

# Der Einsatz von standardisierten Tests in der Logopädie

Ulla Beushausen

**ZUSAMMENFASSUNG.** In der Diagnostik von Kommunikationsstörungen steht eine Vielzahl von standardisierten und normierten Testverfahren zur Verfügung. Im Artikel werden die Tests zu Grunde liegenden Gütekriterien (Objektivität, Reliabilität, Validität) definiert und Kriterien zur Auswahl eines geeigneten Testverfahrens für die therapeutische Praxis erläutert.

Schlüsselwörter: Tests – Kommunikationsstörungen – Objektivität – Reliabilität – Validität – Diagnostik – Logopädie – Testtheorie

## Einleitung

Die Diagnostik von Kommunikationsstörungen und von damit in Verbindung stehenden Störungen ist ein integraler Bestandteil der Logopädie und Sprachtherapie und ein herausfordernder Prozess für die Therapeutin. Das methodische Handwerkszeug umfasst dabei verschiedene Tests und qualitative Analyseverfahren, die zusammen mit anderen klinischen Daten, wie Beobachtungen, Screenings und informellen Prüfverfahren, ein umfassendes Bild der kommunikativen Fähigkeiten eines Klienten ergeben. Nachdem in der sprachtherapeutischen Vergangenheit vorwiegend informelle Prüfverfahren eingesetzt wurden, entstand in den letzten Jahren eine Vielzahl von Testverfahren für unterschiedlichste Störungsbilder, die standardisiert und normiert sind. Die Praktikerin hat oft sogar die Wahl zwischen mehreren Verfahren.

## Was ist ein Test?

Tests erfassen Eigenschaften, Fähigkeiten oder Merkmale bzw. Zustände von Personen. Unter Tests im weiteren Sinne versteht man auch Fragebögen, standardisierte Interviews und standardisierte Beobachtungen. Tests im engeren Sinne bezeichnen Verfahren, die durch die getestete Person möglichst nicht willentlich verfälscht werden können. Tests werden deshalb standardisiert und normiert. Ein standardisiertes Testverfah-

ren ist ein wissenschaftliches Verfahren, das in seiner Durchführung, Auswertung und Interpretation so detailliert beschrieben ist, dass es von verschiedenen Testanwenderinnen in der gleichen Weise durchgeführt, ausgewertet und interpretiert werden kann. Es basiert auf einer theoretischen Grundlage, nach der die einzelnen Testergebnisse interpretiert werden können. In der Regel sollte ein standardisiertes Verfahren an einer größeren Stichprobe hinsichtlich seiner Durchführung und Auswertung erprobt sein (Normierung) sowie die geforderten Gütekriterien erfüllen.

In ihrem Standardwerk zur Testanalyse und Testkonstruktion definieren *Linert & Raatz* (1998, S.1) einen Test als ein „...wissenschaftliches *Routineverfahren* zur Untersuchung eines oder mehrerer empirisch abgrenzbarer *Persönlichkeitsmerkmale* mit dem Ziel einer möglichst *quantitativen Aussage* über den relativen Grad der *individuellen Merkmalsausprägung*“.

Die Sprachfähigkeit eines Kindes, gemessen am individuellen Sprachentwicklungsstand, stellt beispielsweise solch ein Persönlichkeitsmerkmal dar, das mit einem Sprachtest erfasst werden kann. Das Ergebnis ist quantitativ, d.h. die sprachlichen Fähigkeiten werden anhand einer Maßzahl bestimmt (z.B. Summe aller richtig erfüllten Aufgaben), die dann mit dem Durchschnitt aller Kinder dieser Altersklasse verglichen werden kann. Qualitative Daten, beispielsweise welche indivi-

**Dr. Ulla Beushausen** ist Logopädin und Psycholinguistin und als Professorin für Logopädie in den Bachelor- und Masterstudiengängen für Ergotherapie, Logopädie und Physiotherapie an der Hochschule für angewandte Wissenschaft und Kunst (HAWK), Fachhochschule Hildesheim/Holzminde/Göttingen tätig.



duellen Strategien bei der Pluralbildung ein Kind anwendet, sind dabei zunächst nicht von Interesse, weil sich diese nur auf ein einzelnes Kind bezogen beschreiben lassen und nicht als Durchschnittswerte über viele Kinder sinnvoll berechenbar sind.

## Welche Tests gibt es?

Tests werden in drei Gruppen unterteilt: *Leistungstests*, *psychometrische Persönlichkeitstests* und *Persönlichkeits-Entfaltungungsverfahren* (Tab. 1).

■ **Tab. 1: Klassifikationssystem für Testverfahren (in Anlehnung an Brähler et al., 2002)**

### 1. Leistungstests

- Entwicklungstests
- Intelligenztests
- Allgemeine Leistungstests
- Schultests

### 2. Persönlichkeitstests

- Persönlichkeitsstrukturtests
- Einstellungs- und Interessenstests
- Klinische Tests

### 3. Persönlichkeits-/Entfaltungungsverfahren

- Formdeutungsverfahren
- Verbal-thematische Verfahren
- Zeichnerische und Gestaltungsverfahren

Zur Gruppe der *Leistungstests* werden unter der Kategorie Entwicklungstests die bekannten Sprachentwicklungstests (siehe Tab. 2), aber auch Tests zur motorischen oder sensorischen Entwicklung eines Kindes zugeordnet. Bei den Intelligenztests gibt es eine große Bandbreite. Im Zusammenhang mit Sprachstörungen ist zu beach-

ten, ob ein Intelligenztest sprachfrei oder sprachgebunden konstruiert wurde. Allgemeine Leistungstests sind z.B. die Aufmerksamkeits- und Konzentrationstests. Leserechtschreibtests beschreiben den Hauptanteil der Schultests. *Persönlichkeitstests* prüfen verschiedene situationsüberdauernde Eigenschaften einer Person, z.B. den

Grad der Extrovertiertheit oder Introvertiertheit. Einstellungs- und Interessentests finden sich zur beruflichen Eignung und in vielen sozialen Kontexten (Einstellung zum Rauchen, zum Stillen etc.). Tests zu sprachtherapeutischen Störungsbildern wie Aphasie und Dysarthrie zählen zu den klinischen Tests (Tab. 3).

■ Tab. 2: Beispiele für standardisierte Sprachentwicklungstests (vgl. *Beushausen, 2007*)

Testname	Geltungsbereich	Erfasste Merkmale	Gütek. ▼	Norm. ▼
<b>Aktiver Wortschatztest (AWST-R 3-5)</b>	3;0 bis 5;5 J.	Wortschatz	▼	▼
<b>Elternfragebogen für die Früherkennung von Risikokindern (ELFRA)</b>	ELFRA 1: 12 Monate ELFRA 2: 24 Monate	Früherkennung von Risikokindern (Sprachstörungen)	▼	▼
<b>Heidelberger Sprachentwicklungstest (HSET)</b>	4;0 bis 9;11 J.	Entwicklungsstand: Sprachverständnis, Produktion, Syntax, Morphologie, Lexikon	▼	▼ veraltet
<b>Kindersprachtest für das Vorschulalter (KISTE)</b>	3;3 bis 6;11 J.	Entwicklung: Grammatik/Lexik, Syntax/Morphologie, Semantik/Kommunikation	▼	▼
<b>Marburger Sprachverständnistest für Kinder (MSVK)</b>	Kinder ab 5;0 und Erstklässler	Sprachverständnis	▼	▼
<b>Patholinguistische Diagnostik bei Sprachentwicklungsstörungen</b>	Kinder ab 3 J.	Sprachentwicklung im Profil: Phonetik/Phonologie, Semantik/Lexik., Sprachverständnis, Morphologie/Syntax	○	●
<b>Psycholinguistischer Entwicklungstest (PET)</b>	3;0 bis 9;11 J.	Hörmerkspanne, phonologische Bewusstheit, Syntax/Morphologie, Semantik/Lexik., kognitive Leistungen	●	▼ veraltet
<b>Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5)</b>	3;0 bis 5;11 Jahre	Sprachentwicklung rezeptiv/expressiv: Morphologie/Syntax, Semantik/Lexik., phonologisches Arbeitsgedächtnis/Gedächtnisspanne	▼	▼
<b>Sprachentwicklungstest für zweijährige Kinder (SETK-2)</b>	2;0 bis 2;11 Jahre	Sprachentwicklung: rezeptiv/expressiv: Morphologie/Syntax, Semantik/Lexik.	▼	▼
<b>Sprachscreening für das Vorschulalter (SSV)</b>	3 bis 6 Jahre	phonologisches Arbeitsgedächtnis/Plural/Satzgedächtnis	▼	▼
<b>Teddytest</b>	3;6 bis 6;11 Jahre / SSES bis 9;6 Jahre	semantische Relationen	▼	▼
<b>Test zur Überprüfung des Grammatikverständnisses (TROG-D)</b>	3;0 bis 10;11 Jahre	Grammatikverständnis	●	▼
<b>Wortschatz- und Wortfindungstest für 6- bis 10-Jährige (WWT 6-10)</b>	6;0-9;11	Wortschatz, Wortfindung	●	▼

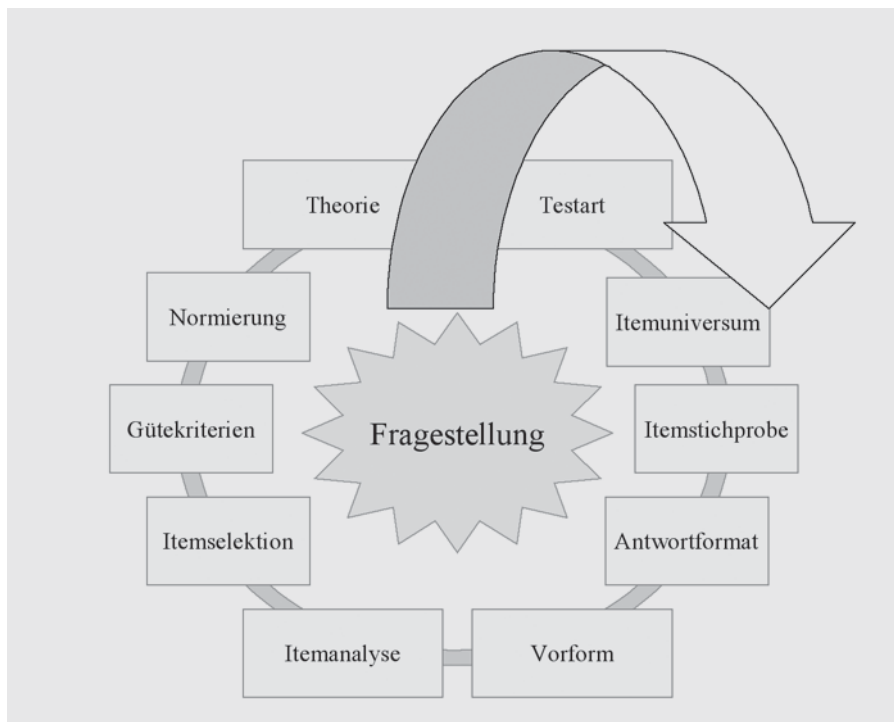
Legende: ● = teilweise, ○ = nein, ▼ = ja

■ Tab. 3: Beispiele für standardisierte Aphasiediagnostik (vgl. *Beushausen, 2007*)

Testname	Geltungsbereich	Erfasste Merkmale	Gütek. ▼	Norm. ▼
<b>Aachener Aphasie Test (AAT)</b>	Aphasiker/postakute und chronische Phase	Feststellung Aphasie, Bestimmung/Schweregrad und Klassifikation in Standardsyndrome	▼	▼
<b>Aphasie-Check-Liste (ACL)</b>	Aphasiker/alle Erkrankungsphasen	Feststellung Aphasie, Schweregrad	●	●
<b>Aphasie-Schnell-Test (AST)</b>	leichte bis mittelschwer betroffene Aphasiker, Akutphase	Feststellung Aphasie und Schweregrad	●	▼
<b>Bogenhauser Semantikuntersuchung (BOSU)</b>	aphasische und nicht aphasische Personen	semantische Störungen	●	●
<b>Kurze Aphasie Prüfung (KAP)</b>	Aphasiker/Akutphase und chronische Phase	Feststellung Aphasie, Bestimmung Syndromklassifikation und Schweregrad	●	●
<b>Regensburger Wortflüssigkeitstest (RWT)</b>	8 bis 15 Jahre, Erwachsene ab 18 Jahren	divergentes Denken (formallexikalische und kategorial-semantische Wortflüssigkeitsleistungen)	▼	▼

Legende: ● = teilweise, ○ = nein, ▼ = ja

■ Abb. 1: Testentwicklung



*Persönlichkeits- und Entfaltungungsverfahren* sind z.B. der Rohrschach-Test mit der Interpretation von Tintenklecksen, verbal-thematische Verfahren, bei denen Kinder z.B. Geschichten zu Bildern mit Konfliktsituationen erzählen, oder zeichnerische und Gestaltungsverfahren wie der Mann-Zeichen-Test oder der Baum-Test.

## Wie entsteht ein Test?

Tests basieren auf der sogenannten Testtheorie. Die Testtheorie ist das Teilgebiet der Messtheorie, das sich mit den differenzialpsychologischen Messproblemen beschäftigt. Die klassische Testtheorie befasst sich unter anderem mit den unterschiedlichen Bestandteilen von Messwerten, zum Beispiel dem Verhältnis des Messfehlers zum wahren Wert. Es wird angenommen, dass Messwerte fehlerbehaftet sind, dass aber im Falle objektiver und standardisierter Messbedingungen diese Fehler lediglich Zufallsschwankungen aufweisen und für den theoretisch denkbaren Fall unendlich vieler Messungen im Mittel 0 betragen. 95 Prozent aller standardisierten Tests basieren auf der klassischen Testtheorie.

Eine Testentwicklung ist eine aufwändige Sache (s. Abb. 1) und erfordert in der Regel ein ganzes Team, das über mehrere Jahre an einem solchen Entwicklungsprojekt arbeitet. Am Anfang steht immer ein Bedarf. Ein Verhalten tritt auf und soll bestimmt werden,

z.B. das sprachlich-prosodische Verhalten von Kindern. Darauf folgt die Bestimmung der zu Grunde gelegten Theorie. Was ist Prosodie? Welche Bereiche umfasst sie? Wie entwickelt sie sich? Wie hängt die Entwicklung der Prosodie von der jeweiligen Muttersprache oder von Dialekt-Einflüssen ab? Hierzu werden Theorien und Studienergebnisse herangezogen, um die theoretische Basis des Tests – das „Konstrukt“ – zu beschreiben. Anschließend muss die „Testart“ festgelegt werden: Soll es ein Entwicklungstest oder ein allgemeiner Leistungstest oder ein klinischer Test werden? Wer ist die Zielgruppe (z.B. Erwachsene oder Kinder)? Wie soll der „Antwortmodus“ sein? Soll Prosodie dabei imitiert oder frei produziert werden oder soll lediglich die Richtigkeit prosodischer Äußerungen bestimmt werden? Steht die Testart fest, steht der Testkonstrukteur vor der Wahl der einzelnen Testaufgaben, auch „Items“ genannt. Aus der Summe aller möglichen Testaufgaben (Itemuniversum) werden möglichst viele repräsentative Vertreter ausgewählt, die dann im Rahmen der Testdurchführung systematisch mit Hilfe statistischer Verfahren reduziert werden. Dies geschieht in der Regel in einem „Prätest“, bei dem bereits Testungen mit einzelnen Probanden, die zur späteren Zielgruppe passen, durchgeführt werden. Diejenigen Items, die vom „Schwierigkeitsgrad“ und von der „Trennschärfe“ her die besten Ergebnisse zur Unterscheidung z.B. zwischen prosodisch auffälligen und un-

auffälligen Kindern mitbringen, gehen in die Endform ein. Danach müssen die Gütekriterien berechnet und die Normdaten an einer größeren Stichprobe erhoben werden.

## Die Gütekriterien eines Testverfahrens

Es gibt verschiedene anerkannte Kriterien, nach denen die Güte eines Tests bestimmt werden kann. Sie werden in den Manualen, also den Handbüchern der Tests, dargestellt und ermöglichen eine Einschätzung der Testgüte. Man unterscheidet Haupt- und Nebengütekriterien. Die Hauptgütekriterien sind „Objektivität“, „Reliabilität“ und „Validität“ (Tab. 4).

### Objektivität

Ein Test soll in der *Durchführung, Auswertung und Interpretation* der Ergebnisse unabhängig vom Testleiter sein, das heißt, die Testergebnisse eines Probanden sollten dieselben sein, egal wer ihn durchführt. In der Regel wird dies durch genaue Testinstruktionen zu den einzelnen Items und durch konkrete Anleitungen zur Auswertung und Interpretation im Handbuch sichergestellt. Man spricht dann von einem standardisierten Test. Das Manual sollte Angaben darüber machen, ob und welche Hilfestellungen gegeben werden dürfen, ob und wie oft eine Instruktion wiederholt werden darf, ob Zeitbegrenzungen für die Testdurchführung existieren,

■ Tab. 4: Hauptkriterien der Testgüte

#### 1. Objektivität (Ist der Test unabhängig vom Testenden?)

- Durchführung, Auswertung und Interpretation

#### 2. Validität (Gültigkeit: Misst der Test, was er messen soll?)

- Kriteriumsvalidität
  - a. Übereinstimmungsvalidität
  - b. Vorhersagevalidität
- Konstruktvalidität
  - a. konvergent
  - b. divergent
  - c. Faktorenanalyse

#### 3. Reliabilität (Zuverlässigkeit: Wie genau misst der Test?)

- Innere Konsistenz
  - Itemanalyse
  - Stabilität

## Glossar

**C-(Centil) Werte:** Transformationsmaß, transformiert Rohwerte in Werte mit einem Mittelwert von 5 und einer Standardabweichung von 2.

**Cut-off-Werte** sind Trennpunkte bzw. kritische Werte. Personen mit Werten oberhalb des Cut-off-Points erhalten eine andere diagnostische Zuweisung (und Behandlung) als die Personen mit Werten unterhalb des Cut-off-Wertes; zum Beispiel feststehende Testwerte in Entwicklungstests, unterhalb derer die Entwicklung einer Testperson als gestört angesehen wird.

**Faktorenanalyse,** ein Verfahren der Datenreduktion. Es wird angenommen, dass hinter einer Reihe von Messwerten, zum Beispiel Ergebnissen eines psychologischen Tests, eine grundlegende, nicht direkt messbare, hypothetische Variable steht, zum Beispiel eine Eigenschaft oder eine Einstellung. Eine solche hypothetische Variable wird als „Faktor“ bezeichnet. Wenn angenommen wird, dass die erzielten Messergebnisse auf einen einzigen Faktor zurückgehen, so bedeutet das, dass die betreffenden Variablen untereinander in hohem Maße korrelieren müssten. Die entsprechende Korrelationsmatrix ist daher Ausgangspunkt der Faktorenanalyse. Sie wird häufig auch eingesetzt, wenn man annimmt, dass eine Serie von Messwerten verschiedene Variablen repräsentiert; dann soll die F. festlegen, welcher Messwert zu welchem Faktor gehört, bzw. entsprechende Vorab-Hypothesen testen. Man versucht nun, aus der Korrelationsmatrix Faktoren zu „extrahieren“; diese Faktoren sollen voneinander unabhängig sein. Nach bestimmten Verfahren werden solange Faktoren ermittelt, bis ein bestimmtes Stopp-Kriterium erreicht ist, bei dem die Annahme weiterer Faktoren keinen Erklärungsgewinn mehr verspricht. Im Allgemeinen sucht man anschließend die Faktoren so mit den Messwerten „abzugleichen“, dass alle Messwerte mit einem der Faktoren sehr hoch zusammenhängen (oder hoch auf ihm „laden“) und mit allen anderen nicht oder nur äußerst niedrig. Zu diesem Zweck werden in einem zweiten Schritt die ermittelten Faktoren „rotiert“ (Faktorenrotation). Die so ermittelten Faktoren müssen nunmehr „interpretiert“ werden, das heißt, man inspiziert die Zusammenhänge zwischen den einzelnen Messwerten und den Faktoren darauf hin, ob sich sinnvolle Ergebnisse gezeigt haben.

**Item:** Eine einzelne Frage oder Aufgabe eines Tests.

**Koeffizient:** Ergebnis eines statistischen Rechenverfahrens als Maß für die Stärke eines Zusammenhanges zwischen zwei Variablen.

**Korrelation:** Ein Maß für die Stärke und die Richtung des Zusammenhanges zwischen zwei Variablen auf einer Skala von  $-1,00$  bis  $+1,00$ .

**Kriteriumsbezogene Validität:** Ein Ansatz zur Validierung eines Tests, bei dem die Testwerte mit vorliegenden oder erst zu erhebenden Daten (etwa der Schulleistung) korreliert werden. Die gewählten Kriterien werden dabei als Ausdruck des Merkmals, das getestet werden soll, betrachtet.

**Mittelwert:** auch „M“ abgekürzt. Das arithmetische Mittel einer Wertereihe oder auch der Durchschnittswert.

**Motorikquotient:** Ein in einem Testverfahren errechneter Motorikquotient (MQ) beschreibt den Entwicklungsstand relativ zur Vergleichsstichprobe, z.B. zu Altersnormen.

**Normierung:** Erhebung von Daten zur Erstellung von Statistiken, die die Testleistungen bestimmter Gruppen zusammenfassen, zum Beispiel nach Alters- oder Klassenstufen. Dabei wird davon ausgegangen, dass Normen repräsentativ für größere Grundgesamtheiten sind, zum Beispiel alle Schüler Deutschlands.

**Objektivität:** Objektivität ist ein Maß dafür, wie weit der diagnostischen Situation eine Standardisierung des gesamten Testvorganges gelingt. Die Objektivität eines Tests gibt den Grad der Unabhängigkeit der Testergebnisse von den Untersuchern (Durchführungsobjektivität), bzw. den auswertenden Personen (Auswertungs- und Interpretationsobjektivität) an. In der Regel ist davon auszugehen, dass ein Test, der genaue einzuhaltende Instruktionen zur Anleitung und Auswertung vorgibt, eine hohe Objektivität aufweist. Dies wird dann auch statistisch mit einem Kennwert versehen, indem verschiedene Testanwender in ihrer Übereinstimmung bei Durchführung, Auswertung und Interpretation eines Tests/bzw. Testergebnisses beurteilt werden. Den Grad der Übereinstimmung (Interrater-Reliabilität) beschreibt dann ein Korrelationskoeffizient.

**Prozentränge (PR)** geben die relative Position eines Probanden innerhalb einer Häufigkeitsverteilung an. Die Rohwerte werden in eine 100er-Skala umgerechnet. So informiert ein PR darüber, wie viel Prozent der Kinder der Vergleichsgruppe in der Normstichprobe einen Testwert erzielten, der unter bzw. über dem Testwert des untersuchten Kindes liegt. Mit Prozenträngen darf nicht mathematisch operiert werden, weil

die Prozentrangskala in den Extrembereichen einen größeren Werteabstand aufweist als im Mittelbereich

**Pseudorandomisiert:** Das Zufallsprinzip der Zuordnung der Probanden zur Interventions- und Kontrollgruppe wird in einigen Parametern bewusst eingeschränkt.

**Randomisiert:** Die Interventions- und die Kontrollgruppe werden jeweils nach dem Zufallsprinzip zusammengesetzt.

**Reliabilität:** Messgenauigkeit eines Instrumentes. Auch: „Zuverlässigkeit, Verlässlichkeit“. Die Reliabilität ist neben der Validität eines der beiden wichtigsten Gütekriterien von Messungen, Tests und Assessments. Bei der Reliabilität geht es darum, mit welcher Verlässlichkeit bzw. in welchem Maß mehrfache Durchführungen einer Messung (wiederholt oder gleichzeitig) oder eines Tests zu gleichen Ergebnissen führen, vorausgesetzt natürlich, dass das zu Messende sich nicht verändert.

**Rohwerte** nennt man die Testergebnisse eines einzelnen Probanden, zum Beispiel in Form einer Punktgesamtschumme für alle richtigen Antworten in einem Test. Der Rohwert ist die Ausgangsbasis für die Umrechnung (Transformation) in andere Verteilungsmaße, beispielsweise in Prozentränge oder T-Werte. Diese Umrechnungen dienen der Vergleichbarkeit des Einzelwertes mit der Norm bzw. des Testwertes mit den Ergebnissen von anderen Tests mit ähnlichem Validitätsbereich.

**Schwierigkeitsindex:** Der Schwierigkeitsindex eines Items ist die prozentuale Häufigkeit der auf das Item entfallenden richtigen Lösungen. Leichte Aufgaben werden durch hohe Indizes gekennzeichnet, schwierigere durch niedrigere. Der Index sollte möglichst breit streuen, um auch in den Extrembereichen eine Differenzierung zu ermöglichen.

**Signifikanz:** Die Wahrscheinlichkeit, dass in einer Studie gemessene Gruppenunterschiede zufällig sind, wird mit dem p-Wert ausgedrückt, wobei eine Wahrscheinlichkeit von kleiner 0,05 (5 Prozent) per Konvention als signifikant galt. Heute tendiert man wieder zurück zu Fishers (einer der „Erfinder“ der Signifikanztests) Auffas-

sung von statistischer Signifikanz, der die Verantwortung über die Signifikanzentscheidung beim Forscher selbst sieht, die im Licht der jeweiligen Studie zu treffen ist (Fisher, 1935). Je kleiner die Wahrscheinlichkeit, dass das Ergebnis zufällig war, umso besser für den Studienleiter.

**Standardabweichung:** Auch SD abgekürzt. Eine Maßeinheit für die Verteilungsbreite einer Menge von Werten. Rechnerisch: die Quadratwurzel aus dem Durchschnitt der quadratischen Abweichungen der Werte vom Mittelwert.

**Standardisierung:** Festlegung der Durchführung eines Tests (Instruktionen, Vorgehen bei der Auswertung, usw.). Dieser Begriff wird von manchen Autoren auch als Bezeichnung für die Sammlung von Daten zur Normierung benutzt.

**Standardschätzfehler** (siehe auch Vertrauensintervall): Da jeder statistische Kennwert mit einem Messfehler behaftet ist, werden bei der Testentwicklung für Rohwerte oder T-Werte die Standardschätzfehler berechnet. Der Standardschätzfehler gibt an, welche Abweichung nach oben oder unten (Vertrauensintervall) man bei der Interpretation jeden Testwertes infolge der Unzuverlässigkeit des Testes berücksichtigen muss. Der gesuchte wahre (Test-) Wert einer Person entspricht also nicht einfach ihrem Prozentrang oder T-Wert, sondern liegt innerhalb des angegebenen Vertrauensintervalls.

**Stanine-Werte** sind eine Transformation der Rohwerte in neun Kategorien (standard nine). Diese Art der Transformation wird bei stetig steigenden Werten benutzt, zum Beispiel bei der Beurteilung der durchschnittlichen Äußerungslänge eines Kindes.

**Stichprobe:** Teil einer Menge, die repräsentativ für das Ganze sein soll.

**Trennschärfe.** Der Trennschärfekoeffizient eines Items ist die Korrelation zwischen dem Itempunktwert (meist 0 oder 1) und dem Gesamtpunktwert eines jeden Probanden der Analysestichprobe. Ein hoher Trennschärfekoeffizient trifft eine Aussage darüber, ob das Item erfolgreiche Probanden (mit hoher Punktzahl) von weniger erfolgreichen Probanden (mit niedriger Punktzahl) trennt. Ein niedriger Trennschärfekoeffizient besagt also, dass eine Aufgabe von guten und schlechten Probanden gleich häufig richtig gelöst wird, was sie als unbrauchbar für einen Test ausweist. Gleicherma-

ßen unerwünscht sind negative Trennschärfekoeffizienten, die demonstrieren, dass eine Aufgabe von guten Probanden eher verfehlt und von schlechten eher richtig gelöst wird. Erstrebenswert für den Testentwickler sind also hohe positive Koeffizienten, die der Erhöhung der Messgenauigkeit dienen.

**T-Werte** sind Standardmaße mit einem Mittelwert von 50 und einer Standardabweichung von 10. T-Werte werden durch Transformation eines Prozentranges oder eines Rohwertes errechnet. Auf der T-Wert-Skala entsprechen gleiche Abstände gleichen numerischen Differenzen. Mit T-Werten darf also gerechnet werden. T-Werte werden in der Regel folgendermaßen verbalisiert:

T-Wert 20-29: weit unterdurchschnittlich

T-Wert 30-39: unterdurchschnittlich

T-Werte 40-60: durchschnittlich

T-Wert 61-69: überdurchschnittlich

T-Wert > 70: weit überdurchschnittlich

**Validität:** Gültigkeit eines Messinstrumentes oder Studienergebnisses. In Bezug auf Studien wird zwischen der internen Validität und der externen Validität unterschieden. Eine Studie bzw. ihre Ergebnisse ist dann „intern valide“, wenn die Ergebnisse wenig oder gar keine anderen Erklärungen zulassen, als die von den Forschern vorgetragenen. Unter „externer Validität“ versteht man die Verallgemeinerungsfähigkeit einer Studie. In Bezug auf Tests ist die Validität das Ausmaß, in dem ein Test das misst, was er messen soll. Man unterscheidet hier Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität.

**Varianz.** Das Quadrat der Standardabweichung.

**Vertrauensintervall / Konfidenzintervall:**

Vertrauensintervalle werden herangezogen, um abschätzen zu können, wie nahe der ermittelte T-Wert dem wahren Wert kommt. Das heißt, dass somit Messfehler der Testung mit berücksichtigt werden. Auf der Basis der aus dem Standardschätzfehler ermittelten Vertrauensintervalle werden Prozentrang- oder T-Wertbänder ermittelt. Der gesuchte wahre (Test-)Wert einer Person liegt also innerhalb des angegebenen Vertrauensintervalls.

-Schutz

in welcher Reihenfolge die Items präsentiert werden sollen und wie die Materialanordnung sein soll, ob Abbruchkriterien bestehen, wenn mehrere Aufgaben hintereinander nicht gelöst werden.

Verschiedene Auswertende sollten bei einem Probanden zu gleichen Testergebnissen, zum Beispiel in Form von Punktwerten, kommen. Dazu sind genaue Auswertungsvorschriften notwendig, z.B. bei Sprachtests Tabellen mit Wortlisten und Beispielauswertungen. Ebenso sollte die Interpretation des Ergebnisses bei jedem Auswertenden gleich ausfallen, was durch standardisierte Interpretationen und Profilübersichten erleichtert wird.

## Reliabilität

Die Zuverlässigkeit oder Reliabilität eines Tests gibt an, wie genau ein Test misst. Die Frage ist hier: Führt der Test, wenn er wiederholt angewendet wird, zu denselben Ergebnissen? Man kann in Bezug auf die Messgenauigkeit die „innere Konsistenz“ eines Tests und seine „Stabilität“ schätzen. Folgende statistische Analysen werden dabei angewandt:

- Berechnung des mittleren Zusammenhangs aller Items einer Skala unter Berücksichtigung der Testlänge („innere Konsistenz“ und „Halbierungsreliabilität“)
- wiederholte Durchführung desselben Tests an den gleichen Probanden („Retest-Reliabilität“ oder „Stabilität“) und
- der Vergleich zweier Testformen mit ähnlichen Items („Paralleltest-Reliabilität“).

Jedes Testinstrument muss in einer sogenannten „Konsistenzanalyse“ auf seine Messgenauigkeit geprüft werden. Dazu werden Berechnungen zur inneren Konsistenz durchgeführt. Dabei wird jedes Item als eigenständiger Testteil angesehen und der mittlere Zusammenhang zwischen den Items in Bezug zur Testlänge bestimmt. Die errechneten Zusammenhänge geben Aufschluss über die Itemzusammensetzung des Tests.

Bei der Halbierungsmethode („Halbierungsreliabilität“, „Split-half-Methode“) wird ein Test in zwei Teile aufgeteilt. Dabei können verschiedene Halbierungstechniken – z.B. nach geraden oder ungeraden Itemnummierungen oder einfach per Zufall – angewandt werden. Eine Halbierung kann auch aufgrund der Analysedaten erfolgen. Man bildet hierbei Aufgabenpaare von annähernd gleicher Schwierigkeit und gleicher Trennschärfe. Dann wird der Zusammenhang beider Testhälften berechnet, je höher er ausfällt, umso

reliabler ist das Verfahren. Die Methode hat zur Voraussetzung, dass der Test aus homogenen Niveaueaufgaben besteht.

Reliabel ist ein Test dann, wenn die oben genannten Berechnungen hohe Zusammenhänge ergeben. In verschiedenen statistischen Verfahren wird dieser Grad ermittelt und in Form von „Konsistenzkoeffizienten“ beschrieben. Dazu existieren verschiedene Rechenverfahren, die je nach Art der Ausgangsdaten ausgewählt werden. Koeffizienten können Werte zwischen 0 und 1 annehmen. Ein Koeffizient von 0 bedeutet, dass kein Zusammenhang zwischen den Werten bzw. Variablen besteht. Ein möglichst hoher Wert (z.B. .9) bedeutet, dass das Verfahren reliabel ist.

## Validität

Valid ist ein Test dann, wenn er auch wirklich das Merkmal erfasst, das er vorgibt zu messen. Somit ist das Kriterium der Validität bei Befundinstrumenten das wichtigste Gütekriterium. Man unterscheidet drei verschiedene Validitätsarten.

► **Inhaltliche Validität.** Bei der inhaltlichen Validität folgt man der – nur durch psychologische Einsicht begründbaren – Annahme, dass die Testaufgaben selbst das bestmögliche Kriterium für das zu untersuchende Persönlichkeitsmerkmal darstellen. Für manche Tests ist diese Übereinstimmung unmittelbar einsichtig, wie zum Beispiel bei Rechtsschreibtests. Bei anderen Tests wird durch Experten beurteilt, ob die gewählten Testaufgaben das Konstrukt, also die zu messende Fähigkeit oder Eigenschaft, repräsentieren.

► **Kriteriumsvalidität.** Hierbei handelt es sich um die Berechnung des Zusammenhangs der Testleistung mit anderen Merkmalen bzw. Kriterien, wie zum Beispiel der Schulnote, dem Intelligenzquotienten oder Ähnlichem, von denen angenommen wird, dass sie ausreichend valide Repräsentanten für das zu messende Merkmal sind. Bei einem zeitlich koexistenten Außenkriterium spricht man von „Übereinstimmungsvalidität“ („konkurrente Validität“) – zum Beispiel der Zusammenhang IQ und Sprachentwicklung. Bei einem zukünftigen Außenkriterium spricht man von „Vorhersagevalidität“ („prädiktive Validität“), zum Beispiel das Ausmaß der Sprachentwicklungsstörung als Vorhersage der späteren Deutschnote. Die Testergebnisse lassen dann Prognosen auf die künftige schulische Entwicklung eines Kindes zu.

► **Konstruktvalidität.** Bei einer Konstruktvalidierung geht man von der dem Test

zu Grunde liegenden Theorie aus und leitet daraus Hypothesen ab, die unter Einsatz des zu validierenden Tests empirisch überprüft werden. Bei der „diskriminanten“ oder „divergenten Validität“ werden Korrelationen mit Tests anderer Gültigkeitsbereiche ermittelt. Beispielsweise wird der Zusammenhang eines Konzentrationstests mit einem Lese-Rechtschreib-Test ermittelt, da der LRS-Test möglichst rein das Konstrukt Lese- und Rechtschreibfähigkeit abbilden soll und nicht zum Beispiel die Konzentrationsfähigkeit. Erwartet würden hier eher niedrige bis mittlere Zusammenhänge. Bei der „konvergenten Validität“ hingegen sollen positive Zusammenhänge des zu validierenden Tests mit Tests mit ähnlichem Validitätsanspruch aufgezeigt werden, z.B. der Zusammenhang eines neuen Sprachentwicklungstests mit einem bereits etablierten. Neben der Korrelationsanalyse kommt auch häufig eine „Faktorenanalyse“ zum Einsatz. Dies ist ein statistisches Verfahren zur Zusammenfassung homogener Inhaltsbereiche und zur Aussonderung von konstruktfernen Bereichen. Das Ergebnis dieser Analyse ist die Angabe der Anzahl der im Test enthaltenen Faktoren (bei mehrdimensionalen Tests) zusammen mit der Wichtigkeit des einzelnen Faktors im Test.

► **Normierung.** Ein psychometrischer Test ist in der Regel normiert. Man möchte die Leistung einer zu testenden Person mit dem Durchschnitt der Leistungen von anderen Personen vergleichen, um zu erfahren, ob sie gleichwertig, über- oder unterdurchschnittlich abgeschnitten hat. Um diesen Vergleich vornehmen zu können, müssen Normen für einen Test vorliegen. Diese werden in einer Untersuchung erhoben, in der der Test an einer hinreichend großen Stichprobe durchgeführt wurde. Aus diesen Testergebnissen werden Kennwerte in Form von Mittelwerten und Streuungsmaßen berechnet, die als Vergleichsbasis für die Einzelergebnisse dienen. Die Normierungsstichprobe sollte „randomisiert“, also zufällig zusammengestellt sein.

Zur Stichprobengewinnung gibt es mehrere Möglichkeiten, zum Beispiel die Zufallsstichprobe, die Quotenstichprobe oder die Klumpenstichprobe. Bei der Zufallsstichprobe werden die Probanden zufällig aus der Grundgesamtheit aller möglichen Personen ausgewählt. Bei der Klumpenstichprobe wird ein natürlich vorhandener „Klumpen“, zum Beispiel eine Kindergartengruppe, untersucht. Bei der am häufigsten angewandten Methode, der Quotenstichprobe, wird anhand von verschiedenen Kriterien, beispielsweise

■ Tab. 5: Beurteilungskriterien für Tests

1. Allgemeiner Zweck
2. Testmaterial/ Testaufbau
3. Allgemeines Grundkonzept
4. Testdurchführung / -auswertung
5. Gütekriterien des Tests
6. Normen
7. Interpretation der Ergebnisse

se Alter oder Geschlecht, die Grundgesamtheit in möglichst homogene Gruppen eingeteilt, zum Beispiel Altersgruppen in Halbjahresabständen. Je nach Aufbau der Stichprobe sollten mehr als 100 Personen getestet werden. Bei Untergruppen, zum Beispiel nach Altersschritten, muss jede Untergruppe entsprechend vertreten sein. Die Stichprobe sollte repräsentativ für diejenige Population sein, für die der Test entwickelt wurde, das heißt, sie sollte in ihrer Zusammensetzung die wahren Gegebenheiten widerspiegeln.

Soll beispielsweise ein Test zur Erfassung des aktiven Wortschatzes von vier- bis sechsjährigen deutschsprachigen Kindern entwickelt werden, sollte sich die Stichprobe aus Kindern aus Deutschland zusammensetzen, Jungen und Mädchen umfassen und auch in verschiedenen Sprachregionen gezogen worden sein, um als repräsentativ für alle deutschsprachigen Kinder dieser Alterstufe gelten zu können. Aber auch andere Variablen könnten die Testergebnisse verfälschen. So hängt sicherlich die soziale Schichtzugehörigkeit mit der Größe des Wortschatzes eines Kindes zusammen und müsste bei der Stichprobensammensetzung berücksichtigt werden.

► **Vergleichbarkeit.** Ein Test ist dann vergleichbar, wenn eine oder mehrere Parallelformen oder Tests mit gleichem Gültigkeitsbereich vorliegen.

► **Ökonomie.** Von Tests erwartet man, dass der Aufwand der Testung im Verhältnis zum Nutzen steht. Ein Test ist dann ökonomisch, wenn er eine kurze Durchführungszeit beansprucht, wenig Material verbraucht, einfach zu handhaben sowie schnell und bequem auswertbar ist. Weiterhin gelten Tests als ökonomisch, die als Gruppentest durchführbar sind. Dies trifft auf Sprachentwicklungstests oder klinische Tests häufig jedoch nicht zu.

► **Nützlichkeit.** In Bezug auf die Auswahl eines geeigneten Testverfahrens bezeichnet die Nützlichkeit eines Tests den Wert eines Testergebnisses im Vergleich zu anderen möglichen Testergebnissen ähnlicher Testverfahren.

## Auswahlkriterien für Tests

Um zu beurteilen, welcher Test für welche Fragestellung geeignet ist, empfiehlt es sich, in gängigen Testhandbüchern (z.B. *Brähler et al., 2002; Beushausen, 2007*) einen Überblick über die wichtigsten Tests und deren Beurteilungskriterien (siehe Tab. 5) zu gewinnen.

Ist ein Test einmal ausgewählt, kommt die Testdurchführende nicht umhin, das gesamte Manual zu lesen. Dieses sollte so gestaltet sein, dass es verständlich und vollständig Auskunft zu den einzelnen Beurteilungskriterien gibt. Hilfreiche Beurteilungskriterien für Tests werden nachfolgend erläutert.

► **Allgemeiner Zweck.** Um was für einen Test handelt es sich? Der Test sollte in das in Tab. 1 dargestellte Klassifikationssystem eingeordnet werden. Welchen Geltungsbereich hat der Test? Was lässt sich mit dem Test erfassen (Ziel der Testung?), was nicht?

► **Testmaterial/Testaufbau.** Wie ist das Testmaterial gestaltet? Welche Materialien werden mitgeliefert, welche müssen eventuell selbst besorgt werden? Wie viele Subtests mit wie vielen Items enthält der Test? Handelt es sich um einen Gruppen- oder Einzeltest, liegt eine Kurzform des Tests vor?

► **Allgemeines Grundkonzept.** Welches theoretische Grundkonzept liegt dem Test zugrunde? Entspricht dies der gängigen Auffassung von Sprache bzw. Spracherwerb? Werden die dem Test zugrunde liegenden theoretischen Annahmen und Modelle erläutert, durch Zitate der Autoren belegt und wird beschrieben, inwieweit das Testkonzept die theoretischen Annahmen realisiert?

► **Testdurchführung und -auswertung.** Wie wird der Test durchgeführt und ausgewertet? Wird die Durchführung der Testung in allen Einzelheiten erläutert und werden die Anforderungen an Testleiterin und Probanden genannt? Wird der Auswertungsvorgang hinreichend beschrieben? Wird die Art der Kennwerte angegeben sowie auf in der Praxis auftretende Auswertungsprobleme hingewiesen?

► **Gütekriterien des Tests.** Entspricht der Test den testtheoretischen Gütekriterien (Objektivität, Reliabilität, Validität)? Welche statistischen Verfahren wurden mit welchem Ergebnis durchgeführt? Wie werden diese Ergebnisse von den Testautoren interpretiert?

► **Normen.** Wurde der Test normiert? Welche Transformationsmaße stehen zur Umrechnung der Rohwerte zur Verfügung? War die Normstichprobe ausreichend groß und repräsentativ für die Zielgruppe des Tests?

► **Interpretation.** Was sagt das Testergebnis aus? Ergeben sich Aussagen zum Vorliegen einer Störung oder lediglich Hinweise auf das Risiko, eine Störung zu entwickeln (wie z.B. beim SETK-2)? Können Profile über verschiedene Bereiche der Sprachfähigkeit erstellt werden, die bereits Hinweise auf die Therapieplanung zulassen?

In der Regel haben Logopädinnen und Sprachtherapeutinnen das Bedürfnis, mit einem Test (a) eine eindeutige Klassifikation einer Störung zu erlangen, (b) Hinweise auf mögliche Ziele und Inhalte der Therapie von Kommunikationsstörungen abzuleiten und (c) sprachliches Verhalten auf verschiedenen Ebenen zu messen. Dies kann jedoch selten mit einem einzelnen Verfahren abgedeckt werden. Die zentrale Frage bei der Testauswahl ist deshalb zunächst: Was ist der Hauptzweck der Testung? Soll eine Diagnose gestellt oder die Therapieplanung für eine bereits diagnostizierte Störung erstellt werden? Was soll mit dem Testergebnis geschehen und welche Relevanz hat es für den Probanden? Es macht wenig Sinn, einen Motoriktest durchzuführen, wenn mit dem Ergebnis nicht gearbeitet wird. Eine Orientierung für Therapeutinnen gibt hier der Leitfaden für Testanwenderinnen (siehe Tab. 6). Neben der therapeutischen Relevanz bildet auch die Ökonomie ein wichtiges Auswahlkriterium für ein Testverfahren aus Sicht einer Therapeutin: Wie ökonomisch ist der Test hinsichtlich Durchführungs- und Auswertungsdauer sowie Einarbeitungsaufwand und Anschaffungspreis?

## Wer testet?

Innerhalb der Heilmittel-Richtlinien und der Rahmenverträge mit den Krankenkassen wird die logopädische Diagnostik als eigenständige Tätigkeit und Abrechnungsposition

■ Tab. 6: Leitfaden für TestanwenderInnen (aus *Beushausen, 2007*)

- Welche Fragestellung habe ich?
- Wie sind meine Interessen bei der Testanwendung?
- Welche Konsequenzen hat die Testung?
- Welche alternativen Informationsquellen gibt es?
- Wie ökonomisch ist der Test?
- Welche Relevanz für die Therapie hat der Test?

explizit beschrieben und auch erste Testverfahren werden dort benannt, die zum Einsatz kommen sollen. Wer im klinischen und sonderpädagogischen Bereich Tests durchführen darf, entnehmen wir den „Standards für pädagogisches und psychologisches Testen“ (Häcker et al., 1998). Dort heißt es in Standard 6.6, testen sollte nur, wer „eine ausreichende Ausbildung und Erfahrung vorweisen (kann), um dieser Verantwortung fachlich und technisch in angemessener Weise zu entsprechen. Alle besonderen Qualifikationen für die Testdurchführung oder -interpretation, die das Handbuch angibt, sollten erfüllt werden“ (ebd., S. 50). In der Regel wird im Handbuch zu einem Test die Zielgruppe derjenigen, die den Test durchführen sollen, beschrieben, oder es werden die notwendigen Voraussetzungen für die Erfüllung der Funktion des Testleiters genannt. Wer einen Sprachentwicklungstest durchführt, sollte zum einen „fachlich“ über Sprachentwicklungstheorien und Abläufe der Sprachentwicklung informiert sein. Zum anderen sollte eine Testanwenderin Wissen über testtheoretische Grundlagen haben.

## Darstellung der Testergebnisse

Die Testergebnisse einer einzelnen Person werden anhand der vergleichbaren Gruppe der Normstichprobe in verschiedene Normwerte umgerechnet, zum Beispiel „T-Werte“ oder „Prozentränge“. Die Berechnung der aus der Normierung gewonnenen Normwerte erfolgt anhand der Normalverteilung, der Gauß'schen Glockenkurve, die eine wichtige Bezugsgröße für statistische Berechnungen darstellt. Ziel ist es, die Rohwerte in ein aussagekräftiges Vergleichsmaß zu transformieren. Alle Transformationsmaße haben Vor- und Nachteile in ihrer Genauigkeit. Die Prozentrangskala differenziert beispielsweise in den Randbereichen (hohe und niedrige Prozentränge) eher ungenau. Deshalb empfiehlt es sich, bei der Angabe von verschiedenen Normskalen in einem Test die Ergebnisse zu vergleichen.

## Resümee

Standardisierte Testverfahren erlauben nicht nur eine präzise, objektive Diagnostik von Kommunikationsstörungen, sie sind auch unverzichtbare Bestandteile von Therapieevaluation und Dokumentation. Durch die Möglichkeit, ein einzelnes Testergebnis eines Patienten mit den erhobenen Durchschnitts-

werten der Normdaten zu vergleichen, lassen sich therapiebedingte Veränderungen in ihrer relativen Größe quantifizieren (Wie groß war die Verbesserung dieses Klienten? Ist sie überzufällig?) und auch die subjektive Wahrnehmung einer einzelnen Therapeutin lässt sich objektivieren. Sinnvoll kombiniert mit informellen Prüfverfahren und qualitativen Erhebungsverfahren sind Tests die Basis jeder logopädischen Tätigkeit.

Der Einsatz standardisierter Tests sollte jedoch stets mit der kritischen Auseinandersetzung über die Vor- und Nachteile der einzelnen Verfahren erfolgen sowie in dem Wissen, wie ein Test entsteht, was einen guten Test ausmacht und wie man sinnvoll testet.

## Literatur

Beushausen, U. (2007). *Testhandbuch Sprache. Diagnostikverfahren der Logopädie und Sprachtherapie*. Bern: Hans Huber

Brähler, E., Holling, H., Leutner, D. & Petermann, F. (2002). *Brickenkamp Handbuch psychologischer und pädagogischer Tests, Band 2*. Göttingen: Hogrefe

Fisher, R.A. (1935). *The design of experiments*. Edinburgh: Oliver & Boyd

Häcker, D., Leutner, D. & Amelang, M. (1998). *Standards für pädagogisches und psychologisches Testen*. Göttingen: Hogrefe

Linert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (5. Auflage). Weinheim: Beltz

## Testquellen

*Aachener Aphasie Test (AAT)*. W. Huber, K. Poeck, D. Weniger, K. Willmes. Göttingen: Hogrefe 1983

*Aktiver Wortschatztest (AWST 3-5)*. C. Kiese-Himmel. Göttingen: Hogrefe 2005

*Aphasie-Checkliste (ACL)*. E. Kalbe, N. Reinhold, U. Ender, J. Kessler, unter Mitarbeit von M. Brand. Köln: ProLog 2002

*Aphasie-Schnell-Test (AST)* (2. verbesserte Auflage). C. Kroker. Idstein: Schulz-Kirchner 2002

*Bogenhauser Semantikuntersuchung (BOSU)*. R. Glinde-mann, D. Klintwort, W. Ziegler, G. Goldenberg. München: Urban & Fischer/Elsevier 2002

*Elternfragebogen für die Früherkennung von Risikokindern (ELFRA)* (2. Aufl.). H. Grimm, H. Doil. Göttingen: Hogrefe 2006

*Heidelberger Sprachentwicklungstest (HSET)* (2. verbesserte Auflage). H. Grimm, H. Schöler. Göttingen: Hogrefe 2001

*Kindersprachtest für das Vorschulalter (Kiste)*. D. Häuser, E. Kasielke, U. Scheidereiter. Göttingen: Hogrefe 1994

*Kurze Aphasieprüfung (KAP)*. C. Lang, A. Dehm, B. Dehm, T. Leuschner. Frankfurt/AM.: Harcourt 1999

*Marburger Sprachverständnistest für Kinder (MSVK)*. C. Elben, A. Lohaus. Göttingen: Hogrefe 2000

*Patholinguistische Diagnostik bei Sprachentwicklungsstörungen*. C. Kauschke, J. Siegmüller. München: Urban & Fischer/Elsevier 2002

*Psycholinguistischer Entwicklungstest (PET)* (2. korrigierte Auflage). M. J. W. Angermaier. Göttingen: Hogrefe 1977

*Regensburger Wortflüssigkeitstest (RWT)*. S. Aschenbrenner, O. Tucha, K. W. Lange. Göttingen: Hogrefe 2000

*Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5)*. H. Grimm, unter Mitarbeit von M. Aktas und S. Frevert. Göttingen: Hogrefe 2001

*Sprachentwicklungstest für zweijährige Kinder (SETK-2)*. H. Grimm, unter Mitarbeit von M. Aktas und S. Frever. Göttingen: Hogrefe 2000

*Sprachscreening für das Vorschulalter (SSV)*. H. Grimm, unter Mitarbeit von M. Aktas und U. Kießig. Göttingen: Hogrefe 2003

*Teddytest*. G. Friedrich. Göttingen: Hogrefe 1998

*Test zur Überprüfung des Grammatikverständnisses (TROG-D)*. A. Fox. Idstein: Schulz-Kirchner 2006

*Wortschatz- und Wortfindungstest für 6- bis 10-Jährige (WWT 6-10)*. Chr. Glück. München: Urban & Fischer/Elsevier 2007

## Internetquellen zum Selbststudium

*ILMES – Internet-Lexikon der Methoden der empirischen Sozialforschung*

[www.lrz-muenchen.de/~wlm/ilmes.htm](http://www.lrz-muenchen.de/~wlm/ilmes.htm)

*Wörterbuch Wissenschaft*

[www.physio-akademie.de/Woerterbuch\\_Wissenschaft.82.0.html](http://www.physio-akademie.de/Woerterbuch_Wissenschaft.82.0.html)

## Autorin

Prof. Dr. Ulla Beushausen  
Bachelor- und Masterstudiengang für Ergotherapie, Logopädie und Physiotherapie  
Bereich Logopädie  
HAWK Hochschule für Angewandte Wissenschaft und Kunst  
Fachhochschule Hildesheim/Holzminde/Göttingen  
Goschentor 1  
31134 Hildesheim  
Beushausen@hawk-hhg.de

## SUMMARY. The use of standardized tests in speech and language therapy and pathology.

The diagnosis of communication disorders is based on a range of standardized and norm-referenced tests. Criteria for quality in tests are defined (objectivity, reliability, validity) and selection criteria for an appropriate test in therapeutic practice are discussed.

Key Words: Tests – communication disorders – objectivity – reliability – validity – diagnosis – speech and language therapy – test theory