



Stand 10.09.2021

---

Paulina Cordero Donoso & Gerolf Renner

**Zitationsempfehlung:**

Cordero Donoso, P. & Renner, G. (2021). *Testinformation zu den CELF-5. Deutsche Fassung Clinical Evaluation of Language Fundamentals – Fifth Edition* (Dia-Inform Verfahrensinformationen 009-01). Ludwigsburg: Pädagogische Hochschule Ludwigsburg.

Paulina Cordero Donoso arbeitet als Psychologin innerhalb der sozialpsychiatrischen Versorgung von Kindern und Jugendlichen. Des Weiteren ist sie als Dozentin in der Ausbildung von Kinder- und Jugendlichenpsychotherapeut\*innen und als Seminarleiterin im Bereich der Intelligenz- und Leistungsdiagnostik tätig.

**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>

## ***CELF-5*** ***Clinical Evaluation of Language Fundamentals – Fifth Edition***

Wiig, E. H., Semel, E. & Secord, W. E. (2020). *CELF-5 Clinical Evaluation of Language Fundamentals – Fifth Edition. Deutsche Fassung*. Frankfurt: Pearson.

### **Vorbemerkung**

---

Diese Testinformation befasst sich ausschließlich mit der manuellen Durchführung und Auswertung der CELF-5 und geht nicht auf die digitalen Testoptionen ein.

### **Altersbereich**

---

6;0 bis 16;11 Jahre.

### **Preis des Testverfahrens und der Verbrauchsmaterialien laut Verlagsseite**

---

Test komplett € 800,00; 25 Protokollbogen 6–8 Jahre € 75,00; 25 Protokollbogen 9–16 Jahre € 75,00; 50 Beobachtungsskalen € 50,00. Alle Angaben zuzüglich Mehrwertsteuer.

### **Diagnostische Zielsetzung lt. Testmanual**

---

Als diagnostische Zielsetzung der CELF-5 nennen die Autor\*innen eine umfassende Sprachdiagnostik, die eine normorientierte Diagnosestellung von Sprachstörungen ermöglichen will. Die CELF-5 soll zu Entscheidungen über Therapie- und Förderbedarfe beitragen und diene der Therapie- und Interventionsplanung sowie der Therapieverlaufskontrolle.

Neben der Frage, ob eine Sprachstörung vorliegt, sollen mit der CELF-5 spezifischere Fragestellungen beantwortet werden können:

- „Liegen Unterschiede zwischen Sprachverständnis und Sprachproduktion vor?
- Bestehen Schwächen in den Bereichen Morphologie, Syntax oder Semantik?
- Inwiefern beeinträchtigt die Sprachstörung die sozialen Interaktionen der Testperson“ (Durchführungsmanual, S. 17).

### **Theoretischer Hintergrund**

---

Das Technische Manual hält sich nicht mit einer längeren theoretischen Einführung auf, sondern beginnt mit einer Vorstellung der Indizes und Untertests des Verfahrens. Dabei wird jeweils definiert, was das Ziel des Untertests ist, und es werden – meist in knapper Form – ausgewählte empirische Befunde vorgestellt, die die Relevanz des Untertests für die Diagnostik von Sprachstörungen begründen. Ein umfassendes theoretisches Modell, das der Struktur der CELF-5 zugrunde liegt, wird nicht vorgestellt. Die altersabhängig wechselnden Testzusammenstellungen werden nicht entwicklungspsychologisch begründet.

Inhaltliche Schwerpunkte des Verfahrens sind Semantik, Morphologie, Syntax und die „Interaktion aus Anforderungen an Semantik, Sprachstruktur und Gedächtnis“ (Technisches Manual, S. 15).

### **Setting, Durchführungsdauer und Auswertungsmöglichkeiten**

---

Die CELF-5 ist ein Einzeltest. Die Durchführungsdauer der CELF-5 ist von der gewählten Testzusammenstellung, vom Alter der Testperson und dem Testeinstieg abhängig. Im

Manual wird betont, dass prinzipiell jeder einzelne Untertest unabhängig von anderen Untertests durchgeführt werden kann. Die vom Manual angegebenen durchschnittlichen Durchführungszeiten der Untertests sind „ungefähre Richtwerte, die in der Praxis stark variieren können“ (S. 22). In einer Validitätsstudie, in der Testpersonen mit einer diagnostizierten Sprachstörung untersucht wurden, betrug beispielsweise die Spanne der Durchführungsdauer bei *Textverarbeitung* 7 bis 25 Minuten, bei einer mittleren Testzeit von 11 bis 12 Minuten. Für die vier Kerntests, die die *Allgemeine Sprachskala* bilden, werden im Manual durchschnittlich 30 bis 36 Minuten Durchführungszeit angegeben. In der eigenen (P. C.) Anwendung betrug die Spanne der Durchführungszeit der vier Kerntests 33 bis 42 Minuten (reine Testzeit ohne Vorbereitung und Einführung).

Eine Auswertung der CELF-5 ist manuell oder internetbasiert über das Testsystem Q-global möglich.

### **Testaufbau, Testmaterial und Durchführung**

---

Die CELF-5 besteht aus 11 Untertests (UT), die in unterschiedlichen Kombinationen der Untertests folgenden Skalen zugeordnet werden (s. Tabellen 1 & 2):

- *Allgemeine Sprachskala* (ASS) (4 UT; altersabhängige Testzusammenstellung)
- *Index für Sprachverständnis* (ISV) (3 UT; altersabhängige Testzusammenstellung)
- *Index für Sprachproduktion* (ISP) (3 UT; altersabhängige Testzusammenstellung)
- *Index für Sprachinhalt* (ISI) (3 UT; altersabhängige Testzusammenstellung)
- *Index für Sprachstruktur* (ISS) (nur für die Altersgruppen 6;0 bis 8;11 Jahre; 4 UT)
- *Index für Sprachgedächtnis* (ISG) (nur für die Altersgruppen 9;0 bis 16;11 Jahre; 3 UT)

Ergänzend kommen drei Beobachtungs- bzw. Fremdeinschätzungsskalen zum Einsatz:

- die *Beobachtungsskala* (BOS) für Bezugspersonen mit 40 Items zur Beurteilung der alltäglichen Sprach- und Kommunikationsfähigkeiten
- das *Pragmatische Fähigkeitsprofil* (PFP) mit 49 Aussagen zur Einschätzung des kulturell angemessenen Sprachgebrauchs der Testperson
- die *Aktivitätencheckliste zur Pragmatik* (ACP), die nach der Beobachtung des Kommunikationsverhaltens der Testperson in natürlichen Interaktionssituationen durch die Testleiter\*innen ausgefüllt wird.

Einige Untertests werden mehrfach unterschiedlichen Indizes zugeordnet, wie z. B. die Untertests *Sätze Nachsprechen* und *Sätze Formulieren*, die sowohl bei der *Allgemeinen Sprachskala* als auch bei den Indizes für *Sprachproduktion*, *Sprachstruktur* und *Sprachgedächtnis* berücksichtigt werden. So ergeben sich etliche Überschneidungen zwischen den Indizes. Zum Teil ändert sich die Zugehörigkeit der Untertests zu den Skalen auch in Abhängigkeit vom Alter: Beispielsweise wird *Sätze nachsprechen* bei den 6- bis 8-Jährigen u. a. dem *Index für Sprachstruktur* zugeordnet, bei den 9- bis 16-Jährigen dagegen dem *Index für Sprachgedächtnis*.

**Tabelle 1**

*Aufbau der CELF-5*

6–8 Jahre	9–12 Jahre	13–16 Jahre
Allgemeine Sprachskala (ASS)		
Standardmäßig durchgeführte Kerntests		
Sätze Formulieren	Sätze Formulieren	Sätze Formulieren
Sätze Nachsprechen	Sätze Nachsprechen	Sätze Nachsprechen
Satzverständnis	Wortbedeutung	Textverarbeitung
Wortstruktur	Semantische Beziehungen	Semantische Beziehungen
Index für Sprachverständnis (ISV)		
<i>Wortbedeutung</i> <sup>a</sup>	Wortbedeutung	<i>Wortbedeutung</i> <sup>a</sup>
<i>Anweisungen Befolgen</i> <sup>a</sup>	<i>Anweisungen Befolgen</i> <sup>a</sup>	Textverarbeitung
Satzverständnis	Semantische Beziehungen	Semantische Beziehungen
Index für Sprachproduktion (ISP)		
Sätze Formulieren	Sätze Formulieren	Sätze Formulieren
Sätze Nachsprechen	Sätze Nachsprechen	Sätze Nachsprechen
Wortstruktur	<i>Sätze Zusammensetzen</i> <sup>a</sup>	<i>Sätze Zusammensetzen</i> <sup>a</sup>
Index für Sprachinhalt (ISI)		
<i>Wortbedeutung</i> <sup>a</sup>	Wortbedeutung	<i>Wortbedeutung</i> <sup>a</sup>
<i>Konzepte Erfassen</i> <sup>a</sup>	<i>Textverarbeitung</i> <sup>a</sup>	Textverarbeitung
<i>Anweisungen Befolgen</i> <sup>a</sup>	<i>Wortdefinitionen</i> <sup>a</sup>	<i>Sätze Zusammensetzen</i> <sup>a</sup>
Index für Sprachstruktur (ISS)	Index für Sprachgedächtnis (ISG)	
Sätze Formulieren	Sätze Formulieren	Sätze Formulieren
Sätze Nachsprechen	Sätze Nachsprechen	Sätze Nachsprechen
Satzverständnis	<i>Anweisungen Befolgen</i> <sup>a</sup>	<i>Anweisungen Befolgen</i> <sup>a</sup>
Wortstruktur	-	-
Beobachtungs- und Fremdeinschätzungsskalen		
Beobachtungsskala		
Pragmatisches Fähigkeitsprofil		
Aktivitätencheckliste zur Pragmatik		

*Anmerkung.* <sup>a</sup> Diese Untertests sind keine Kerntests und müssen zusätzlich durchgeführt werden, um den entsprechen Index berechnen zu können.

Die Inhalte der verschiedenen Skalen werden im Durchführungsmanual folgendermaßen beschrieben:

- Die *Allgemeine Sprachskala* soll einen Indikator der Gesamtleistung im sprachlichen Bereich darstellen.
- Der *Index für Sprachverständnis* erfasst Fähigkeiten in den Bereichen Zuhören und auditives Verständnis.
- Der *Index für Sprachproduktion* gilt als Maß der expressiven Sprachfertigkeiten (Aktivsprache).
- Der *Index für Sprachinhalt* erfasst verschiedene „Aspekte der Sprachentwicklung im semantischen Bereich“ (Durchführungsmanual, S. 171)
- Der *Index für Sprachstruktur* bezieht sich auf die Fähigkeit, „Satzstrukturen korrekt zu interpretieren und zu bilden“ (Durchführungsmanual, S. 171). Er

umfasst allerdings ein- und dieselben Untertests wie die *Allgemeine Sprachskala*.

- Der *Index für Sprachgedächtnis* erfasst die Fähigkeit, verbal dargebotene Informationen zu behalten und umzusetzen, verlangt aber auch die Bildung von Sätzen auf Basis vorgegebener Worte.

**Tabelle 2**

*Aufgabenstellung und inhaltliche Schwerpunkte der Untertests der CELF-5*

<b>Untertest</b>	<b>Aufgabenstellung</b>	<b>Erfasste Fähigkeiten</b>
Satzverständnis	Aus einer Auswahl von vier Bildern soll das Bild gezeigt werden, das zur Aussage eines verbal dargebotenen Satzes passt.	Verständnis von Sätzen zunehmender Länge und Komplexität
Konzepte Erfassen	Auf der Grundlage verbal dargebotener Informationen, die bestimmte Konzepte umfassen, sollen abgebildete Objekte gezeigt werden.	Verständnis grundlegender Konzepte (z. B. „anders“, „viele“) und logischer Operationen
Wortstruktur	Ein zu Bildern passender unvollständiger Satz soll grammatikalisch und inhaltlich korrekt beendet werden.	Erkennen und Anwendung morphologischer Regeln (Numerus, Kasus, Tempus)
Wortbedeutung	Aus drei bzw. vier Worten sollen zwei konzeptuell zusammenpassende ausgewählt werden.	Fähigkeit, Beziehungen zwischen Wörtern auf der Grundlage semantischer Merkmale zu erfassen
Anweisungen Befolgen	Entsprechend mündlichen Anweisungen, die allmählich länger und komplexer werden, sollen abgebildete Objekte gezeigt werden.	Interpretation und Umsetzung mündlicher Anweisungen zunehmender Länge und inhaltlicher Komplexität
Sätze Formulieren	Zu einem Bild soll ein Satz formuliert werden, der ein oder zwei mündlich vorgegebene Wörter enthält.	Formulierung semantisch und syntaktisch korrekter Sätze von zunehmender Länge und Komplexität
Sätze Nachsprechen	Mündlich dargebotene Sätze sollen nachgesprochen werden.	Merkfähigkeit für sprachliche Inhalte (Sätze)
Textverarbeitung	Fragen zu einem vorgelesenen Text sollen beantwortet werden.	Fähigkeit, Inhalte eines Textes zu speichern, abzurufen und daraus Informationen abzuleiten
Wortdefinitionen	Ein Wort wird vorgelesen und beispielhaft in einem Satz angewendet. Anschließend soll das Kind dieses Wort definieren.	Fähigkeit Wörter zu definieren
Sätze Zusammen setzen	Aus mündlich und schriftlich ungeordnet dargebotenen Wörtern bzw. Wortgruppen sollen zwei Sätze gebildet werden.	Formulierung grammatikalisch richtiger und semantisch bedeutungsvoller Sätze.

**Tabelle 2 (fortgesetzt)**

Semantische Beziehungen	Eine Frage zu einem zuvor vorgelesenen Satz bzw. Text soll beantwortet werden, indem aus vier mündlich und schriftlich vorgegebenen Antwortalternativen die zwei richtigen ausgewählt werden.	Interpretation von Sätzen, die Vergleiche herstellen, Position oder Richtung vorgeben, zeitliche Relationen angeben, eine Reihenfolge beinhalten oder im Passiv ausgedrückt sind.
-------------------------	---	---

Als Ergänzung zur standardisierten Auswertung der Untertests wird noch eine optionale Itemanalyse angeboten, die eine nähere qualitative Beschreibung der Fehlermuster ermöglichen soll (z. B. bestimmte semantische Kategorien bei *Wortdefinitionen*).

Zur CELF-5 gehören zusätzlich drei Beobachtungs- und Fremdeinschätzungsinstrumente:

- Die *Beobachtungsskala* besteht aus 40 Items, mit denen anhand einer vierstufigen Skala die Kommunikationsleistung der Testperson sowohl im schulischen als auch im häuslichen Alltag von Eltern/engen Bezugspersonen, Lehrer\*innen und der Testperson selbst (bei älteren Kindern und bei Jugendlichen) eingeschätzt werden kann. Die Items berücksichtigen die Bereiche Zuhören, Sprechen, Lesen und Schreiben. Die beurteilende Person wird gebeten, die 5 bis 10 bedeutsamsten Schwierigkeiten zu markieren und ggf. weitere Auffälligkeiten einzutragen. Die *Beobachtungsskala* kann vor oder nach der Testuntersuchung mit der CELF-5 ausgefüllt werden. Normwerte stehen für die *Beobachtungsskala* nicht zur Verfügung.
- Das *Pragmatische Fähigkeitsprofil* dient der Erfassung kulturell angemessener kommunikativer Verhaltensweisen. 49 Items sollen Defizite im Sprachgebrauch und in der nonverbalen Kommunikation identifizieren. Die Beurteilung erfolgt durch die Testleiter\*innen, wobei zusätzlich Informationen von Eltern bzw. wichtigen Bezugspersonen oder Lehrer\*innen eingeholt werden sollen. Die Items beschreiben soziale und kommunikative Kompetenzen in drei Verhaltensbereichen: Rituale und Konversation; um Informationen bitten, auf diese reagieren und selbst Auskunft geben; nonverbale Kommunikation. Verwendet wird eine 4-stufige Likert-Skala. Das Ausfüllen des *Pragmatischen Fähigkeitsprofils* soll erst erfolgen, nachdem die Testleiter\*innen ausreichende Zeit für die Beobachtung des Kommunikationsverhaltens der Testperson hatten. Die Anwendung des *Pragmatischen Fähigkeitsprofils* setzt Vertrautheit „mit der Kultur und den Lebensumständen der Testperson“ voraus (Durchführungsmanual, S. 141). Für den Gesamtwert des *Pragmatischen Fähigkeitsprofils* stehen Normwerte zur Verfügung.
- Die *Aktivitätscheckliste zur Pragmatik* stellt die Beurteilung von Verhaltensweisen der Testperson in realitätsnahen Kommunikationssituationen in den Vordergrund. Nach der Durchführung von drei der sechs vom Manual vorgeschlagenen Aktivitäten (z. B. Ein Spiel erklären/lernen/spielen) werden Beobachtungen der Testleiter\*innen anhand der Checkliste festgehalten. 32 Items beschreiben verbale (Kommunikationsstil, Relevanz der Kommunikation sowie Qualität und Quantität der Kommunikation) und nonverbale (Blick, Gesten, Ausdruck und Körpersprache) Verhaltensweisen. Ein auf amerikanischen Daten basierender Cut-Off-Wert, der für alle Altersgruppen Gültigkeit beansprucht, kann zur Beurteilung eines auffälligen Kommunikationsverhaltens herangezogen werden, sollte aber

auch nach Ansicht der Testautor\*innen nur „mit Vorsicht“ (Durchführungsmanual, S. 178) und im Kontext weiterer diagnostischer Befunde interpretiert werden.

Der Testkoffer enthält ein Technisches sowie ein Durchführungsmanual, zwei Stimulusbücher, die der Präsentation der Items dienen, jeweils 25 Protokollbogen für die Altersbereiche 6 bis 8 Jahre und 9 bis 16 Jahre und die 50 Exemplare der Beobachtungsskala.

## **Ergebniswerte**

---

Zentrale Ergebniswerte:

- Wertpunkte ( $M = 10$ ,  $SD = 3$ ) für die Untertests und das *Pragmatische Fähigkeitsprofil*
- Indexwerte ( $M = 100$ ,  $SD = 15$ ) mit 68%-, 90%- und 95%-Konfidenzintervallen sowie Prozentränge für die *Allgemeine Sprachskala* und die fünf Indizes

Weitere Ergebniswerte:

- Testalteräquivalente (Entwicklungsalter) für die Rohwertsummen der Untertests im Altersbereich 6;0 bis 16;10 Jahre
- Cut-Off-Wert für die *Aktivitätencheckliste zur Pragmatik*
- Wachstumswerte zur Beurteilung von intraindividuellen Veränderungen der Leistungen. Wachstumswerte sind nicht altersnormiert und stellen ein Maß für die absolute Leistungsfähigkeit dar. Sie können bei einer Testwiederholung Leistungszuwächse aufzeigen, auch wenn sich der Standardwert gegenüber der Ersttestung verschlechtert hat.

Daten zur Profilinterpretation

- Kritische Differenzen auf dem 5%- und dem 15%-Fehlerniveau für Vergleiche zwischen den Indexwerten
- Häufigkeiten der Differenzen zwischen Indexwerten in der Normstichprobe

## **Objektivität**

---

Die allgemeinen Richtlinien und Regeln für die Testuntersuchung mit der CELF-5 geben ausreichende Hinweise zur Vorbereitung der Testung, zur Gestaltung der Testumgebung, zur Beziehungsgestaltung, zum Umgang mit Unterbrechungen und zu Anpassungen der Testdurchführung bei Menschen unterschiedlicher sprachlicher und kultureller Hintergründe. Auf Besonderheiten der Testung und der Beziehungsgestaltung bei jüngeren Kindern und deren Familien wird eingegangen. Es fehlen genauere Hinweise zum Sprechtempo, insbesondere bei den Nachsprechaufgaben und zur Betonung beim Vorlesen der Testinstruktionen. Ebenfalls fehlen Angaben zum Umgang mit schwierigen Testsituationen, wie z. B. Verweigerung der Mitarbeit und zum Umgang mit jüngeren Kindern, die aufgrund von unzureichender Konzentration oder geringer Ausdauer mit der Bearbeitung der Aufgaben von Untertests überfordert sind.

Nicht gesichert ist die Objektivität für das *Pragmatische Fähigkeitsprofil*, da bei der Beurteilung situative Einflüsse und die Auswahl der zusätzlich befragten Personen eine bedeutsame Rolle spielen können. Bei der *Aktivitätencheckliste zur Pragmatik* geben die Instruktionen einen Rahmen für die Beobachtung vor, der den Anwender\*innen viel Spielraum bei der konkreten Umsetzung einräumt. Bei beiden Verfahren ist die geforderte Einbeziehung des kulturellen und sprachlichen Hintergrunds der Testperson nicht objektiv realisierbar.



Die spezifischen Testinstruktionen sind weitgehend klar. Testeinstieg, Durchführung der Demonstrations- und Übungsitens, Startpunkte, zulässige Wiederholungen der verbalen Stimuli, Nachfragen, Testabbruch und Umkehregeln sind genau festgelegt.

Eine objektive Bewertung der Antworten anhand festgelegter Vorgaben kann bei den meisten Untertests der CELF-5 eindeutig erfolgen. Bei den Untertests *Wortstruktur*, *Sätze Formulieren*, *Textverarbeitung* und *Wortdefinitionen* gibt es einen gewissen Interpretationsspielraum. In diesen Fällen sind nach Einschätzung der Autor\*innen die Testleiter\*innen gefordert, eine eigenständige Entscheidung über noch zulässige und nicht mehr zulässige Antworten zu treffen. Für die oben erwähnten Untertests sind ausführliche Regeln und Kriterien entwickelt worden, die eine präzise Bewertung dieser Items ermöglichen sollen.

Daten zur Beurteilerübereinstimmung liegen für die deutsche Version der CELF-5 nicht vor. In einer Untersuchung der amerikanischen Version zeigten zwei unabhängig vorgenommene Bewertungen von zufällig ausgewählten Testungen folgende Korrelationen: *Wortstruktur* .99, *Sätze Formulieren*.89, *Textverarbeitung*.97 und *Wortdefinitionen*.96 (Technisches Manual, S. 57).

Die Interpretationsobjektivität ist durch den Bezug auf Normwerte gesichert.

### **Reliabilität**

---

Als Reliabilitätskennwerte werden getrennt nach Altersjahrgängen die internen Konsistenzen der Untertests und Skalen angegeben (Tabelle 3). Die Schätzungen der Reliabilität beruhen somit auf maximal 63 Kindern und Jugendlichen – eine für Reliabilitätsbestimmungen eher geringe Stichprobengröße (vgl. Charter, 2003).

Die Werte der Untertests liegen mit wenigen Ausnahmen im befriedigenden bis sehr guten Bereich. Werte unter .80 kommen bei *Sätze Formulieren*, *Wortbedeutung*, *Textverarbeitung* und – in einer einzigen Altersgruppe – bei *Sätze Zusammensetzen* vor.

Die internen Konsistenzen für die Skalen fallen ganz überwiegend in den guten bis sehr guten Bereich, nur 2 von 55 Werten liegen unter .90.

Reliabilitätswerte wurden auch an einer klinischen Stichprobe ermittelt. Hier fand sich für *Satzverständnis* mit .48 ein deutlich niedrigerer Wert als in der Normstichprobe, die weiteren Werte zeigen keine gravierenden Abweichungen. Es handelte sich um eine sehr kleine Stichprobe ( $n = 32$  für die Altersgruppe 6-8;  $n = 36$  für die Altersgruppe 9-16), mit der eine zuverlässige Bestimmung von Reliabilitätskennwerten sicherlich nicht möglich war.

Die Retest-Stabilität wurde an einer altersgemischten Stichprobe von 70 Kindern und Jugendlichen bestimmt. Die Spannbreite des Intertestintervalls reichte von 10 bis 263 Tagen. Die Werte der Untertests liegen überwiegend unter .80, die der Skalen mindestens bei .87, womit zumindest auf Skalenebene eine recht hohe Stabilität gegeben ist.

Angegeben werden auch die Effektstärken (Cohens  $d$ ) der Veränderungen vom ersten zum zweiten Messzeitpunkt. Auf Skalenebene liegen diese zwischen .19 (*Index für Sprachverständnis*; Zunahme um 2.2 Indexwerte) und .38 (*Index für Sprachproduktion*; Zunahme um 5.6 Indexwerte).

Zur *Beobachtungsskala* und zur *Aktivitätencheckliste zur Pragmatik* werden keinerlei Reliabilitätsdaten vorgelegt.

**Tabelle 3***Reliabilitätskennwerte der CELF-5 Untertests und Skalen*

Untertest / Skala	Interne Konsistenz <sup>a</sup>	Retest
Satzverständnis	.93 - .95	.73
Konzepte Erfassen	.89 - .97	.64
Wortstruktur	.87 - .92	.77
Wortbedeutung	.67 - .94	.70
Anweisungen Befolgen	.89 - .96	.81
Sätze Formulieren	.70 - .91	.67
Sätze Nachsprechen	.89 - .95	.83
Textverarbeitung	.70 - .90	.71
Wortdefinitionen	.72 - .91	.81
Sätze Zusammensetzen	.75 - .96	.75
Semantische Beziehungen	.91 - .96	.75
Pragmatisches Fähigkeitsprofil	.96 - .98	.92
Allgemeine Sprachskala	.93 - .97	.91
Index für Sprachverständnis	.90 - .97	.89
Index für Sprachproduktion	.89 - .97	.89
Index für Sprachinhalt	.88 - .97	.90
Index für Sprachstruktur	.96 - .97	.92
Index für Sprachgedächtnis	.92 - .96	.87

*Anmerkung.* <sup>a</sup> Angegeben ist jeweils die Spannweite der internen Konsistenzen in den Altersjahrgängen (Technisches Manual, S. 51).

### **Validität**

Grundsätzlich ist die *inhaltliche Validität* der Untertests der CELF-5 als Indikatoren verschiedener Aspekte der Sprachentwicklung gegeben. Bei keinem Untertest lassen sich in der inhaltlichen Analyse Hinweise auf merkbliche konstruktirrelevante Einflüsse feststellen.

Auf Ebene der Indizes ist die Frage der inhaltlichen Validität weniger eindeutig zu beantworten. Die *Allgemeine Sprachskala* enthält in allen Altersgruppen Untertests, die unterschiedliche Facetten sprachlicher Kompetenzen ansprechen, was den Anspruch, eine Aussage über die „allgemeine“ Sprachentwicklung zu treffen, begründen kann.

Nachvollziehbar ist auch die Zusammenstellung der Untertests im *Index für Sprachverständnis*. Inhaltlich weniger plausibel erscheint die inhaltliche Zuordnung von *Sätze Nachsprechen* zum *Index für Sprachproduktion*, handelt es sich doch um eine reproduktive und keine produktive Anforderung. Der *Index für Sprachinhalt* soll laut Durchführungsmanual verschiedene Aspekte der Semantik erfassen, enthält bei den 13- bis 16-Jährigen mit *Sätze Zusammensetzen* jedoch auch einen Untertest, bei dem die Syntax im Vordergrund steht. In keiner Weise nachvollziehbar ist die Bildung des *Index*

für *Sprachstruktur* bei den 6- bis 8-Jährigen, da sich dieser Index nicht im Geringsten von der *Allgemeinen Sprachskala* unterscheidet.

Dem Begriff *Sprachgedächtnis* fehlt eine gedächtnispsychologische Fundierung. Er erweckt, den Eindruck, es gäbe es ein einheitliches sprachbezogenes Gedächtnissystem. Dabei sollte zumindest zwischen der phonologischen Schleife des Arbeitsgedächtnisses und dem semantischen Gedächtnis (einem Subsystem des Langzeitgedächtnisses) unterschieden werden. Ein weiterer Aspekt ist die Differenzierung von Speicherung und Abruf, die in Gedächtnistests z. B. durch einen Vergleich von Erinnern und Wiedererkennen beim Lernen von Wortlisten erfolgen kann. Der *Index für Sprachgedächtnis* erfasst teilweise das unmittelbare Reproduzieren verbaler Inhalte (*Sätze nachsprechen*). Die Effizienz des Abrufs und der Speicherung verbaler Inhalte im Langzeitgedächtnis wird nicht berücksichtigt. Die Aufgaben der Subtests *Anweisungen Befolgen* und *Sätze formulieren*, die ebenfalls in den *Index für Sprachgedächtnis* einfließen, beanspruchen durchaus die phonologische Schleife. Die Testergebnisse werden jedoch nicht allein durch die Gedächtnisleistungen, sondern auch durch das Sprachverständnis bzw. syntaktische Fähigkeiten bestimmt.

Die Bezeichnungen der Indizes und die entsprechenden Beschreibungen im Manual legen nahe, dass jeweils spezifische Aspekte der Sprachentwicklung erfasst werden. Die Teststruktur weist jedoch deutliche Überschneidungen auf, die an der Spezifität der Indizes zweifeln lassen: Beispielsweise haben der *Index für Sprachverständnis* und der *Index für Sprachinhalt* bei den 6- bis 8-Jährigen sowie der *Index für Sprachverständnis* und der *Index für Sprachgedächtnis* bei den 9- bis 16-Jährigen jeweils zwei von drei Untertests gemeinsam.

Einige Untertests wechseln beim Übergang von einer Altersgruppe zur anderen die Skalenzugehörigkeit. Inhaltliche Argumente hierfür liefert das Manual nicht. Warum sollten z. B. *Sätze Formulieren* und *Sätze Nachsprechen* mit 8;11 Jahren „Sprachstruktur“ erfassen, mit 9;0 Jahren jedoch „Sprachgedächtnis“?

Die *Beobachtungsskala* will „mögliche Probleme und Verhaltensweisen der Testperson in den Bereichen Zuhören, Sprechen, Lesen und Schreiben“ (Durchführungsmanual, S. 165) erfassen. Einige der erfragten Verhaltensweisen können diesen Bereichen jedoch nicht eindeutig zugeordnet werden und z. B. auf Aufmerksamkeitsstörungen oder kognitive Entwicklungsstörungen hinweisen. Auch bei der *Aktivitätencheckliste zur Pragmatik* und beim *Pragmatischen Fähigkeitsprofil* ist damit zu rechnen, dass Personen mit deutlichen kognitiven Entwicklungsstörungen auffällige Werte erzielen, die jedoch nicht als spezifische pragmatische Störung eingeordnet werden sollten.

Bei der Diskussion der inhaltlichen Validität wurden bereits die deutlichen inhaltlichen Überschneidungen einiger Indizes angesprochen. Dementsprechend ergaben sich in der Normstichprobe – leider nicht nach Altersgruppen differenziert – durchgehend hohe Interkorrelationen ( $r_{tt} \geq .75$ ), z. B. *Index für Sprachstruktur* und *Index für Sprachproduktion* .97, *Index für Sprachproduktion* und *Index für Sprachgedächtnis* .94, *Index für Sprachverständnis* und *Index für Sprachinhalt* .91. Dies ist ein erster deutlicher Hinweis, dass die Differenzierung auf Ebene der Indizes empirisch nicht gestützt wird.

Zur faktoriellen Validität berichten die Autor\*innen in sehr knapper und kaum nachvollziehbarer Form (es fehlen z. B. alle Angaben zu Faktorladungen) sechs konfirmatorische Faktorenanalysen. Dabei wurden für die drei Altersgruppen jeweils zwei unterschiedliche hierarchische Modelle geprüft, mit der *Allgemeinen Sprachskala* als Faktor zweiter Ordnung und zwei Faktoren erster Ordnung. In allen drei Altersgruppen berücksichtigt

Modell 1 die Unterscheidung zwischen rezeptiver Sprache (Untertests des Index ISV) und expressiver Sprache (Untertest des Index ISP). Bei den 6- bis 8-Jährigen wurden im Modell 2 Sprachinhalt (Index ISI) und Sprachstruktur (Index ISS) als Faktoren erster Ordnung gebildet, bei den 9- bis 16-Jährigen Sprachinhalt (Index ISI) und Sprachgedächtnis (Index ISG). In die Faktorenanalysen wurden laut Technischem Manual (S. 75) auch die Untertests der jeweiligen Indizes aufgenommen, die nicht in der *Allgemeinen Sprachskala* enthalten sind. Die geprüften Modelle entsprechen daher nicht der für die standardmäßige Testauswertung herangezogenen Struktur. Eindeutige Kriterien für die Bewertung der Fit-Indizes werden im Manual nicht angegeben. Nach üblichen Standards ergibt sich eine gute Passung nur für Modell 1 bei den 6- bis 8-Jährigen und für Modell 2 bei den 9- bis 12-Jährigen. Es fällt daher schwer, die Aussage der Autor\*innen, dass die Beibehaltung der vorgeschlagenen Indizes „gerechtfertigt und sinnvoll“ (Technisches Manual, S. 76) sei, nachzuvollziehen.

Zur *konvergenten Validität* liegen erste Daten zum Zusammenhang mit einem Wortschatztest, dem Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2015), vor. Bei einer Stichprobe von 51 Kindern und Jugendlichen im Alter von 6 bis 15 Jahren fanden sich auf Untertestebene positive Korrelationen zwischen .42 (*Konzepte erfassen*) und .61 (*Wortbedeutung*) mit tendenziell höheren Werten für Untertests, die semantische Leistungen ansprechen. Auf Skalenebene lagen die Korrelationen zwischen .58 (*Index für Sprachproduktion*) und .72 (*Index für Sprachstruktur*). Daten zum Zusammenhang mit anderen Sprachentwicklungstests liegen nur für die amerikanische Version vor. Zusammenhänge mit Intelligenztests, Schulleistungstests und spezifischen Leistungstests wurden nicht untersucht. Zur *divergenten Validität* finden sich ebenfalls keine empirischen Befunde.

Bei einem Testverfahren, das klinische Diagnosestellungen ermöglichen will, sind Daten zu relevanten klinischen Gruppen von besonderer Bedeutung. Die Ergebnisse von 68 Kindern und Jugendlichen im Alter von 6 bis 16 Jahren mit Umschriebenen Sprachentwicklungsstörungen wurden mit einer nach Alter und Geschlecht parallelisierten Teilstichprobe der Normierung verglichen. Für alle Untertests und Skalen fanden sich signifikante Unterschiede zwischen den beiden Gruppen. Auf Skalenebene lagen die Effektstärken (Cohens d) zwischen 1.40 (*Index für Sprachinhalt*) und 2.12 (*Index für Sprachgedächtnis*). Die Niveauunterschiede fallen also deutlich und erwartungsgemäß aus.

Darüber hinaus stellt sich die Frage, ob die Testergebnisse der CELF-5 eine Unterscheidung zwischen Kindern mit und ohne Sprachentwicklungsstörungen erlauben, bzw. idealerweise auch eine Unterscheidung verschiedener Sprachstörungen und eine Abgrenzung von Sprachentwicklungsstörungen und weiteren Entwicklungsbeeinträchtigungen (z. B. kognitive Entwicklungsstörungen) ermöglichen. Wichtige Kennwerte für solche Klassifikationsentscheidungen sind u. a. Sensitivität und Spezifität. Für die eben genannte Stichprobe erfolgte die Berechnung von Sensitivität und Spezifität erstaunlicherweise nicht. Unter der Überschrift „Studie zur Sensitivität“ wird im Abschnitt zur Testentwicklung auf eine weitere Untersuchung an 37 Kindern mit Jugendlichen mit einer diagnostizierten Sprachentwicklungsstörung verwiesen. Auch hier hätten sich im Vergleich mit parallelisierten Probanden der Erprobungsstudie signifikante Unterschiede ergeben, wiederum erfolgen keine Angaben zur Sensitivität und Spezifität.

Weitere Daten zu klinischen Gruppen liegen nur für die amerikanische Fassung vor. Dabei fanden sich u. a. auf Skalenebene durchgehend signifikant niedrigere Werte bei Kindern und Jugendlichen mit Beeinträchtigungen beim Lesen oder Schreiben

(Effektstärken zwischen 0.90 und 1.36) und bei Kindern und Jugendlichen mit Autismus-Spektrum-Störungen (Effektstärken zwischen 1.72 und 2.57).

Zur *prognostischen Validität* enthält das Manual keine Daten. Die Normstichprobe umfasst einen hohen Anteil von Kindern und Jugendlichen mit Migrationshintergrund, gesonderte Daten werden für diese Gruppe jedoch nicht dargestellt. Von Interesse wären auch Daten zur Frage, ob mit der CELF-5 zwischen Kindern mit Sprachentwicklungsstörungen und Kindern mit einer noch geringen Exposition in der Testsprache unterschieden werden kann.

Die *Beobachtungsskala* und die *Aktivitätencheckliste zur Pragmatik* werden im Abschnitt „Validität“ des technischen Manuals nicht berücksichtigt.

## **Normierung**

---

Die Normierung der CELF-5 erfolgte zwischen September 2018 und November 2019. Die Stichprobe umfasst 652 Kinder und Jugendliche (zum Vergleich: die US-amerikanische Version wurde an 3 250 Personen im Alter von 5 bis 21 Jahren normiert; Coret & McCrimmon, 2015). Die Normierung wurde vom Testverlag an ein „renommiertes Sozialforschungsinstitut“ (Technisches Manual, S. 42) vergeben. Die Qualitätssicherung bei der Normierung wird detailliert beschrieben und umfasste u. a. intensive Schulungen, Probetestungen und Kontrollen während der Erhebung.

Kinder mit Sprachentwicklungsstörungen, Lese-Rechtschreib-Schwäche sowie Kinder mit Hör- und Sehstörungen wurden nicht berücksichtigt. Bei der Rekrutierung wurde eine den bevölkerungsstatistischen Daten entsprechende Verteilung des Bildungsniveaus der Eltern (vor Besuch einer weiterführenden Schule) bzw. des besuchten Schultyps (ab Besuch einer weiterführenden Schule) angestrebt. Diese Angaben werden nach Altersjahrgängen differenziert aufgeschlüsselt und sprechen dafür, dass diesbezüglich eine repräsentative Stichprobe gewonnen werden konnte. Der Anteil von Kindern und Jugendlichen mit Migrationshintergrund lag bei 31.9%.

Die Berechnung der Normen erfolgte nach einem komplexen mathematischen Modell („inferential norming“), das anhand der Angaben im Manual nicht im Detail nachvollzogen werden kann.

In der klinisch-psychologischen und sonderpädagogischen Diagnostik sollten Testverfahren auch im unterdurchschnittlichen Bereich differenziert messen. Eine Differenzierung im unteren Bereich wird durch Bodeneffekte (es fehlen sehr einfache Items), die im oberen Bereich durch Deckeneffekte (es fehlen sehr schwierige Items) eingeschränkt.

Die CELF-5 ist in keiner Altersgruppe völlig frei von Boden- und Deckeneffekten. Auf Subtestebene finden sich nach den Kriterien von Renner (2017) in der Altersgruppe 6;0 bis 6;5 Jahre sechs leichte und zwei deutliche Bodeneffekte (*Anweisungen Befolgen*, *Sätze Formulieren*). Bei den 9-Jährigen finden sich noch ein deutlicher (*Wortdefinitionen*) und drei leichte Bodeneffekte. Deckeneffekte sind in allen Altersgruppen zu finden, bei den 16-Jährigen können teilweise (*Sätze Zusammensetzen*, *Semantische Beziehungen*) maximal 13 Wertpunkte erzielt werden.

In den Normtabellen finden sich einzelne problematische Itemgradienten. Eine Veränderung des Rohwerts um 1 geht in wenigen Fällen mit einer Veränderung des Standardwerts (Wertpunkte) um 2 einher. Nur in einem Fall (*Sätze Zusammensetzen*) bei den 15-Jährigen) führt die minimal mögliche Veränderung des Rohwerts sogar zu einem

Sprung von drei Wertpunkten – ein einziges mehr oder weniger gelöstes Item führt in diesem Fall zu Ergebnissen, die eine ganze Standardabweichung auseinanderliegen.

Die Altersgruppen der CELF-5 umfassen mit Ausnahme der 6-Jährigen ein volles Jahr, z. B. werden Kinder im Alter von 7;0 und 7;11 Jahren nach der gleichen Normtabelle bewertet. Es stellt sich daher die Frage, ob es an den Grenzen der Normgruppen zu deutlichen Veränderungen in den Standardwerten kommen kann. Eine Überprüfung war nicht für alle denkbaren Rohwerte, sondern nur stichprobenweise möglich. Dabei zeigte sich beispielsweise beim Übergang von der Normgruppe 6;0 bis 6;5 Jahre zur Normgruppe 6;6 bis 6;11, dass in 25% der ausgewählten Beispiele die Überschreitung der Altersgrenze zu einem um 2 Wertpunkte niedrigeren Testwert führte. Vergleichbare Ergebnisse fanden sich auch für den Übergang zur Normgruppe der 7-Jährigen und zur Normgruppe der 9-Jährigen. Zwischen 10 und 11 Jahren kam es maximal zu einer Veränderung von einem Wertpunkt.

### **Weitere Gütekriterien**

---

Die *Ökonomie* der CELF-5 kann nicht pauschal bewertet werden. Mit der *Allgemeinen Sprachskala* können unterschiedliche Aspekte der sprachlichen Entwicklung recht zügig erfasst werden. Der Zeitbedarf für die Auswertung sollte allerdings nicht unterschätzt werden, z. B. wenn Audioaufnahmen, die bei einigen Untertests empfohlen werden, analysiert werden müssen.

Kommen weitere Untertests hinzu, werden Explorationen durchgeführt, um das *Pragmatische Fähigkeitsprofil* auszufüllen und wird die *Aktivitätencheckliste zur Pragmatik* eingesetzt – einschließlich der notwendigen Abklärung des sprachlichen und kulturellen Hintergrunds der Testpersonen – ergibt sich ein erheblicher Zeitaufwand.

### **Zielgruppengerechte Gestaltung**

---

Die Mitarbeit der Kinder ist bei der CELF-5 auf Zeigen und verbale Antworten beschränkt. Nach eigenen (P. C.) Erfahrungen wird das Testverfahren von älteren und leistungsstarken Kindern gut angenommen. Demonstration- und Übungsaufgaben erweisen sich als hilfreich, um das Aufgabenverständnis zu sichern.

Für junge Kinder sowie Kinder mit eingeschränkter Konzentration oder kognitiven Entwicklungsstörungen ist die Testuntersuchung nicht kontinuierlich anregend. Vielfältigere Handlungsmöglichkeiten (z. B. Hantieren mit Gegenständen) und ein abwechslungsreicheres Material könnten sich positiv auf die Kooperation und Motivation der Kinder auswirken.

Untertests, die keine Abbruchregel vorsehen (*Wortstruktur* und *Textverarbeitung*) beeinträchtigten in der eigenen (P. C.) Testpraxis die Motivation leistungsschwächerer Kinder. Jüngere Kinder und Kinder mit kognitiven Einschränkungen zeigten bei einigen Instruktionen (z. B. beim Untertest *Sätze Formulieren*) teilweise ausgeprägte Verständnisprobleme.

### **Zugangsfertigkeiten / Einsatz bei behinderten Kindern / Testfairness**

---

Die teilweise vorhandenen Bodeneffekte (s. Abschnitt Normierung) führen nicht nur zu einer schlechteren Differenzierung im unteren Leistungsbereich, sie können sich auch negativ auf die Testmotivation wirken, da vermehrt Misserfolgserlebnisse auftreten.

Die handmotorischen Anforderungen beschränken sich auf Zeigereaktionen, die in drei bis vier Untertests der CELF-5 verlangt werden. Bei Vorliegen deutlicher Beeinträchtigungen des zielgerichteten Zeigens (z. B. bei dyskinetischen cerebralen Bewegungsstörungen) wird es z. B. bei *Konzepte Erkennen* schwerfallen, die Bewegung exakt auf die Stimuli auszurichten, da der Abstand zwischen den Bildern der Itemauswahl nur ca. 1.5 cm beträgt.

Acht Untertests setzen visuelles Material ein und sind daher für Kinder und Jugendliche mit Sehbehinderungen nicht oder nur eingeschränkt geeignet. Das visuelle Material ist überwiegend ansprechend gestaltet, bei wenigen Items könnten relevante Details bei Vorliegen von Sehbeeinträchtigungen jedoch zu klein dargestellt sein.

Bei Vorliegen sprechmotorischer Störung (z. B. Dysarthrie, Anarthrie) können drei Untertests (*Anweisungen befolgen*, *Konzepte erfassen*, *Satzverständnis*) problemlos durchgeführt werden, wenn die Testpersonen in der Lage sind, zielgerichtet zu zeigen. Bei den anderen Untertests ist zumindest mit einer erhöhten Beanspruchung der getesteten Kinder und Jugendlichen zu rechnen, bei schweren Sprechstörungen wird eine Durchführung nicht möglich sein. Zur Sicherung der Testfairness sind ggf. vermehrt Pausen einzuplanen, oder es ist auf die Durchführung der entsprechenden Untertests zu verzichten.

Da Instruktionen ausschließlich verbal gegeben werden, ist die CELF-5 nicht zur Testung von Kindern und Jugendlichen mit Hörbehinderungen geeignet.

Die Testfairness bei der Untersuchung von Kindern und Jugendlichen, die nicht mit Deutsch als Muttersprache aufgewachsen sind, ist nicht gesichert. Die Normstichprobe der CELF-5 hat zwar einen hohen Anteil an Personen mit Migrationshintergrund, es finden sich jedoch keinerlei Daten zu den Testergebnissen dieser Gruppe.

Im Durchführungsmanual wird die Frage von Testadaptationen vor allem im Hinblick auf kulturelle Vielfalt diskutiert, es finden sich auch in knapper Form allgemeine Hinweise zu möglichen Anpassungen bei der Untersuchung von Kindern und Jugendlichen mit motorischen, sensorischen und kognitiven Beeinträchtigungen.

### **CHC-theoretische Einordnung**

---

Die Cattell-Horn-Carroll Intelligenztheorie (Schneider & McGrew, 2018; deutschsprachige Darstellung Mickley & Renner, 2019) stellt im Sinne eines hierarchischen Intelligenzmodells einen Ordnungsrahmen für die Beschreibung kognitiver Fähigkeiten dar, bei dem die Allgemeinintelligenz (g-Faktor, Schicht-III) in 18 „breite“ (Schicht-II) und mehr als 90 „schmale“ Fähigkeitsbereiche (Schicht-I) untergliedert wird. Sie stellt damit auch eine einheitliche Terminologie für die inhaltliche Beschreibung von Testverfahren zur Verfügung. Auch wenn es sich bei der CELF-5 nicht um einen Intelligenztest handelt, können deren Untertests in das CHC-Modell eingeordnet werden und ggf. auch als ergänzende Verfahren im Rahmen eines Cross-battery-assessments (vgl. Flanagan, Ortiz & Alfonso, 2013) herangezogen werden. In Tabelle 4 sind die entsprechenden Zuordnungen der CELF-5 Untertests nach Ortiz, Flanagan und Alfonso (2019) zu finden.

**Tabelle 4***CHC-theoretische Einordnung der Untertests der CELF-5*

Untertest	CHC Schicht-II	CHC Schicht I
Satzverständnis	Verständnis & Wissen / Kristalline Intelligenz (Gc)	Fähigkeit zum Zuhören/ Sprachverständnis (LS)
Konzepte Erfassen	Verständnis & Wissen / Kristalline Intelligenz (Gc)	Kommunikative Fähigkeiten (CM)
Wortstruktur	Verständnis & Wissen / Kristalline Intelligenz (Gc)	Fähigkeit zum Zuhören/ Sprachverständnis (LS) Grammatikalische Sensitivität (MY)
Wortbedeutung	Verständnis & Wissen / Kristalline Intelligenz (Gc)	Lexikalisches Wissen (VL)
Anweisungen Befolgen	Verständnis & Wissen / Kristalline Intelligenz (Gc) Kurzzeitgedächtnis / Arbeitsgedächtnis (Gwm)	Fähigkeit zum Zuhören/ Sprachverständnis (LS) Arbeitsgedächtniskapazität (MW)
Sätze Formulieren	Verständnis & Wissen / Kristalline Intelligenz (Gc)	Kommunikative Fähigkeiten (CM)
Sätze Nachsprechen	Kurzzeitgedächtnis / Arbeitsgedächtnis (Gwm)	Gedächtnisspanne (MS)
Textverarbeitung	Verständnis & Wissen / Kristalline Intelligenz (Gc)	Fähigkeit zum Zuhören/ Sprachverständnis (LS)
Wortdefinitionen	Verständnis & Wissen / Kristalline Intelligenz (Gc)	Lexikalischen Wissen (VL)
Sätze Zusammensetzen	Verständnis & Wissen / Kristalline Intelligenz (Gc)	Grammatikalische Sensitivität (MY)
Semantische Beziehungen	Verständnis & Wissen / Kristalline Intelligenz (Gc)	Fähigkeit zum Zuhören/ Sprachverständnis (LS)
Pragmatisches Fähigkeitsprofil	Verständnis & Wissen / Kristalline Intelligenz (Gc)	Kommunikative Fähigkeiten (CM)

**Stärken**

- Erfassung vielfältiger sprachlicher Leistungen unter Berücksichtigung von Semantik, Syntax und auditiven Gedächtnisleistungen
- Auf die Fragestellung zugeschnittene modulare Anwendung möglich
- Die Überprüfung sprachlicher Leistungen in der Testsituationen kann mit der *Beobachtungsskala* und dem *Pragmatischen Fähigkeitsprofil* durch einen stärkeren Bezug zu alltäglichen kommunikativen Kompetenzen erweitert werden
- Fast durchgehend hohe Reliabilität der Skalen und mit wenigen Ausnahmen befriedigende bis sehr gute Reliabilitätswerte der Untertests
- Aktuelle Normierung
- Breiter Altersbereich
- Testmaterial für leistungsstärkere Kinder anregend
- Demonstrations- und Übungsaufgaben zur Sicherung des Aufgabenverständnisses



- Unzulässige Testmodifikationen, die eine Verwendung von Normdaten ausschließen, werden klar formuliert
- Vorschläge für „vertiefendes Testen“ geben Anregungen für weitere Beobachtungen und Untersuchungen
- Ausführliche Hinweise zur Bedeutung unterschiedlicher kultureller Hintergründe für die Testung
- Hinweise zum Zusammenhang zwischen Testinhalten und Unterrichtsanforderungen im Manual unterstützen bei der Testinterpretation.
- Optionale Itemanalyse zur Erkennung von Fehlermustern und Stärken innerhalb eines Untertests
- Die Gefahren, die mit einer Interpretation von Altersäquivalenten verbunden sind, werden im Manual deutlich benannt.

### **Einschränkungen / Schwächen**

---

- Die Skalenbildung wird theoretisch und empirisch nicht ausreichend untermauert. Die vorliegenden Daten sprechen gegen die faktorielle Validität der CELF-5.
- Bei den 6- bis 8-Jährigen werden zwei Skalen gebildet (*Allgemeine Sprachskala* und *Index für Sprachstruktur*), die aus exakt den gleichen Untertests bestehen. Trotzdem sollen sie „im Rahmen der Diagnostik einen unterschiedlichen Zweck erfüllen“ (Technisches Manual, S. 14) – eine höchst merkwürdige und nicht nachvollziehbare Behauptung.
- Die mehrfache Verwertung von Untertests zur Bildung der Indizes führt dazu, dass Indizes, die unterschiedliche Aspekte der Sprachentwicklung erfassen sollen, deutliche Überschneidungen aufweisen und sehr hoch miteinander korrelieren.
- Die Datenbasis zur konvergenten Validität ist sehr begrenzt.
- Es fehlen Angaben zu Sensitivität und Spezifität im Hinblick auf die kategoriale Diagnostik von Sprachentwicklungsstörungen
- Die Option der modularen Anwendung wurde bereits bei den Stärken der CELF-5 genannt. Einschränkend ist allerdings zu vermerken, dass es keine Daten zu der Frage gibt, ob die isolierte Vorgabe von Untertests im Vergleich zu einer vollständigen Testung die Ergebnisse beeinflussen kann.
- Bodeneffekte schränken den Einsatz bei Testpersonen im weit unterdurchschnittlichen Leistungsbereich ein.
- Der mehrfache altersabhängige Wechsel der Testzusammenstellung erschwert den längsschnittlichen Vergleich bei Wiederholungsmessungen
- Eher kleine Normstichprobe
- Es finden sich keine Daten zu den Leistungen von Kindern und Jugendlichen mit Migrationshintergrund.
- Eher knappe Hinweise zur Gestaltung der Testsituation bei Kindern und Jugendlichen mit sensorischen und motorischen Beeinträchtigungen, für diese Personengruppe fehlen klare Aussagen zu zulässigen und unzulässigen Adaptationen.
- Länge und Wortwahl einiger Instruktionen (*Sätze Formulieren* und *Textverarbeitung*) sind weniger geeignet für Kinder im unteren Leistungsbereich und/oder für Kinder mit beeinträchtigter Konzentration.
- Der doppelseitige Druck der Stimulusbücher lenkt beim Umblättern insbesondere jüngere Kinder ab („Da ist noch ein Bild, ich möchte es sehen“).

- Die Fallbeispiele im Manual thematisieren obligatorische differenzialdiagnostische Überlegungen nicht hinreichend. Im Fallbeispiel „Andreas“ (Manual, S. 190–193) könnten die Schwierigkeiten, aufmerksam zu bleiben, verbal auf Fragen zu reagieren und mehrschrittigen Aufgaben zu folgen u. a. Symptome einer kognitiven Entwicklungsstörung sein.

### **Sinnvolle Ergänzungen**

---

Je nach Ausgangslage und Fragestellung bieten sich folgende Ergänzungen an:

- Bei Verdacht auf kognitive Entwicklungsstörungen einschlägige Intelligenztestbatterien (KABC-II, WPPSI-IV, WISC-IV, RIAS, SON-R 2-8 u. a. m.), um spezifische Sprachstörungen von allgemeinen kognitiven Entwicklungsstörungen abgrenzen zu können.
- Bei Auffälligkeiten im Sprachgedächtnis ergänzende Überprüfungen zum Kurzzeit- und Arbeitsgedächtnis mit geringeren sprachlichen Anforderungen (z. B. Nachsprechen von Zahlenreihen aus der KABC-II oder der WISC-V, Subtest Wortreihe der KABC-II, Überprüfungen des visuellen Kurzzeitgedächtnisses)
- Überprüfungen aus dem Bereich der auditiven Verarbeitung (z. B. BAKO 1-4, TEPHOBE, H-LAD)
- Überprüfung schriftsprachlicher Kompetenzen durch entsprechende Schulleistungstests

### **Anforderungen an die Testleiterinnen & Testleiter**

---

Wie jeder psychometrische Test sollte die CELF-5 nur von Fachkräften eingesetzt werden, die über fundierte Kenntnisse in Testtheorie und Teststatistik verfügen. Das Manual nennt explizit Sprachtherapeut\*innen, Schulpsycholog\*innen, Sonderpädagog\*innen und andere Diagnostiker\*innen, „die in der Anwendung und Interpretation von standardisierten, sprachbezogenen Individualtestverfahren geschult und erfahren sind und zudem über fundierte Kenntnisse der deutschen Sprachstrukturregeln verfügen“ (Durchführungsmanual, S. 17). Warum in dieser Aufzählung klinische Psycholog\*innen und Neuropsycholog\*innen nicht genannt werden, bleibt ein Rätsel. Testanwender\*innen sollten auch über ein solides Grundwissen im Bereich der Sprachentwicklung verfügen.

Die Durchführung der CELF-5 kann als durchaus anspruchsvoll bezeichnet werden. Allgemeine Hinweise zur Gestaltung der Testsituation und die spezifischen Instruktionen müssen intensiv erarbeitet werden. Eine verantwortungsvolle Anwendung der CELF-5 erfordert ein sicheres und routiniertes Vorgehen, das nur durch mehrere Probetestungen und eine regelmäßige Testpraxis zu erreichen sein wird. Bei der Einarbeitung ist eine Kontrolle des eigenen Testleiterverhaltens durch kollegiales Feedback oder Videoaufnahmen sehr zu empfehlen. Testleiter\*innen müssen auch damit rechnen, dass schon nach wenigen Wochen ohne Testpraxis vermehrt Unsicherheiten und Durchführungsfehler auftreten können.

Auch die Auswertung der CELF-5 erweist sich als anspruchsvoll. Der Untertest *Sätze Nachsprechen* erfordert beispielweise das Erlernen und Einprägen von komplexen Bewertungs- und Sonderregeln; *Textverarbeitung* beinhaltet Fragen, die einen gewissen Interpretationsraum in der Bewertung offenlassen; die Bewertungsregeln von *Sätze Formulieren* umfassen im Durchführungsmanual mehr als sechs Seiten. Dementsprechend sind eine gründliche Einarbeitung sowie regelmäßige Übungen und Diskussionen schwieriger Bewertungen mit Kollegen\*innen zu empfehlen.

Testleiter\*innen „sollten mit dem Dialekt oder der Sprache vertraut sein, die bei der Testperson zu Hause und in ihrem Umfeld gesprochen wird“ (Durchführungsmanual, S. 39), da bei der Bewertung von Antworten beurteilt werden muss, ob diese unter Berücksichtigung des sprachlichen Hintergrunds der Testperson als korrekt gewertet können, auch wenn sie von der Standardsprache abweichen. Angesichts der potentiellen Vielfalt von sprachlichen Milieus ist dies ein sehr hoher Anspruch. Besondere Bedeutung haben Kenntnisse des sprachlichen und kulturellen Hintergrunds für die Beurteilung der Pragmatik. Das Durchführungsmanual (S. 42) fordert, dass Testanwender\*innen „Informationen über die Kultur einholen und das Verhalten mit dem anderer junger Menschen aus demselben regionalen, kulturellen und sprachlichen Umfeld vergleichen“. Außerdem sollten sie bei Eltern oder anderen Bezugspersonen in Erfahrung bringen, welche Verhaltensweisen im Umfeld der Testperson als angemessen gelten.

Zu den grundlegenden Kompetenzen beim Einsatz der CELF-5 gehört die Fähigkeit, die Untersuchungssituation kindgerecht zu gestalten und durchgehend ein freundliches, zugewandtes und flexibles Vorgehen zu realisieren. Insbesondere bei Kindern und Jugendlichen mit Entwicklungsstörungen müssen Testleiter\*innen deren besondere Bedürfnisse erkennen und sich angemessen darauf einstellen.

### **Zusammenfassende Bewertung von Paulina Cordero**

---

Die folgende Beurteilung bezieht sich auf die Anwendung der CELF-5 im Rahmen der sozialpsychiatrischen Versorgung von Kindern und Jugendlichen. Wesentliche Merkmale der CELF-5 sind die flexible und bedarfsorientierte Durchführung von Untertests sowie die Erhebung und Integration alltagsnaher Informationen über die Sprach- und Kommunikationsfähigkeiten der Testpersonen in unterschiedlichen Lebenskontexten. Drei Fremdeinschätzungsbögen sollen Diagnostiker\*innen unterstützen, sprachbezogene Stärken und Schwächen einzuschätzen und als Ausgangspunkt dienen, Interventionsplanung und Entwicklungsfortschritte zu beurteilen.

Es handelt sich um kein zeitökonomisches Verfahren, eine vollständige Durchführung mit der CELF-5 nimmt für Untersucher\*innen und Kinder mehrere Termine in Anspruch. Die Möglichkeit der CELF-5, nur diejenigen Untertests durchzuführen, die für die Beantwortung einer gezielten diagnostischen Fragestellung notwendig sind, kommt solchen Testpersonen entgegen, die nicht länger als 35 bis 40 Minuten konzentriert arbeiten können. Zu einer sinnvollen Auswahl weiterer Untertests, die über die Bestimmung der *Allgemeinen Sprachskala*, des *Index für Sprachverständnis* und des *Index für Sprachproduktion* hinausgehen, macht das Manual keine nachvollziehbaren Vorschläge. Die Durchführung aller Untertests für eine Altersgruppe während *eines* Termins erfordert eine deutlich höhere Konzentrationsleistung der Testpersonen und ist aus meiner Sicht nur bei leistungsstarken Kindern ab 9 Jahren zu befürworten.

Für die Altersgruppe 9 bis 16 Jahre erweist sich die CELF-5 als motivierendes und ansprechendes Untersuchungsverfahren. Instruktionen und Durchführungshinweise sind größtenteils praxistauglich. Bei sehr jungen Kindern, solchen mit beeinträchtigter Konzentration, mit gravierenden Sprachverständnisproblemen und bei Kindern mit kognitiven Einschränkungen stellen die teilweise komplexen und langen Testinstruktionen einiger Untertests (z. B. bei *Sätze Formulieren*, *Textverarbeitung*) eine Überforderung dar. Eine weitere Einschränkung ergibt sich durch den Verzicht auf eine Abbruchregel bei *Wortstruktur* und *Textverarbeitung*; leistungsschwache Testpersonen erleben hier zu viele Misserfolge. Die Aufgaben der CELF-5 werden in dieser Gruppe als wenig abwechslungsreich erlebt. Aktivere Handlungsmöglichkeiten durch vielfältigere Materialien wie

Bildkarten und Spielfiguren würden die Motivation der Kinder unterstützen. Durch die fehlende positive Rückmeldung der Testleiter\*innen bei richtigen Antworten der Übungsaufgaben kann es bei jüngeren Kindern und selbstunsicheren Probanden zu Motivationsverlust und negativen Leistungserwartungen kommen.

Die im Testmaterial benutzten Abbildungen sind meistens gut zu erkennen, einzelne Abbildungen des Untertests *Satzverständnis* waren für manche Kinder in der Darstellung unklar und sorgten für Nachfragen. Die Benutzung des beidseitig bedruckten Aufgabenordners kann jüngere Kinder, die beim Umblättern die Abbildungen der Rückseiten sehen, ablenken.

Durchführung und Auswertung der CELF-5 stellten sehr hohe Anforderungen an mich als Testleiterin, die erst nach einer intensiven Einarbeitung und mehreren Anwendungen erfüllt werden konnten. Kinder mit schweren Artikulationsproblemen stellten mich – trotz ausführlicher Auswertungshinweise – bei der Auswertung der Untertests zum *Index für Sprachproduktion* vor schwierige Entscheidungen. Einen Überblick über Lautfehlbildungen bietet die CELF-5 nicht.

Die Möglichkeit, im Rahmen einer qualitativen Itemanalyse Fehler in ihrem inhaltlichen Muster näher zu analysieren und anschließend eine nicht standardisierte, vertiefende Untersuchung durchzuführen, begrüße ich sehr. Hierzu vermittelt das Manual hilfreiche Anregungen. Hierbei ist jedoch zu berücksichtigen, dass schlechte Leistungen bei schwierigen Aufgaben einzelner Untertests auch durch andere Faktoren bedingt sein können, als es die qualitative Itemanalyse vorschlägt (z. B. Untertest *Anweisungen Befolgen*).

Im Manual der CELF-5 fehlt mir eine klare sprachtheoretische Fundierung, die die teilweise mehrmalige Zuordnung der gleichen Untertests zu verschiedenen übergeordneten Indizes nachvollziehbar macht. Die fehlende theoretische Fundierung der CELF-5 hat zur Folge, dass die Interpretation der Indizes und damit auch die differenzierte Vermittlung der Testergebnisse an Eltern erschwert wird. Mir fehlt eine nachvollziehbare Erklärung für die identische Untertestzusammensetzung der *Allgemeinen Sprachskala* und des *Index für Sprachstruktur* in der Altersgruppe 6 bis 8 Jahre.

Die Einschätzungen der Spontansprache eines Kindes mittels der *Beobachtungsskala* sind aus meiner Sicht eine hilfreiche Ergänzung, sie sind aber unspezifisch. Ihre Ergebnisse müssen im Kontext weiterer anamnestischer Daten und ggf. weiterer Testergebnisse eingeordnet werden. Dann können die Items der Beobachtungsskala einen guten Beitrag zur differenzialdiagnostischen Einordnung der Symptomatik leisten, denn viele der aufgelisteten sprachlichen Schwierigkeiten können bei mehreren entwicklungsbedingten Störungsbildern auftreten.

Das *Pragmatische Fähigkeitsprofil* stellt Testleiter\*innen vor die Anforderung, das Umfeld und die damit verbundenen kulturellen Einflüsse der Testperson gut zu kennen, um kulturelle sowie weitere Aspekte (sozioökonomischer Status, Alter, Geschlecht, Bildungsniveau usw.) in der Beurteilung berücksichtigen zu können. Dieser Anforderung kann bei vielen Patienten, die zu einer Sprachdiagnostik kommen, sowohl aus zeitlichen als auch aus anderen Gründen (z. B. mangelnde Deutschkenntnisse der Eltern) nicht erfüllt werden.

Das Testverfahren setzt bei mehrsprachig aufwachsenden Kindern einiges an Sprachbeherrschung im Deutschen voraus. Die Anwendung bei Kindern mit mangelnden Kenntnissen der deutschen Sprache erfordert eine kompetente Einschätzung und

Abdeckung ihres Unterstützungsbedarfs während der Testung (z. B. Umformulierung von Testinstruktionen, die allerdings eine standardisierte Auswertung unmöglich machen). Die Diagnose und Interventionsplanung bei Kindern mit nicht-deutscher Muttersprache macht die Unterscheidung zwischen einer Sprachverzögerung, einer Sprachstörung und/oder mangelnden Deutschkenntnissen unabdingbar. Hierzu äußert sich das Manual nicht.

Untersuchungen der Sprachentwicklung werden in meinem beruflichen Alltag größtenteils im Vorschulalter veranlasst. Die Diagnostik und Betreuung hat das Ziel, Rückstände möglichst bis zum Schulstart weitgehend zu beseitigen. Bei Jugendlichen stellt sich für mich die Frage einer Überprüfung der Sprachentwicklung eher selten.

Eine Therapieverlaufskontrolle durch einen Längsschnittvergleich der Testwerte wird durch die in den verschiedenen Altersgruppen wechselnde Testzusammenstellung erschwert.

Insgesamt fehlen mir hinreichende Belege, dass die CELF-5 tatsächlich zu Entscheidungen über Therapie- und Förderbedarfe beitragen und der Interventionsplanung dienen kann.

### **Zusammenfassende Bewertung von Gerolf Renner**

---

Meine Bewertung der CELF-5 fällt zwiespältig aus. Es handelt sich sicherlich um ein Verfahren, das vielfältige Einblicke in den Stand der sprachlichen Entwicklung bei Kindern und Jugendlichen erlaubt. Eher selten wird man einen Test finden, dessen Gesamtkonzeption nicht nur auf die reliable Erfassung der relevanten Konstrukte abzielt, sondern der auch Befragungen und gezielte Beobachtungen integriert. Die Hinweise zum „vertiefenden Testen“ und die Anregungen zur Itemanalyse zeigen den Anwender\*innen weitere Möglichkeiten auf, diagnostisch bedeutsame Informationen zu sammeln, die über die quantitativen Testwerte hinausgehen. Diese Herangehensweise kommt Diagnostiker\*innen entgegen, die einerseits Wert auf objektivierbare Daten legen, andererseits aber auch die Notwendigkeit einer individualisierten und differenzierten Betrachtung der Fähigkeiten von Kindern und Jugendlichen sehen. Positiv hervorzuheben ist auch, dass zahlreiche Demonstrations- und Übungsaufgaben zur Sicherung des Instruktionsverständnisses beitragen.

Die Teststruktur der CELF-5 legt den Schwerpunkt auf komplexere sprachliche Leistungen. Damit ist die CELF-5 deutlich breiter aufgestellt als einfache Wortschatztests. Damit verbunden ist der Nachteil, dass – wie auch die Bodeneffekte belegen – die Testung für leistungsschwache Testpersonen sehr anspruchsvoll sein kann. In der sonderpädagogischen Diagnostik kann sich das in fast allen Förderschwerpunkten als Nachteil erweisen. Zumindest für die jüngste Altersgruppe und für Testpersonen mit kognitiven Beeinträchtigungen würden übliche Testformate zum aktiven und passiven Wortschatz und ein spielerischeres Herangehen an das Sprachverständnis die CELF-5 bereichern. Die CELF-5 beansprucht die Aufmerksamkeit, insbesondere die auditive Aufmerksamkeit, der Testpersonen und bietet wenig aktive Handlungsmöglichkeiten.

Die CELF-5 deckt einen weiten Altersbereich ab. Sprachdiagnostische Testverfahren beschränken sich meist auf das Vorschul- und Grundschulalter. Die CELF-5 erlaubt auch die Untersuchung von Jugendlichen. Die wechselnde Testzusammenstellung und Skalenzuordnung erschweren allerdings den längsschnittlichen Vergleich von Testergebnissen.

Für die klinisch-psychologische und sonderpädagogische Diagnostik bei Kindern und Jugendlichen mit Behinderungen sind die im Abschnitt „Zugangsfertigkeiten“ genannten Hinweise zu beachten, wenn der Einsatz der CELF-5 in Erwägung gezogen wird.

Validität ist *das* zentrale Gütekriterium eines Testverfahrens. In den Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014) wird gefordert, dass für jeden intendierten Anwendungsbereich eines Tests Validitätsbelege vorgelegt werden. Das ist bei der CELF-5 erst ansatzweise der Fall (s. Abschnitt Validität). Es klafft eine deutliche Lücke zwischen dem Anspruch, Sprachstörungen differenziert zu diagnostizieren, Förder- und Therapiebedarfe feststellen zu können und Therapien und Interventionen zu planen, und den empirischen Belegen, dass dies mit der CELF-5 tatsächlich gelingen kann. Einige Darstellungen im Manual sind zudem schlecht nachvollziehbar (z. B. Faktorenstruktur), es fehlen wichtige Kennwerte (Sensitivität und Spezifität bei kategorialer Diagnostik), zu verschiedenen Validitätsaspekten liegen keine Daten vor.

Leider werden auch keine Informationen zur Testleistungen von Kindern und Jugendlichen mit Migrationshintergrund bzw. Deutsch als Zweitsprache bereitgestellt. Dies ist besonders verwunderlich, da im Manual die Beachtung des sprachlichen und kulturellen Hintergrunds mehrfach eingefordert wird. Die CELF-5 liefern damit auch keine Hinweise zur Abgrenzung von Sprachstörungen und sprachlichen Besonderheiten, die auf unterschiedliche kulturelle oder sprachliche Erfahrungen und individuelle Lernmöglichkeiten zurückgehen.

Der modulare Aufbau der CELF-5 ermöglicht eine beträchtliche Flexibilität bei der Zusammenstellung der Untertests. Es bleibt allerdings unklar, ob dies Auswirkungen auf die Leistungen der Testpersonen haben kann, deren Aufmerksamkeit mehr oder weniger lang und intensiv beansprucht wird.

Die Berücksichtigung der Pragmatik ist ein besonderes Merkmal der CELF-5. In diesem Bereich fehlen allerdings Belege für die Validität. Sowohl das *Pragmatische Fähigkeitsprofil* als auch die *Aktivitätencheckliste zur Pragmatik* beruhen auf subjektiven Bewertungen der Anwender\*innen, die eine umfassende Berücksichtigung des kulturellen und sprachlichen Hintergrunds erfordern. Hier stellt sich auch die in den Manualen überhaupt nicht thematisierte Frage, ob und unter welchen Bedingungen es sinnvoll ist, Pragmatik ausschließlich auf dem Hintergrund des individuellen Umfelds zu beurteilen. Die Beherrschung pragmatischer Regeln der vorherrschenden Kultur ist ja nicht bedeutungslos für Schule, Ausbildung und andere Lebensbereiche.

Die Struktur der CELF-5 erscheint mir nachvollziehbar, wenn es um die *Allgemeine Sprachskala* und die Unterteilung in den *Index für Sprachverständnis* und den *Index für Sprachproduktion* geht. Diese Indizes sind inhaltlich durchaus abgrenzbar, korrelieren allerdings sehr hoch (.76). Neben der Absurdität der Identität der *Allgemeinen Sprachskala* und des *Index für Sprachstruktur* bei den 6- bis 8-Jährigen lässt die weitere Testgliederung erhebliche Überschneidungen der Indizes erkennen, die teilweise mit extrem hohen Interkorrelationen eingehen und daher keine inhaltliche Spezifität beanspruchen können. Beim *Index für Sprachgedächtnis* ist die inhaltliche Beschreibung (ein „Messwert für die Befähigung der Testperson [...], Gedächtnisleistungen im Kontext sprachbezogener Aufgaben anzuwenden“; Durchführungsmanual, S. 171) gedächtnispsychologisch nicht nachvollziehbar. Hier wäre eine getrennte Erfassung von Kurzzeit-/Arbeitsgedächtnis und Langzeitgedächtnis sinnvoll. Die Interpretation dieses Index fällt

auch wegen der Konfundierung mit Sprachverständnis und Satzbildung bei zwei der drei zugehörigen Untertests schwer. Diese Konfundierung betrifft auch den *Index für Sprachproduktion*, in den Untertests einfließen, die in erster Linie die phonologische Schleife ansprechen.

An einigen Stellen in den Manualen wird erwähnt, dass die Ergebnisse einer Testung mit der CELF-5 in Verbindung mit weiteren diagnostischen Datenquellen zu bewerten sind. Allerdings wird dieser für den diagnostischen Prozess hoch bedeutsame Hinweis m. E. im Manual nicht ausreichend deutlich gemacht. Unterdurchschnittliche Testleistungen in der CELF-5 müssen *immer* zu differentialdiagnostischen Überlegungen anregen, was in den Fallbeispielen nicht konsequent berücksichtigt wird.

Bei einer aktuellen Literaturrecherche hat sich gezeigt, dass es noch wenig Forschung zur CELF-5 gibt, die in der amerikanischen Version 2013 erschienen ist. Unabhängige Untersuchungen der Testqualität liegen nicht vor, es finden sich kaum klinische Studien, in denen die CELF-5 eingesetzt wurde. So sind Testanwender\*innen vor allem auf die im Manual bereitgestellten Daten angewiesen, die doch etliche Lücken aufweisen.

### Aktuelle Entwicklungen

---

Wiig und Secord (2014) veröffentlichten in den USA eine Erweiterung der CELF-5, die CELF-5 Metalinguistics, die weitere Untertests zur Erfassung komplexerer sprachlicher Leistungen bereitstellt (figurative Sprache, Konversationsfähigkeit, logische Inferenzen, multiple Bedeutungen). Ob eine deutschsprachige Adaptation geplant ist, ist den Verfasser\*innen nicht bekannt.

### Testrezensionen

---

Aktuell liegen nur Rezensionen der amerikanischen Fassung vor:

Coret, M. C. & McCrimmon, A. W. (2015). Test Review: Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals—Fifth Edition (CELF-5)*. *Journal of Psychoeducational Assessment*, 33(5), 495–500.  
<https://doi.org/10.1177/0734282914557616>

LEADERSproject (2014). *Test Review: CELF-5*. Teachers College, Columbia University. Zugriff am 13.08.2021. Verfügbar unter:  
<https://www.leadersproject.org/2014/02/17/test-review-celf-5/>

### Zitierte Literatur

---

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington: American Educational Research Association.
- Charter, R. A. (2003). Study samples are too small to produce sufficiently precise reliability coefficients. *Journal of General Psychology*, 130(2), 117–129.  
<https://doi.org/10.1080/00221300309601280>
- 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.
- Flanagan, D. P., Ortiz, S. O. & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3<sup>rd</sup> ed.). Hoboken, NJ: Wiley.
- 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>
- Ortiz, S. O., Flanagan, D. P. & Alfonso, V. C. (2019). *Cross-Battery Assessment Software System 2.4 (X-BASS 2.4)*. Hoboken, NJ: Wiley.
- Renner, G. (2017). Chancen und Fallstricke der Intelligenzdiagnostik bei kognitiven Entwicklungsstörungen. In V. Mall, F. Voigt & N. Jung (Hrsg.), *Entwicklungsstörungen und chronische Erkrankungen. Diagnose, Behandlungsplanung und Familienbegleitung* (Aktuelle Fragen der Sozialpädiatrie, Bd. 2, S. 76–91). Lübeck: Schmidt-Römhild.
- 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* (4<sup>th</sup> ed., pp. 73–163). New York: The Guilford Press.
- Wiig, E. H. & Secord, W. A. (2014). *CELF5 Clinical Evaluation of Language Fundamentals - Fifth Edition. Metalinguistics*. Pearson.

### **Ausgewählte weiterführende Literatur**

---

- Denman, D., Speyer, R., Munro, N., Pearce, W. M., Chen, Y.-W. & Cordier, R. (2017). Psychometric properties of language assessments for children aged 4-12 years: A systematic review. *Frontiers in Psychology*, 8, 1515.  
<https://doi.org/10.3389/fpsyg.2017.01515>

### **Internet**

---

- Informationsflyer zur [CELF-5 Metalinguistics](#)
- Videoeinführung in die amerikanische Version der CELF-5:  
<https://www.youtube.com/watch?v=SMB3geAkPLs>
- Video zur Interpretation der amerikanischen CELF-5:  
<https://www.youtube.com/watch?v=NBZsqOPX4CI>