

HNO

Deutsche Gesellschaft für Hals-Nasen-Ohren-Heilkunde, Kopf- und Hals-Chirurgie
Deutsche Akademie für Hals-Nasen-Ohren-Heilkunde, Kopf- und Hals-Chirurgie

Elektronischer Sonderdruck für H.A. Euler

Ein Service von Springer Medizin

HNO 2010 · 58:1116–1123 · DOI 10.1007/s00106-010-2089-7

© Springer-Verlag 2010

zur nichtkommerziellen Nutzung auf der
privaten Homepage und Institutssite des Autors

H.A. Euler · I. Holler-Zittlau · S. Minnen · U. Sick · W. Dux · Y. Zaretsky · K. Neumann

Psychometrische Gütekriterien eines Kurztests zur Erfassung des Sprachstands 4-jähriger Kinder

H.A. Euler¹ · I. Holler-Zittlau² · S. Minnen³ · U. Sick² · W. Dux⁴ · Y. Zaretsky^{2,5} · K. Neumann²

¹ Institut für Psychologie, Fachbereich Wirtschaftswissenschaften, Universität Kassel

² Schwerpunkt für Phoniatrie und Pädaudiologie, Goethe-Universität, Frankfurt am Main

³ Institut für Heil- und Sonderpädagogik, FB 03 Abteilung Sprachheilpädagogik, Universität Gießen

⁴ Sprachheilbeauftragter Landkreis Fulda, Hünfeld

⁵ Institut für Medizinische Psychologie, Goethe-Universität, Frankfurt am Main

Psychometrische Gütekriterien eines Kurztests zur Erfassung des Sprachstands 4-jähriger Kinder

Sprachkompetenz im Vorschulalter ist ein Prädiktor für schulische Leistungen [1, 4] und beruflichen Erfolg [5]. Die Lesekompetenz deutscher 15-jähriger Schüler und Schülerinnen lag in den PISA-Studien erheblich unter dem Durchschnitt der teilnehmenden OECD-Länder [2, 21], und die Streuung der Testwerte war inakzeptabel hoch.

Neben Kindern aus bildungsfernen Milieus absolvieren v. a. Kinder mit Migrationshintergrund zu oft eine unbefriedigende Schullaufbahn, wofür eine mangelnde Sprachkompetenz bei Schuleintritt mitverantwortlich sein kann. Deswegen haben verschiedene Bundesländer selektionsdiagnostische Programme für eine flächendeckende Erfassung der vorschulischen Sprachkompetenz aufgelegt [25]. Ziel dieser Programme ist es, Kinder mit unzureichender Sprachkompetenz zu identifizieren und einer Diagnostik sowie möglichst zugeschnittenen Förderprogrammen bzw. Therapiemaßnahmen frühzeitig zuzuführen.

Probleme des Screenings

Allerdings wird die Durchführbarkeit eines flächendeckenden Screenings für den Sprachstand wegen unbefriedigender Sensitivität auch skeptisch beurteilt [17].

Beispielsweise sind Kinder ohne Sprach- oder Sprechdefizite leichter zu identifizieren als Kinder mit Defiziten. Außerdem existieren unterschiedliche Definitionen und uneinheitliche Klassifikationskriterien für Sprachentwicklungsstörungen (SES), sodass Angaben über Prävalenzen im deutschsprachigen Raum enorm variieren, von niedrigen einstelligen bis weit in zweistellige Prozentangaben [12]. Während die International Classification of Diseases (ICD) nach den Diagnosekriterien für spezifische (umschriebene) Sprachentwicklungsstörungen (SSES) von kulturuniversalen Prävalenzen zwischen 6 und 8% ausgeht [18], bleiben die Prävalenzen der SES mit Komorbiditäten (unspezifische bzw. eingebettete SES) unbestimmt. Schließlich ist die Evidenz für eine langfristige Wirkung von flächendeckenden Förderprogrammen noch nicht überzeugend [19].

Kurztest

Vorgaben

Die hessische Landesregierung hat 2006 die Einführung einer Sprachstandserfassung von 4-jährigen Kindern in Kindertagesstätten (Kitas) beschlossen. Eine Arbeitsgruppe unter Leitung der Letztautorin wurde mit der Erstellung eines Kurz-

tests beauftragt, mit folgenden Vorgaben: Er sollte

- psychometrische Gütekriterien erfüllen (kurze Testdauer, einfache Durchführung, Kosteneffizienz, hohe konkurrente Validität, befriedigende Konstruktvalidität, Aussicht auf angemessene prädiktive Validität);
- von Erzieher(inne)n akzeptiert werden und von ihnen nach moderatem Schulungsaufwand durchführbar sein, das Verständnis für die Basis-komponenten der Sprachentwicklung verbessern und intensivierte individuelle Sprachförderung anregen;
- nicht nur sprachauffällige Kinder identifizieren, sondern auch unterscheiden, welche Kinder eine pädagogische Sprachförderung im Rahmen der Kita benötigen („sprachpädagogisch auffällig“) und welche einer weiteren medizinischen Abklärung und ggf. Therapie bedürfen („klinisch auffällig“);
- auch auf Kinder mit Migrationshintergrund anwendbar sein;
- aus dem Marburger Sprach-Screening für 4- bis 6-jährige Kinder (MSS; [10]) entwickelt werden.

Konstruktvalidität

Die konkurrente Validität des zu erstellenden Kurztests sollte bemessen werden

an der Korrelation mit bestehenden Siebtests sowie mit Verfahren, die einzelne linguistische Ebenen ausführlich testen. Die Konstruktvalidität war an folgenden Vorhersagen zu messen:

- Muttersprachlich deutschsprachige Kinder erreichen höhere Kennwerte als Kinder mit Migrationshintergrund und Deutsch als Zweitsprache.
- Ein Altersunterschied von 3 Monaten sollte sich in den Kennwerten widerspiegeln.
- Da der Kompetenzunterschied zwischen muttersprachlich deutschen und nichtdeutschen Kindergartenkindern bei älteren Kindern, vermutlich wegen späteren Eintritts der nichtdeutschen Kinder in die Kita, größer ist als bei jüngeren Kindern [9], sollte ein vergleichbarer Unterschied tendenziell schon im Altersabschnitt von 3 Monaten erkennbar sein.
- Schließlich sollte das Verfahren kein Geschlecht benachteiligen, da im Altersbereich von 4;0–4;5 Jahren Unterschiede zwischen den Geschlechtern nicht durchgängig nachweisbar sind [15].

Nachfolgend werden Konstrukt- und konkurrente Validität des Kurztests dargestellt. Die Gütekriterien des Tests als Sprachscreening (Sensitivität, Spezifität) werden wegen der besonderen Schwierigkeiten von Sprachscreenings ausführlich an anderer Stelle beschrieben [20].

Methode

Stichprobe

Testteilnehmer waren 257 Kinder (120 Mädchen, 137 Jungen) im Alter von 4;0–4;5 Jahren aus 43 Kitas in Frankfurt (n=85), Offenbach (n=24), Gießen (n=41), Landkreis Fulda (n=76) und Korbach (n=31), davon 77 (30%) Kinder mit Migrationshintergrund (Spannweite 5%, Landkreis Korbach, bis 58%, Offenbach). Mit der Auswahl der Regionen sollten sowohl städtische Gebiete, mit hohem Migrantenanteil (Frankfurt, Gießen, Offenbach), sowie ländliche Regionen, mit geringem Anteil (Fulda, Korbach), erfasst werden. Weitergehende Repräsentativität war we-

HNO 2010 · 58:1116–1123 DOI 10.1007/s00106-010-2089-7
© Springer-Verlag 2010

H.A. Euler · I. Holler-Zittlau · S. Minnen · U. Sick · W. Dux · Y. Zaretsky · K. Neumann Psychometrische Gütekriterien eines Kurztests zur Erfassung des Sprachstands 4-jähriger Kinder

Zusammenfassung

Hintergrund. Für eine hessenweite Erfassung kindlicher Sprachkompetenz wurde ein psychometrisch geprüft Verfahren als Voraussetzung für das Kindersprachscreening (KiSS) auf Grundlage einer Revision des Marburger Sprach-Screenings (MSS) entwickelt.

Teilnehmer und Methoden. 257 Kinder im Alter von 4;0 bis 4;5 Jahren aus 43 Kindertagesstätten wurden mit einer Testbatterie auf ihren Sprachstand überprüft, davon 214 Kinder erneut nach einem Jahr.

Ergebnisse. Die Kennwerte des Kurztests korrelierten hoch mit 2 konkurrenten Sprachscreenings (SSV, HASE) und mit einem Kennwert aus 4 diagnostischen Tests einzelner Sprachkompetenzen (Reynell III, Patholinguistische Diagnostik bei Sprachentwicklungs-

störungen, PLAKSS, AWST-R). Zur Validität wurde belegt, dass (1) Kinder mit Deutsch als Familiensprache höhere Testwerte hatten als Kinder mit einer anderen Familiensprache, (2) die 3 Monate älteren Kinder höhere Testwerte hatten als die entsprechend jüngeren Kinder und (3) die älteren Kinder mit nichtdeutscher Familiensprache einen größeren Sprachrückstand hatten als die jüngeren Kinder.

Schlussfolgerung. Der Kurztest erfasst aufwandarm, valide und vergleichsweise umfassend den Sprachstand 4-jähriger Kinder.

Schlüsselwörter

Sprachentwicklung · Sprachentwicklungsstörung · Sprachtest · Sprachscreening · Kindergarten

Validity criteria of a short test to assess speech and language competence in 4-year-olds

Abstract

Background. A psychometrically constructed short test as a prerequisite for screening was developed on the basis of a revision of the Marburger Speech Screening to assess speech/language competence among children in Hessen (Germany).

Participants and Methods. A total of 257 children (age 4.0 to 4.5 years) performed the test battery for speech/language competence; 214 children repeated the test 1 year later.

Results. Test scores correlated highly with scores of two competing language screenings (SSV, HASE) and with a combined score from four diagnostic tests of individual speech/language competences (Reynell III, patholinguistic diagnostics in impaired language development, PLAKSS, AWST-R). Valid-

ity was demonstrated by three comparisons: (1) Children with German family language had higher scores than children with another language. (2) The 3-month-older children achieved higher scores than younger children. (3) The difference between the children with German family language and those with another language was higher for the 3-month-older than for the younger children. **Conclusion.** The short test assesses the speech/language competence of 4-year-olds quickly, validly, and comprehensively.

Keywords

Language development · Language impairment · Language test · Language screening · Kindergarten

gen verschiedener Kooperationserfordernisse nicht erfüllbar.

Nach Elternangabe wurde bei den meisten Kindern mit Migrationshintergrund in der Familie sowohl Deutsch als auch mindestens eine andere Sprache gesprochen ($n=69$), bei einer Minderheit ($n=8$) nur die nichtdeutsche Familiensprache. Insgesamt kamen 20 verschiedene Fremdsprachen vor, am häufigsten Russisch ($n=14$) und Türkisch oder Kurdisch ($n=11$).

Die Erfassungsrate über alle Kitas war 83% und variierte zwischen 100 (13 Kitas) und 50% (eine Kita). Mit wenigen Ausnahmen waren die Gründe für Nichterfassung zu etwa gleichen Teilen fehlende elterliche Zusage und längere Krankheit bzw. Abwesenheit des Kindes.

Testbatterie

Marburger Sprach-Screening (MSS)

Das MSS [10] zur Sprachstandserfassung 4- bis 6-jähriger Kinder in Kitas überprüft phänomenorientiert sprachliche Schlüsselkompetenzen (Spontansprache, Sprachverständnis, Sprachproduktion, Wortschatz, Aussprache, Grammatik). Altersnormen liegen nicht vor, sodass die Auswertung klinisch-intuitiv statt algorithmisch geschieht. Zwar wurde das Verfahren an 759 Kindern eingesetzt, aber Itemanalysen und Bestimmungen von Gütekriterien standen noch aus [11]. Deswegen wurde zuerst anhand dieser Daten eine Itemanalyse (Schwierigkeitsgrad und Eigentrennschärfe) vorgenommen und das MSS überarbeitet. Items mit inakzeptablem Schwierigkeitsgrad (meist zu leichte Items) und nicht-signifikanter oder zu geringer Eigentrennschärfe wurden aussortiert, besonders bei inhaltlicher Redundanz. Das MSS enthält zudem je einen Fragebogen für Eltern und Erzieherinnen, die ebenfalls überarbeitet und mit einem Item zur Sprechflüssigkeit angereichert wurden. In den Fragebogen für Erzieherinnen wurden Items aus dem Beobachtungsbogen zur Sprachentwicklung von Migrantenkindern [27] einbezogen sowie ein Item zu Stimmstörungen.

Parallele Siebttests

Zu Vergleichszwecken wurden 2 etablierte Siebttests in die Testbatterie einbezogen,

- das Sprach-Screening für Kinder im Vorschulalter (SSV; [8]) und
- das Heidelberger Auditive Screening in der Einschulungsuntersuchung (HASE; [26]).

Das SSV überprüft das phonologische Gedächtnis für Nichtwörter und das Satzgedächtnis, das HASE das Nachsprechen von Sätzen, die Wiedergabe von Zahlenfolgen und das Nachsprechen von Kunstwörtern. Beide Siebttests untersuchen damit die Prozessierung auditiv dargebotener lautsprachlicher Strukturen, insbesondere syntaktisch-morphologische Leistungen und das phonologische Arbeitsgedächtnis sowie das phonologische Bewusstsein. Dies erfolgt unter der Annahme, dass Störungen dieser Prozesse eine wichtige Ursache für Störungen des Laut- und Schriftspracherwerbs sind. Beide Tests sind psychometrisch konstruiert und können Gütekriterien belegen.

Referenztests

Die Landschaft deutschsprachiger Verfahren für die Diagnose von Spracherwerbsstörungen ist bunt hinsichtlich der untersuchten Sprachkompetenzen und Angaben zur Testgüte [3]. Ein Goldstandard als externes Validierungskriterium und Maßstab zur Quantifizierung von Sensitivität und Spezifität des zu erstellenden Screenings ist nicht verfügbar. Der „Sprachentwicklungstest für 3- bis 5-jährige Kinder“ [7] ist zwar ein psychometrisch konstruierter Sprachtest mit belegten Gütekriterien, aber er überprüft weder den aktiven Wortschatz noch phonetisch-phonologische Auffälligkeiten, und die angegebenen Cut-off-Kriterien sind in ihrer klinischen Bedeutung ungeklärt. Deswegen wurde ein Bewertungskriterium erstellt aus 4 etablierten Verfahren, die jeweils verschiedene und als wesentlich erachtete Sprachkompetenzen überprüfen und zur konkurrenten Validierung geeignet erschienen.

1. Der Reynell-Test [23] überprüft im rezeptiven Teil das Sprachverständnis.
2. Die Psycholinguistische Analyse kindlicher Sprechstörungen (PLAKSS; [6]) detektiert phonetische und phonologische Auffälligkeiten.
3. Aus der Patholinguistischen Diagnostik bei Sprachentwicklungsstörungen

(PD; [13]) wurden 5 Untertests (18, 20, 21, 22, 23) zu syntaktisch-morphologischen Fähigkeiten einbezogen.

4. Der Aktive Wortschatztest für 3- bis 5-jährige Kinder – Revision (AWST-R; [14]) bewertet semantisch-lexikalische Leistungen.

Durchführung

In den Kitas wurde jedes Kind mit dem revidierten MSS, SSV und HASE entweder durch vorher geschulte Erzieherinnen oder durch Sprachtherapeut(inn)en getestet. Alle anderen Verfahren wurden durch Sprachtherapeutinnen durchgeführt. Die Reihenfolge der Tests wurde variiert. Mit jedem Kind wurden 4–5 Sitzungen von je etwa einer Stunde mit variiertem Testreihenfolge absolviert.

Testkonstruktive Analyse

Externes Validierungskriterium

Die Kennwerte der 4 Referenztests wurden z-transformiert und mit je gleicher Wichtung und Polung zu einem Gesamtkennwert addiert. Dieser Gesamtkennwert zeigte zufriedenstellende konkurrente Validitätsmaße, da er mit einem $r=0,55$ mit dem HASE und mit einem $r=0,61$ mit dem SSV korrelierte. Die Konstruktvalidität dieses Gesamtkennwerts war ebenfalls belegbar. Kinder, bei denen zu Hause nur Deutsch gesprochen wird, zeigten signifikant und effektstark höhere Gesamtkennwerte als Kinder, bei denen zu Hause auch oder nur eine andere Sprache gesprochen wurde [$F(1, 237)=97,69, p<0,001$ einseitig, $\eta^2_{\text{partial}}=0,30$]. Da die 1. Gruppe (Familiensprache Deutsch) signifikant älter war als die 2. Gruppe (Familiensprache nicht bzw. nicht nur Deutsch), war hierbei der Einfluss des Alters als Kovariate herausgerechnet worden. Als zweiten Beleg für die Konstruktvalidität der Referenztestbatterie hatten die 3 Monate älteren Kinder der Stichprobe signifikant höhere Kennwerte (Einfluss der Familiensprache herausgerechnet) als die entsprechend jüngeren Kinder [$F(1, 237)=11,56, p<0,001$ einseitig, $\eta^2_{\text{partial}}=0,05$].

Da bei 4- bis 6-jährigen querschnittlich erfassten Kindergartenkindern der Unterschied in der Sprachkompetenz in Deutsch zwischen Kindern mit Deutsch

als Erstsprache und Kindern mit Deutsch als Zweitsprache bei älteren Kindern größer ist als bei jüngeren Kindern [9], ist ein dritter Beleg für die Konstruktvalidität der Befund, dass schon im kurzen Zeitraum von 4;0–4;5 Jahren der querschnittliche Unterschied zwischen beiden Gruppen (Zweitsprache vs. Erstsprache Deutsch) für die 3 Monate älteren Kinder (4;3–4;5 Jahre) größer ist als für die entsprechend jüngeren Kinder (4;0–4;2 Jahre), was sich in einem signifikanten Interaktionseffekt widerspiegelt [$F(1, 236)=4,47$, $p<0,05$]. Die Validität des Gesamtkennwerts aus den 4 Referenztests ist damit gestützt, sodass dieser Kennwert in Ermangelung eines anderen unabhängigen Diagnosekriteriums selbst als konkurrenthes Validitätskriterium für den zu erstellenden Kurztest dienen kann.

Auswahl der Items

Für den Einschluss in den Kurztest wurden Items des MSS nach Schwierigkeitsgrad und Fremdtrennschärfe ausgewählt. Die ausgewählten Items waren leichter als für optimale Trennschärfe (mittlerer Schwierigkeitsgrad) geboten, um die Motivation der Kinder zu stützen und am unteren Niveau der Sprachkompetenz gut differenzieren zu können. Als Fremdtrennschärfe galt die Korrelation des Items mit dem inhaltlich entsprechenden Referenztest. Die Entscheidung über den Einschluss eines Items wurde zusätzlich nach testinhaltlichem Bedarf und bei relativer Unentschiedenheit nach Eigentrennschärfe (Korrelation des Items mit dem Untertestwert bzw. dem Gesamtwert des MSS) sowie der Korrelation mit dem Gesamtwert aus allen 4 Referenztests vorgenommen.

Insgesamt wurden 44 Items aus dem revidierten MSS als konstituierende Bestandteile des Testbogens des Sprachstandserfassungsverfahrens ausgewählt und bildeten 11 Untertests:

1. Pragmatik (Spontanäußerung zum Bild, 1 Item);
2. Sprachverständnis (4 Items);
3. Sprachproduktion (1 Item);
4. Wortschatz (11 Items);
5. Aussprache (12 Items); Wort- und Satzgrammatik anhand von
6. Pluralbildung (3 Items),
7. Partizipbildung (3 Items),

Tab. 1 Itemkennwerte: Schwierigkeitsgrad (% richtig), Eigentrennschärfe und Fremdtrennschärfen (Korrelation mit einzeltem Referenztest und mit Gesamtscore aus 4 Referenztests)

Nr.	Iteminhalt	% richtig	EigenTS	Fremdtrennschärfe (Korrelation mit Test)	Korrelation Item/Test-batterie
1	Kind spricht spontan/auf Nachfrage über Bild	92,4	0,17	-	0,11
2	Kind stottert	Bewertet nur zusammen mit Angaben von Eltern und Gruppenleiterin			
3	Kind ist heiser				
4	<i>Sprachverständnis</i> : „Zeige mir ... – einen Jungen mit einer blauen Hose“	85,8	0,44	0,37	Reynell 0,35
5	– den Jungen neben dem Klettergerüst“	76,9	0,52	0,47	Reynell 0,52
6	– die grüne Kiste unter dem Brett“	56,2	0,49	0,42	Reynell 0,51
7	– das Mädchen, das Sand in den Anhänger schüttet“	67,9	0,61	0,53	Reynell 0,58
8	<i>Sprachproduktion</i> : Kind benennt auf Aufforderung	90,9	0,28	0,24	Reynell 0,27
9	<i>Wortschatz</i> : „Was ist das?“ – Dach	78,6	0,60	0,57	AWST-R 0,49
10	– Fenster	92,0	0,59	0,48	AWST-R 0,48
11	– Klettergerüst	76,0	0,61	0,58	AWST-R 0,50
12	– Rad	81,0	0,58	0,54	AWST-R 0,51
13	– Trommel	83,4	0,62	0,56	AWST-R 0,51
14	„Welche Farbe hat der Pulli?“ – gelb	95,6	0,27	0,22	AWST-R 0,24
15	– die Hose des Kindes?“ – blau	94,8	0,38	0,28	AWST-R 0,31
16	„Wie fühlt sich das (Teddy) an?“ – weich o. Ä.	75,8	0,48	0,46	AWST-R 0,42
17	– die Rutsche – hart o. Ä.	62,8	0,50	0,41	AWST-R 0,40
18	„Welche Form hat – der Ball?“ – rund	60,4	0,48	0,49	AWST-R 0,46
19	– das Fenster?“ – eckig o. Ä.	40,1	0,39	0,42	AWST-R 0,41
20	<i>Aussprache</i> : – Dach	93,5	0,40	0,32	PLAKSS 0,28
21	– Bücher	82,5	0,37	0,32	PLAKSS 0,27
22	– Sand	88,8	0,22	0,31	PLAKSS 0,11
23	– Korb	88,0	0,34	0,27	PLAKSS 0,21
24	– Schaufel	84,7	0,20	0,42	PLAKSS 0,14
25	– Schmetterling	89,5	0,23	0,43	PLAKSS 0,16
26	– Kiste	81,3	0,29	0,37	PLAKSS 0,20
27	– Brille	86,0	0,37	0,43	PLAKSS 0,34
28	– Klettergerüst	81,8	0,40	0,40	PLAKSS 0,31
29	– Rad	89,9	0,40	0,39	PLAKSS 0,29
30	– Trommel	81,3	0,37	0,47	PLAKSS 0,34
31	– Krone	81,2	0,32	0,32	PLAKSS 0,23
32	<i>Grammatik: Plural</i> : – Äpfel	51,2	0,54	0,33	PD 0,51
33	– Bälle	70,7	0,60	0,44	PD 0,63
34	– Autos	81,2	0,51	0,33	PD 0,45
35	<i>Partizipbildung</i> : „Wie ist der Junge in das Rohr gekommen?“ (gekrochen o. Ä.)	64,9	0,51	0,34	PD 0,41
36	– „Der Junge weint. Was ist passiert?“ (hingefallen o. Ä.)	76,3	0,48	0,32	PD 0,38
37	– „Der Junge hat dem Mädchen eine Geschichte (vorgelesen o. Ä.)“	65,3	0,47	0,24	PD 0,42
38	<i>2. Person Singular</i> : Prüferin evoziert – „Du sollst mich nicht hauen!“ (o. Ä.)	45,1	0,30	0,13	PD 0,22
39	– „Du riechst gut“ (o. Ä.)	54,9	0,33	0,27	PD 0,29

Tab. 1 (Fortsetzung)

40	Präposition im Akkusativkontext: Prüferin evoziert –,inden Sandkasten’	52,4	0,45	0,30	PD	0,40
41	–,auf das Dach’	50,2	0,51	0,33	PD	0,52
42	–,durch das Rohr’	42,1	0,46	0,32	PD	0,47
43	Präposition im Dativkontext: Prüferin evoziert –,aus dem Sandkasten’	40,2	0,45	0,30	PD	0,46
44	–,auf der Wippe’	59,3	0,55	0,42	PD	0,57
45	Nebensatz mit Konjunktion: „Herr M. steht an der Tür, damit ...?“ (die Kinder reinkommen o. Ä.)	59,6	0,51	0,39	PD	0,47
46	–,„Der Junge bringt ein Brett, damit ...?“ (die Kinder bauen können o. Ä.)	64,1	0,49	0,40	PD	0,49

EigenTS Eigentrennschärfe: korrigierte Item-Gesamtscore Korrelation; Reynell Reynell-Test, PLAKSS Psycholinguistische Analyse kindlicher Sprechstörungen; PD Patholinguistische Diagnostik bei Sprachentwicklungsstörungen; AWST-R Aktiver Wortschatztest für 3- bis 5-jährige Kinder – Revision.

Tab. 2 Interkorrelationen und Determinationskoeffizienten (in %) von 4 Sprachtests für Kinder mit nur Deutsch als Familiensprache (n=168)

Skalen	AWST-R	PLAKSS ^a	PD
Reynell	0,45*** (20%)	0,18* (3%)	0,31*** (10%)
AWST-R		0,33*** (11%)	0,41*** (17%)
PLAKSS ^a			0,23** (5%)

^a Umgekehrte Polung; * p<0,05; ** p<0,01; *** p<0,001.

8. Subjekt-Verb-Kongruenz bei der 2. Person Singular (2 Items),
9. Präpositionen im Akkusativkontext (3 Items),
10. Präpositionen im Dativkontext (2 Items), sowie
11. Nebensatzbildung (2 Items).

Die 44 Items sind mit ihren Kennwerten in **Tab. 1** dargestellt.

Der Kurztest hat insgesamt folgende Bestandteile: Testbogen mit Bildvorlage, Elternbogen mit 25 Items, Kitabogen mit 26 Items und Befundbogen. Die Durchführung des Testbogens dauert in der Regel 15–20 min. Das Verfahren ist unter der Webadresse <http://www.hmaf.g.hessen.de> >Gesundheit >Kindersprachscreening (KiSS) einsehbar.

Die Eltern- und Kitabögen wurden zur Erhebung anamnestisch bedeutsamer Befunde verwendet (Familiensprache, Komorbiditäten, familiäre Sprachstörungen, Kommunikationsverhalten, Redefluss- und Stimmstörungen, bisherige oder gegenwärtige Therapien). Bei Diskrepanz zwischen Testergebnis und Fragebogenbefund oder bei wichtigen Hinweisen aus den Fragebögen, z. B. auf eine Hörstörung, wurde eine Nachbeurteilung

des Kindes durch eine Sprachtherapeutin aufgrund eines itemspezifischen Algorithmus veranlasst.

Ergebnisse

Interkorrelationen zwischen den Referenztests

Die 4 Referenztests sollen verschiedene und möglichst voneinander unabhängige linguistische Ebenen erfassen. In **Tab. 2** sind die Interkorrelationen der Referenztests dargestellt. Die höchsten Korrelationen ergaben sich zwischen AWST-R und Reynell ($r=0,45$) sowie zwischen AWST-R und PD ($r=0,41$). Die entsprechenden Determinationskoeffizienten, die den Anteil der gemeinsamen Varianz beider Testkennwerte wiedergeben, waren mit 20 und 17% eher gering.

Validitätsmaximierende Wichtung der Items

Zur Maximierung der konkurrenten Validität des Testbogens des Kurztests, also der Korrelation seines Gesamtkennwerts mit dem Gesamtkennwert aus den Referenztests, wurden verschiedene Wichtungsmodele überprüft, die alle vergleichbare Validitätskoeffizienten erga-

ben ($r=0,83$ bis $r=0,85$). Das beste Modell wichtete Items je nach Fremdtrennschärfe 3-fach (Items mit relativ hoher Trennschärfe), 2-fach (relativ mittlere) oder einfach (relativ niedrige Trennschärfe). Items mit hoher Fremdtrennschärfe traten gehäuft bei Items zu semantisch-lexikalischen und morphologisch-syntaktischen Kompetenzen auf, Items mit relativ niedriger externer Trennschärfe gehäuft bei Aussprache. In der praktischen Auswertung des Kurztests zählen 3-fach gewichtete Items 3 Punkte, entsprechend 2-fach und einfach gewichtete Items.

Der konkurrente Validitätskoeffizient von $r=0,85$ darf nicht prima facie als echter Wert angenommen werden, weil die Items u. a. nach Fremdtrennschärfe ausgewählt wurden, wobei als externes Kriterium diejenigen Referenztests galten, die in der konkurrenten Validitätsbewertung des Kurztests ebenfalls mit eingingen. Dadurch wurden mögliche Zufallseffekte addiert, was eine Überschätzung des Validitätskoeffizienten darstellen dürfte. Deswegen wurden nach gut einem Jahr (12–15 Monate später) dieselben Kinder (nun $n=214$) erneut mit dem Kurztest und den 4 Referenztests getestet. Es ergab sich eine Übereinstimmung zwischen Kurztestwerten und Kombinationskennwert aus den 4 Referenztests von $r=0,77$. Dieser konkurrente Validitätskoeffizient dürfte eher eine leichte Unterschätzung darstellen, da der Kurztest für 4-jährige Kinder konstruiert ist, folglich viele Items für 5-jährige Kinder, insbesondere deutschsprachige, zu leicht sind und deswegen die Trennschärfen der Items und damit die Güte des Tests gedrückt werden.

Konstruktvalidität und Reliabilität des Kurztests

Ein konstruktvalider Sprachtest sollte deutlich zwischen muttersprachlich deutschsprachigen und Kindern mit Migrationshintergrund und Deutsch als Zweitsprache unterscheiden. Dieses Kriterium erfüllt der Kurztest (**Abb. 1**). Seine Unterscheidungsfähigkeit ist nach seinen prüfstatischen Werten [$F(1, 40)=126,61$, $p<0,001$, $\eta^2_{\text{partial}}=0,35$; Alters Einfluss wegen Altersunterschied beider Vergleichsgruppen herausgerechnet] so gut wie der Gesamtkennwert aus den 4 umfassenden Diagnostesttests.

Weiterhin sollte der Kurztest den altersbedingten Fortschritt im Spracherwerb widerspiegeln (■ **Abb. 2**). Er erfüllt auch dieses Kriterium der Konstruktvalidität annähernd so gut wie die Referenztests [$F(1, 240)=8,11, p<0,01, \eta^2_{\text{partial}}=0,03$; Einfluss der Familiensprache herausgerechnet].

Schließlich wird erwartet, dass der Kurztest ebenso wie der Gesamtkennwert aus den Referenztests bestätigt, dass die sprachliche Kompetenzdifferenz zwischen deutschen und nichtdeutschen Kindern bei älteren Kindergartenkindern größer ist als bei jüngeren. Dies trifft quantitativ zu, aber der entsprechende Interaktionseffekt verpasst die Signifikanz [$F(1, 239)=3,29, p=0,07$].

Die Verteilung der Kurztestkennwerte war geschlechtsneutral (Mädchen: $M=60,7, SD=19,0$; Jungen: $M=61,5, SD=20,1$; Unterschiede nach t-Test bzw. Hartley F-Test nicht signifikant). Wenn deutsche Kinder und Migrantenkinder getrennt betrachtet wurden, blieb Geschlechtsneutralität bei Mittelwerten erhalten. Bei Migrantenkindern allerdings war die Varianz bei Jungen signifikant größer als bei Mädchen.

Der Cronbach- α -Koeffizient als Schätzung der inneren Konsistenz und damit Reliabilität war 0,92.

Kriterien für Auffälligkeit

Vorgabe für das Sprachstandserfassungsverfahren war, nicht nur sprachauffällige von sprachunauffälligen Kindern zu unterscheiden, sondern eine Trichotomisierung in sprachunauffällige, sprachpädagogisch förderbedürftige und klinisch abklärungs- und ggf. therapiebedürftige Kinder zu liefern. Dabei waren mehr sprachpädagogisch förderbedürftige als therapiebedürftige Kinder zu erwarten. Als „klinisch auffällig“ sollten Kinder mit Verdacht auf SES, Redefluss- und Stimmstörungen erkannt werden.

Die konkreten Setzungen der Cut-offs in den einzelnen Untertests bestimmen die Gütekriterien des Kurztests als Screening (Sensitivität, Spezifität, RAZ-Index, d. h. relativer Anstieg der Trefferquote gegenüber der Zufallstrefferquote). Die Darstellung der Gütekriterien ist aus 2 Gründen kompliziert:

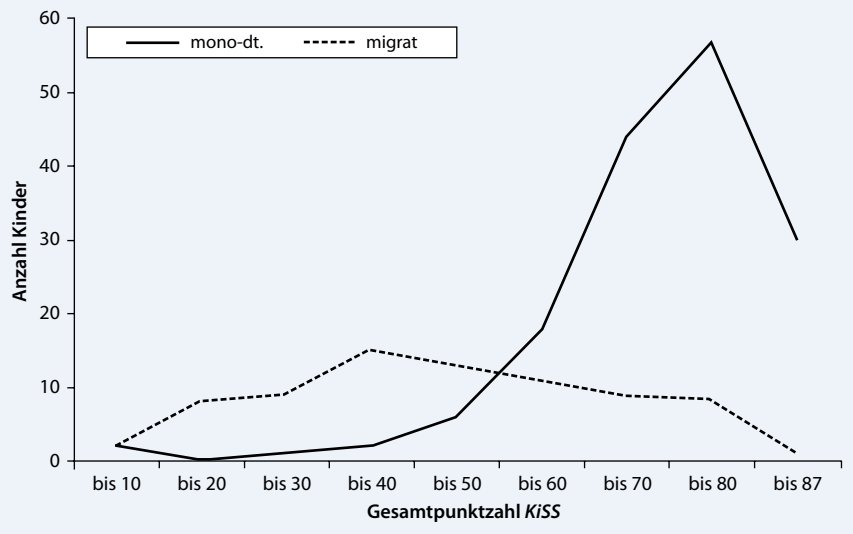


Abb. 1 ▲ Unterschiedlicher Kurvenverlauf im Kindersprachscreening (KiSS) bei Kindern mit nur Deutsch als Familiensprache (*mono-dt.*) und Kindern mit Migrationshintergrund (*migrat.*)

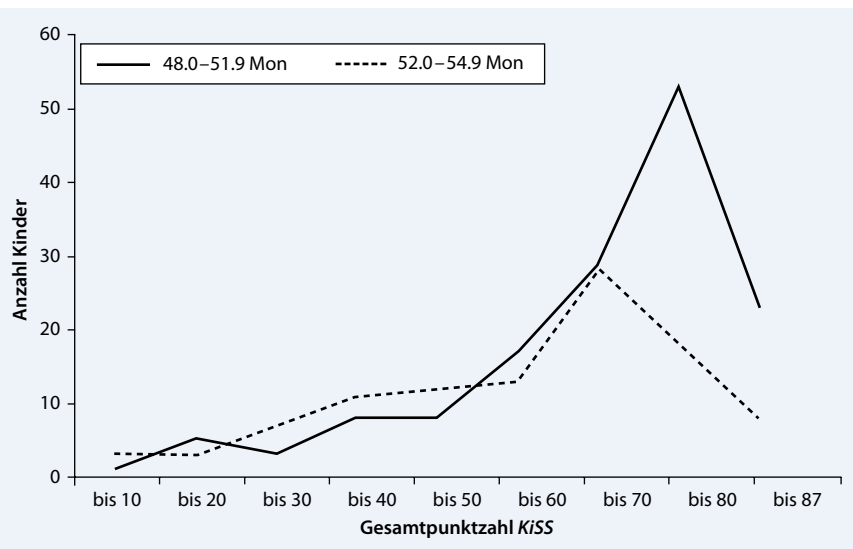


Abb. 2 ▲ Unterschiedlicher Kurvenverlauf im Kindersprachscreening (KiSS) je nach Alter (*Mon Monate*)

- Ein umfassender deutschsprachiger Test mit Goldstandardqualität zur Diagnose von Sprachauffälligkeit in allen linguistischen Ebenen ist nicht oder nur mit inhaltlichen Einschränkungen verfügbar, und
- subjektiven Expertenurteilen auf Grundlage psychometrischer Einzeltests in den verschiedenen linguistischen Bereichen mangelt es an Urteilerübereinstimmung [22].

Zudem müssen verschiedene Validitätsnachweise geliefert werden (s. Diskussion). Deswegen wird hier aus Platzgründen auf die Darstellung der Cut-offs – für

einen Bericht an anderer Stelle [20] – verzichtet.

3 Sprachscreenings im Referenztestvergleich

Da alle Kinder mehrere Sprachstandserfassungs-Kurzverfahren (HASE, SSV, hier vorgestellter Kurztest) sowie weitere Sprachtests absolviert hatten, können die Kurzverfahren miteinander verglichen werden. Das Bewertungskriterium ist dabei die Korrelation der Kurztestwerte mit den 4 Referenztests. Die Ergebnisse zeigt ■ **Tab. 3**. Zu beachten ist, dass die Screenings HASE und SSV nicht wie der hier vorgestellte Kurztest

Tab. 3 Korrelationen zwischen Kennwerten der Siebttests HASE, SSV und dem vorgestellten Kurztest, und 4 Referenztests

Screening	Referenztest				Gesamtkennwert
	Reynell	AWST-R	PLAKSS	PD	
<i>HASE</i>					
Sätze nachsprechen	0,52	0,60	0,39	0,48	0,64
Zahlenfolgen wiedergeben	0,27	0,27	0,27	0,17	0,33
Kunstwörter nachsprechen	0,27	0,21	0,34	0,23	0,33
Gesamtwert	0,44	0,46	0,42	0,36	0,55
<i>SSV</i>					
Phonologisches Arbeitsgedächtnis für Nichtwörter	0,31	0,41	0,48	0,37	0,48
Satzgedächtnis	0,57	0,69	0,28	0,55	0,70
Gesamtwert	0,44	0,56	0,47	0,47	0,61
<i>Hessischer Kurztest</i>					
Bereitschaft für Spontansprache	0,24	0,27	0,18	0,24	0,28
Sprachverständnis	0,61	0,70	0,28	0,52	0,68
Wortschatz	0,56	0,71	0,26	0,57	0,64
Artikulation	0,23	0,29	0,63	0,27	0,40
Grammatik	0,65	0,79	0,43	0,64	0,79
Gesamtwert	0,71	0,85	0,44	0,69	0,85

n=246–252, je nach Vergleich; alle r signifikant; r>0,21 signifikant mit p<0,001.

zur Maximierung der konkurrenten Validität mit den Referenztests konstruiert wurden, sondern zur Maximierung der prädiktiven Validität, und dass diverse Korrelationen zwischen den Kurztestwerten und Referenzwerten inflationiert sein dürften. Der Kurztest ist daher nur bedingt mit den beiden Screenings vergleichbar, zumal HASE für 5- bis 6-jährige Kinder konstruiert wurde. Die Angaben in **Tab. 2** sollen nur der Information dienen, nicht der Evaluation.

Diskussion

Mit dem vorgestellten Kurztest wurde ein Verfahren entwickelt, das bei 4;0- bis 4;5-jährigen Kindern valide den Sprachstand erhebt, mehrere linguistische Ebenen der Sprachentwicklung widerspiegelt, den altersabhängigen Sprachfortschritt abbildet, zwischen Sprachfähigkeiten von muttersprachlich deutschen und bi-/multilingualen Kindern mit Migrationshintergrund unterscheidet, gut im unteren Bereich der Sprachfähigkeiten differenziert, als Ausgangsbasis für eine weiterführende medizinische Diagnostik und Therapie dienen und durch Erzieherinnen durchgeführt werden kann. Seine Eignung für ein flächendeckendes Sprachscreening (Kin-

dersprachscreening, KiSS) wird an anderer Stelle dargestellt [20].

Ein solcher Kurztest unterliegt notwendigerweise psychodiagnostischen Beschränkungen. Das hier vorgestellte Verfahren verzichtet daher u. a. auf eine

- verlässliche Erfassung von individuellen Profilen der verschiedenen sprachlichen Basisqualifikationen;
- differenzierte qualitative Fehleranalyse, um einen echten Entwicklungsrückstand von einem Restrukturierungsprozess in ein höheres Kompetenzniveau und den damit verbundenen temporären „Fehlern“ zu unterscheiden;
- individualbiographische Analyse;
- Sprachstandserfassung in (quasi)natürlicher Sprechsituation, etwa durch Verhaltensbeobachtung, und damit auf eine detaillierte Erfassung von Spontansprache und Sprachpragmatik, und
- Feststellung der sprachlichen Ressourcen, über die Kinder aufgrund von Mehrsprachigkeit verfügen.

In dieser Arbeit wurden die aus kontinuierlichen Werten gewonnene Kriteriums- und konkurrente Validität des Verfahrens belegt. Für seinen Einsatz als Screening ist allerdings zusätzlichen Schwierigkeiten zu

begegnen. Ein Screening liefert üblicher Weise ein dichotomisiertes Ergebnis aus kontinuierlichen Werten unter Anwendung eines Cut-off-Kriteriums „auffällig – unauffällig“. Für das KiSS war eine Trichotomisierung mit 2 Cut-off-Kriterien gefordert. Diese beiden Kriterien unterscheiden sich zudem je nach Spracherwerbshintergrund (muttersprachlich deutsch, bi-/multilingual mit Migrationshintergrund), was einen Validitätsnachweis für alle 6 möglichen Klassifikationsergebnisse („klinisch auffällig – sprachpädagogisch auffällig – unauffällig“ je nach Sprachhintergrund) erfordert. Dieser Validitätsnachweis zusammen mit der Darstellung der Screeningergebnisse wird an anderer Stelle geliefert [20].

Die Unterscheidung in „förderbedürftig“ (Sprachentwicklungsverzögerung, mangelnder Deutschspracherwerb bei Migrationshintergrund, mangelnde Sprachanregung) oder „therapiebedürftig“ (SES, Redeflussstörung, Stimmstörung) können selbst professionelle Sprachtherapeuten für Kinder mit Migrationshintergrund bei mangelnder Beurteilbarkeit des Mutterspracherwerbs nur eingeschränkt und für muttersprachlich deutschsprachige Kinder nur mäßig trennscharf treffen [20]. Daher erhalten im vorgestellten Kurztest phonetisch-phonologische Fähigkeiten ein besonderes Gewicht (annähernd gleich für muttersprachlich deutsche und Migrantenkinder) für die Beurteilung einer klinischen Auffälligkeit, syntaktisch-morphologische Fähigkeiten ein geringeres und semantische gar kein Gewicht.

Den Autoren ist diese Einschränkung bewusst, sie stellt aber den notwendigen Kompromiss eines Siebttests dar, das Migranten- wie deutschsprachige Kinder gleichermaßen beurteilen soll. Dieses Manko wird partiell ausgeglichen, indem Eltern- und Erzieherfragebögen zusätzliche anamnestiche und Beobachtungsdaten liefern und nach einem itemspezifischen Algorithmus in die Bewertung einbezogen werden. Diese Fragebögen liefern weiterhin Aussagen über den Sprachfortschritt (z. B. Deutschsprachfähigkeiten von Migrantenkindern zum Eintritt in den Kindergarten und zum Testzeitpunkt). Schließlich ist die Aufnahme zweier weiterer Untertests beabsichtigt (Nachspre-

chen von Nonsenswörtern und Sätzen), die die prädiktive Validität für den späteren Schriftspracherwerb – auch bei Migrationshintergrund – verbessern könnten. Die Mehrzahl der Kinder mit Migrationshintergrund erhält erst mit Eintritt in den Kindergarten Zugang zur deutschen Sprache, womit sich ihr deutschsprachiger Input von den Spracheindrücken muttersprachlich deutscher Kinder unterscheidet [16]. Als Kompromiss zwischen dem zu erwartenden Stand und den Zielkompetenzen in der Verkehrssprache Deutsch wurden daher für Migrantenkinder die Kriterien morphologisch-syntaktischer und semantisch-lexikalischer Fähigkeiten geringfügig liberaler gesetzt als für muttersprachlich deutsche Kinder.

Der Kurztest ist für eine Durchführung durch verfahrensgeschulte Erzieherinnen ausgelegt. Dabei war streng der Siebtestcharakter zu wahren und zu sichern, dass den Erzieherinnen keine Sprachstandsdiagnostik abverlangt wird. Unsere bisherigen Daten stützen die Annahme, dass Erzieherinnen das Verfahren zuverlässig durchführen und auswerten können. Denn die Ergebnisse der von Erzieherinnen durchgeführten Tests wurden von Sprachtherapeutinnen nachbewertet. Diese Nachbewertung ergab hoch übereinstimmende Werte, und die Durchführung durch Erzieherinnen und durch eigens für das Verfahren geschulte und zertifizierte Sprachexpert(inn)en an verschiedenen Stichproben ergab vergleichbare Verteilungsparameter [20]. Im Übrigen wird jeder durch Erzieherinnen erhobene Befund im hessischen Sprachstandserfassungsprogramm durch Sprachexpertinnen überprüft und autorisiert.

Fazit für die Praxis

Mit dem hier vorgestellten Kurztest wurde ein deutschsprachiges Verfahren entwickelt, das aufwandarm den Sprachstand vier- bis viereinhalbjähriger Kinder unabhängig vom Sprachhintergrund valide erfasst und das aufgrund seiner konkurrenzen und Konstruktvalidität für den Einsatz zur flächendeckenden Sprachstandserfassung im Kindergartenalter geeignet scheint.

Korrespondenzadresse

Prof. H.A. Euler

Institut für Psychologie, Fachbereich Wirtschaftswissenschaften, Universität Kassel
Holländische Straße 36–38, 34127 Kassel
euler@uni-kassel.de

Danksagung. Gefördert durch das Hessische Sozialministerium (seit dem 1.2.2009 Hessisches Ministerium für Arbeit, Familie Gesundheit HMAFG) V4.1–18h26.47.01.

Die Erstellung des Verfahrens wurde vom Hessischen Sozialministerium in Auftrag gegeben. Wir danken Andreas Bodensohn, Barbara Jung, Christiane Kiese-Himmel, Amelie Löffler, Johannes Mertens, Arkadius Neumann, Carla Sack und den Projektverantwortlichen beim Hessischen Sozialministerium, Dr. Angela Wirtz und Dr. Klaus Jahn, für deren Unterstützung und Rat.

Interessenkonflikt. Der korrespondierende Autor gibt an, dass kein Interessenkonflikt besteht.

Literatur

- Aram DM, Ekelman BL, Nation JE (1984) Preschoolers with language disorders: 10 years later. *J Speech Hear Res* 27:232–244
- Baumert J, Klieme E, Neubrand M et al (Hrsg) (2001) PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Leske & Budrich, Opladen
- Bredel U (2007) Sprachstandsmessung – Eine verlässliche Landschaft. In: Ehlich K (Hrsg) Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund. Bundesministerium für Bildung und Forschung (BMBF). Referat Öffentlichkeitsarbeit, Berlin, S 78–119
- DIMDI, Deutsches Institut für Medizinische Dokumentation und Information (Hrsg) (1999) ICD-10-SGB V: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (10. Rev.). Kohlhammer, Köln
- Felsenfeld S, Broen PA, McGue M (1994) A 28-year follow-up of adults with a history of moderate phonological disorder: Educational and occupational results. *J Speech Hear Res* 37:1341–1353
- Fox A (2005) PLAKSS – Psycholinguistische Analyse kindlicher Sprechstörungen, 2. Aufl. Harcourt Test Services, Frankfurt/Main
- Grimm H (2001) SETK 3–5. Sprachentwicklungstest für drei- bis fünfjährige Kinder. Hogrefe, Göttingen
- Grimm H (2003) SSV – Sprachscreening für das Vorschulalter. Hogrefe, Göttingen
- Holler-Zittlau I (2006) Folgeerhebung (2005) Sprachliche Kompetenzen der im Jahre 2003 als sprachauffällig diagnostizierten Kinder. Hessisches Sozialministerium, Wiesbaden
- Holler-Zittlau I, Dux W, Berger R (2003) Marburger Sprach-Screening für 4- bis 6-jährige Kinder (MSS). Persen, Horneburg/Niederelbe
- Holler-Zittlau I, Dux W, Berger R (2004) Evaluation der Sprachentwicklung 4–4 1/2-jähriger Kinder in Hessen. Hessisches Sozialministerium, Wiesbaden
- Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2008) Früherkennungsuntersuchung auf umschriebene Entwicklungsstörungen des Sprechens und der Sprache. Vorbericht S06–01. IQWiG, Köln
- Kauschke C, Siegmüller J (2002) Patholinguistische Diagnostik bei Sprachentwicklungsstörungen. Urban & Fischer, München
- Kiese-Himmel C (2005) AWST-R. Aktiver Wortschatztest für 3- bis 5-jährige Kinder – Revision. Hogrefe, Göttingen
- Klann-Delius G (2005) Sprache und Geschlecht. Metzler, Stuttgart
- Kroffke S, Rothweiler M (2004) Sprachmodi im kindlichen Zweitspracherwerb. *Sprachheilarbeit* 49:18–24
- Law J, Boyle J, Harris F et al (2000) The feasibility of universal screening for primary speech and language delay: findings from a systematic review of the literature. *Dev Med Child Neurol* 42:190–200
- Leonard LB (2000) Specific language impairments across languages. In: Bishop DVM, Leonard LB (eds) *Speech and language impairments in children: causes, characteristics, intervention, and outcome*. Taylor & Francis, Philadelphia, pp 115–130
- Nelson HD, Nygren P, Walker M, Panoscha R (2006) Screening for speech and language delay in preschool children: Systematic evidence review for the US preventive services task force. *Pediatrics* 117:e298–e319
- Neumann K, Sick U, Holler-Zittlau I et al (o J) Katzungoldstandards in der Sprachstandserfassung: Sensitivität und Spezifität des Kindersprachscreening (KiSS). (Eingereicht)
- Prenzel M, Baumert J, Blum W et al (Hrsg) (2004) PISA 2003. Waxmann, Münster
- Records NL, Tomblin JB (1994) Clinical decision making: Describing the decision rules of practicing speech-language pathologists. *J Speech Hear Res* 37:144–156
- Reynell J (1997) The Reynell developmental language scales III. NFER-Nelson, Windsor
- Sarimski K (1985) Sprachentwicklungsskalen nach Joan K. Reynell. Hogrefe, Göttingen
- Schnieders G, Komor A (2007) Eine Synopse aktueller Verfahren der Sprachstandsfeststellung. In: Ehlich K (Hrsg) Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund. Bundesministerium für Bildung und Forschung. Referat Öffentlichkeitsarbeit, Berlin, S 261–342
- Schöler H, Brunner M (2008) Heidelberger Auditives Screening in der Einschulungsuntersuchung (HASE), 2. Aufl. Hogrefe, Göttingen
- Ulich M, Mayr T (2003) Sprachverhalten und Interesse an Sprache bei Migrantenkindern in Kindertageseinrichtungen. Herder, Freiburg