte (SD = 1.65) und die BiTD-Gruppe 13.2 Punkte (SD = 2.29), während es bei der MoDLD-Gruppe im Mittel 9.36 Punkte (SD = 2.92) und bei der BiDLD 9.08 Punkte (SD = 2.99) sind.

Mittels einer Kovarianzanalyse wurde überprüft, ob diese Unterschiede signifikant sind. Als Kovariaten wurden hierbei das Alter und der sozio-ökonomische Status berücksichtigt. Im Gegensatz zum Alter unterscheiden sich die Gruppen zwar hinsichtlich des sozio-ökonomischen Status nicht signifikant (siehe 3.1.2.), jedoch wurde dieser als Kovariate miteinbezogen, da sich schwache Korrelationen zwischen dem LITMUS-CL-NWRT und dem sozio-ökonomischen Status der Teilnehmer:innen zeigten ($r_s = .227$, $p < .05$, $n = 100$). Zwischen dem Alter und dem LITMUS-CL-NWRT zeigte sich ein mittlerer Zusammenhang ($r = .479$, $p < .01$, $n = 103$). Die Kontaktdauer und das AoO wurden nicht als Kovariaten in die Berechnungen miteinbezogen, da für beide gezeigt werden konnte, dass sie keinen Zusammenhang zu den Ergebnissen des LITMUS-CL-NWRT aufweisen, wenn der Einfluss des Alters mit einer partiellen Korrelation kontrolliert wird (Kontaktdauer: $r_p = .061$; AoO: $r_p = -.061$).

Die Ergebnisse der Kovarianzanalyse zeigen, dass sich die Gruppen hinsichtlich ihrer Ergebnisse beim LITMUS-CL-NWRT mit einer großen Effektstärke signifikant voneinander unterscheiden ($F(3, 94) = 12.5$, $p < .01$, $\eta_p^2 = .29$). Durch die Post-hoc-Analysen mit Bonferroni-Korrektur wird deutlich, dass beide DLD-Gruppen signifikant weniger Punkte beim LITMUS-CL-NWRT erreichen als die TD-Gruppen ($p < .01$), während sich weder zwischen der MoTD- und BiTD-, noch zwischen der BiDLD- und MoDLD-Gruppe signifikante Unterschiede zeigen ($p > .05$). Bezogen auf die Kovariaten zeigt sich, dass das Alter einen signifikanten Einfluss auf die Ergebnisse des LITMUS-CL-NWRT hat ($p < .01$), während dies beim sozio-ökonomischen Status nicht der Fall ist ($p > .05$).

Abbildung 3 veranschaulicht die Gruppenunterschiede und zeigt, dass diese mit zunehmender Itemlänge größer werden.

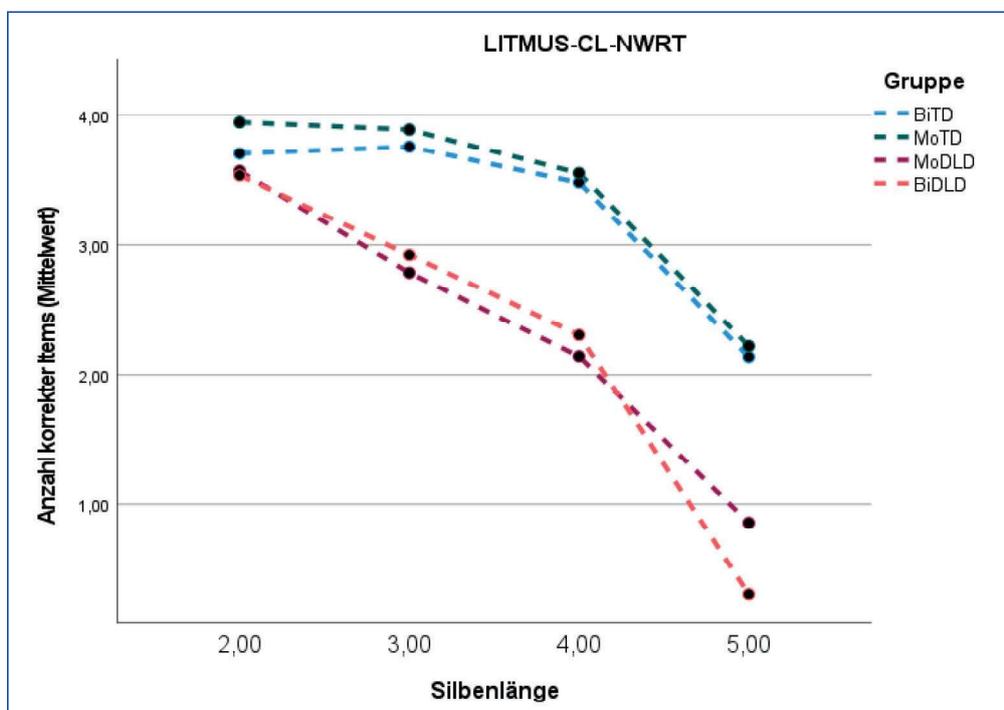


Abb. 3: Diagramm zum Längeneffekt beim LITMUS-CL-NWRT (aus Hinnerichs in Vorb.)

Während beim Nachsprechen von Zweisilbern keine bedeutsamen Unterschiede zwischen den TD- und DLD-Gruppen zu verzeichnen sind, sprechen die Kinder der TD-Gruppen Drei-, Vier-

und Fünfsilber signifikant besser nach als die Kinder der DLD-Gruppen, wobei die Effektstärke bei den Fünfsilbern den höchsten Wert erreicht (Dreisilber $\eta_p^2 = .28$; Viersilber $\eta_p^2 = .24$; Fünfsilber $\eta_p^2 = .30$). Ebenso wie in Bezug auf die Gesamtergebnisse zeigt sich auch für die einzelnen Itemlängen, dass kein bedeutsamer Unterschied zwischen den bilingualen und monolingualen Kindern zu verzeichnen ist.

Zu den Ergebnissen des HASE-NK wurde ebenfalls eine Kovarianzanalyse berechnet und es zeigt sich, dass analog zu den Berechnungen beim LITMUS-CL-NWRT auch hier signifikante Unterschiede zwischen den TD- und DLD-Gruppen zu verzeichnen sind, während der Faktor Bilingualismus keinen Einfluss auf die Ergebnisse hat (Näheres unter Hinnerichs in Vorb.).

Da signifikante Gruppenunterschiede nicht ohne Weiteres mit einer zufriedenstellenden diagnostischen Validität gleichzusetzen sind, wurde in einem nächsten Schritt berechnet, wie viele Kinder die beiden NWRTs korrekt als sprachlich unauffällig bzw. auffällig einordnen (siehe Tabelle 6). Mittels einer ROC-Kurve wurde der ideale Cut-off-Wert berechnet, jeweils für die monolingualen und bilingualen Kinder getrennt. Alle Kinder, die auf oder unter diesem Wert liegen, werden als sprachlich auffällig klassifiziert. Die Sensitivität gibt an, wieviel Prozent der Kinder mit DLD korrekt als sprachlich auffällig erkannt wurden, die Spezifität bezieht sich auf die korrekte Klassifizierung der TD-Kinder. Der AUC-Wert gibt die Vorhersagegenauigkeit einer Testung an. Liegt er unter .60, gilt die Testung als wertlos, bis zu .70 als schwach, bis zu .80 als akzeptabel, bis zu .90 als gut und ein Wert zwischen .90 und 1.0 gilt als sehr gut (Armon-Lotem & Meir 2016).

Tab. 6: Übersicht der Werte zur diagnostischen Validität des LITMUS-CL-NWRT und HASE-NK (nach Hinnerichs in Vorb.)

		LITMUS-CL-NWRT	HASE-NK
Bilinguale Kinder	N	71	68
	AUC	.873	.774
	Cut-off-Wert	12	6
	Sensitivität	85 %	67%
	Spezifität	71%	73%
Monolinguale Kinder	N	32	32
	AUC	.903	.873
	Cut-off-Wert	12	7
	Sensitivität	86 %	93 %
	Spezifität	78 %	61%

Vergleicht man die AUC-Werte der beiden NWRTs, so wird deutlich, dass der LITMUS-CL-NWRT sowohl die monolingualen als auch die bilingualen Kinder besser unterscheidet. So ist die Vorhersagegenauigkeit des HASE-NK bei den bilingualen Kindern im akzeptablen Bereich, während sie beim LITMUS-CL-NWRT im guten Bereich liegt. Bei den monolingualen Kinder liegt der HASE-NK zwar auch im guten Bereich, der LITMUS-CL-NWRT weist jedoch eine sehr gute Vorhersagegenauigkeit auf. In Bezug auf die bilingualen Kinder weist der LITMUS-CL-NWRT eine deutlich höhere Sensitivität auf, wobei der HASE-NK eine etwas höhere Spezifität als der LITMUS-CL-NWRT zeigt. Bei den monolingualen Kindern ist es genau umgekehrt, hier zeigt der HASE-NK eine höhere Sensitivität und der LITMUS-CL-NWRT eine höhere Spezifität.

Legt man diesen Werten die Einschätzung von Plante & Vance (1994) zugrunde, wonach Sensitivitäts- und Spezifitätswerte ab 90 % für eine gute diagnostische Genauigkeit stehen, ab 80 % akzeptabel sind, aber unter 80 % eine zu hohe Rate von Fehldiagnosen anzeigen, so liegen einige der Werte aus Tabelle 6 unterhalb des zufriedenstellenden Bereichs (die Werte, die im zufriedenstellenden Bereich liegen, wurden in Tabelle 6 fett markiert). Beim HASE-NK ist lediglich die Sensitivität bezüglich monolingualer Kinder im guten Bereich, alle anderen Werte liegen deutlich unter 80 %. Die Sensitivität des LITMUS-CL-NWRT liegt immer im akzeptablen Bereich, sowohl bei den monolingualen als auch bei den bilingualen Kindern. Die Spezifitätswerte liegen beide unter 80 %, wobei der Wert der monolingualen Kinder mit 78 % nahe am akzeptablen Bereich liegt.

5 Diskussion und Schlussfolgerungen

Die vorliegende Studie begründet sich durch die Herausforderung, die sprachlichen Fähigkeiten bilingualer Kinder im Hinblick auf das Vorliegen einer Sprachentwicklungsstörung einzuschätzen. Da die Testung einzelner Sprachen viele Hindernisse mit sich bringt, wird das Nachsprechen von Nichtwörtern aufgrund seiner Neutralität in Bezug auf Einzelsprachen als vielversprechendes Testinstrument angesehen.

Jedoch zeigen bisherige Studien vielfach, dass auch NWRs nicht so einzelsprachunabhängig sind wie angenommen.

Der LITMUS-CL-NWRT wurde im Rahmen eines internationalen Forschungsprojekts entwickelt, wobei bei der Itemkonstruktion auf die Kompatibilität mit möglichst vielen Sprachen Wert gelegt wurde.

5.1 Diskussion zu Studie 1

Bevor die Eignung des LITMUS-CL-NWRs anhand von vier Gruppen mit monolingualen und bilingualen Kindern mit und ohne Sprachentwicklungsstörung untersucht wurde, sollten durch eine kleiner angelegte Studie einzelne Auswertungskriterien überprüft werden. So war es das Ziel von Studie 1 herauszufinden, inwiefern sich die Wertung von Additionen, Vokalsubstitutionen und konstanten Konsonantensubstitutionen auf die Differenzierung der BiTD- und MoDLD-Gruppe auswirkt. Die Ergebnisse zeigen, dass es keinen großen Unterschied macht, wenn Additionen und/oder Vokalsubstitutionen von der Fehlerwertung ausgenommen werden. Dennoch sprechen die vorliegenden Daten dafür, diese beiden Abweichungsarten als Fehler zu werten, da sich hierdurch die Differenz zwischen den Kindern mit und ohne Sprachentwicklungsstörung leicht erhöht, ohne dass dabei die bilingualen Kinder benachteiligt werden.

Konstante Substitutionen kamen hingegen weder in der BiTD- noch in der MoDLD-Gruppe vor. Dies ist wohl vor allem durch das Alter der Proband:innen zu erklären. Alle Kinder in Studie 1 sind mindestens 6 Jahre alt, die meisten sogar älter. Wenn konstante Substitutionen im NWRT durch das Vorliegen konstanter phonologischer Prozesse begründet sind, so ist davon auszugehen, dass diese in dieser Altersgruppe nicht mehr beobachtet werden, da phonologische Prozesse im unauffälligen Spracherwerb spätestens am Ende des vierten Lebensjahres überwunden werden (Kannengieser 2019). Sollten also bei Kindern über fünf Jahren im Rahmen eines NWRs konstante Substitutionen beobachtet werden, so ist es ohnehin nicht ratsam, diese aus der Fehlerwertung auszuschließen, da sie das Symptom einer phonologischen Verzögerung oder Störung sein können (Weinrich & Zehner 2017, S.41f.).

Wie verhält es sich nun aber mit jüngeren Kindern? Sollten bei diesen, wie es von einigen Wissenschaftler:innen gehandhabt wird, konstante Substitutionen von der Fehlerwertung ausgeschlossen werden, da diese durch altersgemäße, physiologische phonologische Prozesse zustande kommen könnten?

Es gibt einige Gründe, die dagegen sprechen: Vor allem bei groß angelegten Screeninguntersuchungen, in deren Rahmen NWRs oft eingesetzt werden, kann im Einzelfall wohl kaum eingeschätzt werden, ob der konkrete Substitutionsprozess physiologisch und noch altersadäquat ist. So müsste beim Ausschluss konstanter Substitutionen aus der Fehlerwertung beispielsweise auch die konstante Rückverlagerung von /t/ zu /k/ toleriert werden, obwohl dies als ungewöhnlicher und therapiebedürftiger Prozess gilt.

Ein anderer Aspekt hinsichtlich konstanter Substitutionen ist das Argument, dass diese auch durch phonetische Störungen hervorgerufen werden können. Letztere gelten im Gegensatz zu phonologischen Auffälligkeiten nicht als Symptom einer Sprachentwicklungsstörung, sondern werden zu den Aussprachestörungen gezählt (Kauschke & Vogt 2019). Bei der Durchführung eines Screenings sollten jedoch auch Kinder mit phonetischen Störungen von einem Screeninginstrument herausgefiltert werden, da sie ebenso wie Kinder mit einer Sprachentwicklungsstörung einen Therapiebedarf haben. Auch unter diesem Aspekt scheint es also nicht ratsam, konstante Substitutionen aus der Fehlerwertung auszuschließen. Im Hinblick auf Studien, die einen NWRT gezielt bezüglich seiner Eignung als Testung bei Sprachentwicklungsstörungen untersuchen und somit Kinder mit Aussprachestörungen nicht erfassen wollen, sei angemerkt, dass phonetische und phonologische Störungen oft in Kombination auftreten und in vielen Fällen keine klare Zuordnung zu einem Störungsschwerpunkt möglich ist (Weinrich & Zehner 2017).

Ein weiteres Argument, das für die strengere Auswertung spricht, also für den Einbezug von Additionen, Vokalsubstitutionen und konstanten Substitutionen in die Fehlerwertung, ist die

Praktikabilität. Je mehr Abweichungen aus der Fehlerwertung ausgeschlossen werden, desto fehleranfälliger wird diese in der praktischen Anwendung. Die Abweichungen, die nicht als Fehler gewertet werden dürfen, müssen erst als solche erkannt werden. Im Berufsalltag ist oft wenig Zeit und eine Liste von Abweichungsarten, die nicht als Fehler gezählt werden dürfen, macht die Auswertung von NWRs komplizierter und langwieriger. Eine Ausnahme sollten also lediglich die Abweichungen bilden, die durch die bilinguale Erwerbssituation erklärt werden können. In diesem Fall lohnt es sich die Auswertungsvorgehensweise etwas komplexer zu gestalten, um Fehldiagnosen bezüglich bilingualer Kinder möglichst zu vermeiden. Dies betrifft nach aktuellem Kenntnisstand Abweichungen bei der Stimmgebung und minimale Unterschiede bei der Realisierung von Vokalen. Für Substitutionen durch Vokale, die im Spektrum weiter auseinander liegen, ebenso wie auch für Additionen konnte mit den Daten aus Studie 1 gezeigt werden, dass durch deren Einbezug in die Fehlerwertung keine Benachteiligung bilingualer Kinder entsteht. Da die vorliegende Stichprobe sehr klein ist, sollten diese Erkenntnisse durch weitere Untersuchungen überprüft werden.

5.2 Diskussion zu Studie 2

Die hier als „streng“ definierte Auswertungsvorgehensweise (siehe Tabelle 4) wurde der Gesamtauswertung des LITMUS-CL-NWRs zugrunde gelegt. Beim Vergleich der vier untersuchten Gruppen zeigen die Ergebnisse, dass mit dem LITMUS-CL-NWR ein signifikanter Unterschied zwischen Kindern mit und ohne Sprachentwicklungsstörung gemessen werden kann, ohne dass dabei die bilinguale Erwerbssituation einen nennenswerten Einfluss auf die Nachsprecheleistungen hat. Vor allem der Einsatz längerer Items erhöht die Differenzierungsfähigkeit der Testung. Während das Nachsprechen von Zweisilbern keinen nennenswerten Unterschied zwischen den TD- und DLD-Gruppen hervorruft, können bei Drei-, Vier- und Fünfsilbern signifikante Unterschiede verzeichnet werden.

Da signifikante Gruppenunterschiede nicht automatisch bedeuten, dass ein Großteil der Kinder diagnostisch korrekt erkannt wird, wurden in einem weiteren Schritt die Spezifität und Sensitivität berechnet. Die Werte zur diagnostischen Validität des LITMUS-CL-NWRs liegen hierbei größtenteils im akzeptablen und guten Bereich, wobei die Spezifität sowohl bei den bilingualen als auch bei den monolingualen Kindern nicht zufriedenstellend ist. Letzteres wird durch die große Altersspanne der Proband:innen erklärt. Da sich die Leistungen beim NWR mit zunehmendem Alter verbessern (Buil-Legaz, Aguilar-Mediavilla & Adrover-Roig 2016, Boerma & Blom 2021), ist es nicht empfehlenswert, einen Cut-off-Wert für eine Stichprobe zu berechnen, die 10 Jahre umfasst. Dies wird durch eine nähere Analyse der Fehldiagnosen bestätigt: So ist die Hälfte der Kinder, die durch den LITMUS-CL-NWR fälschlicherweise als sprachlich auffällig eingestuft wurden, obwohl sie der BiTD-Gruppe angehören, deutlich jünger als der Durchschnitt der Gesamtgruppe (Hinnerichs in Vorb.). Durch einen niedrigeren Cut-off-Wert für diese jüngeren Kinder hätte also die Spezifität deutlich verbessert werden können. Die Berechnung separater Cut-off-Werte für einzelne Altersstufen war im Rahmen der vorliegenden Studie nicht möglich, da die Proband:innenanzahl hierfür zu gering war. Die Tatsache, dass die Werte zur diagnostischen Validität trotz dieser ungünstigen Alterszusammensetzung größtenteils zufriedenstellend sind, lässt darauf schließen, dass der LITMUS-CL-NWR ein geeignetes Testinstrument im Kontext von Bilingualismus und Sprachentwicklungsstörung sein könnte. Dies müsste jedoch mit weiteren Studien, in deren Rahmen idealerweise eine große Anzahl von gleichaltrigen Proband:innen getestet wird, überprüft werden.

Des Weiteren müssten auch die Erkenntnisse zum Vergleich des LITMUS-CL-NWR mit dem HASE-NK überprüft werden. Im Einklang mit bisherigen Studien weisen die vorliegenden Ergebnisse ebenfalls darauf hin, dass sich der LITMUS-CL-NWR besser zur Testung bilingualer Kinder eignet als der HASE-NK. Jedoch sind die Unterschiede bezüglich der diagnostischen Validität zwischen dem LITMUS-CL-NWR und dem HASE-NK nicht so groß wie in der Studie aus den Niederlanden: Hier liegt die diagnostische Validität des LITMUS-CL-NWR bei bilingualen Kindern im sehr guten Bereich, während die des LS-NWRs lediglich akzeptabel ist (Boerma et al. 2015). Dies könnte unter anderem an der Itemkonstruktion liegen, da die Nichtwörter des LS-NWRs bei Boerma et al. (2015) im Gegensatz zum HASE-NK komplexer sind. Die Items des HASE-NK sind in ihrer Konstruktion sehr nahe an der des LITMUS-CL-NWR, da sie ebenfalls nur CV-Silben enthalten. Die Unterschiede zwischen den hier vorliegenden Daten und den Ergebnissen aus den Niederlanden könnten also dadurch zustande kommen, dass die bilingualen Kinder bei Boerma et al. (2015) durch den LS-NWR noch stärker benachteiligt

werden, da dessen Items komplexer und dadurch eventuell auch einzelsprachspezifischer sind als die des HASE-NK.

Der Unterschied zwischen dem LITMUS-CL-NWRT und dem HASE-NK besteht in der Itemkonstruktion lediglich aus einer etwas größeren Lautauswahl beim HASE-NK und der Prosodie, die beim LITMUS-CL-NWRT neutral gehalten wird. Ob bilinguale Kinder bei NWRTs jedoch tatsächlich durch den Einsatz von Lauten, die nicht in ihrer Erstsprache vorkommen, benachteiligt werden, wurde bisher kaum untersucht. Erste Ergebnisse deuten darauf hin, dass sich für bestimmte bilinguale Gruppen ein Nachteil ergeben könnte (Hinnerichs in Vorb.).

Die schlechtere diagnostische Validität des HASE-NK könnte aber auch auf andere Faktoren der Testkonstruktion zurückzuführen sein. So beinhaltet der HASE-NK zwar Drei- und Viersilber, jedoch keine Fünfsilber. Da Letztere bezüglich der Unterscheidung der TD- und DLD-Gruppen die größte Effektstärke beim LITMUS-CL-NWRT aufweisen, könnte dies auch zu dessen vergleichsweise höheren diagnostischen Validität geführt haben. Ein weiterer Grund könnte die Itemanzahl sein. Der HASE-NK beinhaltet mit 9 Items fast nur halb so viele Nichtwörter wie der LITMUS-CL-NWRT (16 Items). In testtheoretischen Ansätzen wird davon ausgegangen, dass ein Merkmal besser erfasst werden kann, wenn die entsprechende Testung mehr Items enthält (Brandt & Moosbrugger 2020, S.50; Zaretsky, Lange & Hey 2022, S.11). Welche der genannten Faktoren letztendlich in welchem Ausmaß dazu geführt hat, dass der LITMUS-CL-NWRT eine bessere diagnostische Validität aufweist als der HASE-NK, kann mit den vorliegenden Daten nicht abschließend beantwortet werden. Zieht man die bisherigen Ergebnisse von NWRT-Studien mit bilingualen Kindern heran, so zeigte sich bereits mehrfach, dass der LITMUS-CL-NWRT besser zur Testung bilingualer Kinder geeignet ist als einzelsprachspezifische NWRTs (Boerma et al. 2015; Ortiz 2021; Zaretsky, Lange & Hey 2022). Die hier vorliegenden Ergebnisse legen nahe, dass der HASE-NK aufgrund seiner Itemkonstruktion bilinguale Kinder eventuell weniger benachteiligt als andere LS-NWRTs. Dies müsste mit weiteren Untersuchungen überprüft werden. Zukünftige Studien sollten also gezielt die Auswirkung der Test- und Itemkonstruktion auf die Nachsprehleistungen von monolingualen und bilingualen Kindern untersuchen. Auf der Basis dieser Erkenntnisse könnten dann NWRTs erstellt werden, die Kinder mit und ohne DLD möglichst gut voneinander unterscheiden, ohne dabei bilinguale Kinder zu benachteiligen.

Abgesehen davon, dass der LITMUS-CL-NWRT Kinder mit und ohne DLD ausreichend genau voneinander differenzieren kann, misst er ihre Fähigkeiten unabhängig von ihrem sozialen und sprachlichen Hintergrund.

Dies wird dadurch belegt, dass der sozio-ökonomische Status keinen bedeutsamen Einfluss auf die Nachsprehleistungen hat und auch die Sprachkontaktdauer und das AoO keinen nennenswerten Zusammenhang zu den Ergebnissen des LITMUS-CL-NWRT aufweisen. Diese Erkenntnisse stehen im Einklang mit bisherigen Studien (Antonijevic-Elliott et al. 2020; Chiat & Polišenská 2016; Verhagen & Andringa 2021).

Die große Altersspanne der Proband:innen und die gleichzeitig kleinen Stichproben der DLD-Gruppen schränken die Aussagekraft der vorliegenden Studie ein. Gleichzeitig ist das diagnostische Vorgehen zur Gruppenzuweisung eine Stärke dieser Untersuchung. Während einige Studien die Zuweisungen zur TD- bzw. DLD-Gruppe lediglich auf der Basis eines alleinigen Testverfahrens in der Zweitsprache vornehmen (Zaretsky, Lange & Hey 2022), wurden im Rahmen des BiliSAT-Projekts internationale Empfehlungen eingehalten: In beiden Sprachen wurden mit standardisierten und normierten Testungen mehrere sprachliche Ebenen überprüft, die Normabweichung zur Definition einer Störung wurde an die Sprachdominanz angepasst. Somit kann festgehalten werden, dass die Ergebnisse trotz der großen Altersspanne und der kleinen Stichprobe eine Aussagekraft haben, da die Kinder der DLD-Gruppen eine bestmöglich abgesicherte Diagnose erhalten haben.

6 Fazit und Ausblick

Das Nachsprechen von Nichtwörtern wird in der Diagnostik von Sprachentwicklungsstörungen häufig eingesetzt, sowohl im Rahmen von Screeningverfahren als auch in der sprachtherapeutischen Praxis.

Dabei gelten die eingesetzten NWRTs oft fälschlicherweise pauschal als einzelsprachunabhängig, wodurch bilinguale Kinder benachteiligt werden können.

Es ist daher essenziell, beim Einsatz von NWRs in der Praxis die Konstruktion der Items und die Spracherwerbssituation des jeweiligen Kindes zu berücksichtigen und die Ergebnisse mit Vorsicht zu interpretieren.

Der Einsatz von möglichst einzelsprachunabhängig konstruierten Testungen ist zu empfehlen, bisherige Forschungsergebnisse sind vielversprechend. Weitere Untersuchungen hierzu sind wünschenswert, ein konsistentes Auswertungsvorgehen ist dabei für eine Vergleichbarkeit der Ergebnisse unerlässlich.

Des Weiteren zeigen die Werte zur diagnostischen Validität der vorliegenden Arbeit sowie auch die früherer NWR-Studien, dass selten alle Kinder korrekt als sprachlich auffällig bzw. unauffällig erkannt werden. Deshalb wird immer wieder betont, dass der NWR niemals isoliert, sondern immer in Kombination mit weiteren Testungen zur Einschätzung sprachlicher Fähigkeiten herangezogen werden sollte (Ellis Weismer et al. 2000; Schwob et al. 2021). Dies gilt nicht nur für den sprachtherapeutischen Kontext, wo eine fundierte Therapieplanung ohnehin nur durch die ausführliche Überprüfung der einzelnen sprachlichen Ebenen stattfinden kann. Auch im Rahmen von Screeningverfahren zur Erkennung von Risikokindern sollte der NWR immer durch andere Testungen ergänzt werden (Ortiz 2021).

Die vorliegenden Daten bestärken die bisherigen internationalen Ergebnisse zum LITMUS-CL-NWR und legen nahe, dass dieser auch im deutschsprachigen Raum zu einer verbesserten Erkennung von Sprachentwicklungsstörungen bei bilingualen Kindern führen kann.

Danksagung

Mein Dank geht an alle Familien und insbesondere an die Kinder, die durch ihre Teilnahme an der Datenerhebung diese Studie möglich gemacht haben. Ebenso danke ich Solveig Chilla und Cornelia Hamann und allen anderen Mitarbeiter:innen des BiliSAT-Projekts (DFG-CH-1112/4-1; Prof. Dr. Chilla, Universität Flensburg/Prof. Dr. Hamann, Universität Oldenburg), im Rahmen dessen die Daten erhoben und ausgewertet wurden. Großer Dank gebührt ebenso der Graduiertenakademie der Universität Heidelberg, durch deren finanzielle Unterstützung über verschiedene Stipendienprogramme die Dissertation realisiert werden konnte.

Literatur

- Abed Ibrahim, L. & Fekete I. (2019) What Machine Learning Can Tell Us About the Role of Language Dominance in the Diagnostic Accuracy of German LITMUS Non-word and Sentence Repetition Tasks. *Frontiers in Psychology* 9:2757. <https://doi.org/10.3389/fpsyg.2018.02757>
- Antonijevic-Elliott, S., Lyons, R., O'Malley, M. P., Meir, N., Haman, E., Banasik, N., Carroll, C., McMenamin, R., Rodden, M. & Fitzmaurice, Y. (2020). Language assessment of monolingual and multilingual children using non-word and sentence repetition tasks. *Clinical Linguistics & Phonetics*, 34 (4), 293–311.
- Armon-Lotem, S. & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: evidence from Russian and Hebrew. *International Journal of Language & Communication Disorders*, 51 (6), 715–731.
- Baddeley, A. D. & Hitch, G. J. (1994). Developments in the Concept of Working Memory. *Neuropsychology*, 8 (4), 485–493.
- Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F. & Blom, E. (2015). A Quasi-Universal Nonword Repetition Task as a Diagnostic Tool for Bilingual Children Learning Dutch as a Second Language. *Journal of Speech, Language, and Hearing Research*, 58, 1747–1760.
- Boerma, T. & Blom, E. (2021). Crosslinguistic nonword repetition and narrative performance over time. In S. Armon-Lotem & K. K. Grohmann (Hrsg.), *Language Impairment in Multilingual Settings. LITMUS in action across Europe* (S.301–328). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Brandt, H. & Moosbrugger, H. (2020). Planungsaspekte und Konstruktionsphasen von Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion*, 3. Auflage (S. 39–66). Berlin: Springer.
- Buil-Legaz, L., Aguilar-Mediavilla, E. & Adrover-Roig, D. (2016). Longitudinal trajectories of the representation and access to phonological information in bilingual children with specific language impairment. *International Journal of Speech-Language Pathology*, 18 (5), S. 473–482.
- Bulheller, S. & Häcker, H. O. (2002). *Raven's Coloured Progressive Matrices*. Frankfurt am Main: Pearson Assessment.
- Chiat, S. (2015). Non-Word Repetition. In S. Armon-Lotem, J. De Jong & N. Meir (Hrsg.), *Assessing Multilingual Children – Disentangling Bilingualism from Language Impairment* (S. 125–150). Bristol: Multilingual Matters.
- Chiat, S. & Polišenská, K. (2016). A Framework for Crosslinguistic Nonword Repetition Tests: Effects of Bilingualism and Socioeconomic Status on Children's Performance. *Journal of Speech, Language, and Hearing Research*, 59 (3), 1179–1189.
- Chiat, S., Polišenská, K., Szewczyk, J. & Kapalková, S. (2016). *Scoring Instructions – Crosslinguistic Nonword Repetition Tasks: British English version*. Unpublished material.
- Chilla, S. (2022). *Kindliche Mehrsprachigkeit*. München: Ernst Reinhardt Verlag.
- Dos Santos, C. & Ferré, S. (2016). A Nonword Repetition Task to Assess Bilingual Children's Phonology. *Language Acquisition*, 25, 1–14. <https://doi.org/10.1080/10489223.2016.1243692>
- Ennemoser, M., Pepouna, S. & Hartung, N. (2012). Kulturfaire und prognostisch valide Erfassung des phonologischen Arbeitsgedächtnisses bei Kindern mit Migrationshintergrund. *Unterrichtswissenschaft*, 40 (1), 26–46.

- Fox, A. (2009). *TROG-D. Test zur Überprüfung des Grammatikverständnisses*. Idstein: Schulz-Kirchner.
- Fox-Boyer, A. (2014). *Psycholinguistische Analyse kindlicher Aussprachestörungen (PLAKSS-II)*. Frankfurt am Main: Pearson.
- Fried, L., Briedigkeit, E., Isele, P. & Schunder, R. (2009). Delfin 4 – Sprachkompetenzmodell und Messgüte eines Instrumentariums zur Diagnose, Förderung und Elternarbeit in Bezug auf die Sprachkompetenz. *Zeitschrift für Grundschulforschung*, 2, 13-26.
- Gamper, H., Keller, U., Messerli, N., Moser, M. & Wüst, J. (2012). *Normen für den Mottier-Test bei 4- bis 12-jährigen Kindern. Praxisforschung der Erziehungsberatung des Kantons Bern, Band 12*. Zugriff am 07.09.2022 auf <https://docplayer.org/25529012-Normen-fuer-den-mottier-test.html>
- Gathercole, S. E. & Baddeley, A. D. (1990b). Phonological Memory Deficits in Language Disordered Children: Is There a Causal Connection? *Journal of Memory and Language*, 29, 336-360.
- Glück, C. W. (2011). *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige: WWT 6-10*. München: Elsevier.
- Grimm, A. & Schulz, P. (2014). Specific Language Impairment and Early Second Language Acquisition: The Risk of Over- and Underdiagnosis. *Child Indicators Research*, 7 (4), 821-841.
- Grimm, H. (2015). *Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5). Diagnose von Sprachverarbeitungs-fähigkeiten und auditiven Gedächtnisleistungen*. Göttingen: Hogrefe.
- Grimm, H., Aktas, M. & Kießig, U. (2003). *SSV. Sprachscreening für das Vorschulalter. Kurzform des SETK 3-5*. Göttingen: Hogrefe.
- Hamann, C. & Abed Ibrahim, L. (2017). Methods for identifying specific language impairment in bilingual populations in Germany. *Frontiers in Communication*, 2 (16), 1-19.
- Hasselhorn, M. & Körner, K. (1997). Nachsprechen von Kunstwörtern: Zum Zusammenhang zwischen Arbeitsgedächtnis und syntaktischen Sprachleistungen bei Sechs- und Achtjährigen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 29, 212-224.
- Heeg, K., Bloder, T., Eikerling, M., Frey, N., Starke, A. & Lüke, C. (2022). *Der Mottier-Test in der Anwendung – Überprüfung von Parametern der Durchführung*. Vortrag am XXXIV. Workshop Klinische Linguistik.
- Hinnerichs, J. (2016). *Spezifische Sprachentwicklungsstörungen bei Kindern mit Deutsch als Zweitsprache – Zur Eignung des Nachsprechens von quasi-universalen Nichtwörtern als Differentialdiagnostik*. Unveröffentlichte Masterarbeit, Universität Heidelberg.
- Hinnerichs, J. (in Vorb.). *Ein Testverfahren zur Differenzierung von bilingualen Kindern mit und ohne Sprachentwicklungsstörungen – Untersuchungen zum Nachsprechen einzelsprachunabhängiger Nichtwörter mit dem LITMUS-CL-NWRT bei Kindern mit Deutsch als Erst- und Zweitsprache*. Dissertation: Universität Heidelberg.
- Hinnerichs, J., Polišenská, K. & Chiat, S. (2016). *The Crosslinguistic Nonword Repetition Test: German version*.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26, 55-88.
- Holler-Zittlau, I., Euler, H. A., & Neumann, K. (2011). Kindersprachscreening (KiSS) – das hessische Verfahren zur Sprachstandserfassung. *Spracheheilerarbeit*, 5-6, 263-268.
- Jansen, H., Mannhaupt, G., Marx, H. & Skowronek, H. (2002). *Bielefelder Screening zur Früherkennung von Lese-Recht-schreibschwierigkeiten*. Göttingen: Hogrefe.
- Kannengieser, S. (2019). *Sprachentwicklungsstörungen – Grundlagen, Diagnostik und Therapie*. München: Elsevier.
- Kauschke, C. & Vogt, S. (2019). Positionspapier zur Terminologie und Definition von Sprachentwicklungsstörungen. *Logos*, 27 (3), 174-181.
- Marxen, R., Lüke, C. & Cholin, J. (2022). Kompetenzzempfinden bei der Sprachentwicklungsdiagnostik mehrsprachiger Kinder. *Logos*, 30 (2), 97-106.
- Meir, N. & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic Status (SES) and bilingualism on children's vocabulary and verbal short-term memory. *Frontiers in Psychology*, 8:1442.
- Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache (Hrsg.) (2013): *Qualitätsmerkmale für Sprachstandsverfahren im Elementarbereich. Ein Bewertungsrahmen für fundierte Sprachdiagnostik in der Kita*. Köln.
- Messer, M.H., Leseman, P.P.M., Boom, J. & Mayo, A.Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual pre-schoolers. *Journal of Experimental Child Psychology*, 105, 306-323.
- Parra, M, Hoff, E. & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually developing 2-year-olds. *Journal of Experimental Child Psychology*, 108, 113-125.
- Pierce, L. J., Genesee, J., Delcenserie, A. & Morgan, G. (2017). Variations in phonological working memory: Linking early language experiences and language learning outcomes. *Applied Psycholinguistics*, 38, 1265-1300.
- Plante, E. & Vance, R. (1994). Selection of Preschool Language Tests: A Data-Based Approach. *Language, Speech and Hearing Services in Schools*, 25, 15-24.
- Polišenská, K. & Kapalková, S. (2014). Improving Child Compliance on a Computer-Administered Nonword Repetition Task. *Journal of Speech, Language, and Hearing Research*, 57, 1060-1068.
- Polišenská, K., Chiat, S., Szewczyk, J., Antonijević-Elliott, S., Blom, E., Boerma, T., Bohnacker, U., Nutcher, M., Chan, A., Chilla, S., Chondrogianni, V., Fitzmaurice, Y., Fu, N. C., Gatt, D., Grech, H., Haddad, R., Hamann, C., Hinnerichs, J., Holzinger, D., Jezek, M., Judge-Clayden, F., Kapalková, S., Kunnari, S., Mayer-Crittenden, C., O'Malley, M. P., Öberg, L., Oudgenoeg-Paz, O., Polatidou, A., Schwob, S., Shiyun, M., Skoruppa, K., Tabone, N., van den Berghe, R., Verhagen, J. & White, M. (2022). *Is the crosslinguistic nonword repetition test valid crosslinguistically? Evidence from performance in different language groups and countries*. Vortrag auf der vierten Bi-SLI Konferenz 2022 in Berlin/ZAS.
- Rescorla, L. (2014). Entwicklungswege von Late-Talkern. In S. Ringmann & J. Siegmüller (Hrsg.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen – Jugend- und Erwachsenenalter* (S. 57-77). München: Elsevier.
- Rothweiler, M. (2007). „Mistaken identity“ – Zum Problem der Unterscheidung typischer grammatischer Strukturen bei SSES und bei Mehrsprachigkeit. In U. de Langen-Müller & V. Maihack (Hrsg.). *Früh genug – aber wie? Sprachförderung per Erlass oder Sprachtherapie auf Rezept? Tagungsbericht vom 8. wissenschaftlichen Symposium des Deutschen Bundesverbandes der akademischen Sprachtherapeuten* (S.110-128). Köln: ProLog.
- Scherger, A.-L. (2022). The role of age and timing in bilingual assessment: non-word repetition, subject-verb agreement and case marking in L1 and eL2 children with and without SLI. *Clinical Linguistics & Phonetics*, 36 (1), 54-74, <https://doi.org/10.1080/02699206.2021.1885497>
- Schöler, H. & Brunner, M. (2008). *HASE – Heidelberger Auditives Screening in der Einschulungsuntersuchung*. Wertingen: Westra.
- Schulz, P. & Tracy, R. (2011). *Linguistische Sprachstandserhebung – Deutsch als Zweitsprache (LiSe-DaZ)*. Göttingen: Hogrefe.

- Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Oliveira, P. R. & Skoruppa, K. (2021). Using Nonword Repetition to Identify Developmental Language Disorder in Monolingual and Bilingual Children: A Systematic Review and Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, 64, 3578-3593.
- Sorenson Duncan, T. & Paradis, J. (2016). English Language Learners' Nonword Repetition Performance: The Influence of Age, L2 Vocabulary Size, Length of L2 Exposure, and L1 Phonology. *Journal of Speech, Language, and Hearing Research*, 59, 39-48.
- Stadie, N. & Schöppe, D. (2013). *PhoMo-Kids. Phonologie Modellorientiert für Kinder vom Vorschulalter bis zum dritten Schuljahr*. Köln: Prolog.
- Szewczyk, J. M., Marecka, M., Chiat, S. & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, 179, 23-26.
- Thordardottir, E. (2015). Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Hrsg.), *Assessing Multilingual Children: Disentangling Bilingualism From Language Impairment* (S. 331-358). Bristol: Multilingual Matters.
- Tomblin, J. B., Records, N. L. & Zhang, X. (1996). A System for the Diagnosis of Specific Language Impairment in Kindergarten Children. *Journal of Speech and Hearing Research*, 39, 1284-1294.
- Topbaş, S. & Güven, S. (2011). *Türkçe Erken Dil Gelişim Testi – TEDİL-3:T*. Ankara: Detay.
- Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Hrsg.), *Assessing Multilingual Children: Disentangling Bilingualism From Language Impairment* (S.229-328). Bristol: Multilingual Matters.
- Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morrin, E., Prevost, P., Dos Santos, C., Abed Ibrahim, L. & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication Disorders*, 53, 4, S.888-904.
- Verhagen, J. & Andringa, S. J. (2021). *The role of existing language knowledge in bilingual and multilingual toddlers' repetition of cross-linguistic and language-specific nonwords*. *Linguistic Approaches to Bilingualism*. <https://doi.org/10.1075/lab.20050.ver>
- Vuolo, J. & Goffmann, L. (2020). Vowel Accuracy and Segmental Variability Differentiate Children With Developmental Language Disorder in Nonword Repetition. *Journal of Speech, Language, and Hearing Research*, 63, 3945-3960. https://doi.org/10.1044/2020_JSLHR-20-00166
- Watson, J. C. E. (2002). *The Phonology and Morphology of Arabic*. New York: Oxford University Press.
- Weinrich, M & Zehner, H. (2017). *Phonetische und phonologische Störungen bei Kindern*. Springer: Berlin.
- Zaretsky, E., Lange, B. P. & Hey, C. (2022): Phonological Short-Term Memory: When Bilingualism Matters. *Language Learning and Development*, <https://doi.org/10.1080/15475441.2021.2014846>
- Zebib, R., Henri, G., Khoms, A., Messara, C. & Hreich, E. (2017). *Batterie d'Evaluation du Langage Oral chez l'enfant libanais (ELO-L)*. Kerserwan: LTE.

Zur Autorin

Juliane Hinnerichs ist akademische Sprachtherapeutin (B. Sc. Patholinguistik) und absolvierte den Master „Deutsch als Zweitsprache“ an der Universität Heidelberg. Dort forscht und lehrt sie zu Bilingualismus, Spracherwerb und Sprachstandsfeststellungen. In ihrer Dissertation beschäftigte sie sich mit Sprachentwicklungsstörungen bei bilingual aufwachsenden Kindern.

Korrespondenzadresse

Juliane Hinnerichs, M.A.
Universität Heidelberg
Neuphilologische Fakultät
Institut für Deutsch als Fremdsprachenphilologie
Plöck 55
69117 Heidelberg
juliane.hinnerichs@posteo.de