

Renner, G. & Scholz, M. (2022). Fair oder nicht fair, das ist hier die Frage!. Die Sicherung der Testfairness als Aufgabe der sonderpädagogischen Diagnostik. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 259-274). Regensburg: Universitätsbibliothek. <https://doi.org/10.5283/epub.53149>

Fair oder nicht fair, das ist hier die Frage!

Die Sicherung der Testfairness als Aufgabe der sonderpädagogischen Diagnostik

Gerolf Renner & Markus Scholz

1 Einleitung

»Niemand darf wegen seiner Behinderung benachteiligt werden« (Art. 3 Abs. 3 Grundgesetz). Aber gilt das auch für Kinder und Jugendliche mit Behinderungen, die mit standardisierten testdiagnostischen Verfahren untersucht werden?

Standardisierte Testverfahren werden in der sonderpädagogischen Diagnostik eingesetzt, um Eigenschaften oder Fähigkeiten zu messen, die für die Klärung sonderpädagogischer diagnostischer Fragestellungen relevant sind. Im Gegensatz zu subjektiven und möglicherweise vorurteilsbehafteten Einschätzungen führen wissenschaftlich fundierte und validierte Tests bei fachgerechter Anwendung zu exakteren Ergebnissen. Um diesen Anspruch zu erfüllen, wird von Testverfahren erwartet, dass sie bestimmte Gütekriterien erfüllen (siehe Kapitel »Diagnostische Gütekriterien bei Statustests« in diesem Band).

Eines der drei Hauptgütekriterien ist die Objektivität. Das Testergebnis soll nicht davon beeinflusst werden, welche Person einen Test durchführt, auswertet und interpretiert. Dafür ist die Standardisierung der Testvorgabe und des Testmaterials eine notwendige Voraussetzung. Standardisierung soll verhindern, dass Testergebnisse von willkürlichen und möglicherweise diskriminierenden Vorgehensweisen der Testleiter*innen und nicht von den Fähigkeiten der untersuchten Personen bestimmt werden. Sie hat jedoch einen Preis: Sie kann zu ernsthaften Gefährdungen der Testfairness und damit zu invaliden Testergebnissen führen, wenn das in den Instruktionen vorgesehene Vorgehen Menschen mit Behinderung daran hindert, ihre Fähigkeiten zu zeigen. Die überwiegende Anzahl psychometrischer Leistungstests berücksichtigt diese Thematik jedoch nicht (vgl. Renner & Mickley, 2015a).

Im Folgenden soll nach einer Explikation des Begriffs Testfairness gezeigt werden, (a) warum und wie bei gängigen Testverfahren eine faire Untersuchung von Kindern und Jugendlichen mit Behinderungen gefährdet sein kann und (b) wie in der Praxis der sonderpädagogischen Diagnostik Gefährdungen der Testfairness vermieden oder zumindest reduziert werden können.

Wir beziehen uns dabei in erster Linie auf die Testung von Kindern mit Beeinträchtigungen der körperlichen und motorischen Entwicklung, mit Sinnesbeeinträchtigungen und mit deutlichen kognitiven Beeinträchtigungen.

2 Testfairness: Grundlagen

Nach Moosbrugger und Kelava (2012, S. 24) erfüllt ein Test »das Gütekriterium der Fairness, wenn die resultierenden Testwerte zu keiner systematischen Benachteiligung bestimmter Personen aufgrund ihrer Zugehörigkeit zu ethnischen, soziokulturellen oder geschlechtsspezifischen Gruppen führen«. Zu den Zielgruppen einer fairnessorientierten Diagnostik sind natürlich auch Menschen mit Behinderungen zu zählen (vgl. American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014). Diese Definition impliziert, dass Testwerte ausschließlich von den Fähigkeiten bestimmt werden sollen, auf deren Erfassung der Test abzielt. Personen, die gleiche Fähigkeiten haben, sollen gleiche Testergebnisse erzielen können, und wenn Personen unterschiedliche Testergebnisse erzielen, soll dies nur ihre Fähigkeitsunterschiede und nicht andere Einflussfaktoren widerspiegeln. Diese Definition impliziert weiter, dass Testfairness immer in Bezug auf bestimmte Personengruppen betrachtet werden muss. Ein Test ist nicht grundsätzlich fair oder unfair, sondern eine konkrete Gefährdung der Fairness entsteht aus einer Interaktion zwischen den Eigenschaften des Testverfahrens und den Merkmalen bestimmter Personen.

Testfairness wird allgemein als Nebengütekriterium bezeichnet und in vielen (sonderpädagogischen) diagnostischen Lehrbüchern nur am Rande oder gar nicht behandelt (z. B. Bühner, 2011; Bundschuh & Winkler, 2014; Tröster, 2019; Schmidt-Atzert & Amelang, 2012). Wie die »Standards for Educational and Psychological Testing« (American Educational Research Association et al., 2014) betonen, handelt es sich jedoch um einen fundamentalen Aspekt der Entwicklung und Anwendung von Testverfahren. Bei der Untersuchung von Kindern und Jugendlichen mit Behinderungen muss die Testfairness als gleichwertig mit den klassischen Hauptgütekriterien Objektivität, Reliabilität und Validität angesehen werden (vgl. American Psychological Association, 2012).

2.1 Konstruktrelevanz

Testfairness ist ein fundamentales Thema bei der Beurteilung der Validität. Die Validität eines Tests gilt als wichtigstes Gütekriterium (Schermelleh-Engel, Kelava & Moosbrugger, 2006). Sie bezieht sich auf »das Ausmaß, in dem empirische Befunde und theoretische Argumente die Interpretationen von Testwerten für die beabsichtigten Verwendungen von Tests unterstützen« (American Educational Research Association et al., 2014, S. 11; Übersetzung von Hartig, Frey & Jude, 2020). Validität ist in diesem Sinn keine feststehende Eigenschaft eines Tests, sondern muss für jeden Anwendungszweck des Tests nachgewiesen werden. So folgt aus der Tatsache, dass die Eignung eines Tests zur Messung der aktuellen Leistungsfähigkeit empirisch gut gestützt wird, keineswegs, dass auf Basis des Testwerts zuverlässige Prognosen oder Selektionsentscheidungen getroffen werden können. Ob solche weitergehenden Interpretationen gerechtfertigt sind, muss erst in zusätzlichen Validitätsstudien überprüft werden.

Ein grundlegender Aspekt bei der Beurteilung der Validität ist die **Konstruktrelevanz**. Eine Testung weist dann eine hohe Konstruktrelevanz auf, wenn das Testergebnis weitgehend von dem

Konstrukt bestimmt wird, das der Test erfassen soll. Die Testergebnisse können aber auch von Faktoren beeinflusst werden, die mit dem eigentlichen Konstrukt nichts zu tun haben. Betreffen solche *konstruktirrelevanten Einflüsse* systematisch bestimmte Personengruppen, kann nicht von einer fairen Testung gesprochen werden (vgl. Messick, 2000).

Beispiel: Im Lese und Rechtschreibtest SLRT-II (Moll & Landerl, 2014) sollen innerhalb von 60 Sekunden so viele Wörter wie möglich laut gelesen werden. Ein Kind mit einer Dysarthrie, einer bei Kindern mit cerebralen Bewegungsstörungen häufigen Sprechstörung, wird in diesem Test auch dann keine guten Leistungen erbringen können, wenn es sicher lesen kann. Nicht das Konstrukt Lesefähigkeit bestimmt die Testergebnisse, sondern die konstruktirrelevante Artikulationsgeschwindigkeit.

Für die sonderpädagogische und klinisch-psychologische Diagnostik bedeutet dies, dass Diagnostiker*innen für die Beurteilung der Validität und der Testfairness bei jeder individuellen Testung verantwortlich sind. Die pauschale Aussage von Testautor*innen oder Testverlagen, dass ein Test fair und valide sei, kann in keiner Weise von dieser Aufgabe entbinden (vgl. auch American Educational Research Association et al., 2014). Essentiell für diese Beurteilung ist die Differenzierung von Ziel- und Zugangsfertigkeiten (vgl. Braden, 2003).

2.2 Ziel- und Zugangsfertigkeiten

Zielfertigkeiten (target skills) sind diejenigen Fertigkeiten, die der Test eigentlich erfassen will. Der Bezug zum Begriff Konstruktrelevanz ist evident: Nur wenn die Zielfertigkeiten im Testergebnis abgebildet werden, ist eine konstruktrelevante Testung gegeben.

Tabelle 1: Wichtige Zugangsfertigkeiten bei testdiagnostischen Untersuchungen

- Hörfähigkeit (z. B. Hörschwelle, Ausfälle in bestimmten Frequenzbereichen)
- Sehfähigkeit (z. B. Visus, Farbsehen, Kontrastwahrnehmung, Gesichtsfeld)
- Wahrnehmungsverarbeitung (z. B. Figur-Grund-Unterscheidung, Lautdifferenzierung)
- Sprachverständnis (z. B. Wortverständnis, Satzverständnis)
- Sprechfähigkeiten (z. B. Dysarthrie/Anarthrie, Artikulationsstörung, Stottern)
- Lesen
- Schreiben
- Motorik (z. B. Zeigen, Greifen, Aufnehmen, Grafomotorik, Kopfkontrolle, Rumpfkontrolle, persistierende Reflexe)
- Persönliche Bedürfnisse ausdrücken können (z. B. Hunger, Durst, Schmerz, Toilettengang, Bedürfnis nach Pause)
- Nachfragen können, wenn etwas nicht verstanden wurde
- Aufmerksamkeit (z. B. selektive Aufmerksamkeit, geteilte Aufmerksamkeit, Daueraufmerksamkeit)
- Selbstorganisation / Selbststeuerung (v. a. bei Arbeitsphasen ohne externe Taktung)
- Körperliche Belastbarkeit
- Gedächtnis (z. B. kurz- und langfristiges Behalten von Instruktionen)
- Allgemeines und kulturspezifisches Wissen

Als *Zugangsfertigkeiten* (access skills) werden dagegen alle Fertigkeiten bezeichnet, die vorhanden sein müssen, um eine Testaufgabe überhaupt erfassen und bewältigen zu können. Für sich genommen erlauben diese aber keinen Rückschluss auf das erfasste Konstrukt. Was Zugangs- und was Zielfertigkeiten sind, lässt sich nur in Abhängigkeit des jeweils untersuchten Konstrukts, also erst im konkreten Fall, beurteilen. So ist die Fähigkeit zu Lesen in einem Lesetestverfahren eine Zielfertigkeit, bei einer Testung mathematischer Fähigkeiten, bei der Textaufgaben schriftlich vorgegeben werden, dagegen eine Zugangsfertigkeit. Die Testfairness ist gefährdet, wenn ein Verfahren Zugangsfertigkeiten voraussetzt, über die ein getestetes Kind nicht oder nur eingeschränkt verfügt. Das Testergebnis wird dann nur teilweise oder gar nicht von den Fähigkeiten der Testperson bestimmt.

Beispiel: Menschen mit schwersten körperlichen Beeinträchtigungen und fehlender Lautsprache können über ausgesprochen gute kognitive Fähigkeiten verfügen und akademische Berufe ergreifen. Ihnen fehlen jedoch wesentliche Zugangsfertigkeiten, die in kognitiven Leistungstests mehr oder weniger vorausgesetzt werden. Eine standardisierte Diagnostik, die das Gütekriterium der Fairness nicht beachtet, wird hier – völlig unabhängig von den Fähigkeiten der getesteten Personen – zu dem Ergebnis kommen, dass eine Intelligenzminderung vorliegt. Solche fatalen Unterschätzungen der Fähigkeiten können in der Folge dazu führen, dass angemessene Bildungsangebote vorenthalten werden.

Fehlende oder eingeschränkte Zugangsfertigkeiten können unterschiedliche Auswirkungen haben:

- Die Durchführung der Aufgabe ist grundsätzlich nicht möglich (z. B. Vorgabe von visuellem Testmaterial bei blinden Personen). In diesem Fall ist die mangelnde Eignung des Verfahrens leicht zu erkennen.
- Die Durchführung der Aufgabe ist prinzipiell möglich, aber im Vergleich zu nicht beeinträchtigten Testpersonen erschwert. Die mangelnde Eignung eines Testverfahrens ist dann weniger offenkundig, insbesondere wenn Testleiter*innen wenig Erfahrung in der Untersuchung von Kindern mit Behinderung haben. Die Erschwernis kann sich auf folgende Aspekte beziehen:
 - Die Ausführung der Aufgabe, z. B. wenn motorische Beeinträchtigungen das passgenaue Anlegen von Mosaiksteinen oder Puzzleteilen erschweren oder wenn eine Sprechstörung eine verbale Antwort erschwert.
 - Das Tempo der Aufgabenbearbeitung. Beeinträchtigte Zugangsfertigkeiten führen fast immer dazu, dass Aufgaben nur verlangsamt bearbeitet werden können. Kinder mit einer Dysarthrie müssen z. B. bei Nachsprechaufgaben die dargebotenen Inhalte länger im Arbeitsgedächtnis halten als andere Kinder, da sie die Antwort nur verlangsamt geben können.
 - Erhöhte Anstrengung. Bei einer Sprechstörung strengt Sprechen mehr an, bei einer Sehstörung strengt die visuelle Verarbeitung von kleinen, kontrastarmen und sehr detailreichen Abbildungen mehr an, bei einer motorischen Behinderung kann schon die Einhaltung der Sitzposition Anstrengung kosten, Kinder mit Hörstörungen werden schneller ermüden, wenn sie auditive Stimuli verarbeiten müssen. Die erhöhte Anstrengung wird Ausdauer, Aufmerksamkeit und schließlich auch Motivation negativ beeinflussen.

Zugangsfertigkeiten lassen sich in allgemeine und spezifische Fertigkeiten differenzieren. Allgemeine Zugangsfertigkeiten sind Kompetenzen, die von der Testperson benötigt werden, um überhaupt an einer Testsituation teilnehmen zu können. Hierzu gehören Aufmerksamkeit, Aus-

dauer usw. Spezifische Zugangsfertigkeiten ergeben sich aus der Anforderungsstruktur der Testaufgabe und der Gestaltung des Testmaterials. Sie werden daher für einzelne Items oder bestimmte Untertests beurteilt. Sehfähigkeit, Handhabung eines Stiftes, das Greifen oder Zeigen von Gegenständen oder Abbildungen wären exemplarisch darunter zu fassen (s. Tabelle 1). Der Begriff Zugangsfertigkeiten fokussiert, inwieweit die Testpersonen über die für die Testung erforderlichen Fertigkeiten verfügen. Der Begriff Zugänglichkeit (accessibility; Kettler, Elliott, Beddow & Kurz, 2018) lenkt das Augenmerk auf die Frage, ob ein Test möglichst vielen Testpersonen erlaubt, ihre Fähigkeiten ohne Erschwernisse durch konstruktirrelevante Zugangsvoraussetzungen zu zeigen. Tests mit geringer Zugänglichkeit zeichnen sich etwa durch komplexe sprachliche Instruktionen, schlecht erkennbares visuelles Material (z. B. kleine Abbildungen, schwache Kontraste) oder zusätzliche motorische Anforderungen aus. Gefährdungen der Testfairness werden reduziert, wenn schon in der Phase der Testentwicklung auf einen möglichst einfachen Zugang zur Erfassung der Zielfertigkeiten geachtet wird.

Beispiel: Zur Überprüfung des Sprachverständnisses werden unterschiedliche Aufgabenformate eingesetzt. Eine Variante besteht darin, dass Kinder bestimmte Aufforderungen mit Materialien umsetzen müssen (z. B. Untertest *Sprache Rezeptiv* der Intelligence und Development Scales – Preschool IDS-P; Grob, Reimann, Gut & Frischknecht, 2013). Gerade bei jüngeren Kindern können sich solche aktiven Handlungsmöglichkeiten durchaus motivierend auswirken. Sie führen allerdings auch eine zusätzliche Anforderungskomponente ein, nämlich die Fähigkeit, mit Objekten gezielt zu hantieren. Bei einer motorischen Beeinträchtigung kann dann auch bei vorhandenem Sprachverständnis die intendierte Reaktion nicht so realisiert werden, dass sie für Testleiter*innen als korrekte Antwort erkennbar ist.

3 Beurteilung und Sicherung der Testfairness

Die Beurteilung und Sicherung der Testfairness ist nur möglich, wenn Diagnostiker*innen eine Vorstellung von den Fähigkeiten haben, die sie mittels eines Testverfahrens erfassen wollen, und wenn sie die mit einer Testung verbundenen Ziel- und Zugangsfertigkeiten analysieren und mit den individuellen Voraussetzungen des untersuchten Kindes abgleichen. Da Testmanuale in der Regel Ziel- und Zugangsfertigkeiten nicht systematisch darstellen und kaum auf Gefährdungen der Testfairness hinweisen, sollte das folgende Vorgehen zur Routine der Untersuchungsplanung in der sonderpädagogischen Testdiagnostik werden. Die Schritte sind dabei nicht zwingend in der aufgeführten Reihenfolge zu bearbeiten. Einzelne Überlegungen lassen sich parallel anstellen oder sind im Einzelfall auch in einer abweichenden Abfolge sinnvoll. Zentral ist, dass sich Testleiter*innen über alle Schritte Gedanken machen.

3.1 Schritt 1: Welche Konstrukte sind für die Klärung der Fragestellung von Bedeutung?

Die Untersuchungsplanung beginnt mit der Frage, welche Konstrukte für die diagnostische Fragestellung überhaupt relevant sind. Dieser Schritt ist bei allen diagnostischen Prozessen bedeutsam, nicht nur bei der Untersuchung von Kindern und Jugendlichen mit Behinderungen. Erforderlich ist ein gutes Verständnis des jeweiligen Konstruktes und seiner Bedeutung für die Fragestellung, möglichst unter Bezug auf die wissenschaftliche Theoriebildung. Beispielsweise ist es nicht sinnvoll, pauschal die Erfassung der Lesekompetenz zu planen, ohne vorher zu überlegen, welche Aspekte (Buchstabenkenntnis, Erlesen erster Wörter, Satzverständnis oder

Verständnis komplexer Texte) im Fokus stehen sollen. Darüber hinaus muss bedacht werden, dass bestimmte Verhaltensweisen oder Beeinträchtigungen des Lernens oft die Abklärung unterschiedlicher Konstrukte erfordern. Die Beobachtung, dass bestimmte Schüler*innen im Unterricht häufig verbale Anweisungen nicht umsetzen, kann z. B. auf ein eingeschränktes Sprachverständnis, auf Beeinträchtigungen der Aufmerksamkeit, der Exekutivfunktionen oder des Arbeitsgedächtnisses zurückzuführen sein. Sie kann auch Ergebnis einer fachspezifischen oder allgemeinen schulischen Überforderung (Fehlzeiten im Unterricht, kognitive Beeinträchtigungen usw.) oder durch die pädagogischen Rahmenbedingungen bedingt sein (schlechter oder für die Schüler*innen unpassender Unterricht, mangelnde Differenzierung usw.).

3.2 Schritt 2: Können Testverfahren zur Erfassung der relevanten Konstrukte beitragen?

Danach kann überlegt werden, ob und in welchem Ausmaß Testverfahren zur Erfassung dieser Konstrukte beitragen können. Idealerweise finden sich in Testmanualen klare theoretisch begründete und empirisch belegte Hinweise, welche Konstrukte mit dem jeweiligen Verfahren erfasst werden können. Allerdings orientieren sich Testautor*innen an unterschiedlichen Theorien und diagnostischen Traditionen oder verwenden wenig aussagekräftige, manchmal auch irreführende Bezeichnungen für Skalen und Untertests. Einerseits werden identische Konstrukte häufig unterschiedlich bezeichnet und andererseits werden für theoretisch abgrenzbare Konstrukte zum Teil ähnliche Begriffe verwendet (vgl. Renner, 2017). Weniger erfahrene oder naive Testanwender*innen übernehmen bei der Interpretation ihrer Testergebnisse diese Bezeichnungen oft unkritisch. Diagnostiker*innen sollten immer selbst eine gründliche Analyse der Anforderungsstruktur eines Testverfahrens vornehmen. Dies geschieht am besten unter Bezug auf ein theoretisches Modell des jeweiligen Gegenstandsbereiches. Nur bei einem klar definierten Konstrukt können Zugangs- und Zielfertigkeiten eindeutig voneinander abgegrenzt werden.

3.3 Schritt 3: Liegen beim getesteten Kind Beeinträchtigungen möglicher Zugangsfertigkeiten vor?

Der nächste Schritt besteht darin, den Blick auf die Beeinträchtigungen des Kindes zu werfen, um zu sehen, welche möglichen Zugangsfertigkeiten nicht gegeben sind (vgl. Dembitzer & Kettler, 2018). Dies sollte geschehen durch Explorationen von Bezugspersonen und Fachkräften (z. B. Lehrer*innen, Erzieher*innen, Physio- und Ergotherapeut*innen, Logopäd*innen) und der zu testenden Kindern, durch Sichtung von Vorbefunden (z. B. medizinische Unterlagen) oder durch vorgeschaltete Beobachtungen und Interaktionen, in denen testähnliche Materialien eingesetzt werden, um deren Handhabung durch das Kind zu beobachten.

Beispiel: Viele intelligenzdiagnostische Verfahren verwenden bei der Erfassung des Konstruktes Visuelle Verarbeitung Aufgaben, bei denen das Hantieren mit Klötzen oder Plättchen verlangt wird. Falls beispielsweise die Durchführung des *Mosaik-Tests* aus der Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V; Wechsler, 2017) geplant ist, kann das gemeinsame Hantieren mit Bauklötzen einen Eindruck davon geben, ob motorische Probleme das Nachlegen der Muster beeinträchtigen können. Dabei darf auf keinen Fall das eigentliche Testmaterial zum Einsatz kommen, es dürfen auch keine ähnlichen Anforderungen trainiert werden.

Beispiel: Überprüfungen der phonologischen Schleife des Arbeitsgedächtnisses erfolgen häufig durch Nachsprechaufgaben. Die Behaltensleistung beim unmittelbaren Nachsprechen wird dabei auch von der Sprechgeschwindigkeit beeinflusst (Henry, 2012). Je langsamer ein Kind eine Folge von Zahlen oder Wörtern reproduzieren kann, desto länger muss es die dargebotenen Items im phonologischen Speicher halten. Bei Kindern mit Sprechstörungen könnten Testleiter*innen als erstes die Artikulationsgeschwindigkeit in der Spontansprache beurteilen und dann einige Wörter oder Sätze nachsprechen lassen, die keine Ähnlichkeit mit dem Testmaterial haben. Dabei sollte dann auch auf die mit dem Sprechen verbundene Anstrengung geachtet werden, die sich in einer längeren Testung leistungsmindernd auswirken könnte.

Neben dem Blick auf die für die Testdiagnostik erforderlichen Zugangsfertigkeiten sollten bei diesem Schritt auch weitere Fragen beachtet werden, um eine Testsituation bei Kindern mit Behinderungen angemessen gestalten zu können.

- Welche Hilfsmittel stehen dem Kind zur Verfügung und werden von ihm regelmäßig genutzt? Welche Hilfsmittel sollten in der Testsituation zur Verfügung stehen?
- Welche Kommunikationsmöglichkeiten stehen dem Kind zur Verfügung? Verfügt es über ein sicheres Ja/Nein-Signal?
- Welche Sitzposition ist für das Kind am besten geeignet?
- Kann die Kommunikation mit dem Kind besser gelingen, wenn vor der eigentlichen Testung informeller Kontakt mit dem Kind aufgenommen wird, z. B. um sich in seine sprachlichen Äußerungen einzuhören, die Bedienung eines Kommunikationsgerätes kennenzulernen?
- Wie belastbar ist das Kind? Wie lange kann es mit hoher Wahrscheinlichkeit arbeiten, ohne zu ermüden?
- Welche räumlichen Voraussetzungen sind für die Untersuchung wichtig, z. B. im Hinblick auf Barrierefreiheit, angemessene Ausleuchtung?
- Welche Hilfestellungen benötigt das Kind bei alltäglichen Verrichtungen (z. B. Unterstützung beim Toilettengang)?
- Benötigt das Kind persönliche Assistenz durch eine vertraute Person?
- Bei welchen Personen fühlt sich das Kind sicher und vertraut?
- Wie reagiert das Kind auf fremde Personen und Untersuchungssituationen? Hatte das Kind negative Erfahrungen mit Testuntersuchungen?
- Falls das Kind in nicht vertrauten Umgebungen mit starker Angst reagiert: Ist eine Untersuchung in einer Schule/einer vorschulischen Einrichtung möglich?
- Müssen Besonderheiten bei der Nahrungsaufnahme beachtet werden?
- Muss während des für die Testdiagnostik geplanten Zeitraums die Einnahme von Medikamenten eingeplant werden?
- Zeigt die Leistungsfähigkeit des Kindes deutliche Tagesschwankungen?
- Leidet das Kind regelmäßig unter Schmerzen?
- Gibt es Warnhinweise, auf die zu achten ist, und die bestimmte Maßnahmen des Untersuchers erfordern (z. B. bei gesundheitlichen Risiken oder psychischen Problemen)?
- Was weiß das Kind über Zweck und mögliche Folgen der Testdiagnostik? Könnte es z. B. stark angespannt oder verängstigt sein, weil es gehört hat, dass von der Diagnostik abhängt, ob es auf seiner Schule bleiben darf?

3.4 Schritt 4: Welche Zugangsfertigkeiten verlangen die in Frage kommenden Testverfahren?

Die systematische Analyse der Zugangsfertigkeiten von Testverfahren ist ein entscheidender Schritt in der Untersuchungsplanung. Dabei werden die Aufgabenstellungen und Testmaterialien der in Frage kommenden Testverfahren im Hinblick auf die in Tabelle 1 aufgeführten Zugangsfertigkeiten betrachtet. Es muss beachtet werden, dass einige der aufgeführten Fertigkeiten in bestimmten Kontexten auch Zielfertigkeiten sein können. Von besonderer Bedeutung für die Testdiagnostik sind Hand- und Sprechmotorik, Hör- und Sehfähigkeit sowie das Sprachverständnis – Zugangsfertigkeiten, die in vielen Verfahren ganz selbstverständlich vorausgesetzt werden. Bei der Analyse der Zugangsfertigkeiten sollten Testmaterialien komplett durchgesehen werden, da möglicherweise nur einzelne Items eingeschränkt zugänglich sind (z. B. Abbildungen, die für Kinder mit Farbfehlsichtigkeit schwer zu erkennen sind). Die Auswirkungen verschiedener Sehbeeinträchtigungen können mit Hilfe von Simulatoren (u.a. Coblis – Color Blindness Simulator, Impairment Simulator) veranschaulicht werden.

Beispiel: Der bereits erwähnte SLRT-II setzt neben unbeeinträchtigten Sprechfertigkeiten voraus, dass ein Kind über die nötigen visuellen Wahrnehmungsfähigkeiten verfügt, um die Buchstaben sicher und schnell zu erkennen. Es muss die Testanweisung verstehen (Sprachverständnis, Hörfähigkeit) und in der Lage sein, eine Körperposition einzunehmen und aufrechtzuerhalten, die den Blick auf die Lesevorlage erlaubt (Kopf- und Rumpfkontrolle). Gefordert sind auch die selektive Aufmerksamkeit sowie die Fähigkeit, sich auf der in Spalten angeordneten Lesevorlage zu orientieren. Es bedarf grundlegender exekutiver Funktionen, wie beispielsweise die Fähigkeit zu registrieren, welche Worte schon gelesen wurden und welche noch nicht, um dann das Leseverhalten entsprechend zu steuern.

Auf Basis dieser Analyse ist ein erster Ansatzpunkt zur Sicherung der Testfairness die Verwendung von zugänglichen Testverfahren. Es werden Tests oder Untertests herangezogen, zu deren Bewältigung die beeinträchtigten Zugangsfertigkeiten nicht erforderlich sind.

Beispiel: Wie bereits erwähnt, stellen Nachsprechaufgaben zur Erfassung der phonologischen Schleife des Arbeitsgedächtnisses für Kinder mit Sprechstörungen eine Erschwernis dar. Im Subtest *Wortreihe* der KABC-II ist es dagegen nicht erforderlich, dass die getesteten Personen sprechen. Die Antwort wird durch Zeigen gegeben. Ähnliche Entscheidungsmöglichkeiten haben Testanwender*innen bei Wortschatz- und Sprachverständnistests. Hier gibt es zahlreiche Varianten, die Lautsprache voraussetzen, aber auch Alternativen, bei denen die richtige Antwort aus mehreren Alternativen durch Zeigen ausgewählt werden kann.

Beispiel: Wenn ein Kind Schwierigkeiten im Bereich der aktiven Lautsprache sowie im Sprachverstehen hat, ist es zur Erhebung von fluiden Fähigkeiten oder der visuellen Verarbeitung sinnvoll auf einen Test zurückzugreifen, der nonverbal durchgeführt werden kann (z. B. der Nonverbale Intelligenztest SON-R 2-8; Tellegen, Laros & Petermann, 2018), anstatt ein Verfahren zu wählen, das lautsprachliche Kompetenzen voraussetzt. Bei Beeinträchtigungen der Motorik oder des Sehens wiederum wäre der SON-R 2-8 kein geeignetes Verfahren, da viele Items motorische Grundfähigkeiten benötigen (etwas aufnehmen, halten, ablegen etc.) und alle Materialien Sehfähigkeit voraussetzen.

In der Testpraxis bringt dies in der Regel die Notwendigkeit mit sich, bei der Testzusammenstellung auf unterschiedliche Verfahren zurückzugreifen. Diese gezielte verfahrensübergreifende Testdiagnostik wird als Cross-battery-assessment (XBA) bezeichnet. Bei der Intelligenzdiagnostik bietet sich für die Planung eines XBA die Cattell-Horn-Intelligenztheorie (CHC-Theorie) an,

mit der sich die Struktur kognitiver Leistungen beschreiben lässt. Nähere Informationen zum XBA und zur CHC-Theorie finden sich bei Flanagan, Ortiz und Alfonso (2013) und Schneider und McGrew (2018) sowie deutschsprachig bei Renner und Mickley (2015b) und Mickley und Renner (2019).

Je mehr Zugangsfertigkeiten beeinträchtigt sind, desto geringer wird allerdings die Auswahl an entsprechenden Testaufgaben sein, und es steigt die Gefahr, dass bestimmte Komponenten oder Facetten relevanter pädagogischer oder psychologischer Konstrukte in der Testzusammenstellung nicht mehr berücksichtigt werden. Man spricht dann von mangelnder **Konstruktrepräsentanz**. So würde z. B. der Einsatz des erwähnten SON-R 2-8 keinen umfassenden Überblick über die kognitiven Fähigkeiten erlauben, da das Konstrukt Intelligenz durch den Test nur eingeschränkt repräsentiert wird. Es würden wichtige Intelligenzfaktoren wie Arbeitsgedächtnis, Langzeitgedächtnis, kristalline Intelligenz oder Verarbeitungsgeschwindigkeit nicht berücksichtigt werden. Die Testzusammenstellung muss sich daher immer auch mit der Frage beschäftigen, ob die verwendete Testauswahl die relevanten Konstrukte umfassend und damit aussagekräftig erfasst. Wie bei Schritt 1 ist an dieser Stelle erneut ein gründliches theoretisches Verständnis der Konstrukte gefordert.

3.5 Schritt 5: Sind Adaptationen notwendig und möglich, um eine faire Testung zu gewährleisten?

Nicht immer ist es in der sonderpädagogischen Praxis möglich einen Test zu finden, mit dem sich das relevante Konstrukt in allen Facetten erfassen lässt und der gleichzeitig keine Zugangsfertigkeiten voraussetzt (siehe etwa die Bestandsaufnahme von Capovilla & Kober, 2019, zur Intelligenzdiagnostik bei Sehbehinderung). In diesem Fall muss auf ein Verfahren ausgewichen werden, bei dem die Testperson unter Umständen nicht alle Zugangsfertigkeiten problemlos erfüllen kann. Zur Sicherung der Testfairness ist es dann geboten Anpassungen vorzunehmen. Diese sogenannten Adaptationen sind bewusste Abweichungen von der standardisierten Testdurchführung. Sie werden vorgenommen, um die durch eingeschränkte Zugangsfertigkeiten entstehenden Benachteiligungen zu verhindern oder zumindest zu verringern. Adaptationen sollten so sparsam wie möglich eingesetzt werden: Sie beziehen sich genau auf die eingeschränkten Zugangsfertigkeiten, es werden aber keine darüber hinaus gehenden Veränderungen am Standardvorgehen vorgenommen. Sie unterstützen die Testpersonen bei der Aufgabenbearbeitung, geben aber keine inhaltlichen Hilfestellungen (Salvia, Ysseldyke & Bolt, 2012). Wo immer möglich, werden Adaptationen gemeinsam mit den getesteten Kindern und Jugendlichen oder deren Bezugspersonen geplant und auf ihre Praktikabilität geprüft. Sie sollten nur von Fachkräften eingesetzt werden, die mit der jeweiligen Beeinträchtigungsform vertraut sind, deren Auswirkungen auf die Testsituation beurteilen können und Kenntnisse über typische Hilfsmittel und Unterstützungsmöglichkeiten haben.

Es gibt unterschiedliche Formen von Adaptationen (z. B. CTB/McGraw-Hill, 2004; Lovett & Lewandowski, 2015; Salvia et al., 2012; Thurlow, Elliott & Ysseldyke, 2003):

- *Adaptationen des Settings* beziehen sich auf die Herstellung optimaler Rahmenbedingungen in der Testsituation. Erforderliche Hilfsmittel (z. B. Sehhilfen) müssen bereitstehen. Es sind angepasstes Mobiliar (z. B. unterfahrbare Tisch für Rollstuhlfahrer*innen), eine ausreichende und blendfreie Beleuchtung und eine störungsfreie Akustik notwendig. Bei der Gestaltung des Settings ist auch auf emotionale Aspekte zu achten, wie möglicherweise bestehende Ängste (Beddow, 2018) oder negative Vorerfahrungen mit Testsituationen.

Beispielsweise können Angst und Anspannung bei Kindern und Jugendlichen mit cerebralen Bewegungsstörungen Spastiken verstärken, was die Testung zusätzlich erschwert und das Wohlbefinden der Testpersonen beeinträchtigt. In diesen Fällen ist Wert auf den Beziehungsaufbau zu legen. Botterbusch (1976) schlägt ein sog. »pretesting« vor: Es wird eine testähnliche Situation hergestellt, die – ohne die Verwendung des eigentlichen Testmaterials – den Testpersonen ermöglicht, mit der Testsituation vertraut zu werden.

- *Adaptationen der Item-Präsentationen* bestehen in Veränderungen des Testmaterials und der Instruktionen, z. B. Vergrößerungen von Texten oder Abbildungen, Vereinfachung sprachlicher Anweisungen, Verwendung von Gebärdensprache oder Braille-Schrift, mündliche statt schriftliche Instruktionen und umgekehrt, optimale Positionierung von Testmaterial im Blickfeld des Kindes, das Versehen von Testmaterialien mit Greifhilfen. Diese Adaptationen sind immer individuell auf die Bedürfnisse der Testperson abzustimmen, da Veränderungen je nach individuellen Ausgangsbedingungen unterschiedliche Auswirkungen haben können. So sind Vergrößerungen für Kinder mit einem eingeschränkten Visus sehr hilfreich, dagegen können sie für Kinder mit Störungen des Gesichtsfelds (z. B. Röhrenblick) eine erhebliche Erschwernis darstellen.
- *Adaptationen der Reaktionsform* zielen darauf ab, dass die getesteten Personen Items trotz vorliegender Beeinträchtigungen beantworten können. So können korrekte Lösungen benannt statt gezeigt werden (oder umgekehrt), wobei das Benennen erleichtert werden kann, wenn die Lösungen mit dem Kind bekannten Markierungen (z. B. Farbpunkte) versehen werden. Stehen mehrere Lösungsalternativen zur Auswahl, kann die Antwort über Blickbewegungen erfolgen (Wamba & Marzolf, 1955), was durch Einsatz eines Blickrahmens erleichtert wird (Otto & Wimmer, 2015). Wenn Testpersonen über ein sicheres Ja/Nein-Signal verfügen, kann das sogenannte Scanning-Verfahren angewendet werden, bei dem die Testleiter*innen auf eine Auswahlalternative zeigen und das Kind signalisiert, ob es diese als richtige Lösung mitteilen will oder nicht. Scanning setzt sehr disziplinierte Testleiter*innen voraus, die keine nonverbalen Hinweisreize geben dürfen (z. B. Blick auf die richtige Lösung, zustimmende Mimik vor Bestätigung der Auswahl) und ein gleichmäßiges, auf das Kind individuell abgestimmtes Tempo beim Zeigen der Alternativen realisieren müssen.
- *Adaptationen der Zeitvorgabe* beziehen sich zum einen auf eine den kindlichen Bedürfnissen angemessene Zeitplanung der gesamten Testung, z. B. durch häufigere Pausen, die Verteilung der Untersuchung auf mehrere Tage oder Terminierung von Tests zu Zeiten, in denen die Testpersonen in der Regel die beste Aufmerksamkeit zeigen. Testungen, die das Leistungspotenzial der getesteten Personen aufzeigen wollen, setzen voraus, dass eine Testung schon vor Auftreten von Erschöpfung unterbrochen wird. Zum anderen geht es um die Anpassung der konkreten Zeitvorgaben, die in den Testinstruktionen vorgeschrieben werden. Hier kann die Testzeit pro Item verlängert oder es kann ganz auf Zeitgrenzen verzichtet werden. Insbesondere bei Kindern und Jugendlichen mit motorischen Beeinträchtigungen sind Adaptationen der Zeitvorgaben essentiell.

Durch den Verzicht auf die strenge Beachtung der Durchführungsobjektivität stehen Diagnostiker*innen beim Einsatz von Adaptationen immer vor der Frage, welche Auswirkungen die vorgenommenen Anpassungen auf das Testergebnis haben. Idealerweise werden Adaptationen das Fehlen von Zugangsfertigkeiten kompensieren, ohne die Testergebnisse zu verfälschen. In diesem Fall hat es sich eingebürgert, von *Akkommodationen* zu sprechen (American Educational Research Association et al., 2014). Ein einfaches Beispiel wäre das Tragen einer Brille bei einem Lesetest. Die Brille kompensiert eine mögliche Benachteiligung aufgrund einer Sehbeeinträchtigung, sie verschafft jedoch darüber hinaus keinerlei Vorteile beim Erlesen eines Textes.

Weitere Beispiele wären das Vergrößern des Testmaterials bei Beeinträchtigungen des Sehens, das Anbringen von Greifhilfen oder die Nutzung von Kartenleisten zur leichteren Aufnahme von Testmaterial bei motorischen Beeinträchtigungen (Otto & Wimmer, 2015). Bei den genannten Beispielen ist es relativ klar, dass sich – sofern visuelle oder motorische Fähigkeiten nicht die Zielfertigkeiten sind – keine Veränderung der Aufgabenschwierigkeit ergeben dürfte.

Andere Adaptationen werden jedoch dazu führen, dass sich die Aufgabenschwierigkeit deutlich verändert. Die ist z. B. der Fall, wenn bei verbalen Anforderungen Auswahlalternativen angeboten werden, während die Standarddurchführung das aktive Formulieren der Antwort verlangt. Die Bereitstellung einer Auswahl erleichtert die Aufgabe grundsätzlich, hiervon würden auch Testpersonen profitieren, die über alle erforderlichen Zugangsfertigkeiten (hier: Aktivsprache) verfügen. Die Schwierigkeit eines Tests wird beeinflusst, wenn die in den Testmanualen vorgeschriebenen Zeitvorgaben verändert werden. Adaptationen können auch dazu führen, dass sich der Charakter des erfassten Konstruktes grundlegend verändert, z. B. wenn ein Taschenrechner als Hilfe bei der Bewältigung von Textaufgaben zugelassen wird. In diesen Fällen spricht man von *Modifikationen*. Die Testwerte verändern ihre Bedeutung und sind nicht mehr mit einer Standarddurchführung vergleichbar.

In der konkreten Testpraxis stellen Akkommodationen und Modifikationen zwei Pole eines Kontinuums dar, eine exakte Abgrenzung ist kaum möglich. Dabei ist auch zu berücksichtigen, dass Adaptationen, die auf den ersten Blick keine Veränderung der Aufgabenschwierigkeit mit sich bringen, doch einen Einfluss auf die Testergebnisse haben können, wenn das adaptierte Vorgehen höhere Anforderungen an Ausdauer und Aufmerksamkeit stellt (z. B. ist das oben beschriebene Scanning-Verfahren deutlich zeitaufwändiger und auch anstrengender als eine Standardauswahl durch Zeigen oder Benennen einer Lösungsalternative).

4 Adaptationen in der Praxis: Grenzen und Erfordernisse

Forschungsergebnisse zur Frage, wie sich bestimmte Adaptationen auswirken, liegen vor allem zu standardisierten US-amerikanischen Schul- und Studententests vor (z. B. Dembitzer & Kettler, 2018; Sireci, Scarpati & Li, 2005), wobei die Auswirkungen individualisierter Adaptationen bei Kindern und Jugendlichen mit Behinderungen nicht berücksichtigt werden. Testanwender*innen können daher kaum auf empirisch gesicherte Erkenntnisse zu den Auswirkungen von Adaptationen zugreifen (nähere Hinweise zur Bewertung von Adaptationen finden sich in CTB/McGraw-Hill, 2004 und Salvia et al., 2012). Dies gilt besonders für deutschsprachige Testverfahren. Fragen rund um die Testfairness spielen hierzulande in der empirischen Forschung praktisch keine Rolle, obwohl Adaptationen in der Testpraxis durchaus gebräuchlich sind (Reuner & Renner, 2019). Die wenigen uns bekannten Publikationen in diesem Bereich sind praxisorientiert (z. B. Junglas & Simon, 2019; Otto & Wimmer, 2015) und setzen sich nicht mit den Auswirkungen von Anpassungen auf die Testergebnisse auseinander.

Es ist leider nicht Standard, dass in Testverfahren detailliert auf die Testung von Kindern und Jugendlichen mit Beeinträchtigungen eingegangen wird. Testautor*innen äußern sich zudem in Bezug auf mögliche Adaptationen und die Abgrenzung von Akkommodation und Modifikation in den jeweiligen Handbüchern nicht oder nur sehr diffus. Dementsprechend müssen Testanwender*innen Adaptationen in der Regel eigenverantwortlich und ohne orientierende Hinweise planen. Sie müssen sich bewusst sein, dass adaptierte Testergebnisse von ihren subjektiven

Entscheidungen beeinflusst werden und daher vorliegende Befunde zu den Gütekriterien nicht übertragbar sind.

Werden Adaptationen vorgenommen, ist es notwendig, diese sorgfältig und nachvollziehbar zu dokumentieren. In allen Dokumenten, die in professionellen Kontexten genutzt werden (z. B. sonderpädagogische Gutachten, klinisch-psychologische Befundberichte) muss eindeutig erkennbar sein, welche Änderungen vorgenommen wurden. An keiner Stelle einer Dokumentation darf der Eindruck entstehen, dass die quantitativen Testwerte auf einer standardisierten Durchführung basieren. Nur so können Testergebnisse nachvollzogen und korrekt interpretiert werden.

Die Interpretation von Testergebnissen, die mit adaptierten Verfahren ermittelt wurden, bedarf einer besonderen Sorgfalt. Neben der schwierigen Unterscheidung zwischen Modifikationen und Akkommodationen stellt sich stets auch die Frage nach der Validität von Prognosen. Die adaptierte Testsituation stellt einen bestimmten Rahmen dar, unter dem sich Fähigkeitspotenziale zeigen können. Fehlt dieser Rahmen in anderen Lebenssituationen (z. B. Schulunterricht), werden Kinder ihre Potenziale jedoch nicht realisieren können.

Beispiel: Eine Schülerin mit einer cerebralen Bewegungsstörung erreicht in einem adaptierten Intelligenztest einen Durchschnittswert. Die Adaptationen bezogen sich vor allem auf erweiterte Zeitvorgaben und Anpassungen der Testmaterialien, um diese leichter greifbar zu machen. Intelligenz ist ein relativ guter Prädiktor von Schulleistungen. Kann jetzt angenommen werden, dass die Schülerin diese Leistungen auch in jeglichem schulischen Setting zeigen kann? Diese Folgerung wäre nicht zulässig, da für Prognose des Erfolgs in schulischen Settings noch viele weitere Konstrukte neben der kognitiven Leistungsfähigkeit relevant sind. Es gibt jedoch auch einen besonderen Aspekt: Der adaptierte Test misst die Leistungsfähigkeit der Schülerin eben unter den adaptierten Bedingungen. Fehlen entsprechende Adaptationen im Schulalltag, kann dieser Testwert nur eine geringe prognostische Aussagekraft haben. Es besteht dann die Gefahr, völlig unnötigerweise ein »Schulversagen« zu erzeugen, weil auf die Beeinträchtigungen im Schulalltag nicht eingegangen und von der Schülerin erwartet wird, dass sie alle Anforderungen im üblichen Tempo ohne notwendige motorische Hilfen erfüllt. Ergebnisse adaptierter Tests müssen daher immer in Relation zu den realen Bedingungen, unter denen Kinder und Jugendliche ihr Leistungspotenzial realisieren können – oder auch nicht – gesehen werden.

Dies macht auch deutlich, dass Adaptationen mehr sind als eine Variation des testdiagnostischen Vorgehens. Der erfolgreiche Einsatz von Adaptationen bildet auch eine Brücke zur Förderplanung und zur »advocacy« im Interesse der Kinder.

Die Sicherung der Testfairness und die Planung von Adaptationen setzen voraus, dass die in einem Test verlangten Zugangsfertigkeiten bekannt sind (s. o., Schritt 2). Wir regen daher an, dass Testanwender*innen – wenn möglich im Austausch mit Kolleg*innen – für die von ihnen regelmäßig eingesetzten Testverfahren einen Überblick über erforderliche Zugangsfertigkeiten erstellen. Dabei wird es vor allem um konstruktirrelevante Einflüsse in den Bereichen Motorik (Hand- und Sprechmotorik), Hören, Sehen und Sprachverständnis gehen. Auf Basis dieser Bestandsaufnahme können Testanwender*innen dann eine Liste von (Unter-)Tests erstellen, die z. B. kein visuelles Material verwenden, nur einfache oder keine handmotorischen Anforderungen stellen oder die sprachfrei durchgeführt werden können.

5 Fazit und Ausblick: Auf dem Weg zu einer zugänglichen Testdiagnostik?

Die Sicherung der Fairness ist eine zentrale Aufgabe in der sonderpädagogischen Testdiagnostik. Wir haben hierzu ein praxisbezogenes Vorgehen vorgeschlagen, das Benachteiligungen von behinderten Kindern und Jugendlichen vorbeugen kann, jedoch empirisch nicht ausreichend begründet ist.

Es ist dringend erforderlich, dass bei der Entwicklung von Testverfahren dem Aspekt der Zugänglichkeit wesentlich mehr Aufmerksamkeit geschenkt wird. Die mit dem Einsatz von Adaptationen verbundenen Nachteile und Unsicherheiten werden an Bedeutung verlieren, wenn Testmaterialien so gestaltet werden, dass eine größtmögliche Anzahl von Kindern und Jugendlichen im Rahmen einer standardisierten Testung untersucht werden kann. Diagnostiker*innen in Sonderpädagogik und klinischer Psychologie sollten die Entwicklung entsprechender Verfahren bei Testverlagen regelmäßig einfordern.

Ein Beispiel für eine konsequente Umsetzung des Anspruchs unter Nutzung moderner technischer Hilfsmittel ist das noch nicht in einer deutschen Version adaptierte Computer-Based instrument for Low motor Language Testing (C-BiLLT; Geytenbeek, Morkink, Knol, Vermeulen & Oostrom, 2014), bei dem eine große Vielzahl von Reaktionsmöglichkeiten zur Verfügung steht, die auch von motorisch schwerst beeinträchtigten Kindern genutzt werden können.

Ein weiterer Aspekt bei der künftigen Entwicklung von Testverfahren ist die klare Formulierung von Regeln zu Testadaptationen, die im Rahmen einer standardisierten Diagnostik vorgenommen werden dürfen. Dies gibt es zum Beispiel für die amerikanischen Stanford-Binet Intelligence Scales, Fifth Edition (SB5; Roid, 2003). Für den Test liegt ein Arbeitspapier vor (Braden & Elliott, 2003), das deutlich macht, wie im Rahmen einer standardisierten Testdiagnostik eine systematische Beschreibung von angemessenen und unangemessenen Adaptationen erfolgen kann. Vergleichbare Darstellungen in deutschsprachigen Testverfahren sind den Autoren nicht bekannt.

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.

American Psychological Association (2012). Guidelines for assessment of and intervention with persons with disabilities. *American Psychologist*, 67(1), 43-62.

Beddow, P. A. (2018). Assessing students with autism: Considerations and recommendations. In S. N. Elliott, R. J. Kettler, P. A. Beddow & A. Kurz (Eds.), *Handbook of Accessible Instruction and Testing Practices. Issues, Innovations, and Applications* (2nd ed., pp. 107-118). Cham: Springer International Publishing.

Botterbusch, K. F. (1976). *The use of psychological tests with individuals who are severely disabled*. Stout: University of Wisconsin.

- Braden, J. P. (2003). Accommodating clients with disabilities on the WAIS-III and WMS. In D. S. Tulskey (Hrsg.), *Clinical interpretation of the WAIS-III and WMS-III* (S. 451-486). Amsterdam: Academic Press.
- Braden, J. P. & Elliott, S. N. (2003). Accommodations on the Stanford-Binet Intelligence Scales, Fifth Edition (Stanford-Binet Intelligence Scales, Fifth Edition Assessment Service Bulletin Number 2). Itasca, IL: Riverside Publishing. Zugriff am 09.12.2021. https://www.hmhco.com/~media/sites/home/hmh-assessments/clinical/stanford-binet/pdf/sb5_asb_2.pdf%3Fla%3Den
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3., aktualisierte und erw. Aufl.). München: Pearson.
- Bundschuh, K. & Winkler, C. (2014). *Einführung in die sonderpädagogische Diagnostik* (8. überarb. Aufl.). München: Reinhardt; UTB.
- Capovilla, D. & Kober, A. (2019). Intelligenzdiagnostik bei Kindern mit einer Sehbeeinträchtigung. *Empirische Sonderpädagogik*, 11(1), 31–52.
- CTB/McGraw-Hill. (2004). *Guidelines for inclusive test administration 2005*. Monterey, CA: CTB McGraw-Hill.
- Dembitzer, L. & Kettler, R. J. (2018). Testing adaptations: Research to guide practice. In S. N. Elliott, R. J. Kettler, P. A. Beddow & A. Kurz (Eds.), *Handbook of accessible instruction and testing practices. Issues, innovations, and applications* (2nd ed., pp. 213-230). Cham: Springer International Publishing.
- Flanagan, D. P., Ortiz, S. O. & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3. Aufl.). Hoboken, NJ: Wiley.
- Geytenbeek, J. J., Mokkink, L. B., Knol, D. L., Vermeulen, R. J. & Oostrom, K. J. (2014). Reliability and validity of the C-BiLLT: a new instrument to assess comprehension of spoken language in young children with cerebral palsy and complex communication needs. *Augmentative and Alternative Communication*, 30(3), 252-266.
- Grob, A., Reimann, G., Gut, J. & Frischknecht, M.-C. (2013). *IDS-P. Intelligence and Development Scales – Preschool*. Bern: Huber.
- Hartig, J., Frey, A. & Jude, N. (2020). Validität von Testwertinterpretationen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 529–545). Berlin: Springer.
- Henry, L. (2012). *The Development of Working Memory in Children*. Los Angeles: Sage.
- Junglas, H. & Simon, A. (2019). Diagnostik im FSP motorische Entwicklung. *Lernen konkret* (4), 36-39.
- Kettler, R. J., Elliott, S. N., Beddow, P. A. & Kurz, A. (2018). Accessible instruction and testing today. In S. N. Elliott, R. J. Kettler, P. A. Beddow & A. Kurz (Eds.), *Handbook of accessible instruction and testing practices. Issues, innovations, and applications* (2nd ed., pp. 1–16). Cham: Springer International Publishing.
- Lovett, B. J. & Lewandowski, L. J. (2015). *Testing accommodations for students with disabilities. Research-based practice*. Washington, D.C: American Psychological Association.

- Messick, S. (2000). Consequences of test interpretation and use: the fusion of validity and values in psychological assessment. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment. Honoring Douglas N. Jackson at seventy.* (pp. 3-20). Boston: Kluwer Academic.
- Mickley, M. & Renner, G. (2019). 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.
- Moll, K. & Landerl, K. (2014). *SLRT-II Lese- und Rechtschreibtest (2., korr. Aufl. mit erw. Normen)*. Göttingen: Hogrefe.
- Moosbrugger, H. & Kelava, A. (2012). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion (2. Aufl., S. 7-26)*. Berlin: Springer.
- Otto, K. & Wimmer, B. (2015). Diagnostik in der Sprachtherapie und ihre Anwendungsmöglichkeiten in der Unterstützten Kommunikation bei Kindern. In Isaac – Gesellschaft für Unterstützte Kommunikation e. V. (Hrsg.), *Handbuch der Unterstützten Kommunikation (12. Aufl., S. 14.038.001-14.045.001)*. Karlsruhe: von Loeper-Literaturverlag.
- 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.
- Renner, G. & Mickley, M. (2015a). Berücksichtigen deutschsprachige Intelligenztests die besonderen Anforderungen von Kindern mit Behinderungen? *Praxis Kinderpsychologie und Kinderpsychiatrie*, 64(2), 88-103.
- Renner, G. & Mickley, M. (2015b). Intelligenzdiagnostik im Vorschulalter. CHC-theoretisch fundierte Untersuchungsplanung und Cross-battery-assessment. *Frühförderung interdisziplinär*, 34(2), 67-82.
- Reuner, G. & Renner, G. (2019). Praxis der klinisch-psychologischen und sonderpädagogischen Testdiagnostik bei Kindern und Jugendlichen mit körperlichen und motorischen Beeinträchtigungen – Ergebnisse einer Umfrage unter Anwendern. *Zeitschrift für Heilpädagogik*, 70(2), 84–93.
- Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition*. Rolling Meadows, IL: Riverside Publishing.
- Salvia, J., Ysseldyke, J. & Bolt, S. (2012). *Assessment in Special and Inclusive Education (12. Aufl.)*. Belmont, CA: Wadsworth Cengage Learning.
- Schermelleh-Engel, K., Kelava, A. & Moosbrugger, H. (2006). Gütekriterien. In F. Petermann & M. Eid (Hrsg.), *Handbuch der Psychologischen Diagnostik (S. 420-433)*. Göttingen: Hogrefe.
- Schmidt-Atzert, L. & Amelang, M. (Hrsg.). (2012). *Psychologische Diagnostik (5. Aufl.)*. Berlin: Springer.

- 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.
- Sireci, S. G., Scarpati, S. E. & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75(4), 457–490.
- Tellegen, P. J., Laros, J. A. & Petermann, F. (2018). SON-R 2-8. Non-verbaler Intelligenztest. Göttingen: Hogrefe.
- Thurlow, M., Elliott, J. L. & Ysseldyke, J. E. (2003). *Testing students with disabilities* (2nd. ed.). Thousand Oaks: Corwin.
- Tröster, H. (2019). *Diagnostik in schulischen Handlungsfeldern. Methoden, Konzepte, praktische Ansätze* (1. Auflage). Stuttgart: Verlag W. Kohlhammer.
- Wamba, D. E. & Marzolf, S. S. (1955). Use of eye movements as a response indicator in testing the physically handicapped. *Journal of Clinical Psychology*, 11(4), 405-407.
- Wechsler, D. (2017). *WISC-V. Wechsler Intelligence Scale for Children – Fifth Edition*. Deutsche Bearbeitung Franz Petermann. Frankfurt: Pearson Assessment.

Prof. Dr. Gerolf Renner ist Professor für Psychologie und Diagnostik im Förderschwerpunkt körperliche und motorische Entwicklung an der Pädagogischen Hochschule Ludwigsburg. <https://orcid.org/0000-0003-4345-3619>

Prof. Dr. Markus Scholz ist Professor für Psychologie und Diagnostik im Förderschwerpunkt geistige Entwicklung an der Pädagogischen Hochschule Ludwigsburg. Schwerpunkte seiner Arbeit sind unter anderem die Diagnostik in der Unterstützten Kommunikation sowie die empirische Erforschung zu und Konzeption von Lernmaterialien und Kommunikationshilfen für Personen mit unterschiedlichen Ausgangsbedingungen. <https://orcid.org/0000-0001-8029-5782>