



Stand 18.05.2020

Gerolf Renner & Markus Scholz

Zitationsempfehlung:

Renner, G. & Scholz, M., (2020). *Testinformation ELFE II. Ein Leseverständnistest für Erst- bis Siebtklässler – Version II* (Dia-Inform Verfahrensinformationen 006-01). Ludwigsburg: Pädagogische Hochschule Ludwigsburg.

**Dieser Text wird veröffentlicht unter der Creative-Commons Lizenz
CC BY-NC-ND 4.0 DE**

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

Diese Publikation darf dementsprechend nur unter Nennung der Urheber und ausschließlich zu nicht-kommerziellen Zwecken genutzt und weiterverbreitet werden, Modifikationen des Textes sind nicht zugelassen.

Potenzieller Interessenkonflikt:

Dia-Inform will unabhängige und nicht von kommerziellen Interessen beeinflusste Informationen zur Verfügung stellen. Wir legen deshalb Wert auf eine sorgfältige Dokumentation aller Tatsachen, die auf einen potenziellen Interessenkonflikt hinweisen könnten.

Es liegen keine Interessenskonflikte vor.

Das Projekt Dia-Inform

Projektleitung und Korrespondenz:

Prof. Dr. Gerolf Renner
Pädagogische Hochschule Ludwigsburg
Fakultät für Sonderpädagogik
Reuteallee 46
71634 Ludwigsburg
renner@ph-ludwigsburg.de

Prof. Dr. Markus Scholz
Pädagogische Hochschule Ludwigsburg
Fakultät für Sonderpädagogik
Reuteallee 46
71634 Ludwigsburg
markus.scholz@ph-ludwigsburg.de

Projektinformation:

Ziel des Projektes Dia-Inform an der Pädagogischen Hochschule Ludwigsburg ist es, ein umfangreiches Angebot an Informationsmaterialien zu diagnostischen Verfahren bereit zu stellen, die in der (sonder-)pädagogischen Diagnostik eingesetzt werden.

Informationen zu diagnostischen Verfahren sowie Informations-, Lehr- und Praxisvideos bilden ein abgestuftes System, das für unterschiedliche Informationsbedürfnisse von Lehrenden und Studierenden und verschiedene Einsatzzwecke (Veranstaltungen, Selbststudium, Vorbereitung auf diagnostische Gutachten) flexibel und passgenau verwendet werden kann und in dessen Weiterentwicklung Studierende aktiv mit eingebunden werden können.

Im Rahmen des Projektes soll ein umfassender, nicht von kommerziellen Interessen beeinflusster und anderweitig nicht verfügbarer Informationspool zu diagnostischen Verfahren entstehen.

Dia-Inform Verfahrensinformationen:

Dia-Inform Verfahrensinformationen geben einen Überblick über theoretische Grundlagen, Aufbau und psychometrische Eigenschaften eines diagnostischen Instrumentes. Sie weisen eine einheitliche Struktur auf, die Leser/innen die Orientierung erleichtert. Die Verfahrensinformationen diskutieren Stärken und Schwächen der Verfahren und weisen auf sinnvolle Einsatzmöglichkeiten in der sonderpädagogischen und klinisch-psychologischen Diagnostik hin. Dabei werden insbesondere die spezifischen Anforderungen einer Diagnostik bei Kindern und Jugendlichen mit Behinderungen berücksichtigt.

Dia-Inform Verfahrensinformationen werden von mindestens zwei Autor/innen gemeinsam verfasst, die unterschiedliche professionelle Perspektiven einbringen.

Dia-Inform Verfahrensinformationen stehen Ihnen kostenfrei auf OPUS-PHLB, dem Hochschulschriftenserver der PH Ludwigsburg, unter folgendem Link zur Verfügung:

<https://phbl-opus.phlb.de/solrsearch/index/search/searchtype/collection/id/16235>

ELFE II

Ein Leseverständnistest für Erst- bis Siebtklässler – Version II

Vorbemerkung

Diese Testinformation bezieht sich hinsichtlich der computergestützten Form des ELFE II ausschließlich auf die Angaben im Manual. Eine praktische Erprobung fand nur mit der Paper-Pencil-Version statt.

Autor*innen

Wolfgang Lenhard, Alexandra Lenhard & Wolfgang Schneider

Erscheinungsjahr

2018 (3., korrigierte Auflage)

Anwendungsbereich

Ende der ersten (ab 9. Schulmonat) bis Anfang der siebten Klasse (bis 3. Schulmonat)

Verlag

Hogrefe, Göttingen

Preis des Testverfahrens und der Verbrauchsmaterialien laut Verlagsseite

Test komplett (Paper-Pencil-Version; inklusive Verbrauchsmaterial für 20 Anwendungen): € 145,00; Testhefte: € 45,00 (20 Stück); Auswertungsbogen: € 11,80 (20 Stück).

Computergestütztes Testverfahren: € 200,00. Es stehen weitere Lizenzmodelle zur Verfügung (s. Internetseite des Verlags).

Alle Angaben zuzüglich Mehrwertsteuer.

Diagnostische Zielsetzung lt. Testmanual

Der ELFE II wurde zur „Erfassung der Leseflüssigkeit, der Lesegenauigkeit und des Leseverständnisses auf Wort-, Satz- und Textebene“ (Manual, S. 13) entwickelt. Dabei sollen sowohl „basale als auch höhere Leseprozesse“ (Manual, S. 16) berücksichtigt werden, wobei unter „basal“ hier das Wortverständnis zu verstehen ist, während „höhere Leseprozesse“ die Verarbeitung von ganzen Sätzen und Texten bezeichnen. Die Ergebnisse des ELFE II sollen zur Diagnose von Lese-Rechtschreibstörungen und der „Begleitung“ (Manual, S. 17) von entsprechenden Therapiemaßnahmen dienen.

Theoretischer Hintergrund

Die Autor*innen des Verfahrens verstehen Lesen als aktiven Prozess der Entschlüsselung (Dekodierung) von schriftlichen Informationen. Dabei unterscheiden sie zwischen „hierarchieniederen“ und „hierarchiehohen“ Prozessen. Bei hierarchieniederen Aspekten geht es um die Dekodierung einzelner Worte und der Syntax eines Satzes, bei den hierarchiehohen darum, die erschlossenen Bedeutungen „in eine übergeordnete Makrostruktur oder eine Rahmenhandlung“ (Manual, S. 19) einzuordnen.

Im Manual wird auf aktuelle Modelle der Leseforschung, wie das „Dual-Route Cascaded Model“ des Lesens nach Coltheart, Rastle, Perry, Langdon und Ziegler (2001) auf

Wortebene (Manual, S. 20) und das „Model of Discourse Comprehension“ bzw. das „Construction-Integration Model“ auf Textebene (Manual, S. 23) Bezug genommen.

Beim Satzverständnis betonen die Autoren die Bedeutung der Tiefenstruktur des Satzes als Voraussetzung für das Verständnis und gehen auf die Hypothesenbildung im Leseprozess ein. Im Gegensatz zum Wort- und Textverständnis erfolgt in den Beschreibungen kein Bezug auf eine explizit benannte Theorie.

Auf Aufgabenebene wird daher zwischen Wortverständnis, Satzverständnis und Textverständnis unterschieden.

Setting, Durchführungsdauer und Auswertungsmöglichkeiten

Der ELFE II kann als Einzel- oder als Gruppentest eingesetzt werden. Für die Einzeltestung stehen eine Paper-Pencil-Version und ein Testprogramm zur Verfügung. In der Paper-Pencil-Version finden sich alle Items in einem Testheft, in dem auch die richtigen Lösungen markiert werden. Die Auswertung erfolgt entweder manuell anhand der Normtabellen des Manuals oder durch Eingabe der Rohwerte in das Computerprogramm. In der Computerversion erfolgt die Auswertung automatisch. Die Gruppentestung ist nur mit der Paper-Pencil-Version möglich.

Die Durchführungszeit für die Standardversion wird mit ca. 20 bis 30 Minuten angegeben, davon sind 13 Minuten die reine Bearbeitungszeit. Die Kurzversionen benötigen laut Manual ca. 15 bis 25 Minuten, davon 6 bis 9 Minuten reine Bearbeitungszeit. Wird in der Computerversion die Schwellenmessung der Worterkennung durchgeführt, sind weitere 3 bis 4 Minuten zu veranschlagen.

Informationen für Anwender/innen früherer Auflagen

Die wichtigsten Änderungen gegenüber dem ELFE 1-6 auf einen Blick:

- Überarbeitung aller Untertests (mehr leichte und schwierige Items; Wegfall von Items, die sich auf Geräte beziehen, die einem schnellen technischen Wandel unterliegen)
- Weitgehende Überarbeitung der *Schwellenmessung der Worterkennung* in der Computerversion
- Einsatz jetzt während des gesamten Schuljahres möglich (nicht nur in der Mitte und am Ende des Schuljahres)
- Erweiterung des Einsatzbereiches bis Anfang der 7. Klasse
- Durchführung von Kurzversionen möglich
- Testheft neu gestaltet
- Wegfall der Pseudoparalleltestvariante für Gruppentestungen
- Normwerte der Untertests jetzt als T-Werte und nicht mehr als z-Werte skaliert
- Erweiterte Auswertungsmöglichkeiten: Diskrepanzvergleiche zwischen Untertests (signifikante Differenzen, Grundraten), Analyse des Arbeitsstils (Tempo vs. Genauigkeit)

Testaufbau

Die Standardversion des ELFE II umfasst drei Untertests:

- Die Testung beginnt mit dem *Wortverständnistest* (75 Items; Bearbeitungszeit 3 Minuten). Es soll jeweils erkannt werden, welches der vier zur Auswahl stehenden Wörter am besten zu einer Abbildung passt. Bei den Antwortalternativen sind graphemisch und phonemisch ähnliche Wörter zu finden, die die gleiche Anzahl an

Silben wie die richtige Antwort haben. Die richtige Antwort wird durch Unterstreichen markiert. Die Items (vgl. Abb. 1) sind auf den Seiten des Testhefts in 7 Reihen (auf der letzten Seite in 4 Reihen) mit jeweils 3 Aufgaben angeordnet.

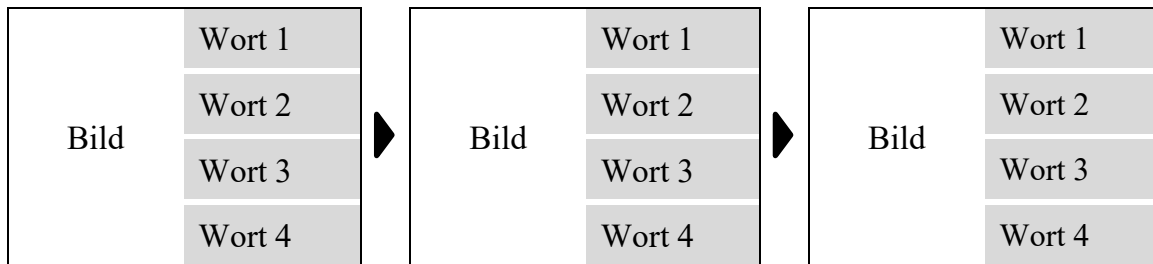


Abbildung 1. Aufbau des Testhefts beim Wortverständnistest

- Im Satzverständnistest (36 Items; Bearbeitungszeit 3 Minuten) werden unvollständige Sätze vorgegeben, in denen ein fehlendes Wort ergänzt werden soll. Es stehen jeweils fünf Antwortalternativen zur Verfügung, die richtige Antwort soll unterstrichen werden (vgl. Abb. 2). Auf den Seiten befinden sich jeweils 5 bis 7 Sätze.

Eine Katze hat zwei	Augen	.
	Füße	
	Schwänze	
	Haare	
	Krallen	

Abbildung 2. Aufbau des Testhefts beim Satzverständnistest (kein Original-Item)

- Beim Textverständnistest (26 Items; Bearbeitungszeit 7 Minuten) müssen auf Basis von Sach- oder Erzähltexten Satzanfänge logisch ergänzt oder Fragen beantwortet werden. Dabei werden jeweils vier Antwortmöglichkeiten angeboten, die richtige soll angekreuzt werden (Abb. 3). Die korrekte Antwort kann laut Manual nur gefunden werden, wenn Informationen aus mehreren Textstellen berücksichtigt werden.

Karl kann heute nicht in die Schule. Er hat Fieber und liegt zuhause im Bett.			
Karl...			
<input type="radio"/>	... macht seine Hausaufgaben.	<input type="radio"/>	... ist krank.
<input type="radio"/>	... geht mit seinen Freunden spielen.	<input type="radio"/>	... malt gerne.

Abbildung 3. Aufbau des Testhefts beim Textverständnistest (kein Original-Item)

Eine Kurzform für erste bis dritte Klassen umfasst den Wortverständnistest und den Satzverständnistest und erlaubt die Ermittlung eines Gesamtwertes auch dann, wenn die Anforderungen des Textverständnistests für das getestete Kind zu schwierig sind. Eine Version mit verkürzter Darbietungszeit kann in den Klassenstufen 4 bis 7 bei leistungstärkeren Schüler*innen eingesetzt werden.

In der Computerform kann als weiterer Untertests die *Schwellenmessung der Worterkennung* durchgeführt werden (36 Items), bei der kurzfristig ein Wort am Bildschirm präsentiert wird. Das Wort soll anschließend einer von drei Kategorien (Pflanze, Gegenstand, Tier) zugeordnet werden. Es handelt sich um ein adaptives Verfahren. Der Test beginnt mit einer Darbietungszeit, bei der ca. 70% aller Personen einer Klassenstufe die ersten Items korrekt bearbeiten. Werden drei Items korrekt gelöst, wird die Darbietungszeit um 30% gekürzt, bei zwei richtigen Lösungen bleibt sie gleich, andernfalls verlängert sie sich um 30%. In den folgenden Dreiergruppen wird die Verkürzung oder Verlängerung der Darbietungszeit zunehmend feiner abgestuft und beträgt am Ende noch 10%. Wenn die Darbietungszeit in drei aufeinanderfolgenden Blöcken gleichbleibt, wird die Schwellenmessung beendet. Bei der *Schwellenmessung der Worterkennung* soll eine in den anderen Untertests mögliche Verzerrung der Ergebnisse durch ein langsames Tempo beim Markieren der Antworten (z. B. aufgrund feinmotorischer Beeinträchtigungen) verhindert werden. Eine Schwelle von 200 Millisekunden wird laut Manual als Indikator des automatisierten Wortlesens angesehen.

Ergebnis(-werte)

Zentrale Ergebniswerte:

- T-Werte ($M = 50$, $SD = 10$) mit 90%- und 95%-Konfidenzintervallen sowie Prozentränge für das Gesamtergebnis.
- T-Werte ($M = 50$, $SD = 10$) und Prozentränge für die Anzahl der richtig gelösten Aufgaben sowie für die Anzahl der bearbeiteten Aufgaben.

Daten zur Profilinterpretation

- Kritische Differenzen auf dem 10%- und dem 5%-Signifikanzniveau für paarweise Vergleiche der Untertests
- Grundraten für die Häufigkeit von Differenzen zwischen Untertests. Angegeben werden die Differenzwerte, die bei weniger als 5% und die bei 5% bis 10% der Normstichprobe auftraten.

Daten zur Analyse des Arbeitsstils:

- Für jeden Untertest werden die Grundraten für Differenzen zwischen den Normwerten der richtig gelösten Aufgaben und der Anzahl der bearbeiteten Aufgaben angegeben.

Die Ergebnisse können auch in einem grafischen Testprofil eingetragen werden. Über das Computerprogramm kann automatisch ein Ergebnisbericht für Eltern in Deutsch, Türkisch, Englisch, Arabisch oder Russisch generiert werden.

Objektivität

Detaillierte Hinweise zur Testdurchführung, wörtliche Instruktionen und allgemeine Hinweise zur Gestaltung der Testsituation sichern die *Durchführungsobjektivität*. Eine objektive Auswertung ist problemlos möglich, die Bewertung der Antworten erfordert keine subjektiven Entscheidungen der Testleiter*innen. Der Bezug auf Normwerte sichert die *Interpretationsobjektivität*.

Reliabilität

Als Reliabilitätsschätzungen werden angegeben die Split-half-Reliabilität, die Retest-Stabilität (bei einem Intervall von 4 Wochen) und die Paralleltest-Reliabilität (Übereinstimmung Papier- und Computerform bei einem Intervall von 3 Wochen).

Die Werte fallen für die Untertests befriedigend bis sehr gut und für den Gesamtwert gut bis sehr gut aus.

Eine im positiven Sinn bemerkenswerte Besonderheit ist, dass die Reliabilitätsberechnungen zusätzlich unter Berücksichtigung des Arbeitsstils der Kinder durchgeführt wurden. Die Ergebnisse zeigen deutlich, dass die Reliabilität bei allen Untertests - besonders ausgeprägt im *Textverständnistest* - bei einem ungenauen Arbeitsstil schlechter ausfällt als für die Gesamtstichprobe und eindeutig im nicht mehr akzeptablen Bereich liegt.

Tabelle 1

Reliabilitätskennwerte des ELFE II

Testteil	Split-half ^a	Retest	Paralleltest
Wortverständnistest:	.98 (.89)	.83	.78
Satzverständnistest:	.94 (.78)	.90	.90
Textverständnistest	.89 (.46)	.81	.88
Gesamtwert	.96	.93	.93

Anmerkung. ^a In Klammern sind die Reliabilitätswerte für die 10 % am ungenauesten arbeitenden Kindern angegeben.

Validität

Inhaltliche Validität: Die inhaltliche Validität als Test des Leseverständnisses ist aufgrund der genau auf diese Fähigkeit abzielenden Aufgabenformate gegeben.

Faktorielle Validität: Berichtet wird eine konfirmatorische Faktorenanalyse, die als Multigruppenanalyse auf Basis der Gesamtstichprobe gemeinsam für die Paper-Pencil-Version und die Computerform durchgeführt wurde. Untersucht wurde ein hierarchisches dreifaktorielles Modell, mit der latenten Variable Leseverständnis als Faktor zweiter Ordnung und drei latenten Variablen erster Ordnung, die die drei Untertests repräsentieren. Die Varianz ist nach diesen Ergebnissen ganz überwiegend durch den Faktor Leseverständnis zu erklären, die untertestspezifische Varianz ist dagegen sehr gering, was trotz der insgesamt guten Modellpassung durchaus als Argument für eine eindimensionale Interpretation der Testwerte gesehen werden könnte. Da laut Manual (S. 136) die Interkorrelationen der Untertests in den Klassenstufen unterschiedlich ausfallen, erscheint die Durchführung einer konfirmatorischen Faktorenanalyse nur auf Basis der Gesamtstichprobe ohnehin wenig aussagekräftig.

Konvergente Validität: Bei einer Stichprobe von 143 Grundschüler*innen wurde die Übereinstimmung mit dem Salzburger-Lese-Screening (SLS 2-9; Wimmer & Mayringer, 2014) überprüft. Das Gesamtergebnis des ELFE II zeigte signifikante Korrelationen mit dem SLS 2-9 von .92 in der zweiten, .75 in der dritten und .67 in der vierten Klasse. In zwei Studien mit 113 und 154 Schüler*innen, ganz überwiegend aus dem Primarbereich, fanden sich signifikante Korrelationen von .70 und .71 mit einer globalen Beurteilung der Lesefertigkeiten durch Lehrer*innen. Korrelationen mit weiteren globalen Beurteilungen durch Lehrkräfte bezüglich Rechtschreibung, Rechnen und allgemeine Sprachfertigkeiten fielen ebenfalls signifikant, aber durchgehend niedriger aus.

Divergente Validität: Als konstruktferne Verfahren, zu denen geringere Zusammenhänge bestehen sollten, wurden der Test d2 - Revision (Brickenkamp, Schmidt-Atzert &

Liepmann, 2010) und der Testteil 2 des Grundintelligenztest Skala 1 – Revision (CFT 1-R; Weiß & Osterland, 2013) jeweils bei 122 Schüler*innen (17 bis 37 je Klassenstufe) eingesetzt. In den Klassenstufen ergaben sich recht inkonsistente Ergebnisse: Die Korrelationen mit dem CFT 1-R reichten von .15 (erste Klasse) bis .60 (vierte Klasse), über alle Klassenstufen hinweg fand sich eine Korrelation von .39. Beim d2 lagen die Korrelationen zwischen .16 (erste Klasse) und .55 (zweite Klasse), für die Gesamtstichprobe fand sich ein Wert von .43.

Diskriminative Validität: Kinder mit einer Lese-Rechtschreibstörung lagen im ELFE II in allen Untertests und im Gesamtwert signifikant unter den Leistungen der Vergleichsgruppe. Mit einem mittlerem T-Wert von 41.8 fiel das Gesamtergebnis noch relativ günstig aus. Die Aussagekraft dieser Daten ist allerdings eingeschränkt, da die Zuordnung zu dieser Diagnosegruppe nur aufgrund von Aussagen der Lehrer*innen und Schüler*innen erfolgte. Unterschiede zwischen Schularten liegen in der erwarteten Richtung, z. B. reichen die mittleren T-Werte in der Sekundarstufe von 36.5 (Förderschule) bis 55.6 (Gymnasium).

Prognostische Validität: Zur prognostischen Validität des ELFE II liegen keine Daten vor.

Normierung

Die Normierungsdaten wurden im Jahr 2015 von speziell geschulten Studierenden (Psychologie, Lehramt) in 9 deutschen Bundesländern erhoben.

Die Ausgangsstichprobe umfasste 5.073 Schüler*innen. Für die Normierung wurde auf Basis von bevölkerungsstatistischen Daten eine nach Schulart und Geschlecht quotierte Stichprobe gebildet, die bei der Papierform 1.520 Schüler*innen und bei der Computerversion 1.287 Schüler*innen umfasst. Die Normen wurden dann auf Basis eines von den Autor*innen entwickelten kontinuierlichen Normierungsmodells berechnet.

Bei der Normierung wurden auch Schüler*innen aus Förderschulen berücksichtigt. Leider ist dem Manual nicht zu entnehmen, um welche sonderpädagogischen Förderschwerpunkte es sich handelte.

Von den Testautor*innen – nicht vom Testverlag – werden auch Zusatznormen für Kinder mit Migrationshintergrund vertrieben. Diese Normen sind Manual des ELFE II nicht enthalten und wurden daher auch in dieser Verfahrensinformation nicht berücksichtigt.

Weitere Gütekriterien

Aufgrund der kurzen Durchführungsdauer sowie der Option der Gruppentestung kann der ELFE II als ökonomisches Verfahren bezeichnet werden.

Zielgruppengerechte Gestaltung

Nach unserer Einschätzung ist der ELFE II so gestaltet, dass eine Testung bei der Mehrheit der Schüler*innen an Regelschulen unkompliziert verlaufen wird. Die Instruktionen sind klar und verständlich. Das Aufgabenformat bietet Schüler*innen, die aus dem Schulalltag selbstgesteuerte Arbeitsphasen und Leistungsüberprüfungen kennen, keine besonderen Herausforderungen. Allerdings fehlen für Schüler*innen mit gravierenden Beeinträchtigungen der Lesefähigkeiten sehr leichte Items. Hier sind Überforderungen zu erwarten.

Die Computerversion sollte laut Manual nur eingesetzt werden, wenn das getestete Kind über Computererfahrungen verfügt und mit einer Computermaus vertraut ist.

Zugangsfertigkeiten / Einsatz bei behinderten Kindern / Testfairness

Das Manual befasst sich nicht mit dem Einsatz des Tests bei Kindern mit Behinderungen. Testfairness wird nur im Hinblick auf Geschlecht und Migrationshintergrund thematisiert.

Eine wichtige Zugangsvoraussetzung des ELFE II ist die visuelle Wahrnehmung. Bei Verdacht auf Einschränkungen sollte dies unbedingt vor einer Testung abgeklärt werden. Die zum Teil sehr schlechten Kontraste bei den Abbildungen (z. B. Himmel, Sand, Raumschiff usw.) und einzelne ungünstige perspektivische Darstellungen (z. B. Maus) führen bei Kindern, die sich z. B. aufgrund einer Wahrnehmungsbeeinträchtigung stärker an Außenkonturen orientieren, vermutlich zu Problemen beim Erkennen der Abbildungen oder zu einem erhöhten Zeitbedarf.

Motorische Anforderungen beschränken sich auf das Unterstreichen der Lösungen und das Umblättern des Testhefts. Bei Schüler*innen mit deutlichen Beeinträchtigungen der Handmotorik kann sich dies durchaus auf Geschwindigkeit und Genauigkeit der Markierung auswirken und damit das Testergebnis negativ beeinflussen. Die Autor*innen geben den Hinweis, dass die *Schwellenmessung der Worterkennung* den Einfluss der Motorik auf die Leistungsgüte reduziert.

Die verbalen Instruktionen können sich für Schüler*innen mit Sprachentwicklungsstörungen und mangelnden Deutschkenntnissen an einigen Stellen als anspruchsvoll erweisen, z. B. wird man die sichere Kenntnis der Begriffe „links“ und „rechts“ in der ersten Klasse nicht durchgehend vorfinden. Allerdings ist das Sprachverständnis auch eine bedeutsame Komponente des Leseverständnisses, so dass es nicht als reine Zugangsfertigkeit bezeichnet werden kann.

Aufgrund der zeitbegrenzten Itemvorgabe könnten Schüler*innen mit einem langsamen Arbeitstempo in ihren Fähigkeiten, Worte und Sätze zu lesen, unterschätzt werden. Da die Testaufgaben im vorgegebenen Zeitrahmen eigenständig bearbeitet werden müssen, werden im Vergleich zu einer sukzessiven Itemvorgabe mit durchgehender Verhaltenssteuerung durch die Testleiter*innen höhere Anforderungen an Aufmerksamkeit und Selbststeuerung gestellt. Das Manual gibt zwar Hinweise, wann eine impulsive Bearbeitung vermutet werden kann, eine genaue Beobachtung des Testverhaltens ist aber stets zu empfehlen, um Fehlinterpretationen der quantitativen Testergebnisse zu vermeiden.

Zielgruppe des ELFE II sind Schüler*innen an Regelschulen. Normen für Förderschulen stehen dementsprechend nicht zur Verfügung.

CHC-theoretische Einordnung

In der Cattell-Horn-Carroll Intelligenztheorie (Schneider & McGrew, 2018; deutschsprachige Darstellung Mickley & Renner, 2019) ist das Leseverständnis ein Schicht-I-Faktor, der auf Schicht-II dem Faktor Lesen und Schreiben (Grw) zugeordnet wird. Aufgrund der zeitbegrenzten Vorgabe kann zudem eine Zuordnung zum Schicht-I-Faktor Lesegeschwindigkeit erfolgen („fluency of reading text with full comprehension“; Schneider & McGrew, 2018, S. 110), der sowohl Grw als auch der Verarbeitungsgeschwindigkeit (Gs) zugeordnet wird.

Stärken

- Plausible theoretische und empirische Grundlage
- Klarer und inhaltlich überzeugender Testaufbau
- Aktuelle Normierung

- Hohe Reliabilität
- Spezifische Daten zur Reliabilität bei Kindern mit unterschiedlichen Arbeitsstilen
- Weitgehend überzeugende Validitätsdaten
- Gruppentestung möglich
- Ökonomische Testdurchführung
- Anwendung während des ganzen Schuljahres möglich
- Erfassung des hochautomatisierten Lesens durch *Schwellenmessung der Worterkennung* möglich (nur in der Computerversion)
- Für Profilanalysen werden sowohl kritische Differenzen als auch Grundraten bereitgestellt. Auf die bei Profilanalysen zu beachteten statistischen Besonderheiten wird im Manual hingewiesen.
- Separate Normen für die Anzahl der richtig gelösten und die Anzahl der bearbeiteten Aufgaben. Diskrepanzen zwischen den entsprechenden Normwerten können auf Besonderheiten des Arbeitsstils hinweisen, die bei der Testinterpretation berücksichtigt werden können.

Einschränkungen / Schwächen

- Weniger bis nicht geeignet für die Testung von Schüler*innen mit erheblichen Beeinträchtigungen der Lesefähigkeiten (sichere Buchstabenkenntnis wird vorausgesetzt)
- Durch bloßes Raten können in den unteren Klassenstufen insbesondere beim *Textverständnistest* schon annähernd durchschnittliche Leistungen erzielt werden
- Schlechte Qualität einzelner Abbildungen (Kontraste, Perspektive)
- Weniger geeignet für den Einsatz bei Schüler*innen mit deutlichen motorischen Beeinträchtigungen
- Nicht geeignet für den Einsatz bei Schüler*innen mit Beeinträchtigungen der visuellen Wahrnehmung
- Weniger geeignet für Kinder mit deutlichen Beeinträchtigungen der Aufmerksamkeit und Selbststeuerung
- Bodeneffekte in den unteren Klassenstufen beim *Satzverständnistest* und beim *Textverständnistest*
- In sehr wenigen Fällen ungünstige Itemgradienten
- In sehr wenigen Fällen ungünstige Altersdifferenzierung
- Konfidenzintervalle werden in den Normtabellen nur für den Gesamtwert, nicht jedoch für die Untertests angegeben.
- Die – grundsätzlich sinnvolle und hilfreiche – Analyse des Arbeitsstils berücksichtigt nur die markierten Lösungen. Wenn deutliche Defizite in Aufmerksamkeit und Selbststeuerung dazu führen, dass Lösungen überhaupt nicht markiert werden, wird nach den Kriterien des Manuals möglicherweise dennoch ein genauer Arbeitsstil festgestellt.
- Bei der Analyse des Arbeitsstils wären Daten zu unterschiedlichen Fähigkeitsniveaus hilfreich. Es kann vermutet werden, dass bei Kindern mit weit unterdurchschnittlichen Lesefähigkeiten unabhängig vom Arbeitsstil größere Diskrepanzen zwischen bearbeiteten und richtigen Lösungen vorkommen.
- Wünschenswert wären Daten zur Frage, ob die Ergebnisse der ELFE II Ende der ersten / Anfang der zweiten Klasse eine Prognose von Leseschwierigkeiten in höheren Klassen erlauben.

- Das Testmanual könnte die Bedeutung von Umfeldfaktoren für den diagnostischen Prozess bei Lese-/Rechtschreibstörungen etwas deutlicher herausstellen.
- Der Einsatz von diagnostisch unerfahrenem Personal wird im Manual als zulässig bezeichnet, ohne dass nähere Vorgaben für die erforderliche Qualifizierung gemacht werden.
- Die Option, eine an Eltern gerichtete Ergebniszusammenerfassung automatisch in mehreren Sprachen erstellen zu lassen, hätte durchaus bei den Stärken des ELFE II eingeordnet werden können, wenn die Darstellung in einfacherer Sprache erfolgen würde. Menschen, die noch keine Erfahrungen mit Testdiagnostik hatten, könnten mit dem Text überfordert sein. Zudem wird es in vielen Fällen nicht sinnvoll sein, Eltern einen Bericht über Ergebnisse eines einzelnen Testverfahrens auszuhändigen, der alle weiteren diagnostisch relevanten Daten nicht berücksichtigt.

Sinnvolle Ergänzungen

Bei unterdurchschnittlichen Testleistungen bieten sich je nach Ausgangslage und Fragestellung folgende Ergänzungen an:

- Abklärung von Hör- und Sehfähigkeiten (vor der Testung mit dem ELFE II)
- Testing-the-limits mit den Aufgabenstellungen des ELFE II, um ergänzende qualitative Hinweise zu den Leseschwierigkeiten zu erhalten (lautes Lesen, Vorgabe ohne Zeitbegrenzung, Einzelvorgabe von Items, sukzessive Hilfestellungen, Befragung des Kindes zu seinen Lösungsstrategien u. a. m.). Testing-the-limits darf nur nach standardmäßiger Vorgabe erfolgen und nicht in die Berechnung der Normwerte eingehen. Wenn Wiederholungsmessungen zur Lernverlaufskontrolle geplant sind, sollte auf Testing-the-limits mit Aufgaben des ELFE II verzichtet werden.
- Lesetests, die lautes Lesen verlangen, um einen näheren Einblick in die Art der Lesefehler zu erhalten und um besser zwischen Schwierigkeiten beim Dekodieren und beim inhaltlichen Verständnis von Texten unterscheiden zu können (z. B. SLRT-II; Moll & Landerl, 2014).
- Lesetests, in denen einfachere Anfangsanforderungen gestellt werden (Phonem-Graphem-Zuordnung, Lesen kurzer und häufiger Wörter), bei weit unterdurchschnittlichen Testleistungen auch unter Berücksichtigung von Vorläuferfertigkeiten.
- Rechtschreibtests, wenn eine entsprechende Beeinträchtigung aufgrund der Vorinformationen nicht ausgeschlossen werden kann.
- Verfahren zur Überprüfung der kristallinen Intelligenz unter Berücksichtigung des lexikalischen Wissens (vgl. die im Manual auf S. 25 dargestellten Befunde zu Determinanten des Leseverständnisses), z. B. Index *Sprachverständnis* des WISC-V (Wechsler, 2017), SET 5-10 (Petermann, 2015), PPVT-4 (Dunn & Dunn, 2015)
- Verfahren zur Überprüfung der auditiven Verarbeitung bei Verdacht auf eine entsprechende Störung oder bei Vorliegen von Risikofaktoren, z. B. BAKO 1-4 (Stock, Marx & Schneider, 2003).
- Bei Verdacht auf kognitive Entwicklungsstörungen Durchführung etablierter Intelligenztestbatterien, die sprachliche Kompetenzen und Arbeitsgedächtnis

berücksichtigen, z. B. WISC-V, AID 3 (Kubinger & Holoher-Ertl, 2014), KABC-II (Melchers & Melchers, 2015).

- Detaillierte Exploration der Schullaufbahn unter Berücksichtigung von Fehlzeiten, (insbesondere länderübergreifenden) Schulwechseln, pädagogischen Rahmenbedingungen (z. B. Lehrerwechsel, Klassenklima, Mobbing), Leselernmethoden, Lehrer*innen-Schüler*innen-Beziehung usw. sowie Abklärung psychosozialer Risikofaktoren, die Lernen beeinträchtigen können.
- Exploration oder standardisierte Befragung zum leistungsbezogenen Selbstkonzept, zum Interesse an Schriftsprache, zur Leistungsmotivation

Anforderungen an die Testleiterinnen & Testleiter

Die Anforderungen an die praktische Durchführung des ELFE II sind aufgrund des klaren Testaufbaus und der eindeutigen Instruktionen relativ gering. Die Testautor*innen sind sogar der Meinung, dass die Computerversion auch von testdiagnostisch weniger geschultem Personal durchgeführt werden kann. Dies verträgt sich jedoch weniger mit der ebenfalls im Manual (S. 62) angesprochenen Notwendigkeit, das Verhalten des Kindes während der Testung kontinuierlich zu beobachten, um nötigenfalls Hilfestellungen zu geben zu können. Ob diagnostisch unerfahrenes Personal immer einschätzen kann, welche Hilfestellungen die Validität der Untersuchung gefährden können, darf bezweifelt werden.

Eine gründliche Vorbereitung ist dennoch erforderlich, mindestens eine Probetestung ist unbedingt zu empfehlen. Gruppentestungen sind als wesentlich anspruchsvoller einzuschätzen und sollten nur erfolgen, wenn man mit dem Testverfahren sehr gut vertraut ist und über Erfahrung in der Leitung von Gruppenuntersuchungen verfügt. Das Manual empfiehlt aus guten Gründen, Gruppentestungen mit zwei oder mehr Testleiter*innen vorzunehmen. Bei Gruppentestungen muss zudem gewährleistet sein, dass alle Störungen des Testablaufs (Hinweise auf mangelndes Instruktionsverständnis, Unaufmerksamkeit usw.) registriert werden können.

Der ELFE II sollte wie jeder psychometrische Test nur von Fachkräften eingesetzt werden, die über fundierte Kenntnisse in Testtheorie und Teststatistik verfügen. Testleiter*innen sollten auch über ein solides Grundwissen über den Schriftspracherwerb und seine Störungen verfügen, um die Bedeutung der Testergebnisse im Kontext weiterer diagnostischer Befunde bewerten zu können.

Im Manual wird leider die Behauptung aufgestellt, dass einige Handbuchkapitel in erster Linie „für Wissenschaftler“ (S. 6) von Bedeutung seien. Während sich manche Passagen tatsächlich nur Leser*innen mit sehr fundierten teststatistischen Kenntnissen erschließen werden, halten wir es jedoch für inakzeptabel, dass auch das Kapitel zu Gütekriterien als weniger wichtig für die praktische Anwendung gehalten wird. Eine verantwortungsvolle Testinterpretation ist ohne ein gründliches Verständnis der Gütekriterien nicht möglich.

Zusammenfassende Bewertung von Gerolf Renner

In der eigenen klinisch-sozialpädiatrischen Praxis hatte ich die Vorläuferversion, den ELFE 1-6, gern eingesetzt, wenn eine psychometrisch fundierte und nicht zu zeitaufwändige Erfassung der Leseleistungen bei Schüler*innen an Regelschulen erforderlich war. Der ELFE 1-6 ließ sich in meiner Testpraxis gut handhaben und wurde von den getesteten Kindern überwiegend gut akzeptiert. Bei Kinder mit gravierenden Beeinträchtigungen der Lesefähigkeiten war der ELFE 1-6 nicht das Verfahren der ersten

Wahl, da in diesem Fall Überforderungen zu erwarten waren und mir die Testung mit einem zeitbegrenzten Verfahren weniger sinnvoll erschien.

Im ELFE II wurde die grundlegende Teststruktur beibehalten. In psychometrischer Hinsicht haben sich einige Verbesserungen ergeben. Die Überarbeitung ist insgesamt gelungen, das Manual ist sehr informativ und liefert zahlreiche für die diagnostische Praxis relevante Daten.

Der ELFE II ist ein sinnvoll aufgebautes, sehr reliables und ökonomisches Verfahren und eignet sich besonders im Regelschulbereich gut zur schnellen und zuverlässigen Identifikation von Schüler*innen mit Beeinträchtigungen des Leseverständnisses.

Das Testergebnis allein gibt jedoch nur wenige Hinweise zu möglichen Ursachen der Problematik. Hier sind zur Abklärung weitere Untersuchungsschritte erforderlich. Nur sehr knapp wird im Manual darauf verwiesen, dass auch Faktoren wie die Unterrichtsqualität einen Einfluss auf das Leseverständnis haben können. Nun ist es nicht Aufgabe eines Testmanuals, eine umfassende Anleitung für die diagnostische Abklärung komplexer Fragestellungen zu geben. Es sollte aber in der diagnostischen Praxis selbstverständlich sein, niedrige Testwerte nicht ausschließlich als Eigenschaft des Kindes zu interpretieren, sondern auch die pädagogischen und familiären Rahmenbedingungen zu berücksichtigen. Vor einer Testung sollten auch organische Faktoren (insbesondere Hör- und Sehbeeinträchtigungen), die im Einzelfall den Erwerb der Schriftsprache beeinträchtigen können, abgeklärt werden.

Die Beurteilung des Arbeitsverhaltens sollte immer auf einer sorgfältigen Beobachtung in der Testsituation basieren. Bei klinischen und sonderpädagogischen Anwendungen werden daher Gruppentestungen kaum in Frage kommen.

Die Analyse des Arbeitsstils halte ich für einen durchaus positiven Aspekt des ELFE II, sie ist aber nicht in der Lage, alle denkbaren Auffälligkeiten im Testverhalten abzubilden. Die Grundrate der Differenz zwischen den beiden Normwerten für richtige und bearbeitete Aufgaben könnte sich zudem auf unterschiedlichen Fähigkeitsniveaus unterscheiden. Bei weit unterdurchschnittlichen Leistungen könnten möglicherweise größere Diskrepanzen zwischen bearbeiteten und korrekt gelösten Aufgaben auftreten. Es ist auch denkbar, dass zumindest bei einem Teil der Kinder eine Überforderung mit den Testaufgaben reaktiv zu einer oberflächlicheren Bearbeitung führt. Die im Manual definierten Kriterien für einen ungenauen Arbeitsstil könnten daher auch dann erfüllt werden, wenn keine primäre Beeinträchtigung des Arbeitsverhaltens vorliegt. Ebenso könnte ein sehr unaufmerksamer Arbeitsstil unerkannt bleiben, wenn Lösungen überhaupt nicht angestrichen werden. Auf jeden Fall ist es sehr zu begrüßen, dass Auswirkungen des Arbeitsstils im Manual diskutiert werden und so die Aufmerksamkeit der Testleiter*innen auf diesen wichtigen Aspekt der Testleistung gerichtet wird.

Bei künftigen Neuauflagen des ELFE II könnte sich die Aufnahme eines kurzen Screenings der Buchstabenkenntnis verbunden mit einem einfachen Vorlesetest (kurze und häufige Wörter) als sinnvoll erweisen, um einen einheitlichen Standard für die Beurteilung der Voraussetzungen einer Testdurchführung zu etablieren.

Als Dozierender will ich Studierenden vermitteln, dass Diagnostiker*innen wesentliche Gütekriterien eines Testverfahrens eigenständig beurteilen sollten. Es hat mich daher geärgert, dass im Vorwort des Manuals angedeutet wird, dass die eingehende Beschäftigung mit den Gütekriterien weniger relevant für die praktische Anwendung ist. Diese Aussage passt m. E. auch gar nicht zu dem sonst im ELFE II deutlich erkennbaren

Bestreben nach einer gründlichen und methodisch anspruchsvollen Fundierung des Verfahrens.

Zusammenfassende Bewertung von Markus Scholz

Ich kenne den ELFE II nicht mehr aus der direkten eigenen diagnostischen Anwendung. Meine Einschätzung beruht daher auf Diskussionen innerhalb diagnostischer Seminare an der Hochschule und der eigenen Analyse und Reflexion des Verfahrens.

Grundsätzlich halte ich den ELFE II für ein Verfahren, welches gut geeignet ist, Leseflüssigkeit, Lesegenauigkeit und Leseverständnis auf den beschriebenen Ebenen (Wort, Satz und Text) bei der angegebenen Altersgruppe zu erfassen. Die Aufgabenformate und die grundsätzliche Strukturierung des Testhefts sind Kindern aus unterrichtlichen Kontexten vermutlich sehr gut bekannt und sollten daher zu keinen größeren Problemen führen.

Anders sieht es bei Kindern mit Beeinträchtigungen aus. Hier ist anzumerken, dass die Gruppe der Kinder mit Beeinträchtigungen bzw. Behinderungen innerhalb des Testmanuals von den Autoren nicht erwähnt und damit explizit nicht berücksichtigt wird. Angesichts der sich im Kontext Inklusion verändernden Schulrealitäten gäbe es hier allerdings einen großen Bedarf. Die Aufgabenformate des ELFE II würden sich schon mit leichten Akkommodationen auch für Kinder mit unterschiedlichsten Ausgangsbedingungen eignen. In der vorliegenden Version führen die Gestaltung des Testheftes, einzelne Abbildungen und die Zeitvorgaben bei Kindern mit verschiedenen Ausgangsbedingungen zu einer Gefährdung der Testfairness (siehe auch Zugangsfertigkeiten / Einsatz bei behinderten Kindern / Testfairness).

Aus Perspektive potenzieller Testleiter*innen sind der klare Testaufbau und die eindeutigen Angaben zu Instruktionen positiv zu bewerten. Die Anforderungen sind nicht sonderlich hoch und sollten von routinierten Fachpersonen leicht und mit einem überschaubaren Einarbeitungsaufwand umgesetzt werden können. Dennoch halte ich es für fraglich, ob ein derartiges Verfahren – auch wenn es nur in der Computerversion ist – tatsächlich wie im Manual angegeben von weniger geschultem Personal durchgeführt (S. 16) werden sollte. Zwar wird dies durch den Hinweis auf die Notwendigkeit von Fachpersonen bei der Auswertung relativiert, Anforderungen ergeben sich aber auch vor und bei der Durchführung eines Verfahrens, insbesondere wenn es zahlreiche Faktoren gibt, die möglicherweise die Testfairness und somit eine valide Bestimmung der Lesefähigkeit gefährden.

Aktuelle Entwicklungen

Der ELFE II ist in erster Auflage 2017 erschienen. Neuauflagen sind daher in nächster Zeit nicht zu erwarten.

Testrezensionen

Schroth, J. (2017). *PSYNDEX Tests Review: ELFE II - Leseverständnistest für Erst- bis Siebtklässler - Version II*. Zugriff am 12.03.2020. Verfügbar unter <https://www.psyndex.de/retrieval/PSYNDEXTests.php?id=9007431>

Zitierte Literatur

- Brickenkamp, R., Schmidt-Atzert, L. & Liepmann, D. (2010). *d2-R. Test d2 - Revision*. Göttingen: Hogrefe.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256.
- Dunn, L. M. & Dunn, D. M. (2015). *PPVT Peabody Picture Vocabulary Test - 4. Ausgabe. Deutsche Bearbeitung von A. Lenhard, W. Lenhard, R. Segerer, S. Suggate*. Frankfurt am Main: Pearson.
- Kubinger, K. D. & Holoher-Ertl, S. (2014). *AID 3. Adaptives Intelligenz Diagnostikum 3*. Göttingen: Beltz Test.
- Melchers, P. & Melchers, M. (2015). *KABC-II. Kaufman Assessment Battery for Children – II von Alan S. Kaufman & Nadeen L. Kaufman. Deutschsprachige Fassung*. Frankfurt a. M.: Pearson.
- Mickley, M. & Renner, G. (2019). Auswahl, Anwendung und Interpretation Auswahl, Anwendung und Interpretation deutschsprachiger Intelligenztests für Kinder und Jugendliche auf Grundlage der CHC-Theorie: Update, Erweiterung und kritische Bewertung. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 68(4), 323–343. <https://doi.org/10.13109/prkk.2019.68.4.323>
- Moll, K. & Landerl, K. (2014). *SLRT-II Lese- und Rechtschreibtest (2., korr. Aufl. mit erw. Normen)*.
- Petermann, F. (2015). *SET 5-10. Sprachstandserhebungstest für Kinder im Alter zwischen 5 und 10 Jahren (3., überarb. & teilw. neu normierte Aufl.)*. Göttingen: Hogrefe.
- Schneider, W. J. & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (4th ed., pp. 73–163). New York: The Guilford Press.
- Stock, C., Marx, P. & Schneider, W. (2003). *BAKO 1–4. Basiskompetenzen für Lese-Rechtschreibleistungen. Ein Test zur Erfassung der phonologischen Bewusstheit vom ersten bis vierten Grundschuljahr*. Göttingen: Beltz Test.
- Wechsler, D. (2017). *WISC-V. Wechsler Intelligence Scale for Children - Fifth Edition. Deutsche Bearbeitung Franz Petermann*. Frankfurt: Pearson Assessment.
- Weiß, R. H. & Osterland, J. (2013). *Grundintelligenztest Skala 1 - Revision (CFT 1-R)*. Göttingen: Hogrefe.
- Wimmer, H. & Mayringer, H. (2014). *Salzburger Lese-Screening für die Schulstufen 2 - 9*. Bern: Hogrefe.

Ausgewählte weiterführende Literatur

- Lenhard, W. (2019). *Leseverständnis und Lesekompetenz. Grundlagen - Diagnostik - Förderung (Lehren und Lernen, 2., aktualisierte Auflage)*.
- 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>

Internet

<https://www.psychometrica.de/elfe2.html>

Internetseite der Testautor*innen mit umfangreichen Informationen zu Hintergrund und Anwendung des ELFE II.