

Gerrit Helm

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

Eine linguistisch-empirische und Tablet-gestützte Analyse

1 | Einleitung

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

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

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

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

eine Sondierung ('stärkere' vs. 'schwächere' Leser*innen) erreichbar wird (vgl. Bredel, Fuhrhop & Noack 2017).

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

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

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

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

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

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

Leistungs(roh-) wert der Proband*innen, welcher zumeist mit Hilfe einer Normtabelle in eine Aussage über altersgemäße Lesefertigkeit überführt werden kann (vgl. etwa Wimmer & Mayringer 2014: 16; Lenhard, Lenhard & Schneider 2018: 63).

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

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

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

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

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

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

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

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

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

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

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

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

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

3 | Forschungsvorhaben und Forschungsfrage

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

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

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

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

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

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

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

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

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

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

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

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

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

(Wimmer & Mayringer 2014)

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

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

4.1 | Kodierung und Analyse auf Segment-Ebene

Ein erster (auf der Hand liegender) Faktor für die Kodierung und Analyse ist sicherlich, die Anzahl der 'Buchstaben' pro Lese-Item zu ermitteln⁷. Rückbinden lässt sich dies z.B. an die

⁷ Ich setze den Terminus 'Buchstaben' hier in einfache Anführungszeichen, um zu markieren, dass mein Beitrag an dieser Stelle eher das Alltagsverständnis dieses Terminus meint und sich nicht konkret auf eine spezifische

‘Buchstaben’-Auffassung von Neef (2005): Dort werden neben dem Inventar des Modernen Römischen Alphabets (MRA) auch die Elemente |ä|, |ö|, |ü| sowie |ß| zu den ‘Buchstaben’ gezählt. Erreicht wird so quasi die Zählung der Tastatur-Anschläge ohne Leerzeichen. Konkret werden z.B. |sch| und |ch| als drei bzw. zwei ‘Buchstaben’ gezählt, das Wort „Schienen“ besteht in der Analyse also aus acht ‘Buchstaben’ (|s|+|c|+|h|+|i|+|e|+|n|+|e|+|n|). Operiert wird mit dem Inventar:

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

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

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

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

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

Definition des Begriffes (die es durchaus gibt) beruft. Aus eben diesem Grund setze ich ferner das Inventar dieser Bezugsgröße in |_|.

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

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

Ein Wort wie „Schienen“ besteht somit aus 4 Einheiten (|sch|+|ie|+|n|+|en|), das Item „Bäume können sprechen“ aus 14 Einheiten (|B|+|ä|+|u|+|m|+|e|+|k|+|ö|+|n|+|n|+|e|+|n|+|s|+|p|+|r|+|e|+|c|+|h|+|e|+|n|).

4.2 | Kodierung und Analyse auf segmentübergreifender Ebene

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

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

Eine dritte segmentübergreifende Größe (noch unterhalb des graphematischen Wortes) ist das Morphem. Morpheme und die morphologische Struktur als Einflussfaktoren auf das Lesen sind an vielen Stellen betrachtet und nachgewiesen worden (z.B. Carlisle & Stone 2005; Bredel, Noack & Plag 2013; Bangel & Müller 2014, 2015; Fix 2015). Aus den genannten Arbeiten jedoch ein Werkzeug zur Kodierung und Analyse herzuleiten, verbleibt eine Herausforderung: v.a. bleibt ungeklärt, in welchen Fällen das Auftreten morphologischer Komplexität als Lesehilfe und in welchen als Herausforderung fungiert. Zwar gilt die morphologische Konstanthaltung bei Stämmen und Affixen („Bäcker“ nicht „*Becker“ wegen „backen“) als Lesehilfe (z.B. Bangel & Müller 2015:21), doch zeigt Fix (2015) auch, dass das Auftreten genauso wie das Nicht-Auftreten morphologischer Komplexität als Leseherausforderung gelten kann. Im anglo-amerikanischen Raum weisen Carlisle & Stone (2005) bzw. Carlisle & Katz (2006) nach, dass Proband*innen Zweisilber aus zwei Morphemen (z.B. „sadly“, „shady“) schneller lesen als Zweisilber aus nur einem Morphem („lady“, „pretty“). Vor diesem Hintergrund soll in der

vorliegenden Kodierung auf eine Betrachtung der morphologischen Strukturen verzichtet werden.

4.3 | Kodierung und Analyse auf wortübergreifender Ebene

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

5 | Ergebnisse der linguistischen Kodierung und Analyse

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

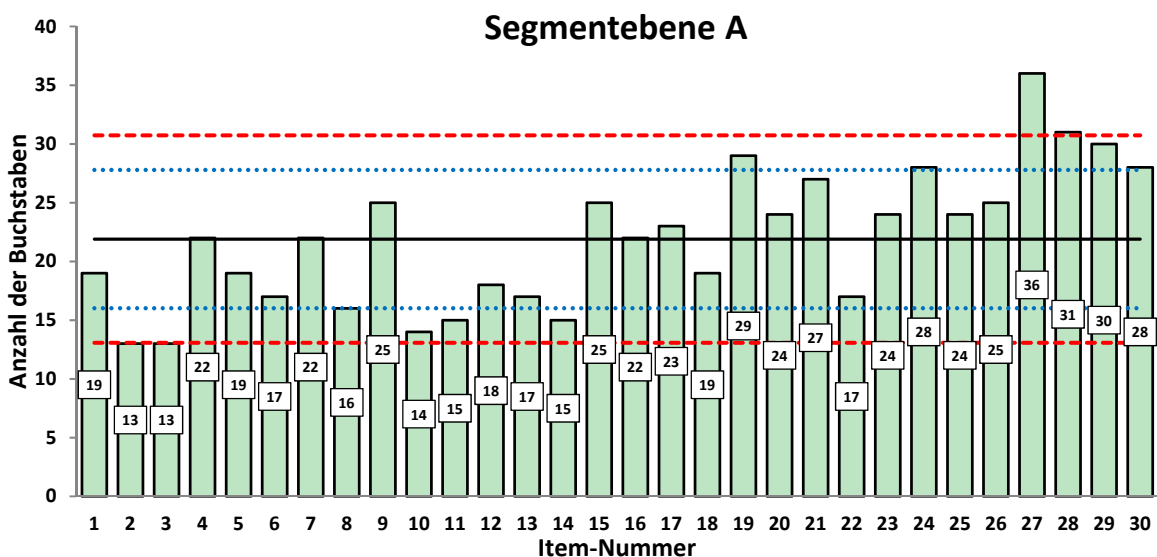


Abb. 1: Die Anzahl der Buchstaben in den 30 Lese-Items des SLS 2–9.

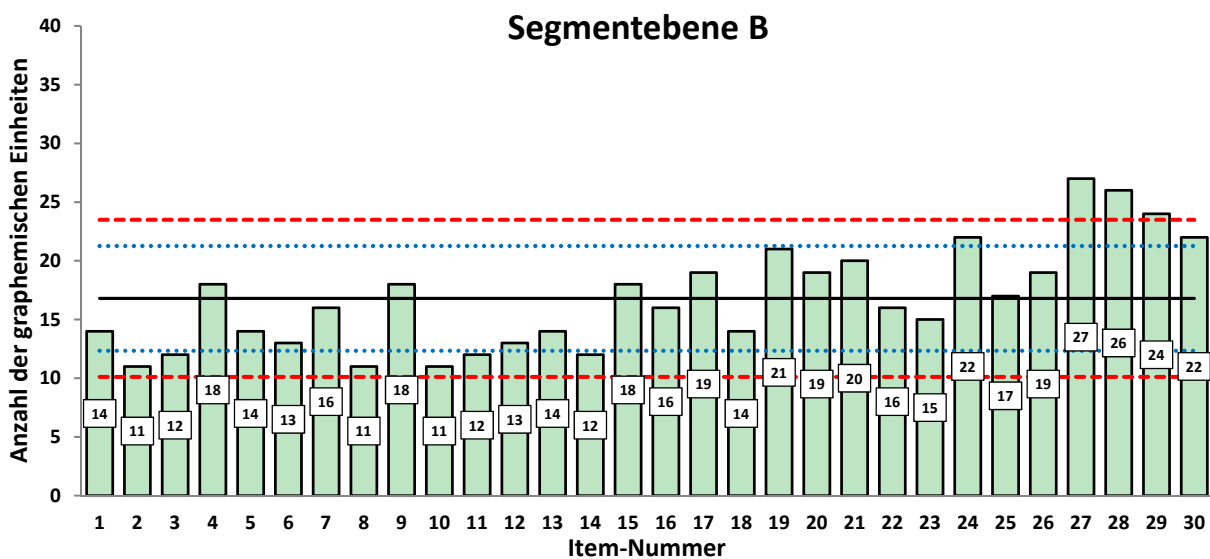


Abb. 2: Die Anzahl der graphemischen Einheiten in den 30 Lese-Items des SLS 2–9 .

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

- Die Items 27 („In einem Krankenhaus arbeiten viele Ärzte“) und 28 („Am Nordpol gibt es viel Eis und Schnee“) weichen sowohl hinsichtlich der Buchstaben als auch hinsichtlich der graphemischen Verarbeitungseinheiten deutlich (nach oben) ab.
- Die Items 2 („Wasser ist nass“) und 3 („Der Löwe brüllt“) weichen hinsichtlich der Buchstaben um mehr als 1,5 Standardabweichungen und hinsichtlich der graphemischen Einheiten um mehr als eine Standardabweichung (nach unten) ab.
- Weitere Items (8, 10, 11, 14, 24 und 29) weichen hinsichtlich beider Kodierungen um eine Standardabweichung vom Durchschnitt ab.

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

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

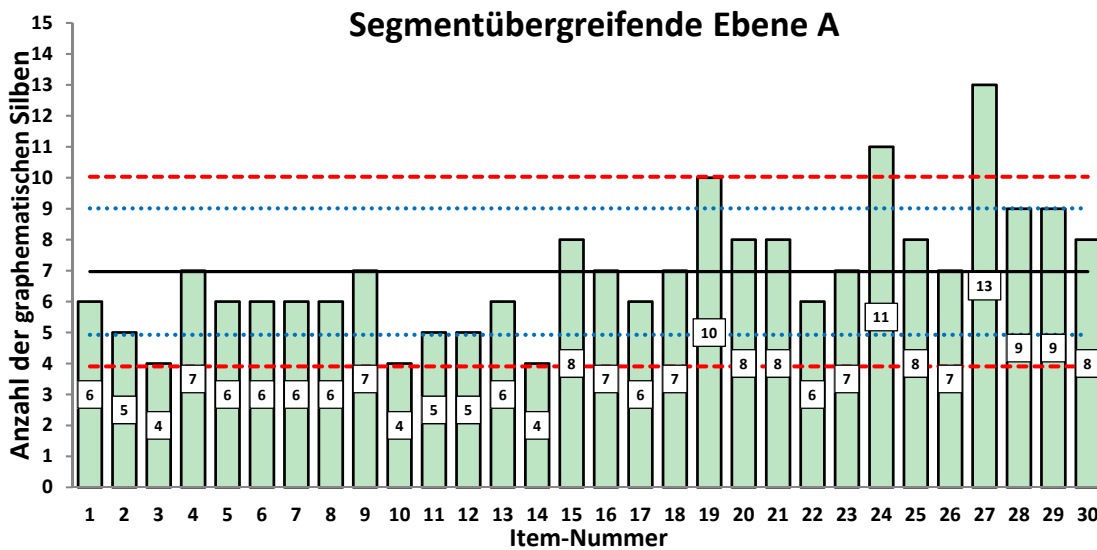


Abb. 3: Die Anzahl der graphematischen Silben in den 30 Lese-Items des SLS 2-9

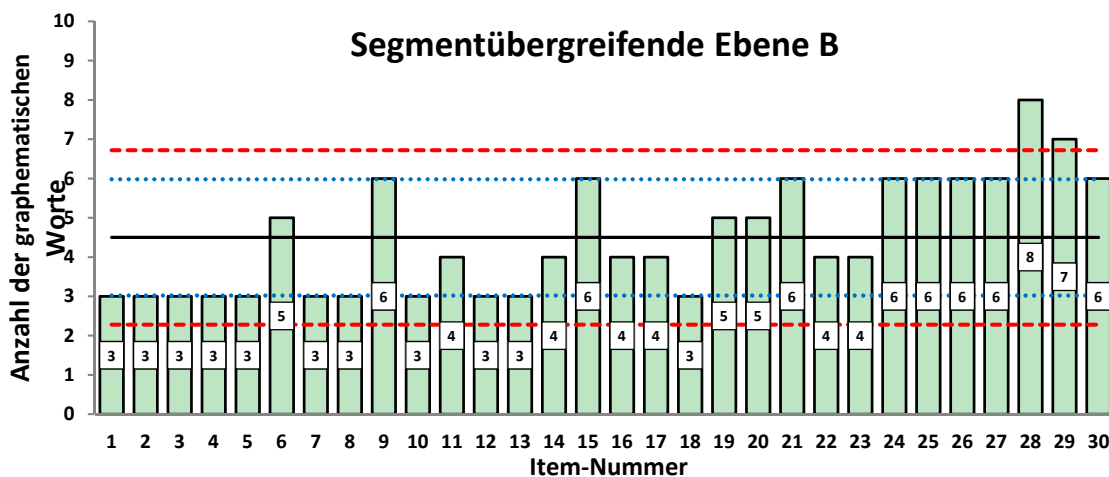


Abb. 4: Die Anzahl der graphematischen Worte in den 30 Lese-Items des SLS 2-9

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

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

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

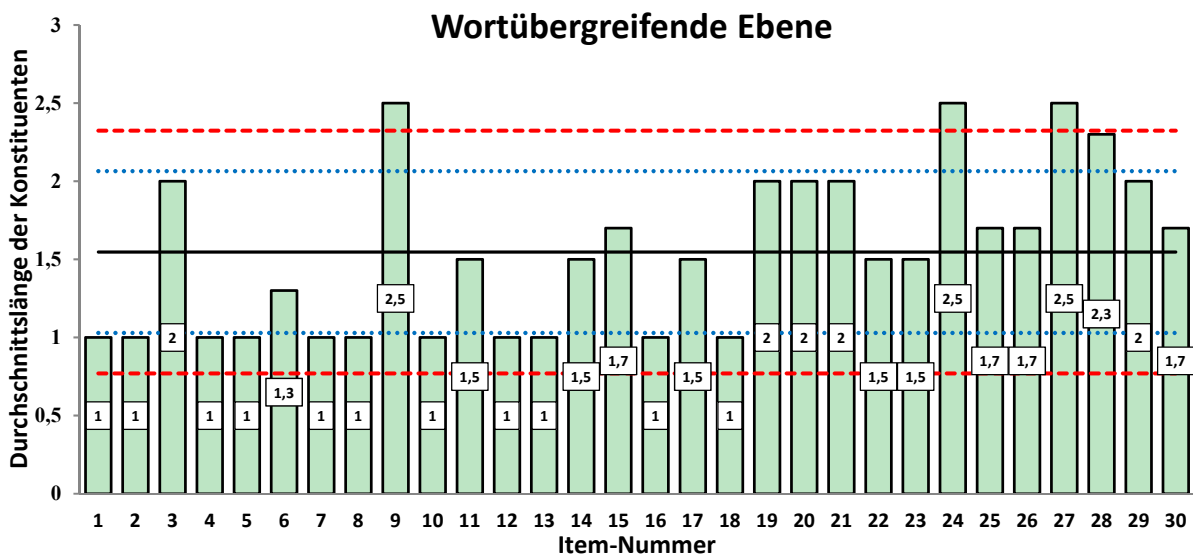


Abb. 5: Die durchschnittliche Länge der syntaktischen Konstituenten in den 30 Lese-Items

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

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

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

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

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

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

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

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

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

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

Auf Android-Tablets (konkret: Lenovo Tab M10) mit kontraststarkem 10,1 Zoll FullHD-Display wurde eine Variante einer (im Rahmen meines Dissertationsprojektes entwickelten) Lese-App mit den 30 Lese-Items aus der Form A1 des SLS 2–9 aufgespielt. Die App präsentiert jeweils ein einzelnes Lese-Item des Sets in schwarzer Fibel-Druckschrift (Fibel Nord) in konstanter Schriftgröße vor gleichbleibend weißem Hintergrund. Ein grüner Haken (=„dieser Satz stimmt“) oder ein rotes Kreuz (=„dieser Satz stimmt nicht“) im unteren Teil des Bildschirms dienen zur Bearbeitung/Beantwortung der Items (durch Antippen) vergleichbar zur Print-Version des SLS. Ferner können Items mit einem Tippen auf ein gelbes Fragezeichen übersprungen werden. Jedes Item wird durch einen Countdown („...“ „:“ „.“) angekündigt. Die App erfasst das Zeitintervall von der Darbietung des Items (Stimulus) am Ende des Countdowns bis zur Eingabe einer Antwort durch die Proband*in⁹. Die Items werden ferner durch die App vor jeder Durchführung in der Reihenfolge randomisiert, sodass jede Proband*in eine individuelle Abfolge der Items bearbeitet.

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

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

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

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

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

7 | Ergebnisse

In einem ersten Schritt sollen Datensätze von Schüler*innen aus dem Gesamtdatensatz herausgefiltert werden, die die Items nicht gelesen, sondern die Antworten ‘geraten’ haben. Wie auch in der Print-Fassung des SLS sind Ergebniswerte solcher Schüler*innen daran zu erkennen, dass ca. 50% der Items, in jedem Fall aber weniger als 75% der Items (23 von 30) korrekt bearbeitet wurden¹¹. Solche Datensätze werden in der hier dargebotenen Auswertung nicht berücksichtigt. Konkret umfasst der Datensatz damit 272 Ergebnissätze aus 17 Schulklassen:

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

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

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

Tab. 2: Zusammensetzung des Datensatzes

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

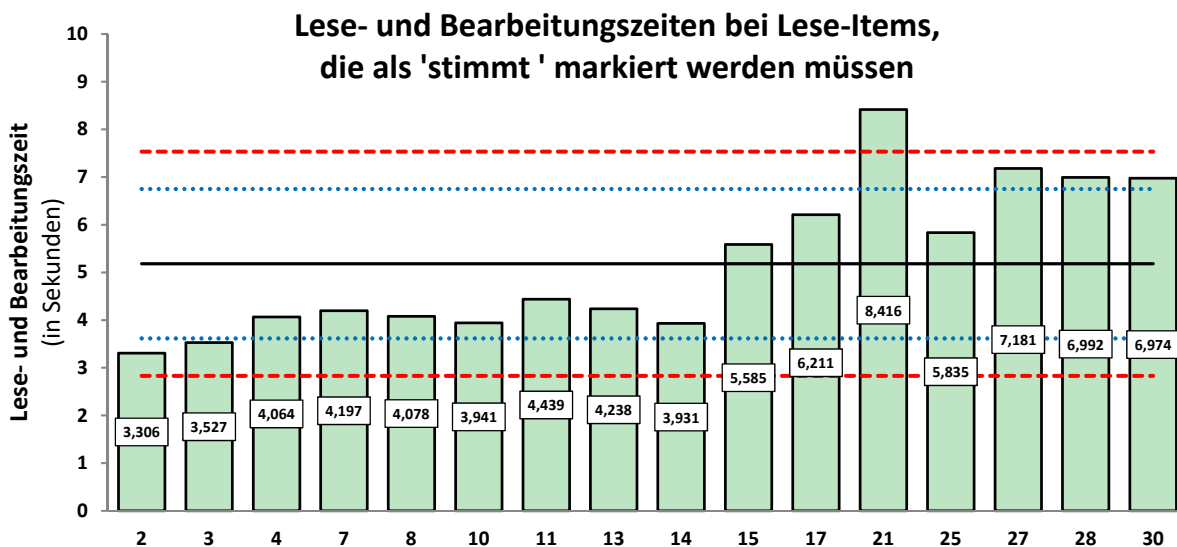


Abb. 7: Durchschnittliche Lese- und Bearbeitungszeiten aller Schüler*innen bei Items, die als 'stimmt' markiert werden mussten.

Erkennbar wird, dass die durchschnittlichen Lese- und Bearbeitungszeiten Item-bezogen durchaus variieren: Sie bewegen sich zwischen 3,3 und 8,9 s. Die Bearbeitung von Items, die als 'stimmt nicht' markiert werden sollen, erfordert durchschnittlich 0,34 Sekunden zusätzliche Zeit gegenüber solchen, die als 'stimmt' markiert werden müssen. Dieser Wert liegt sehr nahe an der von Krueger (1973) mit Tachistoskop-Verfahren festgestellten durchschnittlichen Verzögerung für *false*-Antworten in Klassenstufe 4 von 0,32 Sekunden (vgl. ebd.: 432). Die Datenwerte liegen hier also im Erwartungshorizont.

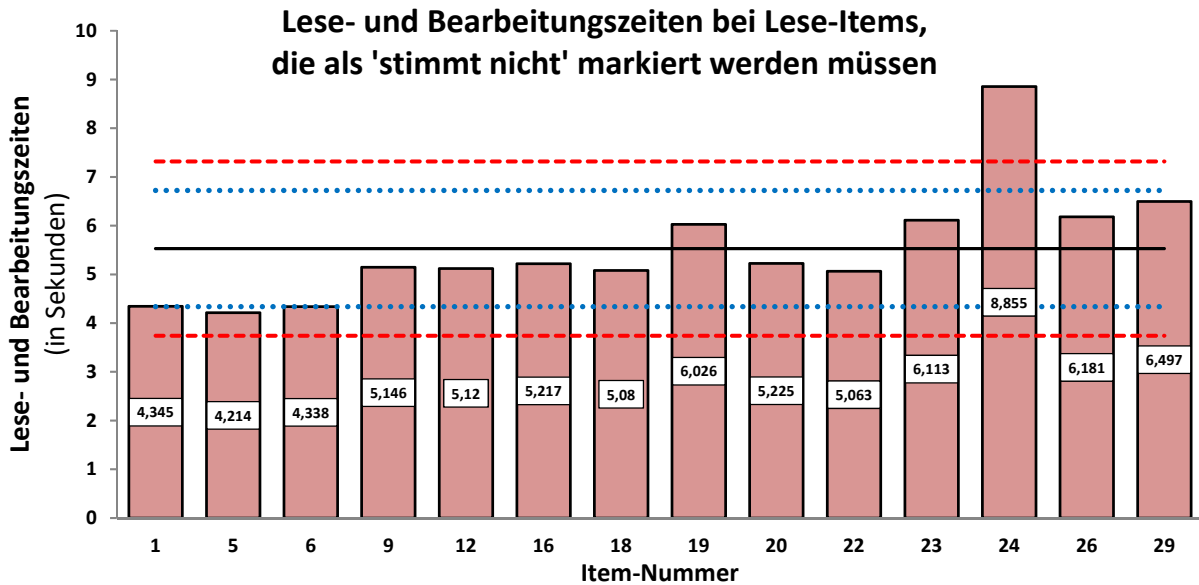


Abb. 8: Durchschnittliche Lese- und Bearbeitungszeiten aller Schüler*innen bei Items, die als 'stimmt nicht' markiert werden mussten.

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

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

Sichtbar wird, dass deutliche Differenzen in den Messwerten zwischen allen Klassenstufen bestehen. Hierbei gilt bei allen Items ausnahmslos: (Klasse 1 >) Klasse 2 > Klasse 3 > Klasse 4. Die Klassenstufe korrelieren signifikant ($p < 0.01$) mit den Lese- und Bearbeitungszeiten. Es zeigt sich hinsichtlich der Items, die als 'stimmt' markiert werden sollten, eine mittlere negative Effektstärke (Eff. -0.491) und hinsichtlich der Items, die als 'stimmt nicht' markiert werden sollten, sogar eine hohe negative Effektstärke (Eff. -0.528) nach Cohen (1988). Dies heißt (erwartungsgemäß): Je höher die Klassenstufe, desto geringer die Lese- und Bearbeitungszeit bei den Items.

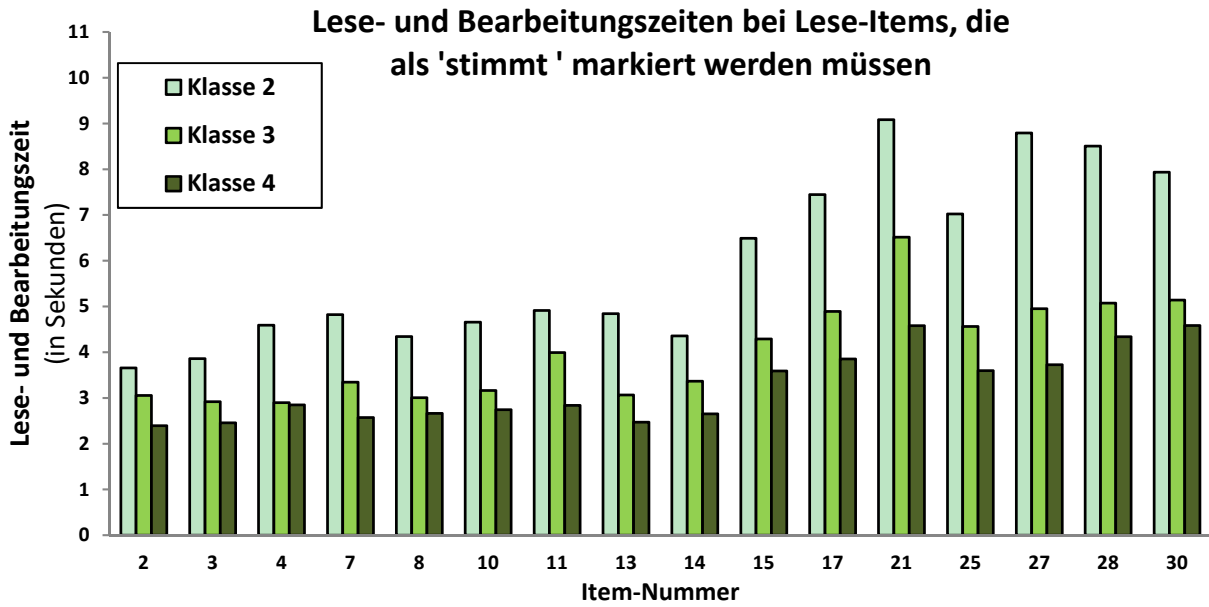


Abb: 9: Durchschnittliche Lese- und Bearbeitungszeiten bei Items, die als 'stimmt' markiert werden mussten nach Klassenstufe (2-4).

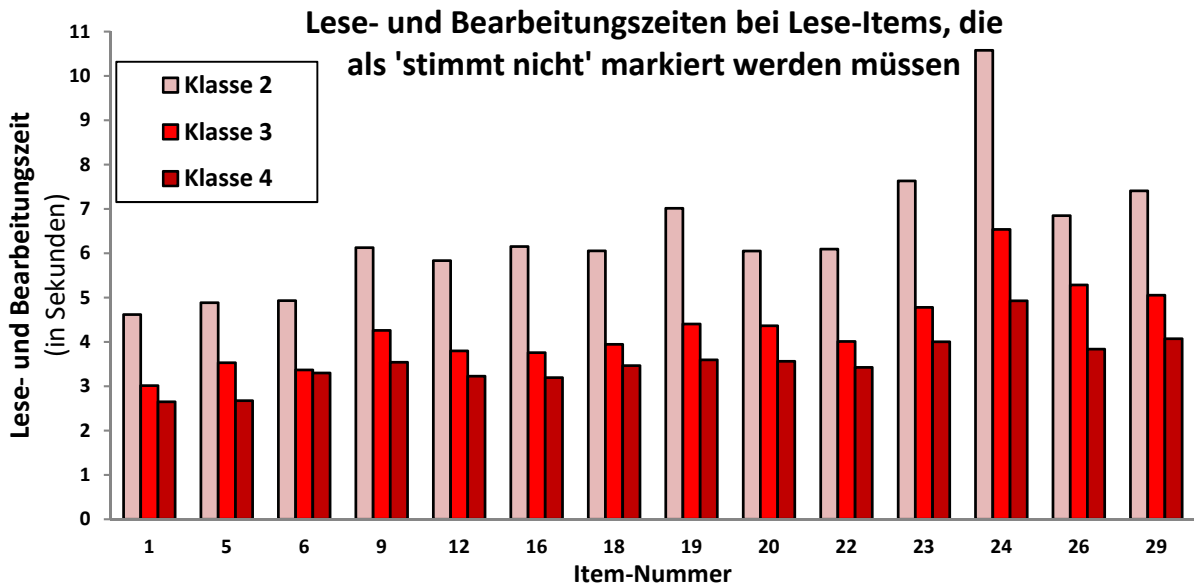


Abb: 10: Durchschnittliche Lese- und Bearbeitungszeiten bei Items, die als 'stimmt nicht' markiert werden mussten nach Klassenstufe (2-4).

Im Durchschnitt entstehen folgende Differenzwerte zwischen den getesteten Klassenstufen:

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

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

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

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

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

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

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

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

8 | Zusammenführung der theoretischen und der empirischen Analysen

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

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

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

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

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

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

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

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

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

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

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

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

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

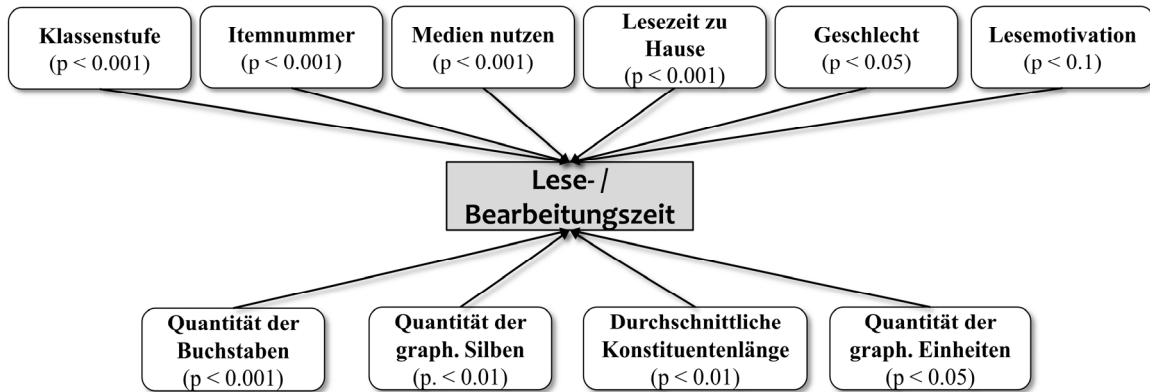


Abb. 11: Lineares Regressionsmodell zu Unterschieden in den erfassten Lesezeit-Werten

Die Analyse liefert damit folgende Ergebnisse:

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

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

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

Residual Standard Error: 4349 on 338 degrees of freedom

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

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

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

Es gilt: Verbleiben alle weiteren Faktoren konstant...

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

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

9 | Zusammenfassung und Diskussion

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

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

tatsächlich annimmt. *Dass* ein Ungleichgewicht festgestellt werden wird, galt bereits im Vorfeld als gesetzt, überraschend ist dennoch das festgestellte Ausmaß.

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

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

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

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

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

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

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

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

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

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

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

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

10 | Literatur

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

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

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