MERCATOR INSTITUT
für Sprachförderung
und Deutsch
als Zweitsprache

Eine Analyse und Bewertung

Die Qualität von Sprachstandsverfahren im Elementarbereich

Uwe Neugebauer
Michael Becker-Mrotzek

Herausgegeben vom
Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache

MERCATOR
INSTITUT
für Sprachförderung
und Deutsch
als Zweitsprache

4 **DAS WICHTIGSTE IN KÜRZE**

6 **RELEVANZ**

9 **VORGEHEN**

12 **ANALYSE UND BEWERTUNG**
12 Berücksichtigung sprachlicher Basisqualifikationen
16 Validität
19 Objektivität
22 Reliabilität
24 Normierung
27 Fehlerquote (Gewichtung der Auswahlfehler)
30 Qualifizierung der pädagogischen Fachkräfte
32 Zeitliche Anforderungen
34 Mehrsprachigkeit
38 Spezifität der Diagnostik

41 **FAZIT UND EMPFEHLUNGEN**

45 **GLOSSAR**

48 **AUTOREN**

51 **QUELLEN**

54 **ANHANG**

56 **IMPRESSUM**

DAS WICHTIGSTE IN KÜRZE

HINTERGRUND

Eine bedarfsgerechte Sprachförderung muss möglichst früh ansetzen, um allen Kindern optimale Startchancen in ihre Schullaufbahn zu ermöglichen. Das Ziel von Sprachstandsverfahren im Elementarbereich ist es daher, sicher und verlässlich die Kinder zu identifizieren, die Unterstützung bei ihrer sprachlichen Entwicklung benötigen. Zwar haben nahezu alle Bundesländer zu diesem Zweck Sprachstandsverfahren eingeführt, die Qualität und Wirksamkeit sind allerdings fraglich: Derzeit sind 21 unterschiedliche Tests, Screenings bzw. Beobachtungsbögen offiziell im Einsatz, die Förderquoten schwanken je nach Bundesland zwischen zehn und 50 Prozent. Bisher gibt es keine vergleichbaren Qualitätsstandards, die Chance auf eine bedarfsgerechte Sprachförderung ist abhängig vom Wohnort und vom sozialen Hintergrund.

Auf Initiative des Mercator-Instituts für Sprachförderung und Deutsch als Zweitsprache hat eine interdisziplinäre Expertenkommission Qualitätsmerkmale entwickelt, die ein Sprachstandsverfahren erfüllen sollte. Für jedes Qualitätsmerkmal ist eindeutig definiert, ab wann ein Verfahren es sehr gut, gut, ausreichend oder nicht ausreichend erfüllt. Nun legt das Mercator-Institut eine Analyse und Bewertung vor, inwieweit die Verfahren diesen Qualitätsmerkmalen entsprechen. Ziel der Studie ist es, eine Diskussionsgrundlage für die Bundesländer und Akteure aus der Testentwicklung zu schaffen, wie der Weg zu einer chancengerechten Erhebung des Sprachstands und anschließender bedarfsgerechter Sprachförderung aussehen kann.

ERGEBNISSE

Die Entwicklung von Sprachstandsverfahren ist insgesamt auf einem guten Weg: Immer mehr Verfahren genügen diagnostischen Standards, wie sie in der Psychologie zugrunde gelegt werden. Nahezu jedes Merkmal wird mindestens einmal gut oder sogar sehr gut erfüllt. Diese Entwicklung ist auch darauf zurückzuführen, dass sich eine interdisziplinäre Testentwicklung durchsetzt. Bei den Ländern und in der Entwicklung ist zudem eine große Bereitschaft zu erkennen, die kritische Analyse und Bewertung ihrer Verfahren zu unterstützen und daraus Konsequenzen zu ziehen.

Bei allen positiven Entwicklungen zeigt diese Analyse und Bewertung jedoch auch, dass nach wie vor großer Handlungsbedarf besteht. Die Hälfte aller Verfahren erfüllt 16 oder weniger der insgesamt 32 Qualitätsmerkmale in den zehn Handlungsfeldern ausreichend. Insbesondere Beobachtungsbögen und Einschätzverfahren erreichten häufig eine nicht ausreichende Bewertung. Der größte Handlungsbedarf

besteht bei den Qualitätsmerkmalen Berücksichtigung sprachlicher Basisqualifikationen, Validität, Objektivität sowie Mehrsprachigkeit. Über die Hälfte der Verfahren kann diese Merkmale nicht ausreichend erfüllen.
In einigen Handlungsfeldern erzielen die Verfahren schon gute Ergebnisse: Bei den fünf formulierten Qualitätsmerkmalen der Normierung erzielen zwischen elf und 14 der Verfahren eine mindestens ausreichende Bewertung. Ähnliches gilt für das Merkmal Zeitliche Anforderungen an die Kinder: 15 Verfahren nehmen weniger als 30 Minuten in Anspruch, viele der 15 Verfahren liegen sogar deutlich darunter mit 15–20 Minuten.

EMPFEHLUNGEN

Konsens über Ziel und Funktion von Sprachstandsverfahren herstellen: Bisher gibt es noch keine hinreichend konsensfähigen Ziel- und Funktionsvorstellungen für die eingesetzten Verfahren. Sie werden teilweise für Zwecke verwendet, für die sie nicht konzipiert und daher nicht geeignet sind. Die Zwecke und Funktionen müssten zunächst formuliert werden, um sich auf Merkmale und Kriterien zur Bewertung der Güte zu einigen, z. B. über die prognostische Validität.

Empirische Belege für die prognostische Qualität ermitteln: Für eine wissenschaftlich fundierte Weiterentwicklung fehlt es nach wie vor an empirischen Belegen, sodass die Erkenntnisse dieser Untersuchung nur vereinzelt unmittelbar in die Weiterentwicklung der eingesetzten Verfahren und in laufende Sprachförderprogramme einfließen können. Ein zentrales Desiderat stellt die prognostische Validität dar: Nur so ist eine Antwort auf die Frage möglich, welche Verzögerungen in der sprachlichen Entwicklung eines Kindes gravierende Auswirkungen auf den späteren Schulerfolg haben. Hier sollte eine Längsschnittstudie ansetzen, die eine überschaubare Anzahl von Verfahren mit sehr unterschiedlichen Funktionalitäten prüft.

Keine Förderung ohne Diagnose – keine Diagnose ohne Förderung: Diagnostik und Förderung gehören zusammen. Die Ableitung konkreter Fördermaßnahmen sollte erst bei den Kindern erfolgen, bei denen ein Sprachförderbedarf festgestellt wurde, da die Ermittlung des spezifischen Sprachförderbedarfs sehr aufwendig ist. Ein zweistufiges Verfahren erscheint also empfehlenswert.

Qualifizierung pädagogischer Fachkräfte: Die zentrale Stellschraube für einheitliche Qualitätsstandards in der Sprachstandsfeststellung ist die Qualifizierung pädagogischer Fachkräfte. Aufseiten der Testentwicklung ergibt sich daraus die Aufgabe, die Verfahren am Bedarf der Praxis auszurichten. Damit Erzieherinnen und Erzieher Sprachstandsdiagnostik und Sprachförderung als Teil ihres Tätigkeitsfeldes akzeptieren, müssen die Verfahren konkrete Hinweise für die Förderung liefern.

Bundesländerübergreifender Transfer einheitlicher Qualitätsstandards: Die Bundesländer sollten in einen Dialog darüber eintreten, ob es zielführend ist, sich gemeinsam auf weniger Verfahren zu verständigen, die jeweils festgelegte Ziele und Funktionen haben. Diese können dann unter Wahrung hoher Qualitätsstandards weiterentwickelt und optimiert werden. Diese Aufgabe sollte ein interdisziplinäres, unabhängiges Gremium übernehmen, der Impuls dafür sollte jedoch aus den Ländern selbst kommen. Das Mercator-Institut hält daher an seiner Empfehlung fest, eine Koordinierungsstelle zu schaffen. Aufgrund der unterschiedlichen Zuständigkeiten für Sprachstandsverfahren auf Länderebene muss es den Bundesländern überlassen bleiben, wie eine solche Stelle strukturell eingegliedert wird.

RELEVANZ

Eine bedarfsgerechte Sprachförderung muss möglichst früh ansetzen, um allen Kindern optimale Startchancen in ihre Schullaufbahn zu ermöglichen. Dass dies auch eine angemessene Sprachdiagnostik im Elementarbereich erfordert, wurde bildungspolitisch schon seit Längerem erkannt: Nahezu alle Bundesländer haben Verfahren eingeführt, um den sprachlichen Entwicklungsstand der vier- bis fünfjährigen Kinder festzustellen und ggf. angemessene Sprachfördermaßnahmen noch vor dem Schuleintritt einzuleiten. In den 1980er-Jahren entwickelten zunächst einzelne Bildungseinrichtungen kleine Beobachtungsbögen (skizziert z. B. in Gogolin, 1988). Untersuchungen aus dem Zeitraum 2002 bis 2007 bescheinigten den Verfahren jedoch noch immer eine hohe Heterogenität und eine zumeist nicht ausreichende Qualität (z. B. Fried, 2004). Seitdem hat es in der Testentwicklung enorme Fortschritte gegeben: Zahlreiche Verfahren sind in interdisziplinären Teams entwickelt worden, wie von Bredel (2007) gefordert. Auch die Empfehlungen von Fried (2004) und Ehlich et al. (2005), welchen Anforderungen Sprachstandsverfahren genügen sollten, finden bei der Entwicklung zunehmend Berücksichtigung.

Dennoch bleiben die Qualität und Funktionalität vielfach weiter offen: Zwar haben alle Bundesländer außer Thüringen flächendeckend einzelne Verfahren eingeführt, in nur elf Bundesländern sind diese jedoch für alle Kinder verpflichtend. In Bayern, Rheinland-Pfalz, Hessen und Mecklenburg-Vorpommern sind die Verfahren freiwillig bzw. nur für einzelne Gruppen obligatorisch (vgl. Tab. 1). Bei einer Umfrage in knapp 2.000 Kitas im Jahr 2007 gab rund ein Drittel an, dass sie selbst erstellte Beobachtungsbögen nutzen (Klich, Wittenborn & Koch, 2010). Diese Heterogenität spiegelt auch der Anteil der Kinder wider, der letztendlich gefördert wird: Je nach Bundesland schwanken die Förderquoten zwischen zehn und 50 Prozent. Bund und Länder haben mehrfach angekündigt, einheitliche Standards für Sprachstandsverfahren zu entwickeln, zuletzt im Nationalen Aktionsplan Integration 2012. Deren Umsetzung steht bisher aus. So kommt auch der Bildungsbericht 2012 zu dem Schluss, dass „die Länder nach wie vor auffallend unterschiedlich vorgehen, was zu einem uneinheitlichen Umgang mit den Kindern führt". Eine umfassende empirische Überprüfung der Verfahren ist bisher nur vereinzelt erfolgt, auch die Vergleichbarkeit wurde bislang nicht übergreifend geprüft.

Sprachstandsverfahren müssen so angelegt sein, dass sie jedem Kind, unabhängig von Wohnort oder sozialem Hintergrund, Zugang zu einer bedarfsgerechten Sprachförderung ermöglichen. Das Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache hat daher eine interdisziplinäre Expertenkommission initiiert, die wissenschaftliche Qualitätsmerkmale für Sprachstandsverfahren im Elementarbereich

entwickelt hat. Ziel des Kriterienkatalogs war es, eine Grundlage für die Analyse und Bewertung der eingesetzten Verfahren zu schaffen. Die (Weiter-)Entwicklung von Sprachstandsverfahren im Elementarbereich muss sich an klar und eindeutig formulierten wissenschaftlichen Anforderungen orientieren, wie auch die Betroffenen und Beteiligten erwarten können, dass diagnostische Entscheidungen auf der Grundlage nachprüfbarer Fakten vorgenommen werden. Hierzu geben die Qualitätsmerkmale möglichst eindeutige Vorgaben, die von jedem Verfahren anzustreben und auch weitgehend zu erreichen sind.

Mit dieser Publikation legt das Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache nun eine Analyse und Bewertung vor, inwieweit die derzeit offiziell eingesetzten Sprachstandsverfahren diese Qualitätsmerkmale erfüllen. Sie zeigt die Diskrepanz zwischen Anspruch und Wirklichkeit auf und gibt den verantwortlichen Akteuren konkrete Hinweise für Veränderungen und Anpassungen. Der Überblick soll eine Grundlage für den Dialog mit den Bundesländern und den Testentwicklern darüber sein, wie der Weg zu einer chancengerechten Erhebung des Sprachstands und anschließender bedarfsgerechter Sprachförderung aussehen kann.

Bundesland	Sprachstandsverfahren	Verpflichtend[f]
Baden-Württemberg	HASE (alle Kita-Kinder), SETK 3-5 (Kinder, bei denen ein Förderbedarf festgestellt wurde, und Nicht-Kita-Kinder), optional LiSe-DaZ	ja
Bayern	Sismik & Seldak (Kita-Kinder), Kenntnisse in DaZ erfassen (mehrsprachige Nicht-Kita-Kinder)	ja[b]
Berlin	Deutsch Plus 4 (Nicht-Kita-Kinder) und QuaSta (Kita-Kinder)	ja
Brandenburg	Meilensteine der Sprachentwicklung, KISTE (Kinder, bei denen ein Förderbedarf festgestellt wurde)	ja
Bremen	Cito Sprachtest Version 3	ja
Hamburg	Vorstellungsverfahren Viereinhalbjähriger (umfasst „Bildimpulse“)	ja
Hessen	KiSS Version 2.0	nein
Mecklenburg-Vorpommern	DESK 3-6	nein
Niedersachsen	Fit in Deutsch	ja
Nordrhein-Westfalen	Delfin 4[c]	ja
Rheinland-Pfalz	VER-ES[d]	ja
Saarland	Früh Deutsch lernen	ja
Sachsen	SSV (Kita-Kinder), S-ENS (alle Kinder)	ja
Sachsen-Anhalt	Delfin 4[e]	ja
Schleswig-Holstein	Sismik & Seldak (Kita-Kinder), HAVAS 5 (Nicht-Kita-Kinder)	ja
Thüringen	Thüringen führt eine alltagsintegrierte Sprachförderung durch und hat keine landesweite Sprachstandserfassung für eine Förderentscheidung vorgesehen.	

a) Neben diesen gesetzlichen bzw. empfohlenen Verfahren wird eine Vielzahl weiterer Verfahren eingesetzt, z. B. um den Entwicklungsverlauf zu dokumentieren.
b) Einsprachig aufwachsende Kinder, die nicht in die Kita gehen, werden nicht erfasst.
c) Laut Koalitionsvertrag wird die Sprachstandsfeststellung in Nordrhein-Westfalen in naher Zukunft weiterentwickelt und Delfin 4 nicht weiter eingesetzt.
d) In Rheinland-Pfalz ist VER-ES nur für Kinder verpflichtend, die nicht in die Kita gehen.
e) In Sachsen-Anhalt gilt seit dem 1. August 2013 ein neues Kinderförderungsgesetz, das keine verpflichtende Sprachstandserfassung zu einem bestimmten Zeitpunkt mehr vorsieht.
f) Die Angabe, ob ein Verfahren verpflichtend ist oder nicht, basiert auf einer Dokumentenrecherche. Abweichungen sind möglich.

Tabelle 1: Übersicht zu den in den Bundesländern eingesetzten Sprachstandsverfahren, um eine Förderentscheidung vorzunehmen.
Quelle: eigene Befragung der 16 Bundesländer. Stand: August/September 2013

VORGEHEN

Insgesamt wurden 21 Verfahren analysiert und bewertet. Die Auswahl beschränkt sich auf Tests, Screenings, Einschätzverfahren und Beobachtungsbögen, die Aufschluss darüber geben sollen, ob ein Kind zusätzliche Sprachförderung benötigt oder nicht (Förderentscheidung).[1] Entscheidend für die Aufnahme in die Studie war, ob das Verfahren gesetzlich vorgeschrieben oder vom jeweils zuständigen Ministerium für den flächendeckenden Einsatz empfohlen wird.[2] [3]

Die Analyse und Bewertung der 21 eingesetzten Verfahren erfolgt anhand des Kriterienkatalogs „Qualitätsmerkmale für Sprachstandsverfahren im Elementarbereich", der im Mai 2013 veröffentlicht wurde. Das Mercator-Institut hatte eine interdisziplinäre Expertenkommission initiiert, die wissenschaftliche Qualitätsmerkmale für Sprachstandsverfahren im Elementarbereich entwickelt hat. Der Bewertungsrahmen definiert aus psychologischer, pädagogischer, sprachtherapeutischer und linguistischer Perspektive, welche Kriterien gelten sollten, um den Sprachstand jedes Kindes vergleichbar und wissenschaftlich fundiert zu ermitteln. Das Bewertungsraster umfasst 32 Merkmale.[4] Für jedes Qualitätsmerkmal ist eindeutig definiert, ab wann ein Verfahren das Merkmal sehr gut, gut, ausreichend oder nicht ausreichend erfüllt.

Sprachstandsverfahren können als Beobachtungsbögen, Screenings, Einschätz- oder Testverfahren konzipiert sein. Um diese unterschiedlichen Verfahren einer einheitlichen Bewertung unterziehen zu können, sind die Qualitätsmerkmale sehr breit angelegt. Daher kann keines der Verfahren in allen Bereichen „gut" oder „sehr gut" abschneiden. Die Berücksichtigung aller sprachlichen Basisqualifikationen beispielsweise ist nur möglich, wenn die Untersuchungsdauer zunimmt. Zwar dienen alle Verfahren dem Zweck, den Sprachstand eines Kindes zu erfassen, allerdings mit jeweils spezifischen Methoden: Während Tests und Screenings standardisierte Aufgaben vorgeben, definieren Beobachtungsbögen Verhaltenskategorien, die bei einem Kind systematisch beobachtet werden sollen. Einschätzverfahren lenken die Wahrnehmung der Erzieherinnen und Erzieher auch auf bestimmte Verhaltens- oder Kompetenzbereiche, beruhen dabei aber weniger stark auf einer systematischen Beobachtung, sondern fordern eine globale Beurteilung. Aus diesem Grund gehört es nicht zu den Zielen dieser Studie, ein Ranking vorzulegen, welches Verfahren am besten abschneidet.

Die Handlungsfelder der Qualitätsmerkmale im Überblick

1. **Berücksichtigung sprachlicher Basisqualifikationen**: Das Verfahren erfasst die zentralen Bereiche der sprachlichen Entwicklung eines Kindes.

2. **Validität**: Das Verfahren ist so konstruiert, dass es exakt den Sprachstand eines Kindes erfasst und nicht andere Bereiche, wie z. B. kulturelles Wissen.

3. **Objektivität**: Das Verfahren ist unabhängig von subjektiven Einflussfaktoren, wie beispielsweise dem Verhalten der pädagogischen Fachkraft.

4. **Reliabilität**: Das Verfahren misst den Sprachstand des Kindes präzise. Wird es wiederholt, steht am Ende das gleiche Ergebnis.

5. **Normierung**: Die Ergebnisse eines Kindes lassen sich mit den Ergebnissen anderer Kinder vergleichen.

6. **Fehlerquote**: Das Verfahren stellt sicher, dass die Kinder, die einen Sprachförderbedarf haben, auch wirklich identifiziert werden.

7. **Qualifizierung der pädagogischen Fachkräfte**: Die Personen, die den Sprachstand der Kinder messen, sind im Vorfeld für Sprachdiagnostik sensibilisiert und qualifiziert worden.

8. **Zeitliche Anforderungen**: Das Verfahren berücksichtigt die Konzentrationsfähigkeit des Kindes und die zeitliche Machbarkeit im pädagogischen Alltag.

9. **Mehrsprachigkeit**: Das Verfahren erfasst die besonderen Rahmenbedingungen von Kindern, die Deutsch als Zweitsprache erlernen.

10. **Spezifität der Diagnostik**: Das Verfahren gibt Hinweise darauf, welche Sprachfördermaßnahmen für das Kind infrage kommen.

QUALITÄTSMERKMALE ZEITLICHE ANFORDERUNGEN

Für eine angemessene zeitliche Durchführbarkeit muss ein Sprachstandsverfahren folgende Kriterien erfüllen:

	Kriterien			
	Nicht ausreichend	Ausreichend	Gut	Sehr gut
Perspektive Kinder: zeitliche Dauer der Durchführung	Mehr als 30 Minuten; keine Angaben	Maximal 30 Minuten	Maximal 20 Minuten	Maximal 20 Minuten, abwechslungsreich gestaltet
Perspektive diagnostischer Aufwand: zeitliche Dauer der Durchführung, Auswertung & Dokumentation	Mehr als 40 Minuten, keine Angaben	Maximal 40 Minuten	Maximal 30 Minuten	Maximal 20 Minuten

Tabelle 2: Beispiel für die Definition der Kriterien, ab wann ein Qualitätsmerkmal erfüllt ist (Quelle: Becker-Mrotzek et al. (2013). Qualitätsmerkmale für Sprachstandsverfahren im Elementarbereich. Ein Bewertungsrahmen für fundierte Sprachdiagnostik in der Kita. Köln: Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache)

Die Analyse und Bewertung sollen einen Impuls für die Testentwicklung geben, wie die Qualität der jeweiligen Verfahren in den einzelnen Bereichen so verbessert werden kann, dass sie einheitliche Standards erfüllen.
Bewertungsgrundlage sind Begleitmaterialien zum jeweiligen Verfahren sowie relevante Untersuchungen, Fachartikel und Bewertungen der jeweiligen Diagnoseinstrumente. Um die Objektivität der Untersuchung zu gewährleisten, haben drei Mitarbeiter des Mercator-Instituts die Analyse und Bewertung unabhängig voneinander durchgeführt.[5] Die Analyse auf Basis des vorhandenen Materials wurde anschließend den Testentwicklerinnen und Testentwicklern sowie den Fachreferaten in den jeweils zuständigen Landesministerien für Ergänzungen, Korrekturen und Stellungnahmen vorgelegt.[6]

1 Die untersuchten Verfahren geben nicht notwendigerweise eine Antwort darauf, welche Fördermaßnahmen für ein Kind sinnvoll sind. Die Diagnostik des Förderbedarfs sollte in einem zweiten Schritt von speziell ausgebildetem Fachpersonal erfolgen.

2 Aufgrund der zahlreich vorhandenen Tests für diesen inhaltlichen und altersbezogenen Bereich musste eine Einschränkung vorgenommen werden. Die Auswahl erfolgte im ersten Schritt auf Grundlage der Angaben im Bildungsbericht 2012 und der aktualisierten Expertise des DJI von 2013 und wurde dann in Rücksprache mit den Bundesländern aktualisiert.

3 Einen Sonderfall stellt LiSe-DaZ in Baden-Württemberg dar. Seit der Fertigstellung im Jahr 2012 wird es vom Ministerium für Kultus, Jugend und Sport in Baden-Württemberg ausdrücklich empfohlen, aber aufgrund der relativ hohen Beschaffungskosten nicht flächendeckend eingesetzt.

4 Die folgenden Betrachtungen umfassen 32 Qualitätsmerkmale. Die beiden Merkmale zum Auswahlfehler „RATZ-Index“ wurden zu einem zusammengefasst, wodurch kein Informationsverlust entstand – Sensitivität und Spezifität waren in diesen Fällen auch immer angegeben.

5 Die exakte Übereinstimmung in den 672 Zuordnungen (21 Verfahren * 32 Merkmale) lag bei drei unabhängigen Personen bei 72 Prozent, die Reliabilität (Cronbach Alpha) bei 0.95, die Intraklassenkorrelation ICC bei 0.947. Die 28 Prozent, bei denen nur leicht divergierende Zuordnungen vorlagen, konnten in einer Diskussionsschleife konsensuell und übereinstimmend aufgelöst werden.

6 Die abschließende Bewertung berücksichtigt die Rückmeldungen zu 20 der 21 untersuchten Verfahren.

ANALYSE UND BEWERTUNG

BERÜCKSICHTIGUNG SPRACHLICHER BASISQUALIFIKATIONEN

Das Verfahren erfasst die zentralen Bereiche der sprachlichen Entwicklung eines Kindes.

Bei den Beobachtungsbögen schneiden *Sismik* und *Seldak* als einzige sehr gut ab: Sowohl der Umfang der berücksichtigten Basisqualifikationen als auch die Unterschiedlichkeit der Aufgabentypen bzw. Beobachtungskategorien werden mit sehr gut bewertet. Beide Verfahren weisen sieben berücksichtigte Basisqualifikationen auf und ermöglichen somit eine umfassende Einschätzung des Sprachstandes. Allerdings können auch *Sismik* und *Seldak* wegen fehlender empirisch belegter Angaben zur Validität nicht gänzlich überzeugen.

Als „Mischform" zwischen Beobachtungsbogen und Test schneidet *DESK 3–6* am besten ab. Bei den Aufgabentypen wird z. B. die semantische Basisqualifikation über drei sehr unterschiedliche Items abgefragt: Gegensätze, Farbbenennungen und Kategoriendefinitionen. Allerdings fehlt ein empirischer Beleg für das Außenkriterium „Bezug zur Basisqualifikation". Zudem wäre die Berücksichtigung instabiler Übergangserscheinungen im Verlauf des Spracherwerbs eines Kindes wünschenswert, um unnötige Verzerrungen in den Ergebnissen zu vermeiden.

Bei den Screenings und Testverfahren erreicht nur *KiSS* eine gute und eine sehr gute Bewertung. *KiSS* berücksichtigt vier Basisqualifikationen mit unterschiedlichen Aufgabentypen. *LiSe-DaZ* beschränkt sich auf die Erfassung von nur zwei Basisqualifikationen. Diese werden aber mit einer Vielzahl von Aufgabentypen abgefragt, sodass das Merkmal „Differenziertheit der Aufgaben" mit sehr gut bewertet wird. *Cito Sprachtest Version 3*, *SETK 3–5*, *Delfin 4* (Stufe 1) und *KISTE* weisen jeweils eine gute und eine ausreichende Bewertung hinsichtlich der Anzahl berücksichtigter Basisqualifikationen sowie der Differenziertheit der Aufgaben auf. Diese Verfahren fokussieren sehr stark auf einzelne Basisqualifikationen mit mindestens zwei unterschiedlichen Aufgabentypen. Wünschenswert wäre es, die erfassten Basisqualifikationen um eine bis drei Weitere zu ergänzen.[1]

HASE berücksichtigt zwei Basisqualifikationen. Die semantische Basisqualifikation wird dabei über drei unterschiedliche Aufgabentypen erfasst (Nachsprechen von Sätzen, Wiedergabe von Zahlenfolgen sowie Nachsprechen von Kunstwörtern). Bei *Kenntnisse in Deutsch als*

Blick in die Praxis: Berücksichtigung der Übergangserscheinungen

Ein gutes Beispiel für die Berücksichtigung von Übergangserscheinungen ist in der Handanweisung des Verfahrens *Meilensteine der Sprachentwicklung* zu finden: „Bei den Pluralformen von Substantiven finden wir in dieser Phase des Spracherwerbs ebenfalls noch viele Übergeneralisierungen, wie z. B. die ‚Onkels', ‚die Fensters', ‚die Löffels' usw. Da es sich hier um zwei herausragende Merkmale der Grammatikentwicklung bei Kindern zwischen dem dritten und vierten Lebensjahr handelt, wurden die Pluralmarkierungen bei Substantiven und die Vergangenheitsbildung mit dem Partizip II im Perfekt bei Kindern im Alter von 36 und 48 Monaten als Beobachtungsitems gewählt." (Quelle: Handanweisungen, S. 7)

Zweitsprache erfassen wird die hohe Aufgabenvielfalt in Stufe 4 gut bewertet: In einem Klassenzimmer werden acht bis 14 unterschiedliche Spielstationen aufgebaut, die von größeren Grundschulkindern betreut werden. Lehrkräfte beobachten die sprachlichen Interaktionen. Wird das Kind nicht aktiv oder äußert es sich überhaupt nicht, wird eine Einschulung in eine Sprachlernklasse empfohlen.

Ein (interdisziplinäres) Expertenrating dazu, ob die Aufgabentypen für die Erfassung der sprachlichen Bereiche geeignet sind, ist nur für *KiSS* bekannt.[2] Drei weitere Verfahren haben hierzu nachvollziehbare Angaben gemacht (*Vorstellungsverfahren Viereinhalbjähriger*, *Cito Sprachtest Version 3* sowie *DESK 3–6*), alle anderen 17 Verfahren wurden mit nicht ausreichend bewertet.

Der Bezug zum Außenkriterium der sprachlichen Basisqualifikation fehlt in allen Verfahren gänzlich. Dieses Ergebnis liegt vor allem darin begründet, dass sich die meisten Testentwicklerinnen und Testentwickler nicht auf diesen Qualifikationenfächer beziehen, da es bisher keine gesicherten empirischen Belege gibt, wie der jeweilige sprachliche Bereich idealtypisch zu erfassen ist. Ein Expertenrating könnte jedoch zeitnah ermitteln, ob die sprachlichen Bereiche plausibel mit den verwendeten Items oder Aufgabentypen erfasst werden, und dabei auch die interdisziplinäre Verankerung eines Verfahrens weiter stärken.

Allgemeine Lesehilfe für die Tabellen

Das Bewertungsschema besteht aus sechs Kategorien: sehr gut (+++), gut (++), ausreichend (+), nicht ausreichend (–), keine Angabe (–*) und nicht bewertet (*/*).

Wenn Eigenschaften unbekannt sind, ist dies mit nicht ausreichend zu bewerten. Zur besseren Nachvollziehbarkeit sind diese Fälle in den Übersichtstabellen mit einem * gekennzeichnet. Aufgrund der unterschiedlichen Anlage einzelner Verfahren kann nicht immer für alle Merkmale eine Bewertung erfolgen. Diese Fälle werden mit / markiert.

Es werden vier Arten von Sprachstandsverfahren unterschieden: Beobachtungsbogen (B), Einschätzverfahren (E), Screening (S), Testverfahren (T) sowie „gemischtes" Verfahren (G), entweder mehrstufig (z. B. *Vorstellungsverfahren Viereinhalbjähriger*, *Fit in Deutsch*) oder Aufgaben, die durchgeführt werden, und andere, die beobachtet werden (z. B. *DESK 3–6*).

SPRACHLICHE BASISQUALIFIKATIONEN	Art des Verfahrens	Bundesland	Ausweisung und Anzahl der berücksichtigten Basisqualifikationen [a]	Differenziertheit der Aufgaben [b]	Berücksichtigung instabiler Übergangserscheinungen [c]	Bezug zur Basisqualifikation: über ein Außenkriterium numerisch-quantifiziert [d]	Expertenrating mit Interrater-Übereinstimmung [e]
Cito Sprachtest Version 3	T	HB	+	++	/	–	+
Delfin 4	S	NW, ST	++	+	–	–	–
DESK 3-6	G	MV	+++	+++	–	–	+
Deutsch Plus 4	G	BE	++	++	–	–	–*
Fit in Deutsch	S	NI	+	++	–	–*	–*
Früh Deutsch lernen	E	SL	+	–	–	–*	–*
HASE	S	BW	–	+++	–	–*	–*
HAVAS 5	T	SH	+++	–	+	–	–
Kenntnisse in DaZ erfassen	G	BY	++	++	–	–	–*
KiSS	S	HE	++	+++	–	–	+
KISTE	S	BB	+	++	–	–	–*
LiSe-DaZ	T	BW	–	+++	++	–*	–*
Meilensteine der Sprachentwicklung	E	BB	++	–	+	–	–
QuaSta	B	BE	++	+	–	–*	–*
S-ENS	S	SN	+	++	–	–	–
Seldak	B	BY, SH	+++	+++	–	–*	–
SETK 3-5	T	BW	+	++	+	–	–*
Sismik	B	BY, SH	+++	+++	–	–*	–
SSV	S	SN	+	–	–	–	–
VER-ES	G	RLP	+	+	–	–*	–*
Vorstellungsverfahren Viereinhalbjähriger	G	HH	+++	+	+	–	+

Tabelle 3: Analyse und Bewertung der Berücksichtigung sprachlicher Basisqualifikationen

Lesehilfe für Tabelle 3

Die sieben sprachlichen Basisqualifikationen lauten: phonische Basisqualifikation, pragmatische Basisqualifikation I und II, semantische Basisqualifikation, morphologisch-syntaktische Basisqualifikation, diskursive Basisqualifikation und literale Basisqualifikation I.

a

Nicht ausreichend: keine Angaben; weniger als 3 Basisqualifikationen (BQ) berücksichtigt.
Ausreichend: BQ werden benannt, mind. 3 berücksichtigt.
Gut: BQ benannt, mind. 4 berücksichtigt.
Sehr gut: BQ benannt, mind. 6 berücksichtigt.

b

Nicht ausreichend: BQ wird mit einem Item abgefragt.
Ausreichend: BQ mit mind. 2 Aufgaben erfasst.
Gut: BQ wird mit mind. 2 unterschiedlichen Aufgabentypen erfasst.
Sehr gut: BQ wird mit mind. 3 unterschiedlichen Aufgabentypen erfasst.

c

Nicht ausreichend: keine Angaben.
Ausreichend: Bei der Interpretation der Ergebnisse wird auf ggf. bestehende instabile Übergangserscheinungen hingewiesen.
Gut: Konkrete sprachliche Bereiche, die instabile Übergangserscheinungen aufweisen, werden benannt.
Sehr gut: Einfluss instabiler Übergangserscheinungen auf die Ergebnisse wird in den einzelnen Bereichen ausgewiesen.

d

Nicht ausreichend: keine Angaben; Validität < 0.6
Ausreichend: Validität des Aufgabentypus für die Basisqualifikation ist $r \geq 0.6$.
Gut: Validität des Aufgabentypus für die Basisqualifikation ist $r \geq 0.7$.
Sehr gut: Validität des Aufgabentypus für die Basisqualifikation ist $r \geq 0.8$.

e

Nicht ausreichend: keine Angaben; wird nur von Testentwickler/-in attestiert.
Ausreichend: von mind. 3 Fachpersonen bewertet; Cohens Kappa ≥ 0.4.
Gut: von mind. 5 Fachpersonen bewertet; Cohens Kappa ≥ 0.6.
Sehr gut: von mind. 7 Fachpersonen bewertet; Cohens Kappa ≥ 0.8.

VALIDITÄT

Das Verfahren ist so konstruiert, dass es exakt den Sprachstand eines Kindes erfasst und nicht andere Bereiche, wie z. B. kulturelles Wissen.

Einen mehr oder weniger expliziten Bezug zu einer Spracherwerbstheorie weisen zehn Verfahren auf, allerdings erreicht nur *LiSe-DaZ* eine bessere Bewertung als ausreichend.[3] Diese fehlende theoretische Fundierung liegt zumeist darin begründet, dass keine Spracherwerbstheorie vollständig empirisch belegt ist. Zusätzlich erweist es sich offensichtlich als schwierig, aus der Theorie Aufgabentypen abzuleiten, die für die Diagnostik von Spracherwerbsstörungen oder -verzögerungen relevant und zugleich in der Praxis anwendbar sind. Das Begleitmaterial zu den jeweiligen Verfahren enthält daher entweder Überblicksarbeiten mit aktuellen Befunden zum (Zweit-)Spracherwerb (z. B. *Delfin 4*, *Kenntnisse in Deutsch als Zweitsprache erfassen*) oder die Autoren und Quellen werden nicht explizit genannt. Einige Verfahren verzichten bewusst darauf, sich auf eine linguistische Perspektive festzulegen. Die Einbindung von Befunden zum Spracherwerb und die konkrete und nachvollziehbare Ableitung von Aufgabentypen führen dazu, dass *LiSe-DaZ* bei diesem Merkmal als einziges mit gut bewertet werden konnte.

Bezüglich der Prüfung, ob die unterschiedlichen Aufgaben auch tatsächlich Unterschiedliches erfassen (empirische Faktorenstruktur), existieren nur für acht der 21 Verfahren geeignete Analysen: *KISTE*, *Früh Deutsch lernen* und *Cito Sprachtest Version 3* wurden mit ausreichend bewertet, das Screening *HASE* sowie *LiSe-DaZ*, *VER-ES* und *HAVAS 5* mit gut. Nur der Beobachtungsbogen *Seldak* erreichte die Bewertung sehr gut.

Die Ermittlung der prognostischen Validität kann in der Regel erst erfolgen, wenn ein Kind, dessen Sprachstand im Alter von vier Jahren erhoben wurde, in die dritte Klasse geht, also nach einem Zeitraum von fünf Jahren. Die meisten Verfahren wurden zwischen dem Jahr 2002 und 2007 entwickelt, entsprechend wäre also im Jahr 2013 damit zu rechnen, dass diese Angaben vorliegen, ohne dass bedeutsamer Mehraufwand damit verbunden wäre. Nur vier Verfahren schaffen daher eine bessere Bewertung als nicht ausreichend – für die übrigen 17 liegen in den meisten Fällen keine Angaben vor.[4]

Die inhaltliche Konstruktvalidität ist unbedingt durch den empirischen Nachweis der konvergenten und diskriminanten Validität abzusichern. Das ist bei zwölf Verfahren nicht der Fall: Steht z. B. der Subtest „Sätze nachsprechen" im Zusammenhang mit der grammatischen Kompetenz eines Kindes oder mit der semantischen Kompetenz? Oder mit beiden? Dies ist im ersten Schritt theoretisch zu ermitteln, im zweiten Schritt ist dieser Nachweis empirisch zu erbringen.

Die Alltagsnähe der Testsituation wird nur bei acht Verfahren als ausreichend bewertet. Die Herausforderung scheint darin zu bestehen, eine Situation zu schaffen, in der die Kinder angstfrei ihre vorhandene Sprachkompetenz auch wirklich einbringen. Allerdings ist die Situation

Blick in die Praxis: Prognostische Qualität

Bei dem *Vorstellungsverfahren Viereinhalbjähriger* stimmt das Ergebnis der Sprachstandserhebung mit den Einschätzungen der pädagogischen Fachkräfte ein Jahr später überein. Für ein valides Ergebnis ist dieser Zeitraum zu kurz, dennoch wurde das Verfahren hier mit gut bewertet. Für *HASE* wurde u. a. der Zusammenhang zwischen der Fähigkeit, Sätze nachzusprechen, mit gut fünf Jahren und dem Wortschatz in der dritten Klasse ermittelt und eine „zufriedenstellende" Korrelation von 0.54 nachgewiesen. Die prognostische Qualität des SETK 3-4 wurde an einer kleinen Stichprobe mit 55 Kindern im Alter von drei Jahren überprüft. Drei Jahre später wurde der Sprachstand der dann sechsjährigen Kinder erhoben. Hier zeigte sich, dass Korrelationen von bis zu 0.64 vorliegen. *DESK 3-6* gibt einen RATZ-Index in Höhe von 73 Prozent an, mit einem Zeitintervall von einer Woche bis sechs Monaten ist die Zeitspanne zwischen der ersten und der zweiten Messung allerdings deutlich zu kurz.

VALIDITÄT	Art des Verfahrens	Bundesland	Bezug zu eindeutig benannter Spracherwerbstheorie[a]	Konstruktvalidität ist von Fachpersonen nachvollziehbar[b]	Empirische Faktorenstruktur[c]	Prognostische Validität[d]	Konvergente oder Kriterienorientierte Validität[e]	Alltagsnähe der Testsituation[f]	Alltagsnähe der Aufgaben/ Beobachtungssituation[g]
Cito Sprachtest Version 3	T	HB	–	+	+	–*	–*	+	++
Delfin 4	S	NW, ST	+	+	–*	–*	+	–	–
DESK 3-6	G	MV	–	+	–*	+++	++	+++	+
Deutsch Plus 4	G	BE	–*	–*	–*	–*	++	+	++
Fit in Deutsch	S	NI	–*	–*	–*	–*	–*	+	+++
Früh Deutsch lernen	E	SL	+	–	+	–*	–	/	+
HASE	S	BW	+	+	++	++	+	–	–
HAVAS 5	T	SH	–	+	++	–*	–*	–	+++
K DaZ*	G	BY	–	–*	–	–*	–*	– 4. Stufe +++	+++
KiSS	S	HE	+	+	–	–	+++	+	++
KISTE	S	BB	+	–*	+	–*	+	–	–
LiSe-DaZ	T	BW	++	+	++	–*	–	+	+
MdS*	E	BB	–	+	–*	–*	–*	/	+
QuaSta	B	BE	–	+	–*	–*	–*	++	+
S-ENS	S	SN	–	+	–*	–*	+++	–	–
Seldak	B	BY, SH	–	+	+++	–*	+	/	+
SETK 3-5	T	BW	+	+	–*	++	–	–	–
Sismik	B	BY, SH	–	+	–*	–*	–	/	+
SSV	S	SN	+	+	–	–	–	–	–
VER-ES	G	RLP	+	+	++	–*	–	+	++
VVV*	G	HH	+	+	–	++	–	–	+

* VVV = Vorstellungsverfahren Viereinhalbjähriger | MdS = Meilensteine der Sprachentwicklung | K DaZ = Kenntnisse in DaZ erfassen

Tabelle 4: Analyse und Bewertung der Validität

Lesehilfe für Tabelle 4

a

Nicht ausreichend: keine Angaben; Verweis auf „aktuelle Spracherwerbstheorien" ohne Quellenangabe.

Ausreichend: wird eindeutig mit Quelle benannt.

Gut: Aufgaben sind aus spezifischen Aussagen der Theorie abgeleitet.

Sehr gut: Güte der Ableitung (Operationalisierung) wird angegeben.

b

Nicht ausreichend: keine Angaben; unbekannt; wird von Testautoren und -autorinnen attestiert.

Ausreichend bis sehr gut: 3 externe Fachpersonen aus unterschiedlichen Fachbereichen (z. B. Sprachwissenschaften, Fachdidaktik, Pädagogik, Psychologie) bescheinigen übereinstimmend „hohe" Konstruktvalidität.

c

Nicht ausreichend: keine Angaben; unbekannt.

Ausreichend: Faktorenstruktur wird mitgeteilt.

Gut: Faktorenstruktur explorativ bestätigt.

Sehr gut: Faktorenstruktur konfirmatorisch gültig.

d

Nicht ausreichend: keine Angaben; unbekannt.

Ausreichend: RATZ-Index > 30.

Gut: RATZ-Index > 40.

Sehr gut: RATZ-Index > 65.

e

Nicht ausreichend: keine Angaben; unbekannt.

Ausreichend: r2 > 50 %.

Gut: r2 > 60 %.

Sehr gut: r2 > 70 %; diskriminante Validität wird mitgeteilt.

f

Nicht ausreichend: Situation verwirrt/verängstigt/setzt Kind unter Stress.

Ausreichend: Testsituation setzt Kind nicht unter Stress.

Gut: Kind kann sich in Beobachtungs- oder Testsituation frei verhalten.

Sehr gut: Kind erlebt Situation als positiv.

g

Nicht ausreichend: Aufgaben/Beobachtungssituation entsprechen in keiner Weise den Alltagserfahrungen des Kindes.

Ausreichend: Aufgaben/Beobachtungen werden in einer dem Kind verständlichen Weise umgesetzt.

Gut: Kind kann Test-/Beobachtungssituation verstehen.

Sehr gut: Kind findet Test-/Beobachtungssituation anregend/bewusstseinsbildend.

für viele Kinder mit Angst besetzt, wie das Phänomen der „Sprachnot" (z. B. in *HAVAS 5* thematisiert) oder auch die Befragungen im Rahmen der Pilotierung von *VER-ES* zeigen: Elf Prozent der Kita-Kinder und 35 Prozent der Nicht-Kita-Kinder waren in der Testsituation durchgängig ängstlich.

Auch die Alltagsnähe der Aufgaben bzw. Beobachtungssituationen wird bei sechs Verfahren als nicht ausreichend bewertet. Dahinter verbirgt sich möglicherweise ein test- und messmethodischer Konflikt. Zwei Testentwickler gaben unabhängig voneinander an, dass sie gerade nicht alltagsnahe Aufgaben nutzen, weil sich in solchen Aufgaben eben nicht nur die Sprachkompetenz zeige, sondern diese durch Welterfahrungen der Kinder überlagert werde – eine plausible Annahme.

OBJEKTIVITÄT

Das Verfahren ist unabhängig von subjektiven Einflussfaktoren, wie beispielsweise dem Verhalten der pädagogischen Fachkraft.

Die Objektivität eines Verfahrens wird dadurch gewährleistet, dass die Durchführung, Auswertung und Interpretation transparent und eindeutig sind. Darüber soll das Manual Auskunft geben. Dieses Kriterium erfüllen 19 der 21 untersuchten Verfahren. Nur für *KiSS* und *Früh Deutsch lernen* liegen keine Testhefte vor.[5] Elf Verfahren verfügen über ein Manual, das die wesentlichen Voraussetzungen zur objektiven Durchführung, Auswertung und Interpretation des Ergebnisses erfüllt. *SSV*, *SETK 3–5* und *HASE* werden gut bewertet, da sie die sprachlichen Interaktionen mehr oder weniger vollständig standardisieren und somit sowohl die Auswertung wie auch die Interpretation objektiv erfolgen können. Alle drei Verfahren beinhalten eine Audio-CD. Damit wird weitestgehend ausgeschlossen, dass die Person, die den Test durchführt, einen Einfluss auf das Ergebnis hat.

Mit sehr gut wurden die Tests *LiSe-DaZ*, *Cito Sprachtest Version 3* und *DESK 3–6* sowie die Beobachtungsbögen *QuaSta* und *HAVAS 5* bewertet. Bei *Cito Sprachtest Version 3* ist durch die vollständige Durchführung am Computer ein subjektiver Einfluss auf die Durchführungsobjektivität nahezu ausgeschlossen. *DESK 3–6* gibt, wie alle mit gut bewerteten Verfahren, im Manual eine klare Standardisierung aller Schritte vor. Zusätzlich finden Schulungen statt, die bereits evaluiert worden sind. Die Beobachtungsbögen *QuaSta* und *HAVAS 5* erreichen ihre gute Bewertung durch die klaren Instruktionen sowie die – insbesondere beim *HAVAS 5* für seine komplexe Auswertung – flankierenden Schulungen.

Im Kontrast zum Merkmal „Manual mit klaren Anleitungen" erfüllen zwölf Verfahren das Merkmal Objektivität nicht ausreichend, da keine empirischen Belege vorliegen. So ist es etwa fraglich, ob Erzieherinnen und Erzieher in ihrer Einschätzung, was „einfache Arbeitsaufträge" sind, übereinstimmen, oder ob Gegenstände des täglichen Lebens „ausreichend" von einem Kind benannt werden können (beide Beispiele aus *Früh Deutsch lernen*).

Objektivität ist für jedes Verfahren ein zentrales Qualitätsmerkmal. Insbesondere Beobachtungsbögen müssen dieses Merkmal erfüllen, da sie nicht in standardisierten Situationen, sondern direkt im pädagogischen Alltag zum Einsatz kommen. Wenn die Objektivität und Reliabilität eines Verfahrens empirisch ermittelt werden, lassen die Ergebnisse nicht nur Rückschlüsse auf das Verfahren und seine Merkmale zu, sondern vor allem auf das Verhalten der Person, die den Test durchführt: Wie objektiv und zuverlässig ist sie in der Lage, das Verfahren durchzuführen? Bei Beobachtungs- und Einschätzverfahren ist die Kompetenz der Durchführenden also sehr wichtig, da hier im Gegensatz zu den Screenings und Testverfahren keine standardisierte Situation vorliegt.[6]

VER-ES erhält hinsichtlich der Objektivität eine gute Bewertung. Die Übereinstimmung (Korrelation) der Bewertung liegt bei r = 0.81, wobei der niedrigste Wert bei der Beurteilung des nonverbalen Kommunikationsverhaltens (r = 0.59), der höchste bei der phonologischen Bewusstheit (r = 0.99) ermittelt wurde. Dies unterstreicht, wie wichtig die Ermittlung der Objektivität gerade bei Beobachtungsverfahren ist. Dies belegt auch das Ergebnis des *DESK 3–6*, der sowohl Durchführungs- als auch Beobachtungselemente aufweist: Beobachtungsaufgaben erweisen sich als deutlich weniger objektiv (38 Prozent Übereinstimmung) als Durchführungsaufgaben (69 Prozent Übereinstimmung). Bei 17 Prozent der Kinder erzielten die beiden pädagogischen Fachkräfte kein übereinstimmendes Screening-Ergebnis.

Beim Hamburger *Vorstellungsverfahren Viereinhalbjähriger* wird der Beobachtungsbogen immer dialogisch im Team besprochen, um die Objektivität zu erhöhen. Bei der Auswertung des Bildimpulses (als Teil des Verfahrens) wurde eine Übereinstimmung von 96 Prozent erzielt.

Bezüglich der Objektivität von *Delfin 4* kann bislang keine Aussage getroffen werden. Die Objektivität wird in Stufe 1 des Verfahrens durch die Art der Testsituation mit vier Kindern als sehr niedrig eingeschätzt. Strukturell gibt es keine Möglichkeit, die Objektivität zu gewährleisten. Die Objektivität für Stufe 2 dürfte dagegen deutlich höher liegen, da es als Individualverfahren durchgeführt wird.

Bei *HAVAS 5* sind Schulungen für die Durchführenden zwingend vorgesehen. Dennoch kann eine geringe Auswertungsobjektivität angenommen werden, da die Auswertung sehr komplex ist. Zudem sind die Anweisungen an einigen Stellen unklar (z. B. „Entscheiden Sie nach Ihrem Eindruck" oder „Den Unterschied kann man in der Regel an der Tonfärbung des Kindes erkennen"). Somit wurde *HAVAS 5* mit nicht ausreichend bewertet.

OBJEKTIVITÄT	Art des Verfahrens	Bundesland	Manual/Testheft des Verfahrens [a]	Objektivität [b]
Cito Sprachtest Version 3	T	HB	+++	+++
Delfin 4	S	NW, ST	+	–*
DESK 3–6	G	MV	+++	+
Deutsch Plus 4	G	BE	+	++
Fit in Deutsch	S	NI	+	–*
Früh Deutsch lernen	E	SL	–	–*
HASE	S	BW	++	+
HAVAS 5	T	SH	+++	–
Kenntnisse in DaZ erfassen	G	BY	+	–*
KiSS	S	HE	+++	–
KISTE	S	BB	+	–*
LiSe-DaZ	T	BW	+++	++
Meilensteine der Sprachentwicklung	E	BB	+	+++
QuaSta	B	BE	+++	–*
S-ENS	S	SN	+	–*
Seldak	B	BY, SH	+	+
SETK 3–5	T	BW	++	–*
Sismik	B	BY, SH	+	–*
SSV	S	SN	++	–*
VER-ES	G	RLP	+	++
Vorstellungsverfahren Viereinhalbjähriger	G	HH	+	++

Tabelle 5: Analyse und Bewertung der Objektivität

Lesehilfe für Tabelle 5

a

Nicht ausreichend: kein Manual.
Ausreichend: Es liegt ein Manual mit eindeutigen Hinweisen zur Durchführung, Auswertung und Interpretation vor.
Gut: Die sprachlichen Interaktionen werden vollständig durch das Manual standardisiert.
Sehr gut: Das Manual wird flankiert durch Schulungen und externe Beratung.

b

Nicht ausreichend: Objektivität ist unbekannt/liegt unter 0.9.
Ausreichend: Objektivität liegt zumindest in einem Bereich bei mind. 0.9 und Durchführungsobjektivität liegt bei mind. 0.8.
Gut: Objektivität liegt global/als Gesamtangabe bei mind. 0.9.
Sehr gut: Objektivität liegt in allen Bereichen über 0.9.

Die empirisch belegte Objektivität ist für zwölf Verfahren unbekannt. Dies ist bei Einschätzverfahren wie *Früh Deutsch lernen* oder *QuaSta* nicht akzeptabel, weil damit unklar ist, wie stark die Förderentscheidung von subjektiven Tendenzen bzw. den sprachlichen Fähigkeiten des Kindes abhängt. Stark unbefriedigend ist es für Verfahren wie *Delfin 4* (Stufe 1) oder *Kenntnisse in Deutsch als Zweitsprache erfassen* (Stufe 4), bei denen mehrere Kinder interagieren. Hier stellt sich zusätzlich die Frage, wie stark Beobachter in der Lage sind, zwischen der Gesamtsituation und den individuellen Leistungen zu differenzieren. Optimierungsbedarf besteht auch bei den Tests, da bereits der Unterschied, ob die Items von der Testleiterin vorgesprochen oder von CD abgespielt werden (s. *HASE*), einen erheblichen Einfluss auf die Einschätzung des Sprachstandes hat.

Blick in die Praxis: Objektivität

Für *Meilensteine der Sprachentwicklung* wurde die Übereinstimmungsgüte (als Korrelationskoeffizient) für jeden Subtest spezifisch angegeben. Die untersuchten Altersgruppen (24, 36, 48 und 60 Monate) wurden von zwei Erzieherinnen unabhängig voneinander eingeschätzt. Bei der Gruppe der 48 Monate alten Kinder (n = 83) wurden die Einschätzungen in den vier Bereichen Lexik, Grammatik, Kommunikation und Literalität betrachtet. Die Übereinstimmung (Objektivität) liegt im Globalurteil bei r = 0.96, während sie in den vier Bereichen zwischen 0.83 und 0.94 schwankt. Ferner wurde der Kappa-Koeffizient für die beiden Bereiche „sprachstrukturelle Kompetenz“ (0.71) und „sozio-kommunikative Kompetenz“ (0.72) ermittelt. Diese Angaben belegen die Objektivität für *Meilensteine der Sprachentwicklung* erfolgreich, auch wenn eine weitere Differenzierung in Durchführungs-, Auswertungs- und Interpretationsobjektivität wünschenswert ist.

RELIABILITÄT

Das Verfahren misst den Sprachstand des Kindes präzise. Wird es wiederholt, steht am Ende das gleiche Ergebnis.

Nur für zwei der 21 Verfahren (*KISTE* und *DESK 3–6*) liegen ideale Kennzahlen (Parallelform-Reliabilität, Retest-Reliabilität) darüber vor, wie reliabel sie sind. Für fünf Verfahren liegen zur Reliabilität keinerlei Angaben vor, für die weiteren 14 Verfahren wurde dies in Form der internen Konsistenz ermittelt. *S-ENS* weist dabei in allen Subtests so niedrige Werte auf, dass es als nicht ausreichend bewertet werden muss.

Sehr gute Reliabilitätskennwerte weisen neben *KISTE* die drei Verfahren *KiSS*, *Vorstellungsverfahren Viereinhalbjähriger* und *Seldak* auf. Gute Reliabilitätskennwerte liegen für *Delfin 4*, *SSV*, *SETK 3–5*, *Deutsch Plus 4* sowie den *Cito Sprachtest Version 3* und das *DESK 3–6* vor. Eine ausreichende Reliabilität weisen *HASE*, *LiSe-DaZ*, *Früh Deutsch lernen* und *VER-ES* auf.

Eine Bewertung von *HAVAS 5* ist wegen der ausgesprochen hohen Heterogenität der Indizes nicht eindeutig möglich: Für das Gesamtverfahren wurde eine interne Konsistenz von 0.63 ermittelt. Werden die einzelnen Bereiche betrachtet, zeigt sich, wie unterschiedlich konsistent die sieben Auswertungsbereiche sind: Drei Bereiche sind als gut, zwei als ausreichend zu bezeichnen sowie zwei Bereiche als nicht ausreichend. Eine Gesamtbewertung wurde deswegen nicht vorgenommen.

Obwohl mit überschaubarem Aufwand ermittelbar, wurde die Retest-Reliabilität nur bei zwei Verfahren angegeben. Wird – wie in der vorliegenden Untersuchung – auch die interne Konsistenz als Schätzer der Messgenauigkeit akzeptiert, so weist immer noch ein knappes Drittel der Verfahren keinen ausreichenden Beleg für die Reliabilität auf. Dies ist bei Verfahren wie z. B. *Fit in Deutsch* kritisch, die auf Basis einer dichotomen Unterscheidung „Förderung: Ja oder Nein?“ festlegen, ob ein Kind Sprachförderung erhält oder nicht (Cut-off-Wert). Dieser Wert müsste umso breiter angegeben werden, je geringer die Messgenauigkeit ist. Ist dies nicht der Fall, ist die Entscheidung rein zufällig.

Blick in die Praxis: Zur Wichtigkeit, die Zuverlässigkeit nicht als interne Konsistenz, sondern als Retest-Reliabilität zu ermitteln

Die interne Konsistenz sagt etwas darüber aus, inwiefern die Items einer Unterskala das Gleiche erfassen. Ob dies eine tagesabhängige Gemütsverfassung ist oder ein stabiles Merkmal, sagt es nicht. Gerade die Erzieherinnen und Erzieher erklären häufig, das Testergebnis würde stark von der jeweiligen Tagesform abhängen. Als Lösung kann die Wiederholungstestung (Retest-Reliabilität) erfolgen. Die Tatsache, dass sich der Sprachstand eines Kindes ständig weiterentwickelt, ist bei einem Intervall von zwei bis drei Wochen nicht gravierend. Bei *DESK 3–6* wurde eine Gruppe von 45 Kindern in einem Abstand von zehn Monaten untersucht. Die Korrelation der standardisierten Werte zwischen beiden Messzeitpunkten liegt bei r = 0.54, was einer mittleren, nicht vollkommen befriedigenden Höhe entspricht. Während aber nur eines der Kinder zum zweiten Zeitpunkt schlechter im Bereich Sprache und Kognition abschnitt, verbesserten sich elf Kinder.

Das Intervall von ein bis zwei Wochen hat *KISTE* eingesetzt und kommt auch auf wesentlich stabilere Schätzungen bzw. günstigere Werte: Die Retest-Reliabilität wird mit r = 0.52 (Subtest Inkonsistenzen Semantik) bis 0.91 (Subtest Aktiver Wortschatz) angegeben und liegt im Mittelwert bei 0.7. Diese Werte weichen teilweise erheblich von den (auch mitgeteilten) internen Konsistenzen ab.

RELIABILITÄT	Art des Verfahrens	Bundesland	Reliabilität des Verfahrens[a]
Cito Sprachtest Version 3	T	HB	++
Delfin 4	S	NW, ST	++
DESK 3–6	G	MV	++
Deutsch Plus 4	G	BE	++
Fit in Deutsch	S	NI	–*
Früh Deutsch lernen	E	SL	+
HASE	S	BW	+
HAVAS 5	T	SH	/
Kenntnisse in DaZ erfassen	G	BY	–*
KiSS	S	HE	+++
KISTE	S	BB	+++
LiSe-DaZ	T	BW	+
Meilensteine der Sprachentwicklung	E	BB	–*
QuaSta	B	BE	–*
S-ENS	S	SN	–
Seldak	B	BY, SH	+++
SETK 3–5	T	BW	++
Sismik	B	BY, SH	–*
SSV	S	SN	++
VER-ES	G	RLP	+
Vorstellungsverfahren Viereinhalbjähriger	G	HH	+++

Tabelle 6: Analyse und Bewertung der Reliabilität

Lesehilfe für Tabelle 6

a

Nicht ausreichend: unbekannt; r < 0.7.
Ausreichend: 0.7 < r < 0.8.
Gut: 0.8 < r < 0.9 oder differenziert nach Subskala > 0.7.
Sehr gut: r > 0.9 oder differenziert nach Subskala > 0.8.

NORMIERUNG

Die Ergebnisse eines Kindes lassen sich mit den Ergebnissen anderer Kinder vergleichen. Nur so lässt ein Verfahren eine Aussage darüber zu, welche Kinder einen Sprachförderbedarf haben und welche Kinder sich sprachlich altersgerecht entwickeln.

Vier Verfahren (*Früh Deutsch lernen*, *Kenntnisse in Deutsch als Zweitsprache erfassen*, *Fit in Deutsch*, *Deutsch Plus 4*) erfüllen keines der fünf Merkmale zur Normierung. Bei den Beobachtungs- und Einschätzverfahren ist dies häufig der Fall, da (eher implizit) von einer Normalitätserwartung ausgegangen wird. So wird in den Auswertungshinweisen von *HAVAS 5* genannt: „HAVAS liefert zunächst einmal nur Aussagen über den Sprachstand (der ‚höher' oder ‚niedriger' sein kann), ohne sich auf Normwerte festzulegen." Dass auch bei Einschätzungs- und Beobachtungsbögen Normierungen möglich und sinnvoll sind, zeigt z. B. *Sismik*, das eine Normtabelle mit Zehn- bzw. 20-Prozent-Abstufungen enthält. Der Umfang der Normierungsgruppe ist mit 545 Kindern im Alter von vier bis fünf Jahren sehr gut, die Art der Auswahl mit ganzen Kindergärten (Klumpenstichprobe) als gut zu bewerten.

Für Tests und Screenings ist eine zuverlässige Normierung die Grundlage, um die Auswahlentscheidung treffen zu können. Als Mindestumfang für die Bewertung ausreichend wurde ein Stichprobenumfang von 200 Kindern in der entsprechenden Altersgruppe (vier bzw. 4,5 Jahre) festgelegt. Der Umfang der Stichprobe schwankt bei den normorientierten Verfahren zwischen 58 bzw. 62 Kindern[7] und 14.859 (*Delfin 4*) bzw. 168.629 Kindern (*HASE*, alle Altersgruppen). Insgesamt wird der Stichprobenumfang bei den o. g. vier Verfahren (*Früh Deutsch lernen*, *Kenntnisse in Deutsch als Zweitsprache erfassen*, *QuaSta*, *Fit in Deutsch*), *Deutsch plus 4*, *Meilensteine der Sprachentwicklung* sowie – bis zur Veröffentlichung der Neunormierung – bei *SSV* und *SETK 3-5* sowie für *HAVAS 5* (n = 80) als nicht ausreichend eingeschätzt.

Das einzige Merkmal, das bei *Sismik* hinsichtlich der Normierung als nicht ausreichend bewertet wurde, ist das Alter der Normierung: Der Erhebungszeitraum ist unklar. Vermutlich erfolgte die Erhebung im Jahr 2001, ist also mit zwölf Jahren eher alt. Ähnliches gilt für *Seldak*, auch wenn der Stichprobenumfang deutlich größer ist (n = 2.419) und eine etwas aktuellere Normierung vorliegt. *Vorstellungsverfahren Viereinhalbjähriger* (2011) sowie *Cito Sprachtest Version 3* (2013) schneiden beim Merkmal Alter der Normierung mit sehr gut ab. Dies trifft auch auf *LiSe-DaZ* zu, das auch alle weiteren Merkmale mindestens gut erfüllt.

Die Art und Weise, wie die Normierungsstichprobe erhoben wird, variiert erheblich. Als beste Methode eignet sich hier die Vollerhebung, die idealerweise jährlich aktualisiert werden sollte. Vollerhebungen wurden bei *Vorstellungsverfahren Viereinhalbjähriger*, *HASE* und dem *Cito Sprachtest Version 3* im jeweiligen Bundesland durchgeführt. Zudem hat *Seldak* hier eine sehr gute Bewertung erzielen können, da die Normierungsstichprobe bei diesem Verfahren – im Gegensatz zur Normierung bei *Sismik* – durch eine Zufallsauswahl ermittelt wurde und damit repräsentativ ist. Sinnvoll erscheint auch das Vorgehen im Rahmen von *DESK 3-6*: Die Normierungsstichprobe besteht aus Vollerhebungen bei Kindergärten, die Einwilligungsquote der Eltern ist bekannt (zwischen 60 und 100 Prozent).

Eine Unterscheidung nach weiteren soziodemografischen Faktoren (neben dem Alter) nehmen fünf Verfahren hinsichtlich des Geschlechts vor, sechs Verfahren hinsichtlich Ein- und Mehrsprachigkeit. Hierbei ist zu beachten, dass *Sismik*, *Seldak* und *Kenntnisse in Deutsch als Zweitsprache erfassen* nur für die Ermittlung des Sprachstands mehrsprachiger bzw. bei *Seldak* einsprachiger Kinder vorgesehen sind. *S-ENS* belegt, dass die Testergebnisse keinen Geschlechtereffekt aufweisen, und verzichtet deshalb auf eine Differenzierung nach Geschlecht. Der *Cito Sprachtest Version 3* differenziert in einer separaten Normtabelle die Ergebnisse für fünf Arten der Ein- bzw. Mehrsprachigkeit und hat eine in die Software integrierte Normtabelle für die Kinder, die mit Türkisch als Erstsprache aufwachsen. *Delfin 4* hat bewusst eine rein einsprachige Normierungsstichprobe vorgenommen. Mehrsprachig aufwachsende Kinder werden anhand dieser Bezugsnorm eingeschätzt. Die Art der Normstichprobe wurde daher bei *Delfin 4* als nicht ausreichend bewertet.

Bei der Normierung zeigt sich, dass Neunormierungen als fortlaufende Verpflichtung für die angemessene Testanwendung aufgefasst werden müssen. Die Dynamik der gesellschaftlichen Veränderung ist nach wie vor so hoch, dass nicht davon ausgegangen werden kann, dass eine vor zwölf Jahren erhobene Referenzstichprobe noch gültig ist. Viel mehr belegen Studien, dass im Laufe der letzten Jahrzehnte erhebliche Veränderungen stattgefunden haben, die weiter fortdauern.

NORMIERUNG	Art des Verfahrens	Bundesland	Bezug zu den Teilbereichen [a]	Alter der Normierung [b]	Umfang der Normstichprobe [c]	Art der Gewinnung der Norm-stichprobe [d]	Unterscheidung unterschiedl. Zielgruppen [e]
Cito Sprachtest Version 3	T	HB	++	+++	+++	+++	++
Delfin 4	S	NW, ST	++	++	+++	–*	–
DESK 3–6	G	MV	++	+	+++	++	+
Deutsch Plus 4	G	BE	–*	–*	–*	–*	–*
Fit in Deutsch	S	NI	–	–*	–*	–*	–*
Früh Deutsch lernen	E	SL	–*	–*	–*	–*	–*
HASE	S	BW	++	+	+++	+++	–
HAVAS 5	T	SH	–	+	–	++	–
Kenntnisse in DaZ erfassen	G	BY	–*	–*	–*	–*	–
KiSS	S	HE	–	+	++	++	+
KISTE	S	BB	++	–	+++	++	+
LiSe-DaZ	T	BW	++	++	+++	++	++
Meilensteine der Sprachentwicklung	E	BB	++	–	–	–*	–
QuaSta	B	BE	–*	–	–*	+	–*
S-ENS	S	SN	++	+	+++	++	+
Seldak	B	BY, SH	++	–	+++	+++	+
SETK 3–5	T	BW	+++	+	–	–	–
Sismik	B	BY, SH	++	–	+++	++	+
SSV	S	SN	+	+	–	–	+
VER-ES	G	RLP	++	++	+++	–	+
Vorstellungsverfahren Viereinhalb-jähriger	G	HH	++	+++	+++	+++	+

Tabelle 7: Analyse und Bewertung der Normierung

Lesehilfe für Tabelle 7

a

Nicht ausreichend: keine Angaben.

Ausreichend: Es wird darauf hingewiesen, welche Bereiche ausreichend normiert sind.

Gut: Normierung ist (auch) für die Teilleistungsbereiche spezifisch.

Sehr gut: Wie „gut" sowie Angabe der spezifischen Konfidenzbänder.

b

Nicht ausreichend: unbekannt; Normierung älter als 12 Jahre.

Ausreichend: Normierung max. 12 Jahre alt.

Gut: Normierung nicht älter als 8 Jahre.

Sehr gut: Normierung wird jährlich/alle 2 Jahre aktualisiert.

c

Nicht ausreichend: $N < 200$.

Ausreichend: $N > 200$.

Gut: $N > 250$.

Sehr gut: $N \geq 300$.

d

Nicht ausreichend: unbekannt.

Ausreichend: Stichprobengewinnung wird beschrieben, keine Randomisierung.

Gut: Stichprobengewinnung wird beschrieben, randomisierte Auswahl, allerdings liegt eine Klumpenstichprobe vor, da Daten in Gruppen (ganze Kindertagesstätten) erhoben wurden.

Sehr gut: Stichprobengewinnung wird beschrieben, vollständig randomisierte Auswahl auf Ebene der Individuen.

e

Nicht ausreichend: wird nicht vorgenommen; unbekannt, für wen die Normierung gültig ist.

Ausreichend: Normierungsstichprobe wird detailliert dargestellt.

Gut und sehr gut: Normierungsstichprobe wird für alle Zielgruppen des Verfahrens detailliert dargestellt.

FEHLERQUOTE (GEWICHTUNG DER AUSWAHLFEHLER)

Das Verfahren stellt sicher, dass die Kinder, die einen Sprachförderbedarf haben, auch wirklich identifiziert werden.

Bei 13 der 21 Verfahren liegen keine Angaben zu Auswahlfehlern vor. Das *Vorstellungsverfahren Viereinhalbjähriger* erhält hinsichtlich Sensitivität eine nicht ausreichende Bewertung, hinsichtlich seiner Spezifität eine sehr gute Bewertung. Ähnliches, aber umgekehrt, ergibt sich für *KISTE*: Hier ist die Sensitivität sehr gut, die Spezifität dagegen nicht ausreichend. Überraschend sind die Werte für *Deutsch Plus 4*: Die Sensitivität konnte nur mit ausreichend bewertet werden, während die Spezifität sehr gut ist – für ein Auswahlverfahren wäre eher eine umgekehrte Reihenfolge sinnvoll (hohe Sensitivität, niedrigere Spezifität, s. u.), um förderbedürftige Kinder sicher zu identifizieren.

In allen sechs weiteren Verfahren wurde für beide Bereiche mindestens eine ausreichende oder eine bessere Bewertung erzielt. *KiSS* ist hinsichtlich seiner Sensitivität gut, hinsichtlich seiner Spezifität sehr gut (75 bis 93 Prozent), ebenso *DESK 3–6* (Sensitivität: 71 Prozent, Spezifität: 88 Prozent). *HASE* ist dagegen sehr gut sensitiv (82 bis 83 Prozent), während seine Spezifität gut ist (61 bis 62 Prozent).

Blick in die Praxis: Minimierung der Fehlerquote

Bei *DESK 3–6* wurden die Erzieherinnen und Erzieher vor der Durchführung gebeten, bei 341 vierjährigen Kindern alle Entwicklungsbereiche auf einer fünfstufigen Skala („deutlich weiterentwickelt", „etwas weiterentwickelt", „altersgemäß entwickelt", „leicht verzögert", „deutlich verzögert") zu beurteilen.[8] „Leicht" und „deutlich verzögert" wurden zu einer Gruppe zusammengefasst. Nimmt man nur die Kinder, die „deutlich" verzögert sind, ergibt sich eine Sensitivität von 100 Prozent bei den vier- bis sechsjährigen Kindern, die Beurteilungen der Erzieherinnen und Erzieher und die Ergebnisse des *DESK 3–6* sind also deckungsgleich. Der Mehrwert des *DESK 3–6* zeigt sich bei den Kindern, deren Sprachentwicklung von den Erzieher/-innen als „leicht verzögert" eingeschätzt wurde: Diese Kinder konnte *DESK 3–6* sicher differenzieren, obwohl sie der pädagogischen Fachkraft zuvor nicht als förderbedürftig aufgefallen waren. Methodisch wäre es wünschenswert, dass eine andere, zweite Fachkraft den *DESK 3–6* im Anschluss durchführt, um zu wirklich unabhängigen Einschätzungen zu kommen. Auch ist, wie die Autoren selbst schreiben, die Validität dieses Außenkriteriums nicht gesichert: Wie sicher können pädagogischen Fachkräfte die förderbedürftigen Kinder identifizieren? Auch dies wird fehlerbehaftet sein. Optimal wäre es, wenn die Erzieherin bzw. der Erzieher einen sehr ausführlichen Test durchführt, um zu einer validen Einschätzung zu kommen.

Für *KiSS* wurde das Außenkriterium wesentlich aufwendiger, aber auch valider, durch die Verwendung von vier Referenzverfahren gebildet.[9] Auf Grundlage dieser Ergebnisse haben Sprachexpertinnen und -experten die Einteilung in sprachlich „auffällig" vs. „nicht auffällig" vorgenommen. Zudem wurden zwei weitere Screenings (*SSV* und *HASE*) mit einbezogen, um die Höhe der Auswahlfehler zu bestimmen (Neumann, Holler-Zittlau, van Minnen, Sick, Zaretsky & Euler, 2010). Dabei zeigte sich, dass selbst Sprachexpertinnen bzw. -experten nicht befriedigend häufig in ihren Einschätzungen übereinstimmen (Konkordanzrate), wenn ihnen unbekannt ist, wie hoch der Anteil von Kindern mit Sprachförderbedarf in der Gesamtbevölkerung ist.

Bei *SSV* wurde die Höhe der Auswahlfehler über das Ergebnis des *SETK 3–5* ermittelt, was methodisch nicht befriedigt, da das SSV eine Kurzform des *SETK 3–5* darstellt: Damit wird lediglich belegt, dass die Kurzform (*SSV*) in der Lage ist, die gleichen Kinder zu identifizieren, die auch nach der Langform (*SETK 3–5*) identifiziert werden, nicht aber, ob dies mit einem unabhängigen zweiten Urteil übereinstimmt. Die Bestimmung der Fehlerquote ist gerade bei mehrstufigen Screenings extrem wichtig, da ansonsten die Selektion immer stärker verzerrt wird. In einer ersten Stufe sollte die Sensitivität sehr hoch sein, während in darauffolgenden Stufen die Spezifität höher liegen sollte. Aus Ökonomiegründen dagegen erscheint es wünschenswert, direkt in Stufe 1 möglichst viele Kinder als nicht förderbedürftig zu klassifizieren. Dabei wird in Kauf genommen, viele förderbedürftige Kinder falsch zu klassifizieren.

Insgesamt ist für 13 Verfahren die Anzahl von Kindern, die durch das Verfahren fälschlicherweise als nicht förderbedürftig klassifiziert werden, nicht bekannt. Dieser Nachweis ist insbesondere für Screening-Verfahren und solche mit mehreren Stufen wesentlich, da ansonsten verzerrte Gruppen entstehen: Wird im ersten Schritt falsch ausgewählt, setzt sich dieser Fehler auf der nächsten diagnostischen Stufe weiter fort.

Lesehilfe für Tabelle 8

a
Nicht ausreichend: unbekannt; < 60 %.
Ausreichend: 60 %–69 %.
Gut: 70 %–80 %.
Sehr gut: > 80 %.

b
Nicht ausreichend: unbekannt; < 50 %.
Ausreichend: 50 %–59 %.
Gut: 60 %–70 %.
Sehr gut: > 70 %.

c
Nicht ausreichend: unbekannt; < 40 %.
Ausreichend: 40 %–54 %.
Gut: 55 %–70 %.
Sehr gut: > 70 %.

FEHLERQUOTE	Art des Verfahrens	Bundesland	Sensitivität [a]	Spezifität [b]	RATZ-Index [c]
Cito Sprachtest Version 3	T	HB	–*	–*	–*
Delfin 4	S	NW, ST	–*	–*	–*
DESK 3–6	G	MV	++	+++	++
Deutsch Plus 4	G	BE	+	+++	+
Fit in Deutsch	S	NI	–*	–*	–*
Früh Deutsch lernen	E	SL	–*	–*	–*
HASE	S	BW	+++	++	++
HAVAS 5	T	SH	–*	–*	–*
Kenntnisse in DaZ erfassen	G	BY	–*	–*	–*
KiSS	S	HE	++	+++	/
KISTE	S	BB	+++	–	+
LiSe-DaZ	T	BW	–*	–*	–*
Meilensteine der Sprachentwicklung	E	BB	–*	–*	–*
QuaSta	B	BE	–*	–*	–*
S-ENS	S	SN	–*	–*	–*
Seldak	B	BY, SH	–*	–*	–*
SETK 3–5	T	BW	++	+++	+++
Sismik	B	BY, SH	–*	–*	–*
SSV	S	SN	–	+	++
VER-ES	G	RLP	–*	–*	–*
Vorstellungsverfahren Viereinhalbjähriger	G	HH	–	+++	/

Tabelle 8: Analyse und Bewertung der Fehlerquote

QUALIFIZIERUNG DER PÄDAGOGISCHEN FACHKRÄFTE

Die Personen, die den Sprachstand der Kinder messen, sind im Vorfeld für Sprachdiagnostik sensibilisiert und qualifiziert worden.

Die Herausforderung, pädagogische Fachkräfte ausreichend für eine diagnostische Situation zu qualifizieren, konnte für *Kenntnisse in Deutsch als Zweitsprache erfassen* nicht überzeugend gelöst werden. Damit wurde dieses Merkmal als nicht ausreichend bewertet, da eine Schulung notwendig ist, um die Durchführung und die Beobachtungsobjektivität zu gewährleisten. Bei *KISTE* und *Meilensteine der Sprachentwicklung* sind keine Angaben zu notwendigen Qualifikationen verfügbar. Für *Delfin 4*, *Fit in Deutsch* und *VER-ES* liegen Materialien vor, es sind aber keine Schulungen vorgesehen. Diese sechs Verfahren erhalten daher ebenfalls die Bewertung nicht ausreichend.

Als gut oder sehr gut wurde u. a. das Vorgehen in Baden-Württemberg bei der Umsetzung von *HASE* bewertet. Hier erhalten die Erzieherinnen und Erzieher einerseits Schulungen. Wird ein Sprachförderbedarf festgestellt, führen Fachkräfte des Gesundheitsamtes das anschließende Verfahren *SETK 3-5* für die Diagnostik des Förderbedarfs durch. Ähnlich intensiv wird in Mecklenburg-Vorpommern *DESK 3-6* mit Schulungen flankiert. Auch *HAVAS 5* soll in Schleswig-Holstein nur von Personen durchgeführt werden, die zuvor an einer Fortbildung teilgenommen haben. In Hessen werden Erzieherinnen und Erzieher für die Durchführung von *KiSS* durch Sprachexperten (Logopäden, Sprachwissenschaftler, Kinderärzte, Sprachheilpädagogen etc.) zertifiziert und erhalten im Anschluss das Material kostenlos. Dieses Vorgehen wird mit sehr gut bewertet.

In mindestens vier Bundesländern (z. B. BW, HB, MV, RLP) werden die Fachkräfte regelmäßig befragt, wie sie die Umsetzung des Verfahrens in der Praxis bewerten und ob sie weitere Qualifizierungen benötigen. Dies ist wünschenswert und wird als gut bewertet. Im Saarland durchlaufen alle Erzieherinnen und Erzieher eine verpflichtende Schulung. Diese wird allerdings nur unsystematisch evaluiert, sodass dieses Merkmal mit ausreichend bewertet wurde. VER-ES wurde gezielt auf die Qualifikation pädagogischer Fachkräfte hin konzipiert. Dennoch muss das Verfahren, bis empirische Belege vorliegen, als nicht ausreichend bewertet werden.

Insgesamt legen die meisten Bundesländer großen Wert darauf, pädagogische Fachkräfte für die Diagnostik zu schulen. Einerseits werden Schulungen durchgeführt und evaluiert, andererseits werden Verfahren gezielt für den Einsatz durch Erzieherinnen und Erzieher konzipiert.

Lesehilfe für Tabelle 9

Die durchführende Person ...
a) kennt die dem Sprachstandsverfahren zugrundeliegende Spracherwerbstheorie und kann diese kritisch reflektieren;
b) kennt die Stärken und Schwächen des verwendeten Verfahrens sowie von Sprachstandsverfahren allgemein, insbesondere im Hinblick auf die zu untersuchenden Zielgruppen;
c) kennt die Durchführung und kann auch auf Besonderheiten und Störungen flexibel reagieren;
d) kennt die Auswertungsanweisungen und kann Besonderheiten einer Diagnostik kompetent und fachlich angemessen berücksichtigen;
e) kennt die Interpretationsregeln und weicht von diesen nur in begründeten Fällen ab;
f) verfügt über die Kompetenz, eine angemessen differenzierte Ergebnisrückmeldung an die Beteiligten/Betroffenen zu geben.

a
Nicht ausreichend: keine Angaben, keine vorgeschalteten Schulungen mit Abschlusstest.
Ausreichend: Die durchführende Person beherrscht mind. 3 Aspekte des verwendeten Verfahrens.
Gut: Die durchführende Person beherrscht mind. 4 Aspekte des verwendeten Verfahrens.
Sehr gut: Die durchführende Person beherrscht alle Aspekte des verwendeten Verfahrens.

QUALIFIZIERUNG DER PÄDAGOGISCHEN FACHKRÄFTE	Art des Verfahrens	Bundesland	Durchführende Person weist Qualifikation für die Durchführung von Sprachstandsverfahren auf. [a]
Cito Sprachtest Version 3	T	HB	+
Delfin 4	S	NW, ST	–
DESK 3–6	G	MV	+++
Deutsch Plus 4	G	BE	++
Fit in Deutsch	S	NI	–
Früh Deutsch lernen	E	SL	+
HASE	S	BW	+
HAVAS 5	T	SH	+
Kenntnisse in DaZ erfassen	G	BY	–
KiSS	S	HE	+++
KISTE	S	BB	–*
LiSe-DaZ	T	BW	+
Meilensteine der Sprachentwicklung	E	BB	–
QuaSta	B	BE	+
S-ENS	S	SN	+
Seldak	B	BY, SH	+
SETK 3–5	T	BW	+
Sismik	B	BY, SH	+
SSV	S	SN	+
VER-ES	G	RLP	–
Vorstellungsverfahren Viereinhalbjähriger	G	HH	+

Tabelle 9: Analyse und Bewertung der Qualifizierung pädagogischer Fachkräfte

ZEITLICHE ANFORDERUNGEN

Das Verfahren berücksichtigt die Konzentrationsfähigkeit des Kindes und die zeitliche Machbarkeit im pädagogischen Alltag.

Hinsichtlich der Frage, ob das Verfahren die Konzentrationsfähigkeit der Kinder berücksichtigt, können die Beobachtungsverfahren als sehr gut bewertet werden: Sie sind in der Regel für die Kinder mit keinem oder nur sehr geringem Aufwand verbunden, da sie alltagsintegriert durchgeführt werden. Dies trifft auf die drei Verfahren *Früh Deutsch lernen*, *Sismik* und *Seldak* zu.
Kein Test konnte die Bewertung sehr gut erreichen. Zwar dauert keines der Verfahren länger als 20 Minuten (*KiSS* ca. 15–20 Minuten, *HASE* ca. 10 Minuten, *SSV* maximal 10 Minuten, *HAVAS 5* 5–10 Minuten, *DESK 3-6* 15–20 Minuten). Die Verfahren sind jedoch wenig abwechslungsreich gestaltet, sodass die Gefahr besteht, dass die Konzentrationsfähigkeit dadurch eingeschränkt wird. Die genannten Verfahren schneiden daher gut ab. *QuaSta* basiert auf den Einschätzungen aus einem strukturierten Bildungsinterview, das mit jedem Kind geführt wird und ca. 15 bis 20 Minuten dauert. Das Bildungsinterview ist sinnvoll in den Kita-Alltag integriert, was zu einer guten Bewertung führt.

Kenntnisse in Deutsch als Zweitsprache erfassen wurde als nicht ausreichend bewertet: Insgesamt dauert die Durchführung ca. zwei Stunden. Auch *KISTE* (35 bis 50 Minuten) und *S-ENS* (ca. eine Stunde) berücksichtigen die Konzentrationsfähigkeit eines Kindes nicht ausreichend, *LiSe-DaZ* (etwa eine halbe Stunde) gerade noch ausreichend.
Der Aufwand für die Auswertung, Interpretation und Ergebnismitteilung dürfte bei *Cito Sprachtest Version 3* am geringsten sein, da die ersten beiden Schritte automatisiert erfolgen. Insgesamt konnten neun Verfahren eine sehr gute Bewertung erzielen, also mit einem Aufwand von maximal 20 Minuten pro Kind umgesetzt werden.
Hinsichtlich der zeitlichen Anforderungen können drei Verfahren nicht überzeugen: *Kenntnisse in Deutsch als Zweitsprache erfassen* weist diese Bewertung wegen eines mehrstufigen, sehr komplexen Aufbaus auf. *KISTE* und *S-ENS* wurden als nicht ausreichend bewertet, weil sie ca. eine Stunde in Anspruch nehmen. Die meisten Verfahren haben dagegen die Konzentrationsfähigkeit der Kinder bei der Konzeption mindestens ausreichend berücksichtigt. Bei dem Aufwand für die Auswertung, Dokumentation und Ergebnismitteilung schneiden dagegen sieben Verfahren schlecht ab – in diesem Bereich besteht bei vielen Verfahren noch Optimierungsbedarf. Sowohl die Auswertung könnte vereinfacht als auch die Ergebnismitteilung besser standardisiert werden, um die zeitliche Belastung zu verringern.

Blick in die Praxis: Effiziente Arten der Auswertung, Interpretation und Ergebnismitteilung

Bei dem computerbasierten Testverfahren *Cito Sprachtest Version 3* erfolgen die Auswertung, Ergebnisinterpretation und Berichtlegung automatisiert, sodass sich die Testleiterinnen und -leiter voll und ganz auf die Reflexion der Ergebnisse konzentrieren können. Diese Art der Umsetzung ist am wenigsten fehleranfällig – wenn Fehler auftreten, sind diese systematisch und damit tendenziell leichter zu beheben.

Auch für konventionelle Papiertests können die Auswertung und Ergebnisinterpretation unterstützt werden, z. B. durch das Layout der Auswertungsbögen. Bei *Deutsch Plus 4* werden nach Stufe 1 die Ja-Antworten für Modul A (zwei Fragen) und Modul B (zehn Fragen) separat gezählt und auf zwei Skalen eingetragen. Das Abbruchkriterium ist dabei visuell als die Unterteilung für beide Skalen eingetragen, somit kann jede Skala bereits alleine als Kriterium für die Förderentscheidung genutzt werden – ohne komplizierte Gewichtungen der einzelnen Skalen vornehmen zu müssen.

Früh Deutsch lernen hat als einziges Verfahren in beiden Kategorien eine sehr gute Bewertung erhalten. Es wird im Rahmen der Schulanmeldung durchgeführt. Es soll ein kurzes Gespräch mit dem Kind stattfinden, um den Sprachstand in sechs Kategorien einschätzen zu können. Der Aufwand für die Dokumentation ist sehr gering und beansprucht weniger als zwei Minuten.

ZEITLICHE ANFORDERUNGEN	Art des Verfahrens	Bundesland	Perspektive Kinder: zeitliche Dauer der Durchführung [a]	Perspektive diagnostischer Aufwand: zeitliche Dauer der Durchführung, Auswertung & Dokumentation [b]
Cito Sprachtest Version 3	T	HB	+	+++
Delfin 4	S	NW, ST	+	+++
DESK 3–6	G	MV	+++	–
Deutsch Plus 4	G	BE	+	+++
Fit in Deutsch	S	NI	+	+++
Früh Deutsch lernen	E	SL	+++	+++
HASE	S	BW	++	+++
HAVAS 5	T	SH	++	+
Kenntnisse in DaZ erfassen	G	BY	–	–
KiSS	S	HE	++	+
KISTE	S	BB	–	–*
LiSe-DaZ	T	BW	+	++
Meilensteine der Sprachentwicklung	E	BB	+++	–
QuaSta	B	BE	++	–
S-ENS	S	SN	–	+++
Seldak	B	BY, SH	+++	+
SETK 3–5	T	BW	+	–
Sismik	B	BY, SH	+++	+
SSV	S	SN	++	+++
VER-ES	G	RLP	+	+++
Vorstellungsverfahren Viereinhalbjähriger	G	HH	+	–

Tabelle 10: Analyse und Bewertung der zeitlichen Anforderungen

Lesehilfe für Tabelle 10

a

Nicht ausreichend: mehr als 30 Minuten; keine Angaben.
Ausreichend: maximal 30 Minuten.
Gut: maximal 20 Minuten.
Sehr gut: maximal 20 Minuten, abwechslungsreich gestaltet.

b

Nicht ausreichend: mehr als 40 Minuten; keine Angaben.
Ausreichend: maximal 40 Minuten.
Gut: maximal 30 Minuten.
Sehr gut: maximal 20 Minuten.

MEHRSPRACHIGKEIT

Das Verfahren erfasst die besonderen Rahmenbedingungen von Kindern, die Deutsch als Zweitsprache erlernen. Nicht jedes mehrsprachige Kind hat per se einen Sprachförderbedarf.

Fit in Deutsch berücksichtigt über einen ausführlichen Elternfragebogen die bisherigen Lerngelegenheiten zur Zweitsprache (L2) und die eingeschätzten Kompetenzen in der Erstsprache (L1).[10] Die Anzahl der Kontaktmonate mit der Zweitsprache wird dabei ebenso abgefragt wie die genaue familiäre Sprachsituation, d. h. wer in welcher Sprache mit dem Kind spricht. Das Verfahren erreicht daher für die meisten Merkmale gute oder sehr gute Bewertungen.

Blick in die Praxis: Erfassung der Sprachbiografie bei Fit in Deutsch

1	Welches ist die Erstsprache Ihres Kindes?			
2	Versteht Ihr Kind zwei (oder mehr) Sprachen?	o ja	o nein	wenn ja, welche?
3	Spricht Ihr Kind zwei (oder mehr) Sprachen?	o ja	o nein	wenn ja, welche?
4	Sprechen Sie selbst zwei (oder mehr) Sprachen?	o ja	o nein	wenn ja, welche?
5	Seit wann lebt die Familie in Deutschland?			
6	Welches ist die wichtigere Sprache?	o Erstsprache		o Zweitsprache
7	Seit wann spricht Ihr Kind Deutsch?			
8	In welcher Sprache spricht Ihr Kind …			
	mit den Eltern?	o Erstsprache	o Zweitprache	o Mischform
	mit den Geschwistern?	o Erstsprache	o Zweitprache	o Mischform
9	Welche Sprachen werden in der Familie außerdem gesprochen? (Z. B. von den Großeltern etc.)			
10	Wie gut beherrscht das Kind die Erstsprache nach Einschätzung der Eltern?			
11	Das Kind erzählt in der Erstsprache …	o viel		o wenig
12	Es kann die wichtigsten Alltagsgegenstände/Situationen benennen	o ja		o nein
13	Es kennt muttersprachliche …	o Lieder	o Spiele	o Geschichten
14	Es versteht Fragen und antwortet darauf:	o ja		o nein
15	Es versteht Aufforderungen:	o ja		o nein
16	Es nimmt in der Erstsprache eigenständig Kontakt auf zu anderen Kindern / zu Erwachsenen	o ja		o nein

MEHRSPRACHIGKEIT	Art des Verfahrens	Bundesland	Werden die Erstsprache sowie mehrsprachige Sprachgebrauchsformen berücksichtigt? [a]	Sprachbiografie, insbes. Kontaktzeit mit Zweitsprache, Aneignungssorte und -qualität [b]	Werden die Eltern gefragt, wie sie die sprachlichen Fähigkeiten ihres Kindes einschätzen? [c]	Kann die Diagnose unterschiedlich ausfallen, je nachdem, ob Mehrsprachigkeit vorliegt oder nicht? [d]	Separate Normwerte für DaZ-Kinder [e]
Cito Sprachtest Version 3	T	HB	++	–	–	+++	+
Delfin 4	S	NW, ST	+	–	–	–	–
DESK 3-6	G	MV	+	–	–	+	+
Deutsch Plus 4	G	BE	+++	–	–	+	–
Fit in Deutsch	S	NI	+++	++	+++	++	–
Früh Deutsch lernen	E	SL	–	–	–	–	–
HASE	S	BW	+	–	–	++	–
HAVAS 5	T	SH	+	–	–	+++	–
Kenntnisse in DaZ erfassen	G	BY	–	–*	–*	/	–*
KiSS	S	HE	+	+	+	++	+
KISTE	T	BB	–	–	–	–	–
LiSe-DaZ	T	BW	+	+	–	+	+++
Meilensteine der Sprachentwicklung	E	BB	–	–	–	–	–
QuaSta	B	BE	+	–	–	+	–
S-ENS	S	SN	++	++	–	–	–
Seldak	B	BY, SH	/	/	/	/	/
SETK 3-5	T	BW	+	–	–	–	–
Sismik	B	BY, SH	+	+	–	/	/
SSV	S	SN	–	–	–	+	–
VER-ES	G	RLP	+	+	+	–	–
Vorstellungsverfahren Viereinhalbjähriger	G	HH	+	–	–	–	+

Tabelle 11: Analyse und Bewertung Mehrsprachigkeit

Lesehilfe für Tabelle 11

a

Nicht ausreichend: keine Angaben; nicht berücksichtigt.

Ausreichend: wird durch Abfrage, ob Deutsch L1 oder L2 ist, berücksichtigt.

Gut: Abfrage, ob Deutsch L1 oder L2 ist. Mind. 5-fach abgestuft werden mehrsprachige sprachl. Handlungsmuster abgefragt.

Sehr gut: Es werden die Sprachen des Kindes und seine Sprachgebrauchsformen abgefragt, und es ist klar geregelt, wie diese in die Diagnose einfließen.

b

Nicht ausreichend: keine Angaben; keine Abfrage.

Ausreichend: Anzahl der Kontaktmonate wird abgefragt und berücksichtigt.

Gut: Anzahl der Kontaktmonate wird abgefragt und berücksichtigt. Aneignungsorte werden abgefragt.

Sehr gut: „gut" und klare Angaben, wie diese Kontaktmonate und Aneignungsorte in die Diagnose einfließen.

c

Nicht ausreichend: keine Angaben; keine Abfrage.

Ausreichend: Es wird die Information abgefragt, ob die Eltern nur, überwiegend oder (gar) nicht in L1 bzw. L2 mit dem Kind sprechen.

Gut: Zusätzlich erfolgt eine Einschätzung durch die Eltern, wie gut das Kind die L1 beherrscht.

Sehr gut: Die Einschätzung der Eltern wird validiert.

d

Nicht ausreichend: keine Angaben; keine Berücksichtigung.

Ausreichend: Hinweis im Manual, ob Diagnose bei Kindern mit DaZ zulässig ist.

Gut und sehr gut: Differenzierung der Diagnose in: (1) mangelnde Sprachbeherrschung Deutsch (L2) trotz ausreichender Beherrschung von L1; (2) mangelnde Sprachbeherrschung Deutsch (L2) und L1; (3) mangelnde Sprachbeherrschung Deutsch (L1); (4) weder noch, d. h. ausreichende Sprachbeherrschung.

e

Nicht ausreichend: liegen nicht vor; unbekannt.

Ausreichend: an kleiner Stichprobe ermittelt ($n < 100$).

Gut: an ausreichender Stichprobe ermittelt ($n \geq 100$).

Sehr gut: an ausreichender Stichprobe ermittelt, gegliedert nach Kontaktmonaten.

Nur separate Normwerte für Kinder mit Deutsch als Zweitsprache fehlen, um dieses Qualitätsmerkmal insgesamt mit einer Bewertung von mindestens ausreichend zu erfüllen. Um die Sprachbiografie vollständig abzudecken, genügen bei *Fit in Deutsch* zwölf Fragen, die überwiegend mit Ja bzw. Nein beantwortet werden können.

Das an *Fit in Deutsch* orientierte Verfahren *Deutsch Plus 4* erreicht hinsichtlich der Berücksichtigung der Erstsprache die gleiche Bewertung. Allerdings wurde die weitere Sprachbiografie nicht berücksichtigt, weshalb es hier keine ausreichende Bewertung erzielt. Eine weitere Variante, die Erstsprache differenzierter als nur durch eine einfache Abfrage zu erfassen, liegt mit *S-ENS* vor. Hier wird gefragt, welche Sprache in den ersten vier Lebensjahren zu Hause mit dem Kind überwiegend gesprochen wurde.

Die Berücksichtigung von Mehrsprachigkeit ist erst dann wirklich gegeben, wenn dies auch in die Diagnoseergebnisse einfließt. Der Idealfall ist die Möglichkeit zur Bestimmung der Sprachkompetenz in der Erst- und Zweitsprache. Diese Möglichkeit bieten nur *HAVAS 5* und *Cito Sprachtest Version 3*. Während *HAVAS 5* in sieben Sprachen vorliegt, besteht diese Möglichkeit bei *Cito Sprachtest Version 3* derzeit nur für Türkisch. Allerdings ist es bei dem Einsatz des *HAVAS 5* in nicht-deutscher Sprache zwingend erforderlich, dass jemand mit dieser Sprache den Test durchführt und v. a. auch auswertet, was organisatorisch einen Mehraufwand bedeutet. Der *Cito Sprachtest Version 3* kann dagegen den türkischen

Sprachstand ermitteln, ohne dass die Person, die den Test durchführt, Türkisch beherrschen muss.
Separate Normwerte für Kinder, die Deutsch als Zweitsprache erwerben, liegen für vier Verfahren vor. Sie sind allerdings teilweise nicht im Manual enthalten, sondern in Veröffentlichungen zur Verwendung des jeweiligen Instrumentes erwähnt. *LiSe-DaZ* gibt als einziges Verfahren Normwerte für Kinder mit Deutsch als Zweitsprache nach der Anzahl der Kontaktmonate an. Für vierjährige Kinder sind diese unterteilt in null bis zwölf Monate und 13 bis 35 Monate. Bei fünfjährigen Kindern wird nicht weiter unterschieden, die Anzahl der Kontaktmonate wird mit 13 bis 47 angegeben. Wünschenswert wäre eine kleinteiligere Skala, dennoch kann diese Merkmalsausprägung mit sehr gut bewertet werden.

Drei der fünf Merkmale bzgl. Mehrsprachigkeit werden von der Mehrzahl der Verfahren nicht ausreichend erfüllt: Bei 14 Verfahren wird die Sprachbiografie nicht ausreichend berücksichtigt, eine Elternabfrage zu deren Einschätzungen der Kompetenzen ihres Kindes in der Erstsprache (L1) fehlt bei 17 Verfahren. Separate Normwerte für Kinder mit Deutsch als Zweitsprache liegen bei 14 Verfahren nicht vor. Während über den Sinn separater Normwerte aus fachlicher Sicht noch Klärungsbedarf besteht, sind Einschätzungen der L1-Kompetenz und zur Sprachbiografie unkontrovers und sollten systematisch in die Verfahren und die Auswertung integriert werden.

SPEZIFITÄT DER DIAGNOSTIK

Das Verfahren gibt Hinweise darauf, welche Sprachfördermaßnahmen für das Kind infrage kommen.

Obwohl *KISTE* ein eher kurzes Testverfahren darstellt,[1] ist die Ableitung geeigneter Sprachfördermaßnahmen mit sehr gut bewertet worden. Neben der klaren Regelung, ab wann ein Förderbedarf besteht,[2] gibt es neun unterschiedliche „Profiltypen" als Ergebniskategorien vor. Darauf aufbauend können vier spezifische Förderhinweise abgeleitet werden.

Bei *LiSe-DaZ* und *KiSS*, die mit gut bewertet werden, überzeugen v. a. die nach Aufgabenbereich differenzierten Förderempfehlungen und -anregungen (*LiSe-DaZ*) bzw. bei *KiSS* die sechs unterschiedlichen Kategorien zur Ergebnisinterpretation und vier unterschiedlichen Empfehlungen zum weiteren Vorgehen. *HAVAS 5* wurde als nicht ausreichend bewertet: Das Verfahren enthält keine klaren Vorgaben für die Förderentscheidung, trotz Hinweisen für die Planung spezifischer Fördermaßnahmen. Aufgrund der geringen Standardisierung ist die Spezifität der Diagnostik jedoch so stark von der Qualifikation der pädagogischen Fachkraft abhängig, das dies nicht in die Bewertung einfließen kann.

Als Minimalanforderung an das Ergebnis eines Verfahrens wurde formuliert, dass es nennt, ob eine Förderung angezeigt ist oder nicht und wie eine spezifische Diagnose zur Erstellung eines Förderplans erfolgen kann. Sowohl Beobachtungsbögen wie *Sismik* und *Seldak* als auch das Screening *SSV* überlassen es dem Nutzer bzw. der Nutzerin, wie mit einem Kind weiter verfahren wird, wenn ein Förderbedarf festgestellt wurde. Während weder *Sismik* noch *Seldak* für einen solchen Zweck konzipiert wurden, ist bei *SSV* zu regeln, wie diese spezifische Diagnose ermittelbar ist. Im Manual ist angedeutet, dass Risikokinder dann den *SETK 3-5* durchlaufen sollten. Noch besser wäre es, wenn auch ein Screening bereits klar regelt, welches Verfahren folgen sollte, und ggf. bereits Profiltypen mitteilt. Insgesamt geben 13 Verfahren zumindest Hinweise darauf, welche Sprachfördermaßnahme für ein Kind in Frage kommt.

Blick in die Praxis: Förderhinweise anhand von Profiltypen bei KISTE

Die Profiltypen wurden clusteranalytisch ermittelt. Die Grundidee ist, dass es Kinder gibt, die in mindestens fünf der sieben Bereiche überdurchschnittliche Leistungen aufweisen, und andererseits Kinder, die in mind. fünf der sieben Bereiche unterdurchschnittliche Leistungen aufweisen.

Wichtig sind Profile, in denen Diskrepanzen zwischen z. B. der kommunikativen und der sprachstrukturellen Leistungsfähigkeit deutlich werden. Somit ergeben sich die **neun Profile**:

1. Ausgeglichen unterdurchschnittliche Leistungen

2. Ausgeglichen durchschnittliche Leistungen

3. Ausgeglichen überdurchschnittliche Leistungen

4. Unterdurchschnittliche kommunikative Befähigung bei durchschnittlicher sprachstruktureller Befähigung

5. Überdurchschnittliche sprachstrukturelle Befähigung bei kommunikativen Defiziten

6. Überdurchschnittliche kommunikative Befähigung bei durchschnittlichen sprachstrukturellen Leistungen

7. Überdurchschnittliche kommunikative Befähigung bei sprachstrukturellen Defiziten

8. Durchschnittliche kommunikative Befähigung bei überdurchschnittlicher sprachstruktureller Befähigung

9. Durchschnittliche kommunikative Befähigung bei sprachstrukturellen Defiziten.

SPEZIFITÄT DER DIAGNOSTIK	Art des Verfahrens	Bundesland	Mit dem Sprachstandsverfahren mögliche Diagnosekategorien[a]
Cito Sprachtest Version 3	T	HB	+
Delfin 4	S	NW, ST	–
DESK 3–6	G	MV	+
Deutsch Plus 4	G	BE	–
Fit in Deutsch	S	NI	+
Früh Deutsch lernen	E	SL	+
HASE	S	BW	+
HAVAS 5	T	SH	–
Kenntnisse in DaZ erfassen	G	BY	+
KiSS	S	HE	++
KISTE	T	BB	+++
LiSe-DaZ	T	BW	++
Meilensteine der Sprachentwicklung	E	BB	–
QuaSta	B	BE	–
S-ENS	S	SN	+
Seldak	B	BY, SH	–
SETK 3–5	T	BW	+
Sismik	B	BY, SH	–
SSV	S	SN	–
VER-ES	G	RLP	+
Vorstellungsverfahren Viereinhalb-jähriger	G	HH	+

Tabelle 12: Analyse und Bewertung Spezifität der Diagnostik

Lesehilfe für Tabelle 12

a

Nicht ausreichend: keine Angaben; Risiko Ja/Nein (dichotom) ohne weiteren Hinweis, wie eine differenziertere Diagnose ermöglicht wird.

Ausreichend: Risiko Ja/Nein (dichotom) sowie Verweis auf eine zweite Stufe, in der spezifisch geprüft wird.

Gut: (1) Differenzierung der Diagnose hinsichtlich mind. 3 sprachlicher Basisqualifikationen sowie (2) Verweis auf eine zweite Stufe, in der spezifisch geprüft wird.

Sehr gut: (1) Differenzierung der Diagnose hinsichtlich mind. 3 sprachlicher Basisqualifikationen; (2) Hinweise auf spezifische Fördermaßnahmen; (3) Differenzierung der Diagnose, ob Förderung oder Therapie nötig ist.

1 Für Cito Sprachtest Version 3, KISTE und den SETK 3–5 z. B. eine Aufgabe zur diskursiven oder zur pragmatischen Basisqualifikation, für Delfin 4 die Berücksichtigung der morphologisch-syntaktischen Basisqualifikation in Stufe 1.

2 Cohens Kappa > 0.4 bei drei Fachpersonen.

3 Verfahren, bei denen der Bezug zu einer Spracherwerbstheorie zumindest angedeutet wird, wurden mit ausreichend bewertet. Vier Verfahren vertreten dabei tendenziell den Nativismus (Früh Deutsch lernen, KiSS und SSV bzw. SETK 3–5). Der Kognitivismus-Interaktionismus wurde bei KISTE benannt. Einige Verfahren orientieren sich am „Interdependenz-Modell" zum Zweitspracherwerb von J. Cummins mit der Unterscheidung in Alltagssprache („basic individual communication skills", BICS) und Bildungssprache („cognitive academic language proficiency", CALP). Das Modell ist keine Spracherwerbstheorie im engeren Sinne, der Bezug wurde jedoch als „ausreichend" bewertet (z. B. Delfin 4 und VER-ES).

4 Für einige Verfahren sind Kennzahlen zur prognostischen Validität seit Längerem angekündigt: Delfin 4 wurde 2009 angekündigt; zum SSV und SETK 3–5 ist eine Veröffentlichung für 2014 angekündigt, VER-ES: „stehen noch aus".

5 Für beide Verfahren gilt allerdings, dass dies vermutlich in der Schulung bzw. Zertifizierung thematisiert wird.

6 Einzig für den Cito Sprachtest Version 3 ist keine Ermittlung der Auswertungs- und Interpretationsobjektivität notwendig (in geringem Maße aber der Durchführungsobjektivität), während eine Äquivalenz von Papier- und Computerversion belegt werden konnte.

7 SETK 3–5 bzw. SSV, wobei für diese beiden Verfahren eine Neunormierung mit n = 300 angekündigt wurde.

8 Zusätzlich wurden 107 dreijährige und 810 fünf- bis sechsjährige Kinder getestet.

9 Reynell Development Language Scales III; PLAKSS, PD, AWST-R.

10 Die Abfrage erfolgt anhand der produktiven und rezeptiven Kompetenz des Kindes sowie familiärer Umgangsformen.

11 KISTE kann als Screening mit zwei oder als Testverfahren mit sechs Subtests durchgeführt werden.

12 „Nach den Vorgaben des Landes Brandenburg (vgl. Klevenz, 2007, S. 10) gelten die Kinder als sprachförderbedürftig, die im KISTE-Untertest ‚Wortschatz' (WO) oder im Untertest ‚semantische und grammatikalische Inkonsistenzen' (IKO) oder im Untertest ‚Satzbildung' (SB) einen c-Wert von 4 nicht erreicht haben" (Quelle: BRB Abschlussbericht 2008, S. 23).

FAZIT UND EMPFEHLUNGEN

Insgesamt ist die Entwicklung von Sprachstandsverfahren auf einem guten Weg. Immer mehr Verfahren genügen diagnostischen Standards, wie sie in der Psychologie zugrunde gelegt werden. Nahezu jedes Merkmal wird mindestens einmal gut oder sehr gut erfüllt. Diese Entwicklung ist auch darauf zurückzuführen, dass sich eine interdisziplinäre Testentwicklung durchsetzt. Bei den Ländern und Entwicklern ist zudem eine große Bereitschaft zu erkennen, die kritische Analyse und Bewertung ihrer Verfahren zu unterstützen und daraus Konsequenzen zu ziehen.

Bei allen positiven Entwicklungen zeigt diese Analyse und Bewertung jedoch auch, dass nach wie vor großer Handlungsbedarf besteht. Die Hälfte aller Verfahren erfüllt 16 oder weniger der insgesamt 32 Qualitätsmerkmale ausreichend. Insbesondere Beobachtungsbögen und Einschätzverfahren erreichten häufig eine nicht ausreichende Bewertung, was aber nicht zwingend an dieser Art des Verfahrens liegt: *Sismik* und *Seldak* beispielsweise liegen im Mittelfeld. Häufig existieren zu diesen Verfahren jedoch nur spärliche Informationen, sodass die nicht ausreichende Bewertung auf fehlende Angaben, nicht unbedingt aber prinzipielle Schwächen des Verfahrens hindeutet. Der größte Handlungsbedarf besteht bei den Qualitätsmerkmalen Berücksichtigung sprachlicher Basisqualifikationen, Validität, Objektivität sowie Mehrsprachigkeit. Über die Hälfte der Verfahren können diese Merkmale nicht ausreichend erfüllen.[1]

In einigen Handlungsfeldern erzielen die Verfahren schon gute Ergebnisse: Bei den fünf formulierten Qualitätsmerkmalen zur Normierung erzielen zwischen elf und 14 der Verfahren eine mindestens ausreichende Bewertung. Ähnliches gilt für das Merkmal Zeitliche Anforderungen an die Kinder: 15 Verfahren nehmen weniger als 30 Minuten in Anspruch, viele der 15 Verfahren liegen sogar deutlich darunter mit 15–20 Minuten.

Für die vier unterschiedenen Verfahrenstypen erscheinen folgende Weiterentwicklungen besonders dringend:

Beobachtungsbögen: Ein häufiges Problem ist, dass eingeschätzt werden soll, ob die Sprache altersgemäß ausreichend entwickelt ist. Damit wird eine subjektive Referenz verlangt. Um dies einheitlich zu gestalten, sind Normen zu ergänzen, die klar benennen, was unter „altersgemäß ausreichend" verstanden wird und welche Leistungen damit verbunden sind.

Einschätzverfahren: Neben der o. g. Herausforderung ist bei Einschätzverfahren die Objektivität zu ermitteln und zu erhöhen, um

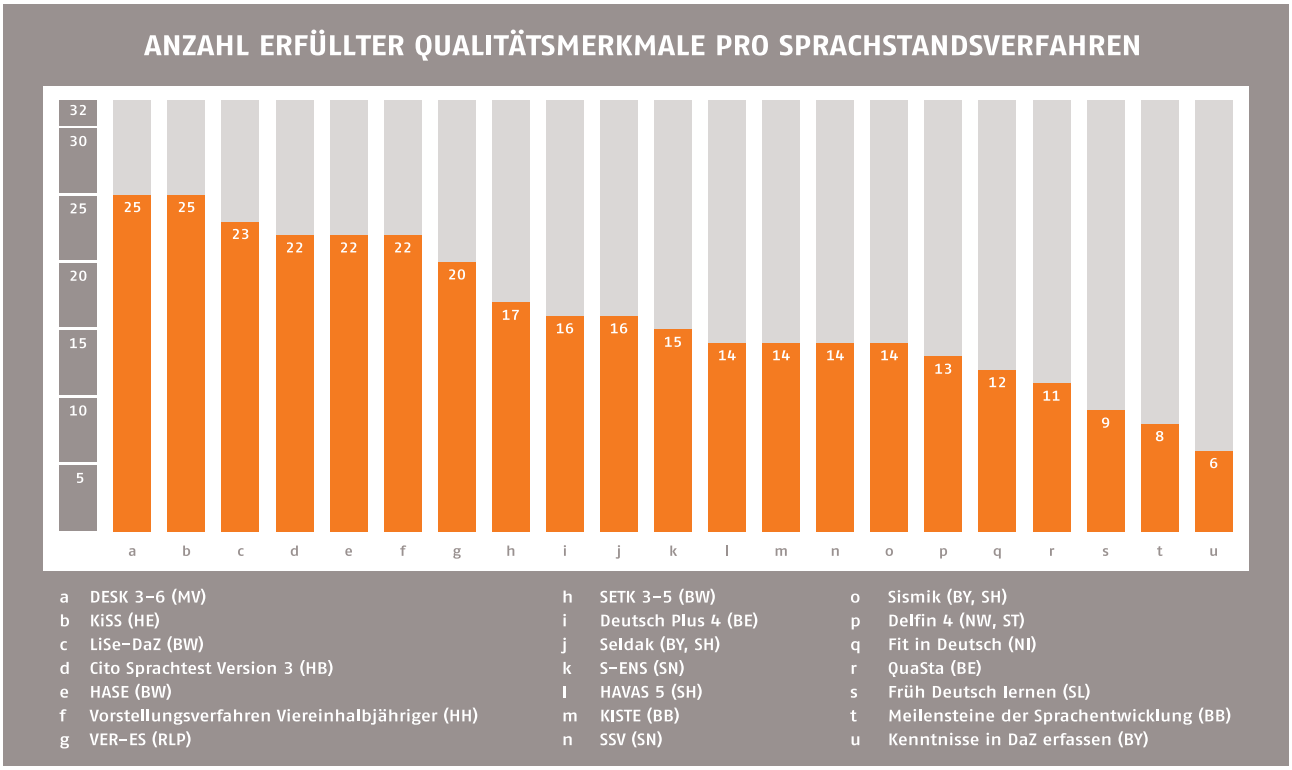

Abbildung 1: Anzahl mindestens ausreichend erfüllter Qualitätsmerkmale pro Sprachstandsverfahren

Lesehilfe für Abbildung 1

Dieses Vorgehen ist rein deskriptiv, die Schlussfolgerung, dass über die Anzahl der erfüllten Merkmale die Qualität des Verfahrens eingeschätzt werden kann, ist nicht zulässig (vgl. Kiese-Himmel & Rosenfeld, 2012, für ein vergleichbares Vorgehen). Nur im Gesamtbild und unter Berücksichtigung des angestrebten Verwendungszwecks (Förderentscheidung vs. Diagnose des genauen Förderbedarfs) ergibt sich das am besten geeignete Verfahren. Eine solche Bewertung kann aber zugunsten einer Gesamtbewertung aller offiziell eingesetzten Verfahren nicht sinnvoll vorgenommen werden.

subjektive Tendenzen zu verringern. Die Verfahren sollten konkrete Verhaltensweisen des Kindes benennen, die der Einschätzung zugrunde liegen sollten.

Testverfahren: Ein wichtiges Maß für die Funktionalität eines Testverfahrens ist die prognostische Validität. Dafür braucht es empirische Belege, ob ein Aufgabentyp Aussagen darüber zulässt, wie sich ein Kind in seiner Bildungsbiografie entwickeln wird. Auch kann über eine solche Analyse bestimmt werden, ab welchem Wert ein Sprachförderbedarf vorliegt (Bestimmung des Cut-off-Wertes).

Screenings: Die Fehlerquote muss für jedes Screening ermittelt werden. Nur so kann die Gültigkeit der Förderentscheidung belegt werden. Solange unbekannt ist, ob die Förderentscheidung auf Grundlage des Screening-Ergebnisses besser als eine Zufallsauswahl ist, kann keine Aussage über den Nutzen getroffen werden.

Über diese spezifischen Möglichkeiten hinaus, einzelne Verfahren weiterzuentwickeln, besteht ein dringender Handlungsbedarf hinsichtlich der Frage, wie der Weg zu einheitlichen Qualitätsstandards für Sprachstandsverfahren im Elementarbereich aussehen kann. Hier ist insbesondere ein abgestimmtes Handeln zwischen Wissenschaft, Politik und Bildungsadministration erforderlich:

Empfehlung 1: Konsens über Ziel und Funktion von Sprachstandsverfahren herstellen
Bisher gibt es noch keine hinreichend konsensfähigen Ziel- und Funktionsvorstellungen für die eingesetzten Verfahren. Sie werden teilweise für Zwecke verwendet, für die sie nicht konzipiert und daher nicht geeignet sind, so z. B. *Sismik*, *Seldak* und *HAVAS 5* für Förderentscheidungen. Die Zwecke und Funktionen müssten zunächst formuliert werden, um sich auf Merkmale und Kriterien zur Bewertung der Güte zu einigen, z. B. über die prognostische Validität. Erst dann können die Verfahren in Studien hinsichtlich des intendierten Zwecks empirisch geprüft werden.

Empfehlung 2: Empirische Belege für die prognostische Validität
Für eine wissenschaftlich fundierte Weiterentwicklung fehlt es nach wie vor an empirischen Belegen, sodass die Erkenntnisse dieser Untersuchung nur vereinzelt unmittelbar in die Weiterentwicklung der eingesetzten Verfahren und laufende Sprachförderprogramme einfließen können. Ein zentrales Desiderat stellt die prognostische Validität dar: Nur so ist eine Antwort auf die Frage möglich, welche Verzögerungen in der sprachlichen Entwicklung eines Kindes gravierende Auswirkungen auf den späteren Schulerfolg haben. Wenn z. B. ein Kind mit acht Jahren aufgrund seiner Sprachkompetenz Probleme hat, dem Unterricht zu folgen – wie waren seine sprachlichen Fähigkeiten im Alter von vier Jahren im Gegensatz zu einem Kind, das diese Schwierigkeiten mit acht Jahren nicht hat? Erst wenn Klarheit über diesen Unterschied besteht, kann sinnvoll definiert werden, wie gut ein Kind mit vier Jahren Sätze nachsprechen oder welche Kunstwörter es identifizieren, segmentieren, erinnern und wiedergeben können sollte. Damit wäre es möglich, die Auswahlkriterien, ab wann ein Kind Sprachförderbedarf hat, fundiert zu bestimmen, und nicht länger mehr oder weniger willkürlich zu setzen. Anstatt nur die Kinder mit dem größten Bedarf zu fördern, könnten diejenigen Kinder in ihrer sprachlichen Entwicklung unterstützt werden, bei denen ansonsten Probleme in ihrer Bildungsbiografie zu erwarten sind – unabhängig davon, ob dies fünf oder 50 Prozent sind.

Hier sollte eine Längsschnittstudie ansetzen, die eine überschaubare Anzahl von Verfahren mit sehr unterschiedlichen Funktionalitäten prüft. Dies ist erforderlich, da alle Sprachstandsverfahren dem Zweck dienen, eine Prognose zu ermöglichen: ob der Sprachstand eines Kindes zum Schuleintritt altersgemäß bzw. ausreichend ist, um dem Unterricht gut folgen zu können. Hierzu sind Daten unverzichtbar, die zeigen, wie sich die Kinder bis zu ihrem neunten Lebensjahr (dritte Klasse) entwickeln, bei denen auf Basis eines Sprachstandsverfahrens im Alter von vier Jahren ein Förderbedarf diagnostiziert worden ist. Im Rahmen von *HASE* wurde dies bereits geleistet, mit dem Projekt MILA (Mehrsprachigkeit und Sprachentwicklungsstörungen als Einflussfaktoren im Spracherwerb) werden weitere Ergebnisse aus Längsschnitt-Erhebungen verfügbar sein. Erst eine umfassende Längsschnittstudie würde es ermöglichen, die derzeit eingesetzten Verfahren zu vergleichen und zu ermitteln, welche Verfahren sich eignen, um den Sprachstand eines Kindes verlässlich und fundiert zu ermitteln.

Empfehlung 3: Keine Förderung ohne Diagnose – keine Diagnose ohne Förderung
Des Weiteren ist es dringend geboten, die Spezifität der Diagnostik zu erhöhen, auch um die Akzeptanz in der Praxis zu verbessern: Diagnostik und Förderung gehören zusammen. Die Ableitung konkreter Fördermaßnahmen sollte erst bei den als grundsätzlich förderbedürftig ausgewählten Kindern erfolgen, da die Ermittlung des spezifischen Sprachförderbedarfs sehr aufwendig sein dürfte. Ein zweistufiges Verfahren erscheint also empfehlenswert. Positive Praxisbeispiele, wie dies in einem Schritt mit der Auswahldiagnostik umgesetzt werden kann, finden sich z. B. bei *KiSS*, während mit *KISTE* und *LiSe-DaZ* gute Beispiele dafür vorliegen, die Diagnostik des genauen Förderbedarfs erst im zweiten Schritt durchzuführen.

Empfehlung 4: Qualifizierung pädagogischer Fachkräfte
Die zentrale Stellschraube für einheitliche Qualitätsstandards in der Sprachstandsfeststellung ist die Qualifizierung pädagogischer Fachkräfte. Aufseiten der Testentwicklung ergibt sich daraus

die Aufgabe, die Verfahren am Bedarf der Praxis auszurichten, wie es viele bereits versuchen (z. B. *VER-ES*). Das betrifft nicht nur die Fortbildung für den Einsatz der Verfahren und die Berücksichtigung der Anforderungen im pädagogischen Alltag, sondern auch die Spezifität der Diagnostik: Damit Erzieherinnen und Erzieher Sprachstandsdiagnostik und Sprachförderung als Teil ihres Tätigkeitsfeldes akzeptieren, müssen die Verfahren konkrete Hinweise für die Förderung liefern.

Empfehlung 5: Bundesländerübergreifender Transfer einheitlicher Qualitätsstandards

Die zentrale Herausforderung besteht in der großen Anzahl der Verfahren, die zu entwickeln und vor allem zu pflegen sind. Die Bundesländer sollten in einen Dialog darüber eintreten, ob es zielführend ist, sich gemeinsam auf weniger Verfahren zu verständigen, die jeweils festgelegte Ziele und Funktionen haben. Diese können dann unter Wahrung hoher Qualitätsstandards weiterentwickelt und optimiert werden. So zeigt sich, dass beispielsweise die frei verfügbaren Verfahren *Sismik* und *Seldak* eine breite, bundesländerübergreifende Verwendung gefunden haben. Für die fortlaufende Beobachtung und Dokumentation im pädagogischen Alltag erscheinen die Verfahren sinnvoll. Ob die Förderentscheidungen damit begründet werden sollten, ist jedoch fraglich.

Diese Aufgabe sollte ein interdisziplinäres, unabhängiges Gremium übernehmen, in dessen Verantwortung es auch läge, den Transfer aktueller wissenschaftlicher Erkenntnisse in die Testentwicklung zu sichern. Dazu wäre es erforderlich, zunächst systematisch die Ziele und Funktionen zu definieren, die mithilfe diagnostischer Verfahren im Elementarbereich erreicht werden sollen. Der Impuls dafür sollte jedoch aus den Ländern selbst kommen. Das Mercator-Institut hält daher an seiner Empfehlung fest, eine Koordinierungsstelle zu schaffen. Aufgrund der unterschiedlichen Zuständigkeiten für Sprachstandsverfahren auf Länderebene muss es den Bundesländern und weiteren zuständigen Stellen überlassen bleiben, wie eine solche Stelle strukturell eingegliedert wird. In einem ersten Schritt sollte die Koordinierungsstelle den Prozess für eine Definition, welche Qualitätsstandards wie erfüllt sein müssen, anstoßen. Daran kann sich die Entscheidung anschließen, welche Verfahren sich für den bundesländerübergreifenden Einsatz eignen.

1 Es fehlt der Nachweis des Bezugs zum Außenkriterium sprachliche Basisqualifikation (21 Verfahren), die empirisch ermittelte konvergente Validität (13 Verfahren), die Alltagsnähe der Testsituation (neun Verfahren) wie auch die der verwendeten Testaufgaben oder Beobachtungskategorien (sechs Verfahren), der empirische Nachweis der Objektivität (zwölf Verfahren), die Anzahl der Fehlklassifikationen (13 Verfahren) sowie bei der Mehrsprachigkeit die Sprachbiografie (14 Verfahren), eine systematische Elternabfrage (17 Verfahren) und separate Normwerte für Kinder, die Deutsch als Zweitsprache erlernen (14 Verfahren).

GLOSSAR

Anamnesebogen [Medizinische Diagnostik] Der Anamnesebogen erfragt eine Beschreibung der aktuellen Situation sowie die Darstellung der (Sprach-)Biografie und andere für eine diagnostische Einordnung relevante Daten. Es sollen persönliche Voraussetzungen erfasst werden, um den Einfluss verschiedener Faktoren festzustellen.

Außenkriterium [Testtheorie] Das Kriterium, das mit dem zu messenden Merkmal zusammenhängt und zur Feststellung der Validität mit dem Ergebnis des Tests übereinstimmen sollte. Wenn z. B. die diskursive Kompetenz eines Kindes mit einem Verfahren erfasst werden soll, so kann das Testergebnis mit dem kommunikativen Geschehen im Kita-Alltag (= Außenkriterium) abgeglichen werden.

Aufgabentyp [Testtheorie] Bezeichnet die Art der Aufgaben. Diese bestehen in diesem Bereich z. B. im Nachsprechen von Lautfolgen (Wörtern, Kunstwörtern, Sätzen), im Bilden von Kategorie-Begriffen, im Benennen von Farbwörtern, im Ausführen von (mehr-schrittigen) Handlungsanweisungen, in der Beschreibung von Bildern oder Bilderfolgen oder auch im Verstehen von Negationen.

Basisqualifikationen [Linguistik] Setzen sich nach Ehlich (2007, S. 14) aus zwei Bestandteilen zusammen: 1. Fähigkeiten, die Kinder während der Sprachaneignung entwickeln müssen, und 2. Wissen, das sie sich aneignen müssen, um mit Sprache kommunizieren zu können.

Cohens Kappa [Testtheorie/Statistik] Ein statistisches Maß der empirischen Forschung, das angibt, inwieweit die Urteile von zwei oder mehr verschiedenen Beobachtern übereinstimmen, also vom Beobachter unabhängig sind. Es berechnet sich aus der gemessenen Übereinstimmung und einem erwarteten Zufallswert der Übereinstimmung. Urteilen die Beobachter immer gleich, ist dieser Wert 1. Landis und Koch (1977) nennen Werte zwischen 0 und 0.2 „etwas Übereinstimmung", 0.21–0.4 „ausreichende Übereinstimmung", 0.41–0.6 „mittelmäßige (moderate) Übereinstimmung", 0.61–0.8 „beachtliche (substantial) Übereinstimmung" sowie > 0,81 „(fast) vollkommene Übereinstimmung".

COTAN-System [Übergreifend] Ein vom Committee On Test Affairs Netherlands entwickeltes Beurteilungssystem, welches ein Gesamturteil über Tests ermöglicht. Anhand von Fragen bezüglich des zu testenden Verfahrens zu verschiedenen Kategorien (Ausgangspunkte der Testkonstruktion, Qualität des Materials, Normen, Reliabilität, Validi-

tät), nach denen ein Test bewertet wird, wird dabei ein Urteil erstellt.

Cut-off-Wert [Testtheorie] Festgelegter Grenzwert, bezogen auf die zu testenden Merkmale, der eine eindeutige Trennung zwischen z. B. förderbedürftigen und nicht förderbedürftigen Kindern ermöglicht. Wird meistens als Punktwert oder Prozentrang mitgeteilt.

Faktorenstruktur [Testtheorie/Statistik] Die Faktorenstruktur eines Testverfahrens oder Beobachtungsbogens ist das Ergebnis der → Faktorenanalyse und beschreibt statistische Zusammenhänge, die in Bezug auf die inhaltlichen Aspekte interpretiert werden können. Die unterschiedlichen inhaltlichen Bereiche sollten auch in der Faktorenanalyse bestätigt werden (konfirmatorische Analyse).

Faktorenanalyse [Testtheorie/Statistik] Eine Gruppe statistischer Analyseverfahren, die z. B. Testergebnisse verschiedener Schüler auf inhaltliche Übereinstimmungen zwischen verschiedenen Testaufgaben untersucht. Die zu beantwortende Frage ist dabei, wie viele unterschiedliche Faktoren an der Lösung von z. B. 30 unterschiedlichen Aufgaben beteiligt sind. Sind alle Aufgaben nur von der Intelligenz abhängig, sollte sich ein Faktor zeigen. Sind die Aufgaben von Intelligenz und Sprache abhängig, zwei.

Indizes [Übergreifend] Vergleichsgrößen, Kennzahlen.

Instabile Übergangserscheinungen [Spracherwerbsforschung] Vorübergehende Phänomene, die während der Sprachaneignung auftreten und z. B. in der unregelmäßigen Anwendung von Regeln bestehen, was zeigt, dass die jeweilige Regel zwar bekannt, aber noch nicht internalisiert ist. Oder auch falscher Einsatz erlernter Grammatikregeln, sog. „Übergeneralisierungen", bevor die neu erlernte Regel nur noch in den zutreffenden Fällen eingesetzt wird.

Interaktionismus [Psychologie] Erklärt menschliches Verhalten durch das Zusammenwirken der jeweiligen Situation, in der beispielsweise ein Testverfahren durchgeführt wird, und persönlichen Eigenschaften.

Interne Konsistenz [Statistik] Die interne Konsistenz gibt an, wie stark die Items einer Skala miteinander zusammenhängen. Sie stellt gewissermaßen einen Umweg dar, die Messgenauigkeit eines Instruments zu erheben, wenn kein Retest oder Paralleltest zur Reliabilitätsbestimmung zur Verfügung steht: Die Reliabilitätsmessung erfolgt Test-intern, wobei jedes Item als Paralleltest behandelt wird.

Interdependenzmodell [Spracherwerbsforschung] Das Modell von J. Cummins (1980; 2000) stellt dar, wie sich die Kompetenzen von Erstprache (L1) und Zweitsprache (L2) wechselseitig beeinflussen. Es geht v. a. von einem starken Einfluss des L1-Niveaus auf das erreichbare L2-Niveau aus.

Items [Testtheorie] Einzelne Testaufgaben (z. B. das Kunstwort „Knafku", welches nachzusprechen ist, ist ein Item). → Aufgabentyp

Kognitivismus [Psychologie] Im Mittelpunkt dieser Lerntheorie stehen Informationsverarbeitungsprozesse. Sie geht davon aus, dass die kognitive Entwicklung die sprachliche Entwicklung beeinflusst, nicht umgekehrt. Prominente Vertreter waren Jerome Bruner und Jean Piaget, derzeit Jerry Fodor.

Konnektionistischer Ansatz [Informatik, Psychologie, Kognitionswissenschaften] Erklärt die Verarbeitung von Informationen im Gehirn durch die Herstellung neuronaler Muster, die Wissen darstellen und Verbindungen zu weiteren Mustern herstellen können. Der Konnektionismus geht von einer parallelen Informationsverarbeitung aus.

Korrelation [Statistik] Bezeichnet den Zusammenhang zwischen einzelnen Merkmalen oder das gemeinsame Auftreten von Merkmalen oder Ergebnissen. Wird zumeist numerisch ausgedrückt, der Wert liegt im Allgemeinen zwischen 0 (kein Zusammenhang) und 1 (perfekter Zusammenhang). Ab einem Wert von 0.6 kann von einem „deutlichen" Zusammenhang gesprochen werden.

Manual [Diagnostik] Beiheft zu einem diagnostischen Verfahren, in dem alle notwendigen Hinweise zur Vorbereitung, Durchführung, Auswertung und Interpretation des Verfahrens bzw. seiner Ergebnisse enthalten sind. Ohne Manual ist kein Verfahren wissenschaftlich bzw. „professionell". Im engeren Sinne gehören zu den notwendigen Angaben auch die ermittelten Gütekriterien (Validität, Objektivität, Reliabilität).

Nativismus [Linguistik] Geht davon aus, dass bestimmte Eigenschaften und Fähigkeiten angeboren sind und nicht erlernt oder entwickelt werden müssen. Ein Vertreter war

Noam Chomsky, der mittlerweile diese Theorie teilweise widerrufen hat.

Normwert [Testtheorie, Diagnostik] Vergleichswert aus einer umfangreichen, möglichst repräsentativen Untersuchungsgruppe, mit der das individuelle Testergebnis einer Person verglichen wird. Das Ergebnis eines Kindes in Höhe von z. B. 60 Punkten ist erst interpretierbar, wenn bekannt ist, dass Kinder im Durchschnitt 50 Punkte erzielen – das untersuchte Kind hat also ein überdurchschnittliches Ergebnis erzielt.

Psychometrisch, Psychometrie [Psychologie, Diagnostik] Die Psychometrie ist das Gebiet der Psychologie, das sich allgemein mit Theorie und Methode des psychologischen Messens befasst. Die Grundfrage, die sich Hermann Ebbinghaus ab 1880 stellte, war: Welche psychischen Eigenschaften sind auf welche Weise wissenschaftlich messbar?

RATZ-Index [Testtheorie] Der von H. Marx 1992 vorgeschlagene RATZ-Index setzt die tatsächliche Trefferquote in Bezug zur Zufallstrefferquote und zur maximal möglichen Trefferquote: Der RATZ-Index gibt an, wie deutlich das Verfahren den Zufall „schlägt".

Retest-Reliabilität [Testtheorie, Statistik] Zeigt den Grad der Übereinstimmung von Testergebnissen nach zwei oder mehr Testungen mit denselben Testpersonen. Damit sind Aussagen über die Zuverlässigkeit des Messvorgangs möglich.

Sensitivität [Testtheorie, Diagnostik] Die Sensitivität (auch Richtig-Positiv-Rate) gibt den Anteil der korrekt als positiv klassifizierten Personen an der Gesamtheit der tatsächlich positiven Personen an. Die Sensitivität bei einer Förderentscheidung gibt den Anteil der förderbedürftigen Kinder an, bei denen auch festgestellt wurde, dass ein Förderbedarf vorliegt.

Spracherwerbstheorie [Linguistik] Stellt Vermutungen/Aussagen darüber an, wie Sprache in natürlicher Umgebung unbewusst erworben wird, im Gegensatz zum bewussten Erlernen.

Spezifität [Testtheorie, Diagnostik] Die Spezifität (auch Richtig-Negativ-Rate) gibt den Anteil der korrekt als negativ klassifizierten Personen an der Gesamtheit der in Wirklichkeit negativen Personen an. Die Spezifität bei einer Förderentscheidung gibt den Anteil der nicht-förderbedürftigen Kinder an, bei denen auch festgestellt wurde, dass kein Förderbedarf vorliegt.

Übergeneralisierung [Linguistik] Liegt dann vor, wenn ein sprachliches (z. B. grammatikalisches) Phänomen auch auf andere Bereiche übertragen wird. Ein gemeinsames Merkmal der betreffenden Wörter kann dazu führen, dass sie z. B. mit derselben Endung verwendet werden, wie bei der Pluralbildung: das Auto – die Autos, der Tisch – die Tischs.

Subtest [Testtheorie, Psychologie, Diagnostik] Unter-Einheit innerhalb eines (Test-)Verfahrens, die entweder über einen anderen Aufgabentyp und/oder einen anderen inhaltlichen Bereich einen Aspekt des vom Verfahren untersuchten Merkmals erfasst.

Validität [Testtheorie] Validität ist die Belastbarkeit der Operationalisierung eines Konstrukts in beobachtbare/prüfbare Sachverhalte: Inwieweit misst das Testinstrument das, was es messen soll? Konvergente Validität ist die Übereinstimmung mit einem ähnlichen Merkmal. Diskriminante Validität bezeichnet den Nachweis, dass es nur wenig Übereinstimmung mit einem anderen (divergenten) Merkmal gibt. Die prognostische Validität gibt an, inwieweit die Vorhersage durch ein Testergebnis tatsächlich zutrifft.

Vollerhebung [Testtheorie, Empirie] Meint eine Untersuchung, für die jedes einzelne Individuum einer Gruppe (z. B. alle vierjährigen Einwohner der Bundesrepublik Deutschland) befragt bzw. getestet wird; im Gegensatz zu einer Stichprobenuntersuchung, bei der nur (zufällig) ausgewählte Personen befragt werden.

AUTOREN

Dr. Uwe Neugebauer ist wissenschaftlicher Mitarbeiter am Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache. Neben der psychologischen Diagnostik und Beratung nach Banküberfällen liegt sein Schwerpunkt in der Evaluation sprachlicher Fördermaßnahmen, so u. a. im Rahmen des BLK-Programms FörMig oder auch des Programms „sag mal was" der Baden-Württemberg-Stiftung.

Prof. Dr. Michael Becker-Mrotzek ist Direktor des Mercator-Instituts für Sprachförderung und Deutsch als Zweitsprache. Er ist Sprecher des Trägerkonsortiums Bildung durch Sprache und Schrift (BiSS), das im Auftrag von Bund und Ländern die Initiative zur Verbesserung der Sprachförderung, Sprachdiagnostik und Leseförderung koordiniert und wissenschaftlich ausgestaltet. Michael Becker-Mrotzek ist Mitglied im Beirat des Zentrums für Mehrsprachigkeit und Integration Köln (ZMI) und stellvertretender wissenschaftlicher Leiter des Zentrums für LehrerInnenbildung der Universität zu Köln.

ÜBER DAS MERCATOR-INSTITUT FÜR SPRACHFÖRDERUNG UND DEUTSCH ALS ZWEITSPRACHE

Das Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache ist ein von der Stiftung Mercator initiiertes und gefördertes Institut der Universität zu Köln. Ziel des Instituts ist es, langfristig die sprachliche Bildung an deutschen Schulen zu verbessern, damit alle Schülerinnen und Schüler gute Chancen auf eine erfolgreiche Bildungskarriere haben. Es berät Hochschulen dabei, Deutsch als Zweitsprache in der Lehrerausbildung zu verankern, fördert, vermittelt und betreibt anwendungsorientierte Forschung und trägt zur Qualifizierung von Lehrenden in Schule und Hochschule bei. Darüber hinaus beobachtet und berät es Bildungspraxis, -verwaltung und -politik.

DANKSAGUNGEN

Die Analyse und Bewertung der eingesetzten Verfahren waren nur aufgrund des Engagements der Verantwortlichen in den jeweiligen Landesministerien und der Testentwicklerinnen und Testentwickler möglich.

Daher danken wir:

Frank Adelhelm
Cito Groep Deutschland

Gerhard Bley
Ministerium für Arbeit, Gleichstellung und
Soziales Mecklenburg-Vorpommern

Detlef Diskowski
Brandenburgisches Ministerium für Bildung, Sport und Jugend

Christoph Eich
Ministerium für Schule und Weiterbildung des Lan-
des Nordrhein-Westfalen

Christa Engemann
Ministerium für Kultus, Jugend und Sport Baden-Württemberg

Dr. Judith Flender
Universität Dortmund

Dr. Marco Franzen
Universitätsmedizin Greifswald

Prof. Dr. Lilian Fried
Technische Universität Dortmund

Thomas Gericke
Ministerium für Arbeit und Soziales Sachsen-Anhalt

Karin Greve
Kultusministerium Sachsen-Anhalt

Prof. Dr. i. R. Hannelore Grimm

Annette Hautumm-Grünberg
Senatsverwaltung für Bildung, Jugend und Wissenschaft Berlin

Dr. Stefan Herb
Sozialministerium Hessen

Lena Hochstein
Senatorin für Bildung und Wissenschaft, Bremen

Daniela Hoffmann
Ministerium für Arbeit, Gleichstellung und
Soziales Mecklenburg-Vorpommern

Prof. Dr. Gisela Kammermeyer
Universität Koblenz-Landau

Anke Matthaei
Kultusministerium Niedersachsen

Dr. Peter May
Landesinstitut für Lehrerbildung und Schulentwicklung

Dr. Bernhard Nagel
Institut für Frühpädagogik (IFP)

Wiebke Richter
Behörde für Schule und Berufsbildung (BSB) Hamburg

Prof. Dr. Susanna Roux
Pädagogische Hochschule Weingarten

Ines Rackow
Senatsverwaltung für Bildung, Jugend und Wissenschaft

Sabine Rutten
Ministerium für Bildung und Wissenschaft Schleswig-Holstein

Arnfried Schlosser
Staatsministerium für Kultus Sachsen

Prof. Dr. Hermann Schöler (im Ruhestand)

Karen Schönenberg
Ministerium für Integration, Familie, Kinder,
Jugend und Frauen Rheinland-Pfalz

Bettina Schultz
Ministerium für Bildung, Wissenschaft und Kultur Thüringen

Prof. Dr. Petra Schulz
Goethe-Universität Frankfurt am Main

Andrea Stuck
Universität Koblenz-Landau

Prof. Dr. Rosemarie Tracy
Universität Mannheim

Prof. Dr. Heinrich Tröster
Technische Universität Dortmund

Jens Vollmar
Regierungsschulrat, Ministerium für Bildung und Kultur Saarland

Besonderer Dank gilt Prof. Dr. Marcus Hasselhorn (Deutsches Institut für Internationale Pädagogische Forschung) und Prof. Dr. Jürgen Wilbert (Universität Potsdam) für die wertvollen Hinweise bei der Analyse und Bewertung der Verfahren.

Adelhelm, F.: Persönliche Mitteilung vom 05.09.2013: Datenblatt zum Cito Sprachtest Version 3.

Aktas, M. (2012, Hrsg.). Entwicklungsorientierte Sprachdiagnostik und -förderung bei Kindern mit geistiger Behinderung. München: Urban & Fischer.

Albers, T. (2009). Sprache und Interaktion im Kindergarten. Bad Heilbrunn: Klinkhardt.

Baddeley, A. (1997). Human Memory. Theory and Practice. East Sussex: Psychology Press Ltd.

Becker-Mrotzek, M., et al. (2013). Qualitätsmerkmale für Sprachstandsverfahren im Elementarbereich. Ein Bewertungsrahmen für fundierte Sprachdiagnostik in der Kita. Köln: Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache.

Berliner Senatsverwaltung für Bildung, Wissenschaft und Forschung (2008). Sprachlerntagebuch für Kindertagesstätten und Kindertagespflege. Verfügbar unter: http://www.berlin.de/imperia/md/content/sen-bildung/bildungswege/vorschulische_bildung/sprachlerntagebuch_handreichung_erzieher.pdf [14.10.2013]

Bredel, U. (2007). Sprachstandsmessung – Eine verlassene Landschaft. In: Ehlich, K. et al., 2007. Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund. Bildungsforschung Band 11. Bonn/Berlin: BMBF.

Breideneichen, J. (2012). Vergleich der Diagnose und Förderansätze von Delfin 4 und LiSe-DaZ. Examensarbeit an der Universität zu Köln.

CITO Deutschland GmbH: www.de.cito.com

De Langen-Müller, U., et al. (2012). Sprachentwicklung – Verlauf, Störung, Intervention. In: Kiese-Himmel, C. (Hrsg.), Diagnostik von (umschriebenen) Sprachentwicklungsstörungen, Band 7. Frankfurt/M.: Peter Lang Verlag.

Demirkaya, S., Gültekin-Karakoc, N., & Riemer, C. (2010). Abschlussbericht MiKi. Wissenschaftliche Begleitforschung der vorschulischen Sprachförderung für Kinder mit Migrationshintergrund in Bielefeld. Universität Bielefeld. Verfügbar unter: http://www.uni-bielefeld.de/lili/-studium/faecher/daf/miki/Abschlussbericht.pdf [13.10.2013]

Döpfner, D., Mersman, S., & Trost-Brinkhuis, G. (2005). Screening des Entwicklungsstandes bei Einschulungsuntersuchungen (S-ENS). Göttingen: Hogrefe.

Duindam, T., Konak, Ö., & Kamphuis, F. (2010). Sprachtest. Wissenschaftlicher Bericht. Butzbach: Cito Deutschland GmbH. Verfügbar unter: http://www.de.cito.com/~/media/cito_de/leistungen_und_produkte/sprachtest/Sprachtest_wissenschaftlicher_Bericht_Internet.ashx [14.10.2013]

Ehlich, K., et al. (2005). Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Sprachförderung von Kindern mit und ohne Migrationshintergrund. Eine Expertise für das Bundesministerium für Bildung und Forschung. Bonn/Berlin: BMBF.

Ehlich, K., Bredel, U., & Reich, H. H. (Hrsg.) (2008). Referenzrahmen zur altersspezifischen Sprachaneignung – Forschungsgrundlagen. Bildungsforschung Band 29. Berlin: BMBF.

Eich, C.: Persönliche Mitteilung vom 22.08.2013.

Eirich, H.: Persönliche Mitteilung vom 04.07.2013.

Euler, H. A., Holler-Zittlau, I., Minnen, S., Sick, U., Dux, W., Zaretsky, Y., & Neumann, K. (2010). Psychometrische Gütekriterien eines Kurztests zur Erfassung des Sprachstands 4-jähriger Kinder. HNO, 58, S. 1116–1123.

Flender, J. (2005). Früherkennung von Entwicklungsstörungen durch Erzieherinnen: Überprüfung der Gütekriterien des Dortmunder Entwicklungsscreening für den Kindergarten (DESK 3-6). Doktorarbeit an der TU Dortmund. Verfügbar unter: http://eldorado.tu-dortmund.de:8080/bitstream/2003/21557/1/Dissertation_Flender.pdf [14.10.2013]

Franze, M., Gottschling, A., & Hoffmann, W. (2010). Das Dortmunder Entwicklungsscreening für den Kindergarten (DESK 3–6) als Basis gezielter individueller Förderung in Kindertageseinrichtungen in Mecklenburg-Vorpommern. Bundesgesundheitsblatt, 53, S. 1290–1297.

Franze, M., Gottschling, A., & Hoffmann, W. (2013a). Gesundheitsförderung und Prävention durch Früherkennung in Kindertageseinrichtungen als sozialmedizinische Aufgabe. In: Gostomzyk, J. G. (Hrsg.), Angewandte Sozialmedizin. Handbuch für Weiterbildung und Praxis. Heidelberg: ECOMED, S. 1–25.

Franze, M., Gottschling, A., & Hoffmann, W. (2013b). Screeningverfahren für Kindertageseinrichtungen als Grundlage zur Förderung der Kindergesundheit. Beitrag im Online-Handbuch „Kindergartenpädagogik". Verfügbar unter http://www.kindergartenpaedagogik.de/2258.pdf [14.10.2013]

Fried, L. (2008). Durchführungsanleitung zu Stufe 2: Besuch im Pfiffikus-Haus. Verfügbar unter: http://www.schulministerium.nrw.de/BP/Schulsystem/Schulformen/Grundschule/Sprachstand/Sprachstandsfeststellung_zwei_Jahre_vor_der_Einschulung/Protokollheft_2_Stufe.pdf [14.10.2013]

Fried, L., et al. (2009). Sprachkompetenzmodell Delfin 4. Sprachtheoretische Fundierung. Unveröffentlicht.

Fried, L. (2004). Expertise zu Sprachstandserhebungen für Kindergartenkinder und Schulanfänger. Eine kritische Betrachtung. Deutsches Jugendinstitut. Verfügbar unter: http://dji.de/bibs/271_2232_ExpertiseFried.pdf (01.10.2013)

Fried, L., Briedigkeit, E., Isele, P., & Schunder, R. (2009). Delfin 4 – Sprachkompetenzmodell und Messgüte eines Instrumentariums zur Diagnose, Förderung und Elternarbeit in Bezug auf die Sprachkompetenz. Zeitschrift für Grundschulforschung, 2/2009, S. 13–26.

Gogolin, I. (1988). Erziehungsziel Zweisprachigkeit. Konturen eines sprachpädagogischen Konzepts für die multikulturelle Schule. Hamburg: Bergmann u. Helbig.

Gottschling, A., Franze, M., & Hoffmann, W. (2012). Entwicklungsverzögerungen bei Kindern in Kindertagesstätten und Einschülern. Ein Screeningverfahren als Grundlage der bedarfsgerechten Planung von gezielter individueller Förderung in Mecklenburg-Vorpommern. Deutsches Ärzteblatt, 7, S. 308–309.

Grimm, H. (2003). Störungen der Sprachentwicklung. Göttingen: Hogrefe.

Grimm, H. (2012). Störungen der Sprachentwicklung. 3. Aufl. Göttingen: Hogrefe.

Grimm, H., Aktas, M., & Frevert, S. (2010a). SETK 3-5. Sprachentwicklungstest für drei- bis fünfjährige Kinder. Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen. Göttingen: Hogrefe.

Grimm, H., Aktas, M., & Frevert, S. (2010b). Protokollbogen zum SETK 3-5, Altersbereich 4;0 – 5;11 Jahre. Aus: Grimm, H., Aktas, M., Frevert, S. (2010). SETK 3-5. Sprachentwicklungstest für drei- bis fünfjährige Kinder. Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen. Göttingen: Hogrefe.

Grimm, H., Aktas, M., & Kießig, U. (2003). SSV. Sprachscreening für das Vorschulalter. Kurzform des SETK 3-5. Göttingen: Hogrefe.

Grimm, H.: Persönliche Mitteilung vom 19.08.2013

Grobe, T. G., Dörning, H., & Schwartz, F. W. (2012). BARMER GEK Arztreport 2012. Schriftenreihe zur Gesundheitsanalyse, Band 12. St. Augustin: Asgard-Verlag.

Häuser, D., & Jülisch, B.-R. (2013a). Meilensteine der Sprachentwicklung. Ein Beobachtungsinstrument. Handanweisung (Meilensteine_Handanweisung.pdf). Verfügbar unter: http://www.mbjs.brandenburg.de/sixcms/detail.php/bb1.c.312823.de [13.10.2013]

Häuser, D., & Jülisch, B.-R. (2013b). Meilensteine der Sprachentwicklung. Sprachbeobachtungsbögen und Auswertungsschema (Beobachtungsboegen.pdf). Verfügbar unter: http://www.mbjs.brandenburg.de/sixcms/detail.php/bb1.c.312823.de [13.10.2013]

Häuser, D., & Jülisch, B.-R., (2013). Meilensteine der Sprachentwicklung – Ein Beobachtungsinstrument. Handanweisung. Berlin: Netzwerk Integrative Förderung, S. 6.

Häuser, D., Kasielke, E., & Scheidereiter, U. (1994). KISTE. Kindersprachtest für das Vorschulalter. Weinheim: Beltz.

Hautumm-Grünberg, A.: Persönliche Mitteilung vom 15.10.2013.

Hessisches Sozialministerium (2011). KiSS Kinderbogen (KiSS_Kinderbogen_210111.pdf). Verfügbar unter: http://verwaltung.hessen.de/irj/HSM_Internet?cid=70cc3aa8b7453bb0b71f6941a7c4b532 [13.10.2013]

Hessisches Sozialministerium (2012). Online-Flyer (Flyer_Sprachscreening_KiSS_2012.pdf). Verfügbar unter: http://verwaltung.hessen.de/irj/HSM_Internet?cid=70cc3aa8b7453bb0b71f6941a7c4b532 [13.10.2013]

Hochstein, L.: Persönliche Mitteilung vom 23.08.2013.

Holler-Zittlau, I., Euler, H. A., & Neumann, K. (2011). Kindersprachscreening (KiSS) – das hessische Verfahren zur Sprachstandserfassung. Sprachheilarbeit, 5-6/2011, S. 263-268.

Hölscher, P. (2002). Kenntnisse in DaZ erfassen. Screening-Modell für Schulanfänger. Herausgeber: Staatsinstitut für Schulpädagogik und Bildungsforschung München. Stuttgart: Klett.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2009). Abschlussbericht: Früherkennungsuntersuchung auf umschriebene Störungen des Sprechens und der Sprache. IQWiG-Berichte, Nr. 57.

Kallmeyer, K. (2007). Vorschulische Maßnahmen zur Sprachstandserhebung in den deutschen Bundesländern. Eine wissenschaftliche Synopse ausgewählter praxisrelevanter Verfahren. Aachen: Shaker.

Kammermeyer, G., Roux, S., & Stuck, A. (2007a). Verfahren zur Einschätzung des Sprachförderbedarfs im Jahr vor der Einschulung. Manual. Mainz: Ministerium für Bildung, Wissenschaft, Jugend und Kultur.

Kammermeyer, G., Roux, S., & Stuck, A. (2007b). Verfahren zur Einschätzung des Sprachförderbedarfs im Jahr vor der Einschulung: Theoretischer Rahmen zur Erprobungsfassung 2006. Landau: Universität.

Kammermeyer, G., Roux, S., & Stuck, A. (2008). Verfahren zur Einschätzung des Sprachförderbedarfs im Jahr vor der Einschulung (VER-ES). Überprüfung der Endfassung 2007. Landau: Universität.

Karakaşoğlu, Y. (2011). Wissenschaftliche Expertise mit Handlungsempfehlungen für einen „Entwicklungsplan Migration und Bildung". Universität Bremen. Verfügbar unter: http://www.bildung.bremen.de/sixcms/media.php/13/migratioon-bildung.pdf [13.10.2013]

Kastner-Koller, U., & Deimann, P. (2005). Testbesprechung. Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 37 (3), S. 168–170.

Keilmann, A., Moein, G., & Schöler, H. (2012). Werden mit dem SETK 3–5 klinisch diagnostizierte Sprachentwicklungsstörungen erfasst? HNO, 60, S. 63-71.

Kiese-Himmel, J., & Rosenfeld, C. (2012). Analyse aktueller Untersuchungsinstrumente zur Früherkennung von Auffälligkeiten in Sprechen und Sprache in der pädiatrischen Vorsorgeuntersuchung U8. Gesundheitswesen, 74, S. 661–672.

Kliche, T., Wittenborn, C., & Koch, U. (2009). Was leisten Entwicklungsbeobachtungen in Kitas? Eigenschaften und Verbreitung verfügbarer Instrumente. In: Praxis der Kinderpsychologie und Kinderpsychiatrie. 58. Göttingen: Vandenhoeck & Ruprecht GmbH & Co. KG, S. 419-433.

Kooperationsinstitut der Freien Universität Berlin (2008). Zum Zusammenhang von pädagogischer Qualität in Kindertageseinrichtungen und Familien und dem Sprachstand von Kindern. Abschlussbericht BRB Sprache. Verfügbar unter:

http://www.mugv.brandenburg.de/sixcms/media.php/4113/BRB_Sprache_Abschlussbericht_2008.pdf [13.10.2013]

Krämer, D., Kompauer, I., & Schöler, H. (2013). Normierung des Screening HASE für 4- bis 4 ½-jährige Kinder. Gesundheitswesen, 75 (08/09), S. 492-495.

Kurzwernhart, P. J. (2009). Sprachstandserhebungsverfahren im Überblick. Diplomarbeit an der Universität Wien. Verfügbar unter: http://othes.univie.ac.at/6699/1/2009-08-10_0203973.pdf [13.10.2013]

Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg (2012a). Einschätzungsbogen zum Vorstellungsverfahren für Viereinhalbjährige (Protokollbogen-2.pdf). Verfügbar unter: http://www.schulenfoerdern.de/schulportal/index.php?page=18 [13.10.2013]

Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg (2012b). Protokollbogen zur Zusammenfassung der Ergebnisse des Vorstellungsverfahrens für Viereinhalbjährige (Protokollbogen-1.pdf). Verfügbar unter: http://www.schulenfoerdern.de/schulportal/index.php?page=18 [13.10.2013]

Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg (2012c). Informationen für Kitas zum Vorstellungsverfahren für Viereinhalbjährige (Infos.pdf). Verfügbar unter: http://www.schulenfoerdern.de/schulportal/index.php?page=18 [13.10.2013]

Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg (2012d). Beobachtungsaufgaben zur Kompetenzeinschätzung im Vorstellungsverfahren (Beobachtungsaufgaben.pdf). Verfügbar unter: http://www.schulenfoerdern.de/schulportal/index.php?page=18 [13.10.2013]

Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg (2012e). Merkblatt für die Sprachstandserhebung im Rahmen des Vorstellungsverfahrens für Viereinhalbjährige (Merkblatt.pdf). Verfügbar unter: http://www.schulenfoerdern.de/schulportal/index.php?page=18 [13.10.2013]

Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg (2012f). Anlage zur Dokumentation der Sprachstandserhebung im Vorstellungsverfahren für Viereinhalbjährige. Unveröffentlicht.

Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg. Bericht zum Vorstellungsverfahren Viereinhalbjähriger, Schuljahr 2005/06, Teil II. Unveröffentlicht.

Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg. Bericht zum Vorstellungsverfahren Viereinhalbjähriger, Schuljahr 2011/12. Unveröffentlicht.

Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg. Gesprächsleitfaden zum Vorstellungsverfahren für Viereinhalbjährige an der Grundschule im Schuljahr 2012/2013. Unveröffentlicht.

Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg. Übersicht: Hamburger Verfahren der Sprachstandserhebung am Übergang Kita-Schule. Unveröffentlicht.

Lisker, A. (2013). Sprachstandsfeststellung und Sprachförderung vor der Einschulung. Aktualisierte Expertise im Auftrag des DJI. München: Deutsches Jugendinstitut e. V.

May, P.: Persönliche Mitteilung vom 03.09.2013.

Mayr, T., & Ulich, M. (2010). Der Beobachtungsbogen „Seldak" – theoretische und empirische Grundlagen. In: Fröhlich-Gildhoff, K., Nentwig-Gesemann, I., & Strehmel, P. (Hrsg.), Forschung in der Frühpädagogik III, Band 5, S. 75-106.

Mayr, T., & Gsottschneider, J. (2013). Zur Objektivität und Validität des Einschätzbogens „Sprache und Literacy bei deutschsprachig aufwachsenden Kindern – SELDAK". Frühe Bildung, 2 (4), S. 203-2011.

Ministerium für Arbeit, Gleichstellung und Soziales Mecklenburg-Vorpommern (2012). Kindertagesförderung in Mecklenburg-Vorpommern - Das Gesetz und seine Verordnungen. Verfügbar unter: http://service.mvnet.de/_php/download.php?datei_id=48819 [14.10.2013]

Ministerium für Bildung, Kultur und Wissenschaft (2004). Früh Deutsch Lernen. Ein Ratgeber für Lehrer und Eltern. Verfügbar unter: http://www.saarland.de/dokumente/thema_bildung/FruehDeutschLernen.pdf [14.10.2013]

Neumann, K., Holler-Zittlau, I., van Minnen, S., Sick, U., & Euler, H. A. (2008). Das Kinder-Sprach-Screening (KiSS) – das hessische Verfahren zur Feststellung des Sprachstandes vierjähriger Kinder. Verfügbar unter: http://www.uni-kassel.de/fb4/psychologie/personal/euler/Sprachscreening.pdf [13.10.2013]

Neumann, K., Holler-Zittlau, I., van Minnen, S., Sick, U., Zaretsky, Y., & Euler, H. A. (2011). Katzengoldstandards in der Sprachstandserfassung. Sensitivität und Spezifität des Kindersprachscreenings (KiSS). HNO, 59, S. 97-109.

Niedersächsisches Kultusministerium, Referat 26 (2006a). Fit in Deutsch. Feststellung des Sprachstandes. Handreichung und Erläuterungen (fit_in_deutsch_01-3.03.06.pdf). Verfügbar unter: http://www.mk.niedersachsen.de/portal/live.php?navigation_id=2016&article_id=6149&_psmand=8 [14.10.2013]

Niedersächsisches Kultusministerium, Referat 26 (2006b). Kopiervorlage für „Fit in Deutsch". Unveröffentlicht.

Redder, A., Schwippert, K., Hasselhorn, M., Forschner, S., Fickermann, D., & Ehlich, K. (2010a). Inhaltliche und organisatorische Erläuterungen zu einem Forschungsprogramm „Sprachdiagnostik und Sprachförderung". ZUSE-Diskussionspapier, 2, S. 62ff.

Redder, A., Schwippert, K., Hasselhorn, M., Forschner, S., Fickermann, D., & Ehlich, K. (2010b). Grundzüge eines nationalen Forschungsprogramms zu Sprachdiagnostik und Sprachförderung. ZUSE-Diskussionspapier, 1.

Reich, H. H., & Roth, H.-J. (2004a). HAVAS 5. Auswertungsbogen Katze und Vogel – Deutsch. Verfügbar unter: http://www.blk-foermig.uni-hamburg.de/cosmea/core/corebase/mediabase/foermig/intranet/evaluation/programmevaluation/pdf/de_ab.pdf [14.10.2013]

Reich, H. H., & Roth, H.-J. (2004b). HAVAS 5. Auswertungshinweise Katze und Vogel – Deutsch. Verfügbar unter: http://www.blk-foermig.uni-hamburg.de/cosmea/core/corebase/mediabase/foermig/intranet/evaluation/programmevaluation/pdf/de_ah.pdf [14.10.2013]

Reich, H. H., & Roth, H.-J. (2007). HAVAS 5 – das Hamburger Verfahren zur Analyse des Sprachstands bei Fünfjährigen. In: Reich, H. H., Roth, H.-J., Neumann, U. (Hrsg.), Sprachdiagnostik im Lernprozess. Verfahren zur Analyse von Sprachständen im Kontext von Zweisprachigkeit. FÖRMIG Edition Band 3. Münster: Waxmann, S. 71-94.

Röhner, C. (2005, Hrsg.). Erziehungsziel Mehrsprachigkeit. Weinheim: Beltz Juventa.

Roos, J., Schöler, H., & Treutlein, A. (2007). Zur prognostischen Validität des Heidelberger Auditiven Screenings in der Einschulungsdiagnostik HASE. Abschlussbericht des Projektes EVER. Verfügbar unter: http://www.ph-heidelberg.de/wp/schoeler/datein/Abschlussbericht_EVER-HASE_Feb-2007.pdf [13.10.2013]

Rothe, J. (2007). Mit Primo durch den Sprachtest. KiTa Aktuell ND, 1/2007, S. 16-19.

Roux, S., Kammermeyer, G., & Stuck, A. (2011). Verfahren zur Einschätzung des Sprachförderbedarfs im Jahr vor der Einschulung (VER-ES). Psychologie in Erziehung und Unterricht, 58, S. 1-14.

Roux, S.: Persönliche Mitteilung vom 28.08.2013.

Rudolph, S., Franze, M., Gottschling, A., & Hoffmann, W. (2013). Entwicklungsgefährdungen im Bereich sozialer Kompetenzen bei 3-6-jährigen Kindern in Kindertageseinrichtungen in Mecklenburg-Vorpommern: Prävalenz und Risikofaktoren. Kindheit und Entwicklung, 22(2), S. 97-104.

Rutten, S.: Persönliche Mitteilung vom 18.09.2013.

Saarländisches Ministerium (2012). Früh Deutsch Lernen – Beobachtungen zum Sprachstand (während des Vorkurses). Unveröffentlicht.

Saarländisches Ministerium (undatiert). Sprachstandsfeststellung.doc. Unveröffentlicht.

Sarimski, K. (2001). Testinformationen zum SETK 3-5. Diagnostica, 48 (4), S. 200-207.

Schöler, H. (2011). Prognose schriftsprachlicher Leistungen und Risiken im Vorschulalter am Beispiel des Heidelberger Auditiven Screening in der Einschulungsuntersuchung (HASE). In: Hasselhorn, M. & Schneider, W. (Hrsg.), Tests & Trends, N. F., Band 9: Frühprognose schulischer Kompetenzen, S. 13-31. Göttingen: Hogrefe.

Schöler, H., & Brunner, M. (2008). HASE – Heidelberger Auditives Screening in der Einschulungsuntersuchung. Wertingen: Westra.

Schöler, H.: Persönliche Mitteilung vom 18.08.2013.

Schulministerium NRW (2007). Normierungstabelle/Entscheidungsmatrix zu Delfin 4 - Stufe 1 - Test „Besuch im Zoo" (Auswertung_und_Matrix_BiZ.pdf). Verfügbar unter: http://www.schulministerium.nrw.de/BP/Schulsystem/Schulformen/Grundschule/Sprachstand/Sprachstandsfeststellung_zwei_Jahre_vor_der_Einschulung/Auswertung_und_Matrix_BiZ.pdf [13.10.2013]

Schultz, J. (2004). Bericht über die Untersuchung zum Vergleich zwischen dem HAVAS und einer ausgewählten Testbatterie aus ausgewählten Verfahren. Im Auftrag von: Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg. Verfügbar unter: http://jensschultz.de/HAVAS_A.pdf [13.10.2013]

Schulz, P., & Tracy, R. (2012). LiSe-DaZ. Linguistische Sprachstandserhebung – Deutsch als Zweitsprache. Manual. Göttingen: Hogrefe.

Schütz, H.-P. (2008). Liste der FAQ zum Kinder-Sprachscreening (FAQ_160708-pdf) des Hessischen Sozialministeriums. Verfügbar unter: http://verwaltung.hessen.de/irj/HSM_Internet?cid=70cc3aa8b7453bb0b71f6941a7c4b532 [13.10.2013]

Senatsverwaltung für Bildung, Jugend und Sport (2005). Das Sprachstandsinstrument DEUTSCH PLUS. IA3/ID12. Verfügbar unter: http://www.berlin.de/imperia/md/content/sen-bildung/schulqualitaet/lernausgangsuntersuchungen/info_sprachstandsfeststellung.pdf?start&ts=1155733834&file=info_sprachstandsfeststellung.pdf [14.10.2013]

Senatsverwaltung für Bildung, Jugend und Wissenschaft (2013, Hrsg.). Erhebungsinstrument Deutsch Plus 4. Druckversion. Unveröffentlicht.

Settineri, J. (2012). Zur Messgüte sprachstandsdiagnostischer Verfahren im Vorschulalter. Erste Ergebnisse der Korrelationsstudie Sismik/Seldak – Delfin 4. In: Ahrenholz, B., Knapp, W., Becker, T., & Jeuk, S. (Hrsg.), Sprachstand erheben – Spracherwerb erforschen. Stuttgart: Klett. S. 35-51.

Stahn, D. (2005). Wie valide kann die Sprachkompetenz im Deutschen bei Vorschulkindern mit türkischer Muttersprache beurteilt werden? Diplomarbeit an der Universität Bielefeld. Verfügbar unter: http://www.bielefelder-institut.de/dateien/dokumente/Stahn-2005.pdf [13.10.2013]

Testzentrale: www.testzentrale.de. Informationen zu S-ENS, SSV. Hogrefe Verlag.

Tippelt, S., Kühn, P., Großheinrich, N., & von Suchodoletz, W. (2011). Diagnostische Zuverlässigkeit von Sprachtests und Elternrating bei Sprachentwicklungsstörungen. Laryngo-Rhino-Otologie, 90, S. 421-427.

Tröster, H., Flender, J., & Reineke, D. (2004). DESK 3-6. Dortmunder Entwicklungsscreening für den Kindergarten. Göttingen: Hogrefe.

Tröster, H., Flender, J., & Reineke, D. (2011). Prognostische Validität des Dortmunder Entwicklungsscreenings für den Kindergarten. Diagnostica, 57 (4), S. 201-211.

Ulich, M., & Mayr, T. (2003). Sismik. Sprachverhalten und Interesse an Sprache bei Migrantenkindern in Kindertageseinrichtungen (Beobachtungsbogen und Begleitheft). Freiburg: Herder.

Ulich, M., & Mayr, T. (2006). Seldak. Sprachentwicklung und Literacy bei deutschsprachig aufwachsenden Kindern (Beobachtungsbogen und Begleitheft). Freiburg: Herder.

Ulich, M., & Mayr, T. (2006). Sismik. Sprachverhalten und Interesse an Sprache bei Migrantenkindern in Kindertageseinrichtungen (Beobachtungsbogen und Begleitheft). Überarbeitete Auflage. Freiburg: Herder.

Ulich, M., & Mayr, T. (2006). Und wie steht es um die sprachliche Bildung deutscher Kinder? Der Beobachtungsbogen SELDAK. Kindergarten heute, 12/2006, S. 22–27.

Universität Hamburg (2010). Hamburger Verfahren zur Analyse des Sprachstands Fünfjähriger (HAVAS 5). Verfügbar unter http://www.blk-foermig.uni-hamburg.de/web/de/all/mat/diag/havas/ [14.10.2013]

Uysal, T., & Röhner, C. (2005). Diagnose von Sprachverhalten und Sprachkompetenzen mit SISMIK und CITO. In: Röhner, C. (Hrsg.), Erziehungsziel Mehrsprachigkeit. Weinheim: Beltz Juventa. S. 105–123.

Vollmar, J.: Persönliche Mitteilung vom 09.09.2013.

von Suchodoletz, W. (2012). Methoden zur Diagnostik und Therapie von Sprech- und Sprachentwicklungsstörungen. Ergänzung zu Kapitel 3 des Leitfadens Kinder- und Jugendpsychotherapie, Band 17: Sprech- und Sprachstörungen. Göttingen: Hogrefe.

Weinert, S., & Grimm, H. (2012). Sprachentwicklung. In: Schneider, W., & Lindenberger, U. (Hrsg.), Entwicklungspsychologie. 7. Aufl. Weinheim: Beltz. S. 448.

Wendt, W., Roick, T., Stanat, P., & Tschackert, K. (2008). Vorschulische Sprachstandsfeststellung ab dem Alter von 4 Jahren: Die Pilotierung von Deutsch Plus 4. Erste Ergebnisse. Institut für Schulqualität der Länder Berlin und Brandenburg e. V. Verfügbar unter: http://www.isq-bb.de/uploads/media/dplus4-bericht-080425-ppt-erlaeutert_05.pdf [14.10.2013]

Wolf, K., Stanat, P., & Wendt, W. (2009). EkoS- Evaluation der kompensatorischen Sprachförderung. Erster Zwischenbericht. Verfügbar unter: http://www.isq-bb.de/uploads/media/ekos-bericht-1-endfassung_02.pdf [13.10.2013]

GESAMTTABELLE	Ausweisung & Anzahl	Differenziertheit der Aufgaben	Berücksichtigung instabiler Übergangserscheinungen	Bezug: über ein Außenkriterium quantifiziert	Expertenrating	Bezug zu eindeutig benannter Spracherwerbstheorie	Konstruktvalidität ist von Fachpersonen nachvollziehbar	Empirische Faktorenstruktur	Prognostische Validität	Konvergente Validität	Alltagsnähe der Testsituation	Alltagsnähe der Aufgaben/Beobachtungssituation	Manual/Testheft des Verfahrens	Objektivität	Reliabilität des Verfahrens
CSV3	+	++	/	–	+	–	+	+	–	–	+	++	+++	+++	++
Delfin 4	++	+	–	–	–	+	+	–	–	+	–	–	+	–	++
DESK 3-6	+++	+++	–	–	+	–	+	–	+++	++	+++	+	+++	+	++
D+4	++	++	–	–	–	–	–	–	–	++	+	++	+	++	++
FiD	+	++	–	–	–	–	–	–	–	–	+	+++	+	–	–
FrDL	+	–	–	–	–	+	–	+	–	–	/	+	–	–	+
HASE	–	+++	–	–	–	+	+	++	++	+	–	–	++	+	+
HAVAS 5	+++	–	+	–	–	–	+	++	–	–	–	+++	+++	–	+
K DaZ	++	++	–	–	–	–	–	–	–	–	+++	+++	+	–	–
KiSS	++	+++	–	–	+	+	+	–	–	+++	+	++	+++	–	+++
KISTE	+	++	–	–	–	+	–	+	–	+	–	–	+	–	+++
LiSe-DaZ	–	+++	++	–	–	++	+	++	–	–	+	+	+++	++	+
MdS	++	–	+	–	–	–	+	–	–	–	/	+	+	+++	–
QuaSta	++	+	–	–	–	–	+	–	–	–	++	+	+++	–	–
S-ENS	+	++	–	–	–	–	+	–	–	+++	–	–	+	–	–
Seldak	+++	+++	–	–	–	–	+	+++	–	+	/	+	+	+	+++
SETK 3-5	+	++	+	–	–	+	+	–	++	–	–	–	++	–	++
Sismik	+++	+++	–	–	–	–	+	–	–	–	/	+	+	–	–
SSV	+	–	–	–	–	+	+	–	–	–	–	–	++	–	++
VER-ES	+	+	–	–	–	+	+	++	–	–	+	++	+	++	+
VVV	+++	+	+	–	+	+	+	–	++	–	–	+	+	++	+++

CSV3 = Cito Sprachtest Version 3 | D+4 = Deutsch Plus 4 | FiD = Fit in Deutsch | FrDL = Früh Deutsch lernen

K DaZ = Kenntnisse in DaZ erfassen | MdS = Meilensteine der Sprachentwicklung | VVV = Vorstellungsverfahren Viereinhalbjähriger

++	++	+	++	+++	++	++	-	++	++	++	-	-	-	++	-	-	-	++	++	++	Bezug zu den Teilbereichen
+++	++	+	-	+	-	+	-	-	++	-	+	-	+	+	-	-	-	+	++	+++	Alter der Normierung
+++	+++	-	+++	-	+++	+++	-	-	+++	+++	++	-	-	+++	-	-	-	+++	+++	+++	Umfang der Normstichprobe
+++	-	-	++	-	+++	++	+	-	++	++	++	-	++	+++	-	-	-	++	-	+++	Art der Gewinnung der Normstichprobe
+	+	+	+	-	+	+	-	-	++	+	+	-	-	-	-	-	-	+	-	++	Unterscheidung unterschiedl. Zielgruppen
-	-	-	-	++	-	-	-	-	-	+++	++	-	-	+++	-	-	+	++	-	-	Sensitivität
+++	-	+	-	+++	-	-	-	-	-	-	+++	-	-	++	-	-	+++	+++	-	-	Spezifität
	-	++	-	+++	-	-	-	-	-	+		-	-	++	-	-	+	++	-	-	RATZ-Index
+	-	+	+	+	+	+	+	-	+	-	+++	-	+	+	+	-	++	+++	-	+	Durchführende Person ist qualifiziert
+	+	++	+++	+	+++	-	++	+++	+	-	++	-	++	++	+++	+	+	+++	+	+	Perspektive Kinder: Dauer der Durchführung
-	+++	+++	+	-	+	+++	-	-	++	-	+	-	+	+++	+++	+++	+++	-	+++	+++	Perspektive Aufwand: Dauer der Durchführung, Auswertung & Dokumentation
+	+	-	+	+	/	++	+	-	+	-	+	-	+	+	-	+++	+++	+	+	++	Werden die L1 sowie mehrsprachige Sprachgebrauchsformen berücksichtigt?
-	+	-	+	-	/	++	-	-	+	-	+	-	-	-	-	++	-	-	-	-	Sprachbiografie
-	+	-	-	-	/	-	-	-	-	-	+	-	-	-	-	+++	-	-	-	-	Elternabfrage bzgl. sprachlicher Fähigkeiten ihres Kindes
-	-	+	/	-	/	-	+	-	+	-	++	/	+++	++	-	++	+	+	-	+++	Diagnose unterschiedlich, je nachdem, ob Mehrsprachigkeit vorliegt oder nicht
+	-	-	/	-	/	-	-	-	+++	-	+	-	-	-	-	-	-	+	-	+	Separate Normwerte für DaZ-Kinder
+	+	-	-	+	-	+	-	-	++	+++	++	+	-	+	+	+	-	+	-	+	Mögliche Diagnosekategorien

Herausgeber
Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache
Innere Kanalstraße 15
50823 Köln
0221 – 470 5718
info@mercator.uni-koeln.de
www.mercator-institut-sprachfoerderung.de

Das Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache ist ein von der Stiftung Mercator initiiertes und gefördertes Institut der Universität zu Köln.

ISBN 978-3-944835-00-6

Ergänzungen und Korrekturen dieser Publikation werden auf www.mercator-institut-sprachfoerderung.de veröffentlicht.

Verantwortlich
Prof. Dr. Michael Becker-Mrotzek

Autoren
Dr. Uwe Neugebauer, Prof. Dr. Michael Becker-Mrotzek

Redaktion
Anna Kleiner

Diese Publikation darf, unter Einhaltung der gängigen Zitierregeln und mit Angabe der Quelle, gern weiterverwendet werden:
Neugebauer, Uwe/ Becker-Mrotzek, Michael (2013): Die Qualität von Sprachstandsverfahren im Elementarbereich. Eine Analyse und Bewertung. Köln: Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache.

Lektorat
TextSchleiferei, Köln
www.textschleiferei.de

Gestaltung
kikkerbillen – Büro für Gestaltung, Köln
www.kikkerbillen.de

Druck
Druckhaus Süd, Köln
www.druckhaus-sued.de