

Diagnostik als Basis für Förderung: Der Grazer Leseverständnistest

Lisa Paleczek* & Susanne Seifert*

Zusammenfassung

Um Kinder angepasst an ihre Lesefähigkeiten fördern zu können, ist es notwendig zu wissen, welche Teilfähigkeiten sie bereits beherrschen. Mittels digitalisierter Lesetests können diese Teilfähigkeiten meist recht unkompliziert erhoben werden und zudem erleichtert die Digitalisierung die Auswertung. Ein sich gerade in Entwicklung befindlicher digitaler Lesetest, der das Leseverständnis auf Wort-, Satz- und Textebene überprüft, ist der Grazer Leseverständnistest (GraLeV). Auf Wortebene sind die Schülerinnen und Schüler aufgefordert zu drei Bildern die passenden Wörter bei einer Auswahl aus sechs Wörtern zuzuordnen. Auf Satzebene wird eine Situation bildlich dargestellt und jener Satz, der sie beschreibt, muss aus vier Sätzen gewählt werden. Das Leseverständnis auf Textebene wird einerseits durch kurze Quatschgeschichten, zu denen jeweils eine explizite und schlussfolgernde Multiple-Choice-Frage beantwortet werden muss, erhoben. Andererseits wird es durch Texte mittels Mazetechnik, bei der jedes siebte Wort aus einer Auswahl von drei Wörtern gewählt werden muss, gemessen. Der vorliegende Beitrag stellt den GraLeV vor und geht auf die Durchführung, Auswertung und Interpretation ein.

Schlüsselwörter: Leseverständnis, Testentwicklung, digitale Messung



Originalbeitrag
open access

* Korrespondenz:

Ass.-Prof.ⁱⁿ Mag.^a Mestre Lisa
Paleczek, PhD
lisa.paleczek@uni-graz.at

Ass.-Prof.ⁱⁿ Dipl.-Patholog.ⁱⁿ
Susanne Seifert, PhD
susanne.seifert@uni-graz.at
Universität Graz
Institut für Bildungsforschung
und PädagogInnenbildung
Abteilung Inklusive Bildung und
Heilpädagogische Psychologie
Merangasse 70/2
8010 Graz, Österreich

Zitation: Paleczek, L. & Seifert, S.
(2021). Diagnostik als Basis für
Förderung: Der Grazer
Leseverständnistest
Sprachtherapie aktuell:
Forschung - Wissen - Transfer 2:
Schwerpunktthema: Perspektiven
auf Beeinträchtigungen der
Schriftsprache: e2021-36; doi:
10.14620/stadbs210736

1 Einleitung

Um Kindern maßgeschneiderte Leseförderung zuteilwerden zu lassen, ist es vorab notwendig, ein möglichst differenziertes Bild der Leseverständnisleistungen der Kinder zu erheben. Basierend darauf kann Leseförderung individualisiert angeboten werden (Paleczek & Seifert, 2020; Wanzek & Vaughn, 2007). Lehrpersonen sind in der Regel in der Lage, die Fähigkeiten ihrer Schülerinnen und Schüler einzuschätzen, jedoch existieren Faktoren, die das Urteil in manchen Fällen beeinflussen (Paleczek, Seifert, & Gasteiger-Klicpera, 2017; Südkamp, Kaiser, & Möller, 2012). Um Lehrpersonen und Personen, die in der Leseförderung tätig sind, zusätzlich zu ihrer Expertise noch die Möglichkeit eines objektiven Verfahrens an die Hand zu geben, ist es sinnvoll, standardisierte Lesetests zum Einsatz kommen zu lassen. Je differenzierter die Aussage des Tests ist, umso differenzierter kann auch die darauf basierende Förderung aufgebaut werden. Wenn die Leseverständnisfähigkeit nicht lediglich über kurze Sätze (z. B. Salzburger Lesescreening 2-9: Mayringer & Wimmer, 2014) erhoben wird, ist eine Testung zwar aufwändiger, gibt jedoch Informationen über mehrere Komponenten der Lesefähigkeiten. Deutschsprachige Testverfahren, die zu solchen eher aufwändigen, jedoch detaillierten Lesetests zählen, sind beispielsweise der HAMLET 3-4 (Hamburger Lesetest für 3. und 4. Klassen: Lehmann, Peek, & Poerschke, 2006) und der ELFE II (Ein Leseverständnistest für Erst- bis Siebtklässler – Version II: Lenhard, Lenhard, & Schneider, 2020). Der ELFE II ist dabei ein auf die Erhebung des Leseverständnisses spezialisiertes Testverfahren. Leseverständnis wird definiert als die Fähigkeit, die Bedeutung von geschriebenen Wörtern, Sätzen oder Texten zu verstehen (Perfetti, Landi, & Oakhill, 2005) und involviert Prozesse auf verschiedenen hierarchischen Ebenen (Mullis & Martin, 2016). Die erste Ebene bildet die Wortebene, der eine Buchstaben-Laut-Zuordnungsfähigkeit (Dekodierfähigkeit) der Kinder zugrunde liegt. Wenn Wörter erlesen werden können und deren Bedeutung den Kindern verständlich ist, ist die erste Grundlage des Leseverständnisses gemeistert. Werden diese Wörter jedoch durch syntaktische Elemente zu Sätzen verbunden, ist die Aufgabe ans Kind bereits eine schwierigere. Auf Satzebene reicht ein Verstehen der einzelnen Wörter nicht mehr aus, sondern es müssen auch komplexere Strukturen verstanden

werden (Syntax), die die Wörter zu einem zusammenhängenden Satz durch sprachliche Einheiten verknüpfen. Die Textebene schließt dann noch weitere Faktoren, wie die Fähigkeit lokale und globale Kohärenzen zu erschließen, mit ein (Mullis & Martin, 2016). Um ein differenziertes Bild über die Fähigkeiten der Kinder und eventuelle Schwierigkeiten zu bekommen und eine gezielte Förderung anzusetzen, ist eine separate Erfassung des Leseverständnisses auf diesen drei Ebenen sinnvoll.

2 Lesetests digital vs. Print

Der Trend zur Digitalisierung macht aufgrund der vielen Vorteile in Bezug auf Anwendung und Auswertung auch vor der Lesediagnostik keinen Halt. Einige Messinstrumente für die Erhebung der Leseverständnisleistung bieten mittlerweile neben der gängigen Printerhebung, bei denen die Kinder in Testheften die Aufgaben lösen, eine digitale Version an, die mittels Tablet oder Computer bearbeitet werden kann (bspw. ELFE II: Lenhard et al., 2020). Es gibt aber auch bereits Erhebungsinstrumente, die ausschließlich eine digitale Bearbeitung anbieten (LEVUMI: Kompetenzorientierte Lernverlaufdiagnostik: Gebhardt, Diehl, & Mühling, 2016; ProDi-L: Prozessbezogene Diagnostik von Lesefähigkeiten im Grundschulalter: Richter, Naumann, Isberner, Neeb, & Knoepke, 2017).

Digitale Lesetests bringen einige Vorteile, aber auch Herausforderungen mit sich. Ein Vorteil digitaler Formate liegt sicherlich in der Erleichterung der Vorbereitung, Durchführung, Auswertung und Interpretation der Ergebnisse (Cheung & Slavin, 2012). Bei Printformaten müssen die durchführenden Personen nach Testdurchführung für jedes Kind einzeln das Testheft auswerten, meist Summenscores aus einzelnen richtig beantworteten Items bilden und dann die Ergebnisse mit einer Tabelle in Normwerte umwandeln, die Aufschluss über die Leistungen des Kindes, gemessen an der passenden Altersstufe, gibt. Das kann mitunter, wenn es sich beispielsweise um Klassentestungen handelt, einige Zeit in Anspruch nehmen und ist zudem durch die einzelnen Schritte, die nötig sind, fehleranfällig. Digitale Instrumente können diese Schritte automatisieren und auch die Dokumentation (für Lernverlaufsmessungen und Evaluierung einer gesetzten Fördermaßnahme) erleichtern. Dadurch, dass Lehrpersonen durch die digitale Testung Zeit einsparen, bleibt ihnen mehr Zeit für Instruktion und Förderung (Gebhardt et al., 2016).

Allerdings ergeben sich für den Einsatz von digitalen Lesetests auch einige Hürden. Eingeschränkte digitale Kompetenzen bzw. Vorerfahrungen sowie eine unzureichende digitale Ausstattung (Tablets, Computer, Internet) können die Anwendung von digitalen Tests erschweren oder sogar gänzlich unmöglich machen. Selbst wenn die Förderung digitaler Kompetenzen im

österreichischen (Bundesministerium Bildung, Wissenschaft und Forschung [BMBWF], 2020) und deutschen Kontext (DigitalPakt Schule 2019 bis 2024: Bundesministerium für Bildung und Forschung [BMBF], 2019) bereits im Grundschulalter verankert ist, sind viele Lehrpersonen noch nicht daran gewöhnt, digitale Geräte im Unterricht einzusetzen (Brandhofer, 2015). Dies ist neben der fehlenden Ausstattung auch nicht selten den fehlenden Kompetenzen der Lehrpersonen geschuldet (Balula, 2016). Auch den Schülerinnen und Schülern fehlen mitunter die Kompetenzen, digitale Geräte selbstständig einzusetzen (Medienpädagogischer Forschungsverbund Südwest, 2018).

Digitale Testinstrumente unterliegen genau wie die Varianten in Print strengen Untersuchungen, damit sie reliabel und valide die Fähigkeiten, die das Instrument zu erfassen intendiert, messen – und zwar genauso gut und zuverlässig wie ein Print-Test. Auch wenn die Forschung im Bereich des Vergleichs von Testungsarten (digital vs. Print) voranschreitet (Kindergarten: Palczek, Seifert, & Schöfl, 2021; 1. bis 6. Schulstufe: Lenhard, Schroeders, & Lenhard, 2017; 4. Schulstufe: Hamhuis, Glas, & Meelissen, 2020; 9. Schulstufe: Jerrim, Micklewright, Heine, Salzer, & McKeown, 2018; Metaanalyse Kindergarten bis 12. Schulstufe: Wang, Jiao, Young, Brooks, & Olson, 2007), sind die Ergebnisse dennoch nicht einheitlich (Blumenthal & Blumenthal, 2020). In einigen Studien wird berichtet, dass die Testergebnisse in Print- und digitaler Form gleich oder mit vernachlässigbaren Unterschieden ausfallen (Hamhuis et al., 2020; Wang et al., 2007), in anderen ist hingegen von Unterschieden die Rede (Lenhard et al., 2017). Beim Lesetest ELFE II (Lenhard et al., 2020) beispielweise, der wie bereits erwähnt sowohl in Printformat als auch in digitalem Format vorliegt, sind die Ergebnisse in den beiden Versionen zwar vergleichbar ausgefallen, dennoch zeigten sich Unterschiede vor allem bei jüngeren Kindern. Diese lösten die Aufgaben im digitalen Format zwar schneller, jedoch weniger akkurat (Lenhard et al., 2017).

Unabhängig von der Vergleichbarkeit der Ergebnisse kann eine allgemeine Akzeptanz digitaler Testformate sowohl bei Lehrpersonen (Förster & Souvignier, 2014) als auch bei Kindern (Blumenthal & Blumenthal, 2020) festgestellt werden. Blumenthal and Blumenthal (2020) konnten für einen Mathematik-Test zeigen, dass Kinder das digitale Testungsformat sogar bevorzugten und meinten, dass sie dabei auch besser abschneiden würden, wobei dieser Eindruck nicht den erhobenen Daten entsprach.

Im Hinblick auf die eindeutigen Vorteile und bereits nachgewiesene Akzeptanz von digitalen Testformaten, wurde der Grazer Leseverständnistest (GraLeV: Palczek, Seifert, Franz, Riedl, & Wohlhart, in Vorbereitung) zuerst

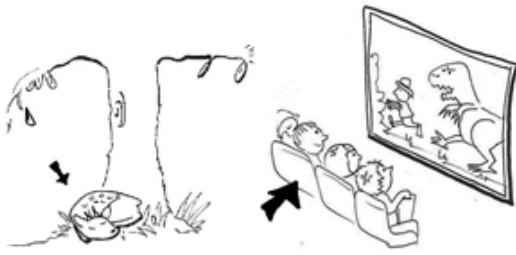


Abbildung 1: Bilder im GraLeV Version 1 mit Zielwort Kitz und Ablenker Sitz (interpretiert als Kids).

als digitaler Lesetest entwickelt. Erste Überprüfungen der Validität und Reliabilität ergaben bereits zufriedenstellende Werte (Palczek & Seifert, in Vorbereitung a; Seifert & Palczek, eingereicht; in Druck). Der vorliegende Beitrag widmet sich der Konzipierung und Entwicklung des Tests und beschreibt den Testaufbau, die Durchführung und Auswertung.

3 Konzipierung, Entwicklung und Pilotierungsuntersuchungen des GraLeV

Der GraLeV ist ein Leseverständnistest, der aus vier Subtests besteht und für die Grundstufe II (3. und 4. Schulstufe) konzipiert wurde. Mit diesem Test kann das Leseverständnis auf Wort-, Satz- und Textebene erhoben werden, wobei die Textebene durch zwei unterschiedliche Subtests (Textebene Subtest I und II) erfasst wird. Der GraLeV wurde im Rahmen mehrerer Vorstudien immer weiterentwickelt (Pilotierung 1, Pilotierung 2/Powertestung, Validierung). In seinen ersten Versionen war der GraLeV auf einer Lernplattform eingebettet und konnte nur online bearbeitet werden.

In seiner ersten Version (Pilotierung 1: Sommer 2019) bestand der GraLeV aus insgesamt 230 Items (Wort: 114 (38 Itemsets mit je 3 Items), Satz: 24, Text Subtest I: 32 (16 Texte mit je 2 Fragen), Text Subtest II: 60 (4 Texte mit je 15 Auswahlklammern)). Der GraLeV wurde in dieser langen Version zehn Kindern auf dem Tablet zur Bearbeitung vorgelegt und die Kinder wurden aufgefordert, während der Bearbeitung laut mitzusprechen (think aloud), welche Gedankengänge sie beim Bearbeiten und Lösen hatten. Dadurch konnten Informationen dazu gewonnen werden, ob (a) der Testaufbau und die Aufgabenstellungen nachvollziehbar waren, (b) es technische Herausforderungen gab und (c) die Items inhaltlich logisch waren. Nach der Rückmeldung der Kinder wurden Items eliminiert bzw. adaptiert, um Missverständnisse zu vermeiden. Um ein Beispiel für eine notwendige Adaptierung zu nennen, wurde das Zielwort Kitz entfernt, da der Ablenker „Sitz“, dargestellt als Kinositze mit Kindern darauf, von einem der Kinder in der Pilotierung 1 als „Kids“ (die Kinder auf dem Sitz) interpretiert wurde (Bilder dazu siehe Abbildung 1).

Daraufhin entstand der GraLeV in seiner zweiten Version mit insgesamt 221 Items (Wort: 105 (35 Itemsets mit 3 Items), Satz: 24, Text Subtest I: 32 (16 Texte mit je 2 Fragen), Text Subtest II: 60 (4 Texte mit je 15 Auswahlklammern)).

Diese zweite Version des GraLeV wurde dann im Rahmen einer Powertestung (Pilotierung 2: 10/2019-12/2019) mit 273 Schülerinnen und Schülern der dritten und vierten Schulstufe im Klassensetting durchgeführt. Das Vorgehen war so gewählt, dass keine Zeitbeschränkung bestand, sondern abgebrochen wurde, sobald 80 Prozent der teilnehmenden Kinder in der Klasse fertig waren. Mit diesem Vorgehen (Powertestung statt Speedtestung) gelang es uns (a) Zeitlimits für die Subtests zu ermitteln, (b) Itemcharakteristika wie Schwierigkeit und Trennschärfe zu berechnen sowie (c) eine Distraktoranalyse durchzuführen. Anhand dieser Daten wurde dann in einem weiteren Schritt entschieden, welche Items in den GraLeV aufgenommen werden und deren Reihenfolge (nach Schwierigkeit geordnet) festgelegt. Items mit Schwierigkeiten unter .5 und über .98 wurden ausgeschlossen, ebenso Items mit einer Trennschärfe unter .2 sowie Items, bei denen ein bestimmter Distraktor besonders oft gewählt wurde (mehr als 20%). Häufig traten diese problematischen Werte in mehreren Bereichen simultan auf (Schwierigkeit, Trennschärfe und Distraktorenanalyse). Dabei war zu berücksichtigen, dass die bestehenden Itemsets auf Wortebene (pro Aufgabe 3 Zielwörter als Itemset) bzw. die Geschichten auf Textebene I (Quatschgeschichten mit einer Frage zu einer explizit im Text angegebenen Information und einer Frage zu implizitem Schlussfolgern) nur dann behalten wurden, wenn alle Einzelitems der Itemsets die genannten Anforderungen erfüllten. Außerdem wurde die Anzahl der Items reduziert, um die Testzeit nicht zu verlängern (bspw. erfüllten auch 11 weitere Itemsets auf Wortebene die Kriterien und hätten in die finale Testversion übernommen werden können), da es auch ein Ziel war, einen ökonomisch und unkompliziert einsetzbaren Test zu entwickeln.

4 Subtests des GraLeV

Basierend auf den Vortestungen (Pilotierung 1 und 2) entstand der GraLeV in der finalen Version mit 98 Items (Wort: 36 (12 Itemsets mit je 3 Items), Satz: 16, Text Subtest I: 16 (8 kurze Texte mit je 2 Fragen), Text Subtest II: 30 (2 Texte mit je 15 Auswahlklammern)). Allen Subtests ist gemein, dass vorab Beispielitems dargeboten werden, um das Instruktionsverständnis der Schülerinnen und Schüler zu sichern. Im Folgenden soll genauer auf den Aufbau der unterschiedlichen Subtests, die das Leseverständnis auf den verschiedenen Ebenen erfassen, eingegangen werden.

Ziehe das richtige Wort zu jedem Bild.

Esc drücken, um das Vollbild zu beenden.

Seite 1/12

172 Sekunden

Tasche Rutsche Lasche Masche Kutsche Flasche

ABBRECHEN

Abbildung 2: Wortebene im GraLeV (Wortart Nomen; Zielwörter: Flasche, Rutsche und Tasche).

4.1 Wortebene

Der Subtest Wortebene besteht aus zwei Beispielitems und 36 Items (in zwölf Itemsets mit je drei Zielwörtern) mit einem Zeitlimit von 3 Minuten. Das Aufgabenformat sind Wort-Bildzuordnungen. Jede Aufgabe besteht aus drei Bildern und sechs schriftlich dargebotenen Wörtern. Abbildung 2 zeigt exemplarisch ein Item aus diesem Subtest des GraLeV. Die Schülerinnen und Schüler müssen aus den sechs dargebotenen Wörtern jene drei auswählen, die zu den drei Bildern passen und mit Drag and Drop den Bildern zuordnen. Die Wörter, die in einer Aufgabe dargeboten werden (Zielwörter und Distraktoren), gehören immer derselben Wortart an.

Insgesamt werden fünfzehn Nomen, fünfzehn Verben und sechs Adjektive als Zielitems präsentiert. Drei aus den sechs dargebotenen Wörtern pro Aufgabe dienen als Distraktoren. Die Distraktoren wurden aufgrund semantischer und/oder phonologischer Ähnlichkeit zu den Zielwörtern ausgewählt. Jedes Zielwort und jeder Distraktor zeichnen sich zudem dadurch aus, dass das Wort in der Altersgruppe, für die der GraLeV entwickelt wurde, Relevanz hat. Die Wörter der Wortebene im GraLeV wurden basierend auf den Frequenzangaben der Datenbank Childlex (Schroeder, Würzner, Heister, Geyken, & Kliegl, 2015) ausgewählt. Anhand dieser Datenbank ist es möglich, die Häufigkeit eingepflegter Wörter basierend auf der aktuellen Schriftsprache von Grundschulkindern zu ermitteln. Der zugrundeliegende Korpus besteht aus Kinder- und Schulbüchern und umfasst zehn Millionen Wörter. Es werden Daten zu den Wörtern in drei verschiedenen Altersklassen angeboten, wobei für den GraLeV auf jene der Altersklasse der Neun- bis Zehnjährigen zurückgegriffen wurde (das Alter also, das Kindern der dritten und vierten Schulstufe und damit der Zielgruppe des GraLeV entspricht). Es wurden nur Wörter verwendet, die für die Altersklasse eine Wortfrequenz von mindestens 1 in 1 Million Wörter aufwiesen, da die Wortfrequenz als ein wesentliches Merkmal, das den Wortabruf auf lexikalischer Ebene beeinflusst, gilt (Griffin & Bock, 1998).

4.2 Satzebene

Der Subtest Satzebene umfasst zwei Beispielitems und 16 Items mit einem vorgegebenen Zeitlimit von drei Minuten. Die Aufgabenstellung zeigt ein Bild, auf dem eine Tätigkeit oder Situation dargestellt wird. Zudem werden vier Sätze präsentiert, wobei nur ein Satz tatsächlich die Tätigkeit oder Situation beschreibt und drei weitere als Distraktoren fungieren. Auch hier wurden die Distraktoren systematisch konstruiert, indem grammatische Ablenker gebildet wurden, die zwar den Regeln der Grammatik entsprechen, aber nicht die Tätigkeit auf dem Bild adäquat beschreiben. Es wurden jeweils ein syntaktischer und ein semantischer Distraktor gewählt sowie eine Kombination der beiden. Die syntaktischen Distraktoren wurden entsprechend folgender syntaktischer Aspekte erstellt:

- Plural vs. Singular
- alternatives Subjekt (mit Pronomen dargestellt: er statt sie)
- aktiv vs. passiv (Subjekt und Objekt vertauscht)
- Negationen (wo im Satz das „nicht“ steht)
- Perfekt vs. Präsens (er pflückt vs. er hat gepflückt)
- Subordination (während/ nachdem)
- Topikalisierung (Den Hund jagt das Pferd vs. Der Hund jagt das Pferd)

Abbildung 3 zeigt exemplarisch ein Item der Satzebene. Der syntaktische Distraktor beruht auf Subjekt-Objekt-Alternierung im Dativ. Der semantische Distraktor bezieht ein semantisch-thematisch verwandtes Objekt mit ein (Weihnachtsmann statt Weihnachtsbaum).

4.3 Textebene Subtest I (Quatschgeschichten)

Die Textebene im Leseverständnis wird im GraLeV mit zwei verschiedenen Subtests erhoben (Seifert & Palczek, in Druck). Subtest Textebene I besteht aus Quatschgeschichten, in denen Informationen über nichtexistierende Dinge, Kreaturen (beides Nomen) oder Tätigkeiten (Verben) dargeboten werden. Jeder Quatschgeschichte folgen zwei Fragen. Die erste Frage bezieht sich auf eine

Welcher Satz passt zum Bild? Klicke auf den Satz.

Seite 7/16

Fullscreen

59 Sekunden



Mit der Weihnachtskugel schmückt der Junge den Weihnachtsbaum.

Mit der Weihnachtskugel schmückt der Junge den Weihnachtsmann.

Mit dem Weihnachtsbaum schmückt der Junge den Weihnachtsmann.

Mit dem Weihnachtsbaum schmückt der Junge die Weihnachtskugel.

ABBRECHEN

Abbildung 3: Satzebene im GraLeV.

explizit in der Quatschgeschichte angegebene Information, wohingegen die zweite Frage verlangt, einfache Schlussfolgerungen zu ziehen, um die Frage beantworten zu können. Diese beiden Prozesse (explizite Informationsentnahme und implizite Schlussfolgerungen) werden auch in internationalen Large-Scale-Studien verwendet, um Textverständnis differenziert zu erfassen (bspw. PIRLS am Ende der vierten Schulstufe: Widauer & Wallner-Paschon, 2017). Beide Fragen haben im GraLeV immer vier Antwortmöglichkeiten, wobei nur eine Antwort jeweils die richtige darstellt und die anderen als Distraktoren dienen. Dies vermindert die Ratewahrscheinlichkeit. Abbildung 4 zeigt exemplarisch ein Item aus dem Subtest Quatschgeschichten mit den beiden Fragen.

Durch die Verwendung von Quatschgeschichten anstelle von realen Dingen, Lebewesen oder Tätigkeiten kann ausgeschlossen werden, dass die Kinder die Fragen bereits aufgrund ihres Vorwissens und nicht aufgrund ihrer Fähigkeit, Informationen aus dem gelesenen Text zu extrahieren, beantworten können. Insgesamt besteht

dieser Subtest aus einem Beispielitem (eine Quatschgeschichte mit zwei Fragen – eine explizite und eine implizite) und acht Quatschgeschichten mit je zwei Fragen, wobei innerhalb von drei Minuten so viele Items wie möglich zu lösen sind. Sechs dieser Quatschgeschichten sind sehr einfache und kurze Geschichten, die aus zwei bis drei kurzen Sätzen bestehen. Die zwei anderen sind längere Geschichten, bestehend aus vier Sätzen. In Tabelle 1 sind die Charakteristika der Quatschgeschichten präsentiert. Die Lesbarkeit der kurzen Texte wurde mittels Regensburger Analyseinstrument Ratte (Wild & Pissarek, k.A.) mit der Lesbarkeitsformel aus dem Regensburg-Index von Wild and Pissarek (2020) berechnet. In diesem Index werden einerseits Charakteristika, wie beispielsweise Silbenanzahl der Wörter und Anzahl der Sätze, und andererseits Schwierigkeitsparameter der Sätze, wie beispielsweise das Vorkommen passiver Satzkonstruktionen und Satzkomplexität, berücksichtigt. Je höher der Index, umso schwieriger ist der Text.

Lies die Geschichte und beantworte die Fragen.

Seite 2/8

Fullscreen

147 Sekunden

Emre freut sich. Die Schule ist aus. Dann kann er jetzt endlich branteln.

Was steht in der Geschichte?

Die Schule beginnt.

Emre freut sich nicht.

Die Schule ist aus.

Emre ärgert sich.

Was steht in der Geschichte?

Emre brantelt gern.

Emre kann noch nicht branteln.

Emre brantelt in der Schule.

Emre geht nicht in die Schule.

ABBRECHEN

Abbildung 4: Quatschgeschichte „branteln“ mit den dazugehörigen zwei Fragen.

Tabelle 1: Textcharakteristika im GraLeV Subtest Textebene I (Quatschgeschichten).

Texttyp	Nummer des Texts	Quatschwörter	Wortart	Anzahl der Wörter	Anzahl der Sätze	RIX
kurz	Beispielitem	Tinatos Kanat	Kreaturen, Ding	17	3	2.13
	1	Relemis	Kreaturen	14	2	2.44
	2	branteln	Ding	13	3	1.86
	3	Rafiza	Ding	12	2	2.22
	4	minnern	Tätigkeit	16	3	2.09
	5	krolken	Tätigkeit	20	2	2.93
lang	6	Basati	Ding	18	3	2.22
	7	Sinalas	Kreaturen	41	4	3.98
	8	Zünglis	Ding	44	4	4.1

4.4 Textebene Subtest II (Maze)

Der zweite Subtest, der die Leseverständnisfähigkeiten der Kinder auf Textebene erfasst (Subtest Textebene II), bedient sich der im englischsprachigen Raum sehr etablierten Maze-Methode (Conoyer et al., 2017; Muijselaar, Kendeou, Jong, & van den Broek, 2017; Wayman, Wallace, Wiley, Tichá, & Espin, 2007). Er besteht aus zwei kurzen Sachtexten, die aus dem für die Altersstufe der dritten Klasse vorgesehenem differenzierten Lesematerial aus dem Projekt DiLu (Differenzierter Leseunterricht; abrufbar unter <https://differenzierter-leseunterricht.uni-graz.at/de/>) entnommen und adaptiert wurden. Bei den für den GraLeV aufbereiteten Texten wurde jedes siebente Wort durch eine Lücke, die im digitalen GraLeV durch eine Drop-down-Auswahl umgesetzt wurde, ersetzt. Die Schülerinnen und Schüler haben die Aufgabe, aus drei Wörtern jenes Wort auszuwählen, das am besten in den Text passt (Seifert & Palczek, 2020, in Druck). Die zwei Distraktoren waren dem Zielwort entweder graphematisch bzw. phonologisch ähnlich oder ähnelten dem Zielwort semantisch-syntaktisch. Zudem wurde darauf geachtet, dass die Auswahl der Distraktoren ebenfalls einen grammatikalisch korrekten Satz bilden würden (in Anlehnung an Ketterlin-Geller, McCoy, Twyman, & Tindal, 2006), um die Schwierigkeit dieses Subtests zu erhöhen. Jeder der beiden Texte beinhaltet 15 Drop-Down-Entscheidungen und ist circa hundert Wörter lang. In Tabelle 2 sind weitere Charakteristika der beiden Texte dargestellt. Um die Lesbarkeit zu bestimmen, wurde einerseits wieder der RIX (Wild & Pissarek, 2020) und andererseits

der gSmog (Simple measure of Gobbledygook - German; Bamberger & Vanecek, 1984), bei dem die Anzahl der mehrsilbigen Worte in Bezug zur Anzahl der Sätze gesetzt wird, verwendet. Die Schülerinnen und Schüler haben 100 Sekunden Zeit, um die Aufgabe (Bearbeitung der zwei Texte mit 30 Lücken) zu lösen. Abbildung 5 zeigt exemplarisch einen Text mit den Drop-Down-Aufgaben.

5 Durchführung

Für die Durchführung des GraLeV benötigen Kinder etwa 15 bis 25 Minuten, abhängig von der Dauer der Instruktion (reine Testzeit ohne Instruktion: 10 Minuten 40 Sekunden). Bei der Durchführung des GraLeV ist es äußerst wichtig, dass auf jeder Ebene abgesichert wird, dass alle Kinder die Instruktion verstanden haben. Dies soll durch die gemeinsame Bearbeitung der Beispiele bei jedem Subtest gewährleistet werden. Hierbei ist es relevant, dass die Testleitung die Aufgabenstellung genau erklärt, denn sonst würde nicht das Leseverständnis der Kinder überprüft werden, sondern andere Fähigkeiten, wie beispielsweise die digitale Kompetenz der Kinder, die sich unter anderem daraus ergibt, ob sie mit Tablets und Spielen, die ähnliche Tätigkeiten (z. B. Drag and Drop oder Drop-Down-Menüs) verlangen, vertraut sind.

Bei der Instruktion der einzelnen Subtests gibt es keine Zeitbegrenzung, so dass die Testleitung in Ruhe sicherstellen kann, dass die Kinder verstanden haben, was zu tun ist und die Beispielitems gemeinsam bearbeitet und verglichen werden können. Unterstützend können Bilder

Tabelle 2: Textcharakteristika im Subtest Textebene II (Maze).

Text	Titel	Anzahl der Wörter	Anzahl der Sätze	gSmog	RIX
1	Was sind Nutztiere?	107	15	3.83	3.87
2	Wie entsteht Tomatensoße?	106	13	5.29	5.29

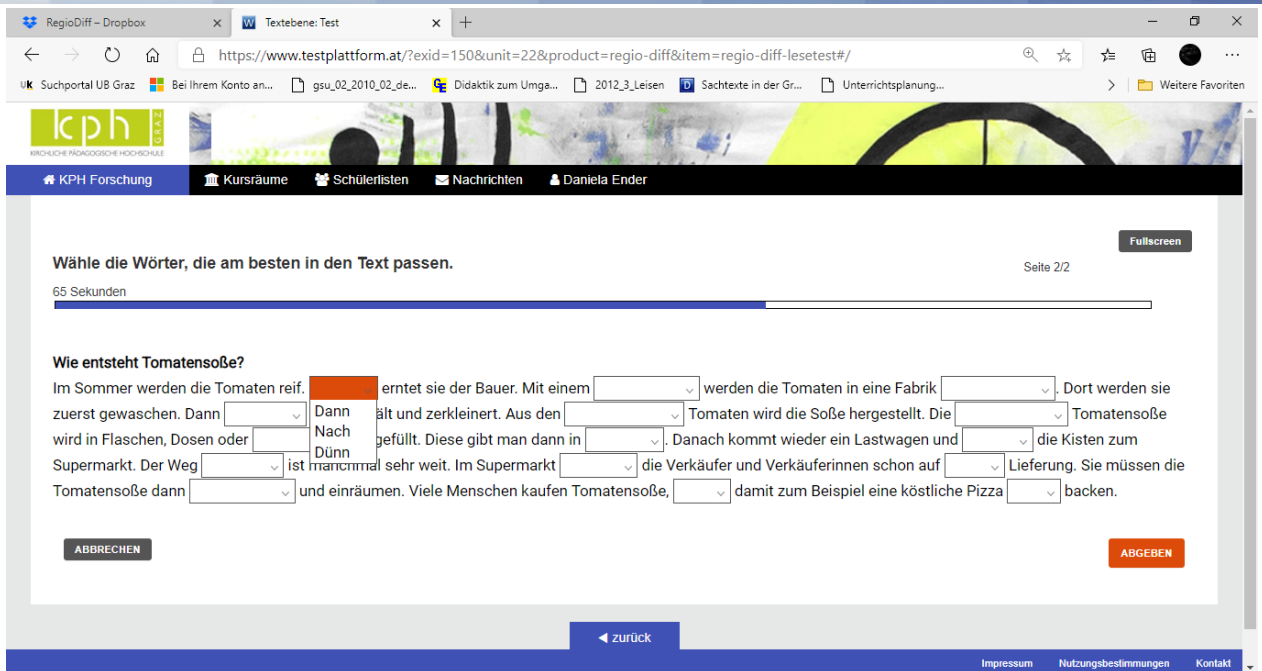


Abbildung 5: Maze-Prozedur im GraLeV Textverständnis Subtest II.

für die Tafel verwendet werden, die die Bilder der Beispielitems und die Lösungen sowie Distraktoren darstellen. Mit den Bildern könnten auch Besonderheiten des digitalen Auswählens einer Antwort, zum Beispiel der Drag-and-Drop-Vorgang beim Subtest Wortebene, gut im Klassenverband vorgestellt und besprochen werden. Die Kinder werden aufgefordert, bei jeder Aufgabe eine Lösung anzugeben und auch wenn sie nicht sicher wissen, was die Lösung ist, sich für die Antwort zu entscheiden, die ihnen am plausibelsten vorkommt. Wenn alle Kinder die Aufgabenstellung verstanden haben, kann der eigentliche Test begonnen werden. Die Zeit des jeweiligen Subtests läuft automatisch, sobald das Kind den Test beginnt, es muss also keine Zeit mitgestoppt werden (im Gegensatz zu einer Testung bei einem vergleichbaren Speedtest im Printformat). Nun bearbeiten die Kinder selbstständig so viele Items, wie sie in der vorgegebenen Zeit des Subtests schaffen. Einige wenige Kinder schaffen eventuell bereits vor Ablauf der Zeit alle Items eines Subtests zu bearbeiten und gelangen somit auf den nächsten Screen, wo aber noch keine Zeit abläuft. Im Klassensetting sind alle Kinder so fast gleichzeitig mit den einzelnen Subtests fertig und es kann nach jedem Subtest gemeinsam mit der Instruktion des darauffolgenden Subtests begonnen werden.

6 Auswertung

Die Auswertung des GraLeV erfolgt digitalisiert und automatisch. Es werden für das jeweilige Kind die Rohwerte (Anzahl der richtig gelösten Items) für jeden Subtest einzeln ausgegeben. Zusätzlich werden bei Falschantworten die Rohwerte bestimmter gewählter Distraktoren erfasst. Somit können Tendenzen der Falschlösungen

(z. B. immer syntaktische Distraktoren beim Subtest Satzebene) festgestellt werden. Das kann eine wichtige Basis für darauf basierende Förderschritte darstellen, um Kinder in weiterer Folge möglichst maßgeschneidert unterstützen zu können.

7 Interpretation

Die Rohwerte, die Kinder in den einzelnen Subtests erreichen, werden zukünftig automatisch in Prozentränge umgewandelt und als Ergebnis dargestellt. Die Interpretation der Testergebnisse kann somit anhand einer Normstichprobe vorgenommen werden und die Leistung der überprüften Kinder anhand von Leistungen anderer Kinder derselben Klassenstufe eingeordnet werden. Es werden die Prozentränge pro Subtest (Subtest Wortebene, Subtest Satzebene, Subtest Textebene I-Quatschgeschichten und Subtest Textebene II-Maze) sowie der Gesamtprozentrang rückgemeldet, wobei eine Rückmeldung des Gesamtprozentrangs nur erfolgen kann, wenn zu allen vier Subtests Ergebnisse vorliegen. So kann, wenn das Kind beispielsweise bei der Bearbeitung einer der Subtests technische Probleme hatte und dieser eine Subtest nicht ausgewiesen werden konnte, auch kein Gesamtprozentrang vorliegen. Ein Prozentrang gibt an, wie viele Kinder schlechtere oder genau gleiche Leistungen innerhalb einer vergleichbaren Normstichprobe erzielten. Ein Prozentrang von 75 heißt beispielsweise, dass 75% der Schülerinnen und Schülern der Normstichprobe in derselben Klassenstufe über eine niedrigere oder gleich hohe Lesekompetenz verfügen. 25% zeigen bessere Leseleistungen.

8 Limitationen und weitere Schritte

Zum Zeitpunkt der Publikation des vorliegenden Beitrags lag der GraLeV als digitale Version auf einer Lernplattform vor. Die beschriebenen Pilotierungsstudien fanden in dieser Version statt, jedoch gab es wiederholt Schwierigkeiten in den Klassentestungen. Diese Schwierigkeiten resultierten vor allem daraus, dass für die Bearbeitung des GraLeV auf der Lernplattform eine stabile und schnelle Internetverbindung vorhanden sein muss, um den Test durchzuführen. An österreichischen Grundschulen ist das allerdings eher selten der Fall. Das Laden der Bilder dauerte in manchen Situationen ein paar Sekunden. Nachdem der GraLeV als Speedtest eingesetzt wird, kann es durch diese Verzögerungen zu Verfälschungen in den Ergebnissen kommen und es wird nicht mehr die Leseleistung des Kindes, sondern die Leistung des Internets gemessen.

Aus diesem Grund wurde nach Möglichkeiten gesucht, den GraLeV in der Durchführung vom Internet zu entkoppeln. Dies gelang uns durch zwei Darbietungsformen. Einerseits wurde eine Printversion des GraLeV entwickelt und andererseits konnte durch eine Kooperation mit der Pädagogischen Hochschule Oberösterreich eine App entwickelt werden.

Die kürzlich entwickelte Printversion hat zum Ziel, möglichst vergleichbare Ergebnisse wie die digitale Form des GraLeV zu erzielen. Hierbei wurde versucht eine möglichst ähnliche Darstellungsform zu wählen, wobei das nur bedingt möglich war (bspw. Drop-Down). Eine Vergleichsstudie zwischen den beiden Testformaten wird derzeit mit ca. 300 Drittklässlerinnen und Drittklässlern durchgeführt (Durchführung der digitalen Version und Printversion des GraLeV im Abstand von ungefähr zwei Wochen, wobei die Hälfte mit der Printversion beginnt und die andere Hälfte zuerst die digitale Testung durchführt, um Lerneffekte zu vermeiden: Seifert & Palczek, eingereicht).

Die App-Version des GraLeV wurde ebenso vor kurzem fertiggestellt. Das Verbinden mit dem Internet ist in der App-Version nur noch notwendig, wenn die Testergebnisse ins System hochgeladen werden, damit anschließend die Auswertung erfolgen kann. Für die Zukunft, wenn Normierungsdaten vorliegen, ist angestrebt, dass die Ergebnisse als Prozentränge innerhalb der App generiert werden können. Bereits im derzeitigen Schuljahr 20/21 wird der GraLeV als App-Version mit über 800 Schülerinnen und Schülern der dritten und vierten Schulstufe in der Steiermark durchgeführt, um erste Normierungsdaten zu erhalten. Eine weitere Normierung in Österreich und Deutschland ist für das Schuljahr 21/22 geplant.

Zukünftig soll der GraLeV Lehrpersonen als Printform und als App kostenfrei zur Verfügung gestellt werden

und die Prozentränge für die jeweilige Klassenstufe abrufbar sein. Dafür sind derzeit jedoch noch einige Studien notwendig.

Es bestehen auch Überlegungen, die Altersspanne für den GraLeV auszuweiten. Einerseits nehmen wir an, dass der GraLeV auch in der fünften Schulstufe eingesetzt werden kann. Eine Zeitanpassung könnte dabei jedoch nötig sein, da die älteren Kinder sicherlich schneller lesen und es somit zu Deckeneffekten kommen könnte. Erste Pilotierungsversuche in einzelnen Klassen werden vielleicht bereits in einigen Monaten durchgeführt. Andererseits nehmen wir an, dass der GraLeV auch bereits am Ende der zweiten Schulstufe ein Messinstrument sein kann, das differenzierte Informationen über die Leseverständniskapazitäten der Kinder liefert. Allerdings müssen auch hierbei zuerst Pilotierungsstudien durchgeführt werden, um zu ermitteln, ob auch die Subtests der Textebene für die Altersstufe geeignet sind (auf unserer Homepage können die aktuellen Entwicklungen nachgelesen werden: <https://regionen-kennenlernen.uni-graz.at/de/test-gralev/der-grazer-leseverstaendnistest-gralev/>).

Literatur

Balula, A. (2016). The use of DigComp in teaching and learning strategies. In Association for Computing Machinery (Chair), *DSAI '2016 : proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*: December 1-3, 2016.

Bamberger, R., & Vanecek, E. (1984). *Lesen-Verstehen-Lernen-Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache*. Wien: Jugend und Volk.

Blumenthal, S., & Blumenthal, Y. (2020). Tablet or Paper and Pen? Examining Mode Effects on German Elementary School Students' Computation Skills with Curriculum-Based Measurements. *International Journal of Educational Methodology*, 6(4), 669–680. <https://doi.org/10.12973/ijem.6.4.669>

Brandhofer, G. (2015). *Die Kompetenzen der Lehrenden an Schulen im Umgang mit digitalen Medien und die Wechselwirkungen zwischen Lehrtheorien und mediendidaktischem Handeln* (unveröffentlichte Dissertation), Technische Universität Dresden.

Bundesministerium Bildung, Wissenschaft und Forschung (2020). *Digitale Schule*. Retrieved from <https://www.bmbwf.gv.at/Themen/schule/zrp/dibi.htm>

1

- Bundesministerium für Bildung und Forschung (2019). *DigitalPakt Schule*. Retrieved from <https://www.bmbf.de/de/wissenswertes-zum-digitalpakt-schule-6496.php>
- Cheung, A., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, 7, 198–215.
- Conoyer, S. J., Lembke, E. S., Hosp, J. L., Espin, C. A., Hosp, M. K., & Poch, A. L. (2017). Getting More From Your Maze: Examining Differences in Distractors. *Reading & Writing Quarterly*, 33(2), 141–154. <https://doi.org/10.1080/10573569.2016.1142913>
- Förster, N., & Souvignier, E. (2014). Learning progress assessment and goal setting: Effects on reading achievement, reading motivation and reading self-concept. *Learning and Instruction*, 32, 91–100. <https://doi.org/10.1016/j.learninstruc.2014.02.002>
- Gebhardt, M., Diehl, K., & Mühling, A. (2016). Online Lernverlaufsmessung für alle SchülerInnen in inklusiven Klassen. *Zeitschrift Für Heilpädagogik*, 66, 444–453.
- Griffin, Z. M., & Bock, K. (1998). Constraint, Word Frequency, and the Relationship between Lexical Processing Levels in Spoken Word Production. *Journal of Memory and Language*, 38, 331–338.
- Hamhuis, E., Glas, C., & Meelissen, M. (2020). Tablet assessment in primary education: Are there performance differences between TIMSS' paper-and-pencil test and tablet test among Dutch grade-four students? *British Journal of Educational Technology*, 51(6), 2340–2358. <https://doi.org/10.1111/bjet.12914>
- Jerrim, J., Micklewright, J., Heine, J.-H., Salzer, C., & McKeown, C. (2018). PISA 2015: how big is the 'mode effect' and what has been done about it? *Oxford Review of Education*, 44(4), 476–493. <https://doi.org/10.1080/03054985.2018.1430025>
- Ketterlin-Geller, L. R., McCoy, J. D., Twyman, T., & Tindal, G. (2006). Using a Concept Maze to Assess Student Understanding of Secondary-Level Content. *Assessment for Effective Intervention*, 31(2), 39–50. <https://doi.org/10.1177/073724770603100204>
- Lehmann, R. H., Peek, R., & Poerschke, J. (2006). *HAMLET 3-4: Hamburger Lesetest für 3. und 4. Klassen* (2., überarbeitete Auflage). Göttingen: Hogrefe.
- Lenhard, W., Schroeders, U., & Lenhard, A. (2017). Equivalence of Screen Versus Print Reading Comprehension Depends on Task Complexity and Proficiency. *Discourse Processes*, 54(5-6), 427–445. <https://doi.org/10.1080/0163853X.2017.1319653>
- Lenhard, W., Lenhard, A., & Schneider, W. (2020). *ELFE II: Ein Leseverständnistest für Erst- bis Siebtklässler – Version II* (4., unveränderte Auflage). Göttingen: Hogrefe.
- Mayringer, H., & Wimmer, H. (2014). *SLS 2-9: Salzburger Lese-Screening für die Schulstufen 2-9*. Göttingen: Hogrefe.
- Medienpädagogischer Forschungsverbund Südwest (2018). *KIM-Studie 2018. Basisstudie zum Medienumgang 6- bis 13-Jähriger in Deutschland*. Retrieved from https://www.mpfs.de/fileadmin/files/Studien/KIM/2018/KIM-Studie_2018_web.pdf
- Muijselaar, M. M. L., Kendeou, P., Jong, P. F. de, & van den Broek, P. W. (2017). What Does the CBM-Maze Test Measure? *Scientific Studies of Reading*, 21(2), 120–132. <https://doi.org/10.1080/10888438.2016.1263994>
- Mullis, I. V., & Martin, M. O. (Eds.) (2016). *PIRLS 2016 Assessment Framework* (2nd ed.). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Palczek, L., & Seifert, S. (in Vorbereitung b). *Differences above and beyond digital presentation: test mode effects in a German Reading Comprehension Test*.
- Palczek, L., & Seifert, S. (2020). Pädagogische Diagnostik und ihre Bedeutung für inklusiven Leseunterricht. In L. Palczek & S. Seifert (Eds.), *Inklusiver Leseunterricht: Leseentwicklung, Diagnostik und Konzepte* (1st ed., pp. 125–147). Wiesbaden: Springer Fachmedien Wiesbaden.
- Palczek, L., Seifert, S., Franz, A., Riedl, S., & Wohllhart, D. (in Vorbereitung). *Grazer Leseverständnistest: GraLeV*.
- Palczek, L., Seifert, S., & Gasteiger-Klicpera, B. (2017). Influences on teachers' judgment accuracy of reading abilities of second and third grade students: A Multilevel Analysis. *Psychology in the Schools*, 54(3), 228–245. <https://doi.org/10.1002/pits.21993>
- Palczek, L., Seifert, S., & Schöfl, M. (2021). Comparing digital to print assessment of receptive vocabulary with GraWo-KiGa in Austrian kindergartens. *British Journal of Educational Technology*. DOI: 10.1111/bjet.13163.

- Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The Acquisition of Reading Comprehension Skill. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 227–247). Oxford: Blackwell.
- Richter, T., Naumann, J., Isberner, M., Neeb, Y., & Knoepke, J. (2017). *ProDi-L: Prozessbezogene Diagnostik von Lesefähigkeiten im Grundschulalter*. Göttingen: Hogrefe.
- Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015). Childlex: A lexical database of German read by children. *Behavior Research Methods*, 47(4), 1085–1094. <https://doi.org/10.3758/s13428-014-0528-1>
- Seifert, S., & Palczek, L. (angenommen). Digitally Assessing Text Comprehension in Grades 3-4: Test Development and Validation. *Electronic Journal of e-Learning*.
- Seifer, S. & Palczek, L. (in Druck). Comparing Digital and Print Mode of a German Reading Comprehension Test in Grade 3: Influence of test order, gender and language.
- Seifert, S., & Palczek, L. (2020). Development of a German Digital Assessment of Reading Comprehension in Grades 3-4. In C. Busch, M. Steinicke, & T. Wendler (Eds.), *Proceedings of the 19th European Conference on e-Learning: A Virtual Conference hosted by University of Applied Sciences HTW Berlin*, Germany, Online.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>
- Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2007). A Meta-Analysis of Testing Mode Effects in Grade K-12 Mathematics Tests. *Educational and Psychological Measurement*, 67(2), 219–238. <https://doi.org/10.1177/0013164406288166>
- Wanzek, J., & Vaughn, S. (2007). Research-Based Implications From Extensive Early Reading Interventions. *School Psychology Review*, 36(4), 541–561. <https://doi.org/10.1080/02796015.2007.12087917>
- Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature Synthesis on Curriculum-Based Measurement in Reading. *The Journal of Special Education*, 41(2), 85–120. <https://doi.org/10.1177/00224669070410020401>
- Widauer, K., & Wallner-Paschon, C. (2017). Entwicklung und Aufbau der Testinstrumente und Kontextfragebögen. In C. Wallner-Paschon & U. Itzlinger-Bruneforth (Eds.), *PIRLS 2016. Technischer Bericht* (pp. 9–22). Salzburg.
- Wild, J., & Pissarek, M. (k.A.). *Ratte. Regensburger Analysetool für Texte*. Retrieved from <http://www.uni-regensburg.de/sprache-literatur-kultur/germanistik-did/ratte/index.html>
- Wild, J., & Pissarek, M. (2020). *RATTE Regensburger Analysetool für Texte: Dokumentation*. Retrieved from https://www.uni-regensburg.de/assets/sprache-literatur-kultur/germanistik-did/ratte_dokumentation_2_.pdf