

Fluktuationen in den psychometrischen Kennwerten des hessischen „Kindersprachscreenings“ während und nach der Validierung

Eugen Zaretsky (Marburg), Benjamin P. Lange (Berlin) und Christiane Hey (Marburg)

Abstract

Most German states assess German language skills of children of the late kindergarten age in large-scale language screening programs. In the state of Hesse, language test “Kindersprachscreening” (“Language screening for children”, KiSS) was developed for this purpose. Its target group are kindergarten children aged four- to four-and-a-half years. In the present study, changes in the psychometric characteristics of KiSS – level of difficulty, item-total correlations, internal consistency – and possible explanations for these changes were scrutinized. For this purpose, seven KiSS samples were re-analysed retrospectively: five from KiSS validation studies and two from the Hessian language screening program. Psychometric characteristics of KiSS improved for the monolingual German-speaking children in the language screening program in comparison with KiSS validation studies. Fluctuations in the item-total correlations and internal consistency were linked to changes in the difficulty level of KiSS tasks. In the data of bi-/multilingual children, no fluctuations of psychometric characteristics were detected. Some findings point at a selection bias – an unexpectedly high proportion of children needing language therapies – in the samples of KiSS validation studies and of the language screening program. This bias can result in incorrect KiSS cut-off values, if these are updated on the basis of data from the language screening program.

1 Einleitung

Aktuell setzen fast alle Bundesländer Sprachstandserfassungsprogramme für Kinder im Kindergarten- bzw. Vorschulalter ein (cf. BMMFI 2019). In sechs davon werden nur bestimmte Gruppen der Kinder getestet, z. B. Kindergartenkinder in Hessen, in acht alle Kinder. Die meisten Bundesländer ließen für ihre Sprachstandserfassungsprogramme neue Sprachtests bzw. -screenings entwickeln, die direkt nach der Validierung und Normierung großflächig bzw. flächendeckend eingeführt wurden. Wirksamkeitsprüfungen, Nachjustierung der Normwerte und andere Verbesserungsmaßnahmen erfolgen erst später, während des breit angelegten Einsatzes dieser Verfahren.

Schon wenige Jahre nach der Einführung der Sprachstandserfassungsprogramme kamen Zweifel auf, ob diese zur Verbesserung des kindlichen Sprachstands beitragen und ihre Finanzierung

rechtfertigen. Zum einen konnte das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen keinen Nutzen der Früherkennungsuntersuchungen für die Identifikation von Kindern mit einem besonders hohen Risiko für Schwierigkeiten in der Schulausbildung nachweisen (cf. IQWiG 2009). Zum anderen zeigten die meisten Studien zur Wirksamkeit der Sprachfördermaßnahmen, darunter auch solcher, die an Sprachstandserfassungsprogramme geknüpft waren, keine nennenswerte Verbesserung des Sprachstands der geförderten Kinder im Deutschen (cf. Schneider et al. 2013). Nicht wirksame, nicht evidenzbasierte oder fehlende Sprachfördermaßnahmen relativieren den Nutzen der Sprachstandserfassungsprogramme, da Sprachtests allein zur Verbesserung der Deutschkenntnisse kaum beitragen können. Auch starke, unerwartete Fluktuationen in Auffälligkeitsraten bei validierten, normierten Sprachscreenings (cf. z. B. Weiland et al. 2019) gaben Anlass, die Glaubwürdigkeit solcher Programme zu hinterfragen. Die Weiterentwicklung der Sprachstandserfassungsprogramme wurde daraufhin in einigen Bundesländern zurückgestellt.

Es ist nicht auszuschließen, dass die unzureichende Aussagekraft der Sprachstandserfassungsprogramme teilweise auf einige Mängel in der Validierung und/oder Normierung bzw. auf Änderungen der psychometrischen Kennwerte bei der flächendeckenden Einführung von Sprachtests zurückzuführen ist. Beispielsweise können nicht repräsentative Stichproben, etwa bei einem zu hohen Anteil von sprachauffälligen Kindern in der Validierungsstudie, dazu führen, dass zu niedrige Normwerte definiert werden. Durch Fluktuationen des Schwierigkeitsindex bei der flächendeckenden Einführung eines Sprachtests können sich auch andere psychometrische Kennwerte ändern. Weitere mögliche Einflussfaktoren auf Testmerkmale bzw. -güteindizes sind mangelhafte Qualitätskontrollen bei der Testdurchführung, ausbleibende (Nach-)Schulungen der Tester und Testerinnen, Neigung zur unzulässigen Kürzung der Testverfahren, eine Missachtung der Normwerte, Einschränkung der Zielgruppe auf Bequemlichkeitsstichproben oder auffälligste Kinder und ausbleibende Nachtstungen der Kinder mit grenzwertigen bzw. fragwürdigen Befunden. Bei späteren Nachjustierungen der Normwerte auf der Grundlage der Daten aus Sprachstandserfassungsprogrammen können solche Einflussfaktoren zu fehlerhaften, und zwar zu eher niedrig angesetzten Cut-off-Werten führen.

Fluktuationen in den Ausprägungen von Testmerkmalen in Validierungsstudien und Sprachstandserfassungsprogrammen wurden bisher kaum untersucht und sollen daher in der aktuellen Arbeit anhand des hessischen „Kindersprachscreenings“ (KiSS) (cf. Euler et al. 2010; Holler-Zittlau/Euler/Neumann 2011) für vier- bis viereinhalbjährige Kinder geschildert werden. Dafür soll zuerst ein Überblick über die Entwicklung des hessischen Sprachscreenings und Sprachstandserfassungsprogramms erfolgen.

Im Jahr 2006 traf die hessische Landesregierung die Entscheidung, ein Sprachscreening für Kindergartenkinder entwickeln zu lassen. Das Screening sollte in hessischen Kindertageseinrichtungen (Kitas) durch speziell dafür geschulte Kita-Erzieher und -Erzieherinnen durchgeführt werden. Als Itempool für den neuen Test, KiSS, diente das *Marburger Sprach-Screening für 4- bis 6-jährige Kinder* (MSS) (cf. Holler-Zittlau/Dux/Berger 2003). MSS enthält Untertests zu Sprachverständnis, Wortschatz, Aussprache und Grammatik. Mit der Durchführungsdauer von ca. 40 Minuten (einschl. Auswertung) entspricht es aber nicht den Anforderungen an ein Screening.

Die Entwicklung von KiSS erfolgte in den Jahren 2007–2012. Dabei wurden nicht nur bereits vorhandene MSS-Untertests komprimiert, sondern auch neue Untertests und Items hinzugefügt (vor allem Nachsprechen von Kunstwörtern und Sätzen zur Prüfung des phonologischen Kurzzeitgedächtnisses). In mehreren prospektiven Kohortenstudien wurde KiSS in den Versionen KiSS.1, KiSS.XL und KiSS.2 validiert und normiert. Seit 2017 wird an der neuesten Version, KiSS.3, gearbeitet (cf. Zaretsky/Hey 2021).

KiSS.2 ist das einzige Sprachscreening in Deutschland, welches alle linguistischen Ebenen – die phonetisch-phonologische, die lexikalisch-semantische, die morphologisch-syntaktische – und zusätzlich die metalinguistische (pragmatisch-kommunikative) innerhalb kürzester Zeit prüfen kann. Im Gegensatz zu den meisten anderen Sprachtests, die ausschließlich dichotom zwischen sprachauffälligen und -unauffälligen Kindern unterscheiden, nimmt KiSS darüber hinaus eine weitere Differenzierung in Form einer Trichotomisierung vor und unterteilt sprachauffällige Kinder in sprachpädagogisch förderbedürftige und klinisch abklärungsbedürftige. Während sprachpädagogisch förderbedürftige Kinder lediglich eine individuell zugeschnittene Sprachfördermaßnahme und damit gezielten Sprachinput benötigen, wird bei klinisch Abklärungsbedürftigen eine Sprachentwicklungsstörung und damit ein Therapiebedarf vermutet.

Eine besondere Herausforderung bei der KiSS-Entwicklung stellte der wachsende Anteil zwei- bzw. mehrsprachiger Kinder in Deutschland dar. Typische Problemfelder bei der Sprachtestentwicklung, wenn man solche Kinder bei der Testauswertung nicht ausklammert, sind die Erarbeitung von speziellen Cut-off-Kriterien für alle zwei-/mehrsprachigen Kinder oder ihre Untergruppen, die Berücksichtigung des Alters des Deutscherwerbs, eingeschränkte Testbarkeit mancher Kinder aufgrund ihrer geringen Deutschkenntnisse, wesentliche Unterschiede in Testgüteindizes (z. B. im Schwierigkeitsindex) zwischen einsprachig deutschen und zwei-/mehrsprachigen Kindern sowie schwere Unterscheidbarkeit der Sprachentwicklungsstörungen und eingeschränkten Deutschkenntnisse bei Letzteren (cf. Grimm/Schulz 2014, 2016). Während die meisten Sprachtests zum Zeitpunkt der KiSS-Entwicklung zwei-/mehrsprachige Kinder in ihren Handbüchern entweder gar nicht oder nur marginal in ergänzenden Hinweisen erwähnten (cf. z. B. Grimm 2001; Kiese-Himmel 2005), wurde im Laufe der Arbeit an KiSS eine spezielle Studie alleine der Präzisierung von (in vorherigen Studien definierten) Cut-off-Kriterien für solche Kinder gewidmet (cf. Neumann/Euler 2010; Zieleniewicz 2014).

Da Kinder mit Migrationshintergrund in Deutschland deutlich seltener sprachtherapeutisch versorgt werden als Kinder ohne Migrationshintergrund (cf. Bahr 2007; Zaretsky et al. 2021), kann davon ausgegangen werden, dass Sprachentwicklungsstörungen bei (in der Regel zwei- bzw. mehrsprachigen) Kindern mit Migrationshintergrund unterdiagnostiziert bleiben (cf. De Jong/Çavus/Baker 2010; Yağmur/Nap-Kolhoff 2010). Daher standen bei der KiSS-Entwicklung sprachentwicklungsgestörte zwei-/mehrsprachige Kinder im Fokus. Um solche Kinder trotz minimaler Deutschkenntnisse bzw. bei eingeschränkter Testbarkeit zuverlässiger zu detektieren, wurden in den KiSS-Fragebögen für Eltern und Kita-Erzieher bzw. -Erzieherinnen Items zur Entwicklung des Kindes in der nicht-deutschen Muttersprache (z. B. später Sprachbeginn), familiären Disposition (z. B. Lese-Rechtschreib-Schwäche in der Familie) und zu typischen Risikofaktoren (z. B. Risiko- bzw. Frühgeburt) integriert (cf. Grimm/Schulz 2014). Obwohl die

Zwei- bzw. Mehrsprachigkeit das Risiko einer Sprachentwicklungsstörung nicht erhöht (cf. Armon-Lotem 2012; Paradis et al. 2003), musste bei der KiSS-Entwicklung davon ausgegangen werden, dass allein schon aufgrund des Rekrutierungsbias sämtliche Validierungsstichproben einen hohen Prozentsatz an sprachentwicklungsgestörten zwei-/mehrsprachigen Kindern aufweisen würden, deutlich höher als in der Gesamtpopulation. Dies war tatsächlich der Fall (cf. Neumann/Euler 2009, 2010), was die Normierung von KiSS zusätzlich erschwerte und auch die Neunormierung immer noch erschwert.

Da bei zwei-/mehrsprachigen Kindern sehr schwache bzw. fehlende Deutschkenntnisse nicht unbedingt auf eine Sprachentwicklungsstörung hinweisen, sondern oft auf eine eingeschränkte Quantität bzw. Qualität des Deutschinputs, wurden separate KiSS-Normwerte für den klinischen Abklärungsbedarf für zwei-/mehrsprachige (ZM) und monolingual deutschsprachige Kinder (MO) entwickelt. Für ZM wurden nur wenige Cut-off-Werte definiert. Diese basieren auf Hinweisen aus KiSS-Fragebögen (z. B. auf den fehlenden Fortschritt im Deutscherwerb), konsequenten Aussprache Fehlern und Unfähigkeit, Sätze (auch teilweise) nachzusprechen. Auch Kunstwörter zum Nachsprechen wurden ursprünglich für die zuverlässigere Unterscheidung zwischen sprachentwicklungsgestörten und nicht gestörten Kindern in KiSS integriert, aber aufgrund ihrer geringen Zuverlässigkeit aus den Cut-off-Kriterien für ZM wieder ausgeschlossen. Das Alter des Deutscherwerbs als einer der wichtigsten Einflussfaktoren auf den Sprachstand in der Zweitsprache (cf. Grimm/Schulz 2012; Meisel 2009; Schulz/Tracy/Wenzel 2008) wird zwar in den KiSS-Fragebögen erfasst, aber in den Cut-off-Kriterien nur marginal berücksichtigt. Bei MO wurden zusätzlich mehrere Cut-off-Kriterien für den klinischen Abklärungsbedarf im Sinne von Punktzahlen in KiSS-Untertests („Wortschatz“, „Grammatik“, „Sprachverständnis“, „Nachsprechen“) definiert.

Für den sprachpädagogischen Förderbedarf gelten dagegen sowohl für ZM als auch für MO gleiche Auffälligkeitskriterien in KiSS. Dies liegt daran, dass KiSS ursprünglich als vorgeschobener Teil der Schuleingangsuntersuchung konzipiert war und der damalige hessische Schuleingangstest *Screening des Entwicklungsstandes bei Einschulungsuntersuchungen* (S-ENS; Döpfner et al. 2005) keine separaten Auffälligkeitskriterien für ZM anbot. Die mehrfach geäußerten Forderungen nach speziellen Normwerten für ZM (cf. Gogolin 2002; Jeuk 2009) wurden zwar von den KiSS-Autoren und -Autorinnen diskutiert, aber im Vordergrund bei der KiSS-Entwicklung standen (a) die Anwendbarkeit von KiSS im Sinne von einfachen, zeiteffizienten Auswertungskriterien, (b) die Vorbereitung der Kinder auf den Schulunterricht, in dem ZM denselben Anforderungen genügen müssen wie MO. Daher wurden doch einheitliche Normwerte für alle Kinder entwickelt.

Bei der Suche nach optimalen KiSS-Güteindizes musste ein Kompromiss gefunden werden, der ein Mindestmaß an Trennschärfe und Reliabilität für alle Kinder unabhängig von ihrem sprachlichen Hintergrund bietet, trotz großer Unterschiede im Schwierigkeitsindex. Dabei wurden KiSS-Güteindizes für ZM optimiert, da Auffälligkeiten im Deutscherwerb eher in dieser Untergruppe erwartet wurden als unter MO (beispielsweise wurden nicht Kunstwörter mit den höchsten Güteindizes in KiSS aufgenommen, sondern solche, die ZM am wenigsten benachteiligten (cf. Neumann/Euler 2009)). Um unter sprachlich schwachen Kindern besser zwischen

sprachauffälligen und -unauffälligen differenzieren zu können, wurden bei den sprachlich starken, zu denen überwiegend MO gehören, leichte Deckeneffekte im Schwierigkeitsindex zugelassen.

Bereits 2007 wurde KiSS im Rahmen eines groß angelegten hessischen Sprachstandserfassungsprogramms (HSP) für vier- bis viereinhalbjährige Kinder eingeführt (cf. Weiland et al. 2019). Ursprünglich wurde KiSS für die Durchführung in allen hessischen Kitas vorgesehen (cf. Hanack 2019). Letztendlich führte aber Hessen als einziges Bundesland von denen, die ihre eigenen Sprachscreenings für das Kitaalter entwickeln ließen, keine Teilnahmepflicht ein – trotz Empfehlung der Beauftragten der Bundesregierung für Migration, Flüchtlinge und Integration (cf. BBMFI 2019) und des Nationalen Integrationsplans der Bundesregierung (cf. Presse- und Informationsamt der Bundesregierung 2007).

Es ist dabei unbekannt, wie zuverlässig sich in Validierungsstudien errechnete psychometrische Kennwerte von KiSS auf andere Stichproben übertragen lassen. Wesentliche Unterschiede zwischen Testkennwerten in Validierungsstudien und bei der späteren Testdurchführung können auf mehrere Problemfelder hinweisen, von einer fehlerhaften Validierung (z. B. wegen veralteter oder wenig aussagekräftiger Referenztests) bis hin zur nicht standardisierten Testanwendung im HSP (z. B. Nicht-Einhaltung der Einschluss- bzw. Auswertungskriterien). Im Fall von KiSS kann u. a. die nicht obligatorische Teilnahme am HSP dazu führen, dass die HSP-Stichprobe für die Gesamtpopulation vier- bis viereinhalbjähriger Kinder in Hessen nicht repräsentativ ist, mit negativen Auswirkungen auf die Aussagekraft dieses Screenings.

Veröffentlichte Daten zum HSP deuten tatsächlich auf mögliche Problemfelder bei der Implementierung von KiSS hin. Dazu zählen vor allem sehr hohe Auffälligkeitsraten. Obwohl die Cut-off-Werte von KiSS so festgelegt wurden, dass etwa 10% der Kinder klinisch abklärungsbedürftig sein sollten, plus maximal 5% für falsch Positive (cf. Neumann/Euler 2010), demonstrierte die Übertragung der Cut-off-Werte auf Kinder im HSP weitaus höhere Auffälligkeitsraten. In den Jahren 2010–2017 erwiesen sich 19% der MO und 21% der ZM als klinisch abklärungsbedürftig (cf. Weiland et al. 2019). Dies übertraf sogar den Anteil der klinisch Abklärungsbedürftigen in der KiSS.2-Validierungsstudie (16% aller Kinder; cf. Neumann/Euler 2010), obwohl beim Festlegen entsprechender Cut-off-Werte ein Rekrutierungs- bzw. Selektionsbias einkalkuliert wurde: Eltern sprachauffälliger Kinder unterschreiben Einverständniserklärungen für Studienteilnahmen häufiger als Eltern sprachunauffälliger Kinder. Es ist daher anzunehmen, dass es aufgrund der freiwilligen Teilnahme am HSP zu demselben Bias in der Rekrutierung kam wie in der Validierungsstudie davor.

Die aktuelle retrospektive Studie hatte zum Ziel, Änderungen in einigen psychometrischen Kennwerten von KiSS.1 und KiSS.2 während und nach Validierungsstudien zu erfassen. Allerdings lassen sich retrospektiv nur die Testmerkmale prüfen, für die keine externen Referenztests notwendig sind, da diese nach dem Abschluss der Validierungsstudien nicht mehr durchgeführt wurden und auch in den Validierungsstudien immer wieder ausgetauscht wurden. Daher wurden in der aktuellen Arbeit nur die folgenden psychometrischen Kennwerte untersucht: Schwierigkeitsindex, Eigentrennschärfe und interne Konsistenz (Reliabilität). Zudem wurde auf der Grundlage der Daten der Schuleingangsuntersuchung exemplarisch geprüft, ob indirekte Hinweise auf einen Selektionsbias bei der Durchführung der KiSS-Studien bestehen.

2 Methoden

2.1 Testverfahren

Das „Kindersprachscreening“ (sowohl KiSS.1 als auch KiSS.2) enthält folgende Bestandteile:

- Test- bzw. Wimmelbild,
- „Screeningbogen“: der eigentliche Sprachtest,
- „Kitabogen“: Fragebogen für Kita-Erzieher und -Erzieherinnen,
- „Elternbogen“: Fragebogen für Eltern bzw. Sorgeberechtigte in mehreren Sprachen,
- „Kinderbogen“: Formular für eine komprimierte Zusammenfassung der KiSS-Ergebnisse (diese Formulare werden an hessische Gesundheitsämter weitergeleitet),
- Elterninformation zur Teilnahme am HSP bzw. KiSS-Validierungsstudien in mehreren Sprachen,
- Einverständniserklärung der Eltern bzw. Sorgeberechtigten in mehreren Sprachen,
- „Elternbrief“: Formular für die Erfassung der Sprachtestergebnisse und Empfehlungen.

KiSS.1 und KiSS.2 unterscheiden sich in der Item-Konstellation (siehe Elektron. Supplement 1) und im Umfang der Fragebögen für Eltern und Kita-Erzieher und -Erzieherinnen. Der wesentliche Unterschied besteht darin, dass in KiSS.2 Nachsprechaufgaben zur Prüfung des phonologischen Kurzzeitgedächtnisses implementiert wurden. Andere Untertests wurden dagegen um einige Items gekürzt.

Mit einer Durchführungsdauer von 12–15 Minuten können KiSS.1 und KiSS.2 als zeitökonomische Verfahren bewertet werden. Der Zeitaufwand für das Ausfüllen der Eltern- und Kitafragebögen beträgt je Fragebogen ca. fünf Minuten, der für die KiSS-Auswertung ca. zwei Minuten.

2.2 Testdurchführung

Nahezu alle Sprachtestungen in den KiSS-Validierungsstudien und alle im HSP wurden in Kitas durchgeführt. Als Tester und Testerinnen wurden in den Validierungsstudien geschulte Medizin-Doktoranden und -Doktorandinnen, Praktikanten und Praktikantinnen (Linguistikstudenten und -studentinnen) und Honorarkräfte (Medizin- und Linguistikstudenten und -studentinnen) eingesetzt, während im HSP geschulte Kita-Erzieher und -Erzieherinnen die Testungen durchführten. Abgebrochene Testungen wurden sowohl in KiSS-Validierungsstudien als auch im HSP soweit wie möglich durch Nachtestungen der Kinder vervollständigt, weil KiSS über keine Abbruchkriterien verfügt.

Es wurde Wert darauf gelegt, inkorrekte und fragwürdige Antworten zu dokumentieren, damit bei Bedarf Korrekturen in der Ergebnisdokumentation vorgenommen werden können. Alle Testbatterien wurden kontrolliert: in Validierungsstudien durch wissenschaftliche Mitarbeiter und Mitarbeiterinnen (klinische Linguisten und Linguistinnen), im HSP durch Sprachexperten und Sprachexpertinnen hessischer Gesundheitsämter.

2.3 Stichproben

Sowohl in den KiSS-Validierungsstudien als auch im HSP wurden vier- bis viereinhalbjährige Kinder ohne weitere Ein- oder Ausschlusskriterien mit KiSS.1 oder KiSS.2 getestet. Für alle Kinder mussten allerdings Einverständniserklärungen der Eltern bzw. Sorgeberechtigten vorliegen und auch die Kitas mit der Testdurchführung einverstanden sein.

Im Rahmen der aktuellen Studie wurden mehrere Stichproben retrospektiv ausgewertet. Aus den KiSS-Validierungsstudien lagen fünf Stichproben vor (zwei für KiSS.1, drei für KiSS.2), aus dem HSP nur KiSS.1-Daten, i. e. eine Stichprobe. Wichtigste Stichprobencharakteristika sind Tabelle 1 zu entnehmen.

KiSS.1-Stichproben			
	KiSS.1-Validierungsstudie	KiSS.XL-Validierungsstudie	Sprachstandserfassungsprogramm
Zeitraum	2007–2008	2008–2009	2008–2010
Alter: Mittelwert/Median (Monate)	50,6/51,0	50,5/51,0	unbekannt
N	222	181	6144
Männlich	122 (55,0%)	91 (50,3%)	3116 (50,7%)
Weiblich	100 (45,0%)	90 (49,7%)	3028 (49,3%)
MO	143 (64,4%)	94 (51,9%)	4280 (69,7%)
ZM	79 (35,6%)	87 (48,1%)	1864 (30,3%)
KiSS.2-Stichproben			
	KiSS.2-Validierungsstudie	BASS-Validierungsstudie	KiSS.3-Validierungsstudie
Zeitraum	2009–2010	2011–2012	2017–2019
Alter: Mittelwert/Median (Monate)	50,6/50,0	50,4/50,0	50,5/51,0
N	311	631	885
Männlich	180 (57,9%)	336 (53,2%)	474 (53,6%)
Weiblich	131 (42,1%)	295 (46,8%)	411 (46,4%)
MO	131 (42,1%)	279 (44,2%)	357 (40,3%)
ZM	180 (57,9%)	352 (55,8%)	528 (59,7%)

Tabelle 1: Stichprobenmerkmale

Anmerkung: BASS = „Bochum-Aachener Stotterscreening“ (Neumann/Euler/Schneider 2014), MO = monolingual deutschsprachige Kinder, ZM = zwei-/mehrsprachige Kinder.

Zudem konnten für einige Berechnungen Daten von Tomasik et al. (2020) – KiSS.2-Ergebnisse aus dem HSP – herangezogen werden. Insgesamt wurden von Tomasik et al. (2020) 43.556 Datensätze (vier- bis viereinhalbjährige Kinder) ausgewertet. Die Testungen fanden 2010–2017 statt. Es liegen keine Rohwerte zu deskriptiven Statistiken vor, aber 68% der Kinder waren MO und 51% waren männlich. Das Durchschnittsalter betrug $M = 4,30$ Jahre, i. e. ca. 52 Monate.

Angaben zu Familiensprachen wurden in allen Stichproben Eltern- und Kitafragebögen entnommen, die zu KiSS.1 und KiSS.2 gehören. Auf diese Weise wurden alle Kinder in MO und

ZM eingeteilt. Zwei- bzw. mehrsprachige Kinder wurden sowohl in prospektiven KiSS-Validierungsstudien als auch in der aktuellen retrospektiven Auswertung unabhängig vom Alter des Deutscherwerbs als ZM klassifiziert, dementsprechend gelten KiSS-Normwerte für ZM für alle zwei-/mehrsprachigen Kinder (auch für solche, die Deutsch seit Geburt erwerben). Angaben zum Alter des Deutscherwerbs aus dem HSP lagen dem Autorenteam nicht vor, entsprechende Angaben aus KiSS.2-Validierungsstudien wurden dagegen weiter unten für Berechnungen zu Güteindizes von KiSS.2 bei Untergruppen der ZM herangezogen.

Aufgrund möglicher Unterschiede in den KiSS-Güteindizes von simultanen und sukzessiven ZM (zur Terminologie cf. Genesee/Paradis/Crago 2004; Grimm/Schulz 2014; Paradis/Genesee/Crago 2011) wurden zusätzliche Berechnungen für diese Untergruppen durchgeführt. Als simultane ZM (SiZM) wurden Kinder klassifiziert, die in den ersten zwei Lebensjahren mind. zwei Sprachen erwarben (einschl. Deutsch), als sukzessive ZM (SuZM) solche, die erst ab dem dritten Lebensjahr mit dem Erwerb der deutschen Sprache begonnen haben. Einheitlichkeitshalber wurde für diese Berechnungen nur die aktuellste und größte Stichprobe herangezogen, in der KiSS.2 zum Einsatz kam (KiSS.3-Stichprobe, siehe Tabelle 1 oben). SiZM ($n = 216$) unterschieden sich nicht von SuZM ($n = 227$) im Alter in Monaten laut Mann-Whitney U-Test ($M = 50,4$ vs. $50,3$ Monate, $p > 0,05$). Von weiteren 85 Kindern lagen keine Angaben zum Alter des Deutscherwerbs vor (16,1% der ZM in der KiSS.3-Stichprobe).

Für die exemplarische Prüfung der Repräsentativität der KiSS-Stichproben wurden 151 Kinder aus der KiSS.2-Validierungsstudie (siehe Tabelle oben) mit 216 Kindern, die mit KiSS.2 nicht getestet worden waren, verglichen. Beide Untergruppen wurden zeitgleich in den Jahren 2010–2011 in Gesundheitsämtern mit dem Schuleingangstest S-ENS (cf. Döpfner et al. 2005), einschl. Zusatzaufgaben zur Sprache (S-ENS-Add-on, cf. Neumann/Zaretsky/Euler 2011) getestet. Der Zeitabstand zwischen KiSS.2 und S-ENS lag durchschnittlich bei $M = 17,6 \pm 7,5$ Monaten. Das Alter der Kinder mit und ohne KiSS.2-Teilnahme betrug zum Zeitpunkt der Schuleingangsuntersuchung in beiden Fällen $M = 67,8$ Monate (Altersspanne 5;0–6;11 Jahre) ohne signifikante Unterschiede laut Mann-Whitney U-Test ($p > 0,05$). Die Anteile der ZM waren in beiden Gruppen vergleichbar laut Chi-Quadrat-Test ($p > 0,05$): 53/151 (35,1%) unter KiSS-Kindern vs. 82/216 (38,0%) unter anderen Kindern.

2.4 Statistische Analyse

Sowohl für KiSS.1 als auch für KiSS.2 wurden folgende psychometrische Kennwerte geprüft:

- 1) Schwierigkeitsindex: Anteil der richtigen Antworten, geteilt durch 100,
- 2) Eigentrennschärfe (part-whole-korrigiert): punktbiseriale Korrelationen,
- 3) Interne Konsistenz (Reliabilität): Cronbachs Alpha.

Zuerst wurden deskriptive Statistiken zum Schwierigkeitsindex berechnet: (a) als Roh- und Prozentwerte der richtigen Antworten für einzelne KiSS-Items, (b) als Schwierigkeitsindex-Mittelwerte für KiSS-Untertests und den Gesamttest sowie (c) als Häufigkeit der mittleren Werte (Schwierigkeitsindex 45–55%) und extremen Werte (Schwierigkeitsindex 0–10% und 90–100%). Diese Berechnungen wurden getrennt für jede der KiSS-Stichproben (siehe Tabelle 1 oben) durchgeführt.

Die Eigentrennschärfe wurde als punktbiseriale Korrelationen zwischen einzelnen KiSS-Items und KiSS-Untertestscores bzw. dem -Gesamtscore bestimmt. Die interne Konsistenz wurde anhand von Cronbachs Alpha für einzelne KiSS-Untertests und für den Gesamttest berechnet.

Im nächsten Schritt wurden nicht-parametrische Unterschiedstests durchgeführt, um Fluktuationen in den Ausprägungen der o. g. Kennwerte zu quantifizieren. Sowohl für KiSS.1 als auch für KiSS.2 wurden Unterschiede im Schwierigkeitsindex und in der Eigentrennschärfe zuerst anhand von Kruskal-Wallis H-Tests geprüft (für jeweils drei KiSS.1- bzw. KiSS.2-Stichproben) und dann anhand von Mann-Whitney U-Tests für paarweise Vergleiche zwischen einzelnen Stichproben. Der Vergleich der Cronbachs Alpha-Werte wurde mit dem Online-Rechner „Cocron“ durchgeführt (cf. Diedenhofen/Musch 2016). Auch hier wurden zuerst Unterschiedsanalysen für alle verfügbaren Stichproben zusammen (getrennt für KiSS.1 und KiSS.2) und danach paarweise Vergleiche vorgenommen. Als Fälle dienten in den Unterschiedstests nicht getestete Kinder, sondern einzelne KiSS.1- bzw. KiSS.2-Items.

Da eigene HSP-Daten zu KiSS.2 nicht vorlagen, wurden solche aus dem Elektronischen Supplement 1 des Artikels von Tomasik et al. (2020) herangezogen. KiSS.2-Kennwerte aus Validierungsstudien wurden im Folgenden mit denen aus dem HSP verglichen. Für Angaben zum Schwierigkeitsindex und zur Eigentrennschärfe erfolgte dieser Vergleich anhand von Mann-Whitney U-Tests. Zudem wurden Angaben zur internen Konsistenz von KiSS.2 (Cronbachs Alpha) aus HSP und KiSS-Validierungsstudien nach dem Cocron-Berechnungsverfahren gegenübergestellt. Wegen der Nicht-Unterscheidung zwischen MO und ZM in der Arbeit von Tomasik et al. (2020) waren ihre Ergebnisse nicht direkt mit denen aus der KiSS.2-Validierungsstudie vergleichbar. Um trotzdem einen Vergleich zu ermöglichen, wurde in der aktuellen Studie das Verhältnis der MO und ZM durch Gewichtung angepasst. Da dieses Verhältnis in den HSP-Daten von Tomasik et al. (2020) bekannt war (MO:ZM 3:1), wurden die Ergebnisse der MO aus der KiSS.2-Validierungsstudie bei Mittelungen dreifach gewichtet.

Um den Zusammenhang zwischen dem Schwierigkeitsindex und der Eigentrennschärfe von KiSS zu verdeutlichen, wurden Spearman-Korrelationen zwischen diesen zwei Güteindizes berechnet, getrennt für MO und ZM. Dafür wurden Mittelwerte dieser Güteindizes aus den Tabellen E2 im Elektronischen Supplement 1 und E5 im Elektronischen Supplement 2 herangezogen.

Um mögliche Einflussfaktoren auf die Eigentrennschärfe zu bestimmen, wurden KiSS.1- und KiSS.2-Daten zusammen betrachtet. Dafür wurde eine lineare Regression mit der Eigentrennschärfe „Item vs. Gesamtscore“ als abhängige Variable gerechnet (Methode „Einschluss“). Als unabhängige Variablen wurden (a) Schwierigkeitsindex, (b) Lingualität (MO vs. ZM), (c) einzelne KiSS-Stichproben (siehe Tabelle 1 oben) und (d) die dichotome Unterscheidung KiSS.1- vs. KiSS.2-Stichproben herangezogen.

Vor dem Vergleich der KiSS.2-Güteindizes in den Untergruppen SiZM und SuZM wurde anhand einer Chi-Quadrat-Analyse geprüft, ob zwischen SiZM und SuZM signifikante Unterschiede im Anteil der Sprachauffälligen nach KiSS.2-Kriterien vorliegen. Nur bei einem statistisch signifikanten Ergebnis konnte davon ausgegangen werden, dass KiSS.2-Güteindizes sich wesentlich unterscheiden können. Es wurde angenommen, dass SuZM besonders schwache

KiSS.2-Ergebnisse im Sinne eines hohen Anteils an sprachauffälligen Kindern erzielen würden (cf. Grimm/Schulz 2012, 2014; Paradis 2010). Unterschiede zwischen SiZM und SuZM in den Schwierigkeitsgraden und in der Trennschärfe von KiSS.2 wurden anhand von Mann-Whitney U-Tests untersucht, die in der Reliabilität anhand von Cocron-Verfahren (siehe oben).

Als Letztes wurde geprüft, ob Hinweise auf einen Rekrutierungsbias, der für Fluktuationen psychometrischer Kennwerte mitverantwortlich sein könnte, in den KiSS-Validierungsstudien vorliegen. Dafür wurde exemplarisch eine Untergruppe der Kinder aus der KiSS.2-Validierungsstichprobe ausgewertet, die im Rahmen der Schuleingangsuntersuchung mit dem Einschulungstest S-ENS-Add-on nachgetestet wurde. Die entsprechenden Daten (Anteile sprachpädagogisch förderbedürftiger und klinisch abklärungsbedürftiger Kinder laut Sprachexperten- und Sprachexpertinnen-Urteil) wurden mit den Daten von Kindern verglichen, die an der KiSS.2-Validierungsstudie nicht teilgenommen hatten. Durch Kreuztabellierung und Chi-Quadrat-Berechnung wurde geprüft, ob die Anteile sprachauffälliger Kinder in beiden Untergruppen vergleichbar waren.

Für alle Mann-Whitney U-Tests wurde die Effektstärke „probability of superiority index“ (PS-Index, \hat{p} ; cf. Grissom/Kim 2012) bestimmt. Effektstärken nah an 0,5 gelten als niedrig, die nah an 0,0 als hoch.

Falls möglich, wurden die oben beschriebenen Berechnungen getrennt für MO und ZM durchgeführt, da sich ihre Schwierigkeitsindex-Mittelwerte in KiSS signifikant unterscheiden (cf. Zaretsky/Hey 2021; Zaretsky/Lange/Hey 2022). Als statistisch signifikant galten p-Werte unter 0,05. Bei multiplen Testungen kam eine Bonferroni-Holm-Korrektur der p-Werte zum Einsatz.

3 Ergebnisse

Zusammenfassende deskriptive Statistiken zum Schwierigkeitsindex, einschl. Visualisierung, sind dem Elektronischen Supplement 1 zu entnehmen (Tabelle E1, E2, Abbildung E1), die zur Eigentrennschärfe dem Elektronischen Supplement 2 (Tabelle E4, E5, Abbildung E2, E3), Angaben zur internen Konsistenz sind im Elektronischen Supplement 3 visualisiert (Abbildung E4).

Änderungen im Schwierigkeitsindex blieben laut Unterschiedstests in allen geprüften KiSS.1- und KiSS.2-Stichproben nicht signifikant (siehe Tabelle 2). Fluktuationen in der Eigentrennschärfe wurden nur bei MO festgestellt.

KiSS.1	Schwierigkeitsindex		Eigentrennschärfe Item vs. Untertestscore		Eigentrennschärfe Item vs. Gesamtscore	
	MO	ZM	MO	ZM	MO	ZM
Kruskal-Wallis H-Tests: $\chi^2_{(2)}$ Mann-Whitney U-Tests (Z/ \hat{p}):						
KiSS.1-V vs. KiSS.XL-V	-0,29/0,48	-0,71/0,46	-0,77/0,45	-0,97/0,44	-0,64/0,46	-1,18/0,43
KiSS.1-V vs. HSP	-0,72/0,46	-0,80/0,45	-1,74/0,39	-2,26/0,36	-3,13*/0,31	-1,47/0,41
KiSS.XL-V vs. HSP	-0,25/0,48	-1,38/0,41	-2,84*/0,32	-0,91/0,44	-3,12*/0,31	-0,14/0,49

KiSS.2	Schwierigkeitsindex		Eigentrennschärfe Item vs. Untertestscore		Eigentrennschärfe Item vs. Gesamtscore	
	MO	ZM	MO	ZM	MO	ZM
Kruskal-Wallis H-Tests: $\chi^2_{(2)}$ Mann-Whitney U-Tests (Z/ \hat{p}):						
KiSS.2-V vs. BASS-V	-0,03/0,50	-0,09/0,49	-2,02/0,38	-0,63/0,46	-2,84*/0,33	-0,97/0,44
KiSS.2-V vs. KiSS.3-V	-0,29/0,48	-1,60/0,40	-3,54**/0,28	-0,48/0,47	-3,82**/0,27	-0,32/0,48
BASS-V vs. KiSS.3-V	-0,55/0,46	-0,13/0,41	-1,85/0,39	-0,93/0,44	-1,25/0,42	-0,55/0,47

Tabelle 2: Unterschiede in Schwierigkeitsindex und Eigentrennschärfe zwischen einzelnen KiSS-Stichproben: Kruskal-Wallis H- und Mann-Whitney U-Tests

Anmerkungen: MO = monolingual deutschsprachige Kinder, ZM = zwei-/mehrsprachige Kinder; KiSS.1-V = KiSS.1-Validierungsstudie, KiSS.XL-V = KiSS.XL-Validierungsstudie, HSP = hessisches Sprachstandserfassungsprogramm, KiSS.2-V = KiSS.2-Validierungsstudie, BASS-V = BASS-Validierungsstudie (BASS = „Bochum-Aachener Stotterscreening“), KiSS.3-V = KiSS.3-Validierungsstudie; \hat{p} = probability of superiority index.

* p < 0,05, ** p < 0,01 (Bonferroni-Holm-Korrektur)

Unterschiedstests nach dem Cocron-Verfahren zeigten signifikante Fluktuationen in den Cronbachs Alpha-Werten nur bei MO (siehe Tabelle 3).

KiSS-Untertests	MO	ZM
KiSS.1-Untertest „Sprachverständnis“	1,21	4,47
KiSS.1-Untertest „Wortschatz“	9,89*	1,20
KiSS.1-Untertest „Aussprache“	10,23*	6,44
KiSS.1-Untertest „Grammatik“	5,98	4,23
KiSS.1: alle Items	15,10*	3,87
KiSS.2-Untertest „Sprachverständnis“	0,82	3,84
KiSS.2-Untertest „Wortschatz“	5,22	1,90
KiSS.2-Untertest „Aussprache“	6,07	0,64
KiSS.2-Untertest „Grammatik“	9,31*	1,25
KiSS.2-Untertest „Nachsprechen“	3,47	6,62
KiSS.2: alle Items	9,91*	3,53

Tabelle 3: Unterschiede in der internen Konsistenz (Cronbachs Alpha) zwischen je drei KiSS.1- und KiSS.2-Stichproben: Cocron-Verfahren (Chi-Quadrat, df = 2)

Anmerkungen: MO = monolingual deutschsprachige Kinder, ZM = zwei-/mehrsprachige Kinder.

* $p < 0,05$ (Bonferroni-Holm-Korrektur)

Angaben zur internen Konsistenz sind Tabelle 4 zu entnehmen. Paarweise Vergleiche der internen Konsistenz zeigten bei KiSS.1 nur Unterschiede zwischen KiSS.1- (MO: $\chi^2_{(1)} = 18,57$, $p < 0,001$) bzw. KiSS.XL-Stichproben (MO: $\chi^2_{(1)} = 12,35$, $p < 0,001$) einerseits und dem HSP andererseits, aber nicht innerhalb der Validierungsstichproben (gerechnet als Item vs. Gesamtscore). Unter den einzelnen KiSS.2-Validierungsstichproben bestanden signifikante Unterschiede nur zwischen KiSS.2- und KiSS.3-Stichproben (MO: $\chi^2_{(1)} = 8,02$, $p = 0,015$). Bei ZM waren Fluktuationen in der internen Konsistenz nicht signifikant.

KiSS.1-Stichproben	KiSS.1-V		KiSS.XL-V		HSP	
	MO	ZM	MO	ZM	MO	ZM
KiSS.1-Untertest „Sprachverständnis“	0,55	0,71	0,48	0,52	0,44	0,62
KiSS.1-Untertest „Wortschatz“	0,59	0,82	0,61	0,79	0,74	0,83
KiSS.1-Untertest „Aussprache“	0,83	0,72	0,73	0,74	0,83	0,82
KiSS.1-Untertest „Grammatik“	0,74	0,83	0,76	0,88	0,82	0,88
KiSS.1: alle Items ^a	0,82	0,89	0,82	0,92	0,89	0,92
KiSS.2-Stichproben	KiSS.2-V		BASS-V		KiSS.3-V	
	MO	ZM	MO	ZM	MO	ZM
KiSS.2-Untertest „Sprachverständnis“	0,31	0,43	0,38	0,57	0,42	0,57
KiSS.2-Untertest „Wortschatz“	0,72	0,79	0,68	0,79	0,60	0,82
KiSS.2-Untertest „Aussprache“	0,84	0,76	0,78	0,78	0,78	0,76
KiSS.2-Untertest „Grammatik“	0,82	0,87	0,75	0,86	0,72	0,85
KiSS.2-Untertest „Nachsprechen“	0,53	0,59	0,64	0,67	0,54	0,53
KiSS.2: alle Items ^a	0,89	0,90	0,85	0,92	0,83	0,91

Tabelle 4: Interne Konsistenz des „Kindersprachscreenings“ (KiSS.1 und KiSS.2): Cronbachs Alpha

Anmerkungen: MO = monolingual deutschsprachige Kinder, ZM = zwei-/mehrsprachige Kinder; KiSS.1-V = KiSS.1-Validierungsstudie, KiSS.XL-V = KiSS.XL-Validierungsstudie, HSP = hessisches Sprachstandserfassungsprogramm, KiSS.2-V = KiSS.2-Validierungsstudie, BASS-V = BASS-Validierungsstudie (BASS = „Bochum-Aachener Stotterscreening“), KiSS.3-V = KiSS.3-Validierungsstudie.

^a Bei der Berechnung der internen Konsistenz innerhalb der KiSS-Untertests wurden Items zur Spontansprache bzw. Sprachproduktion nicht berücksichtigt, da diese keinem Untertest angehören. Bei der Berechnung der internen Konsistenz von allen KiSS-Items waren aber diese zwei Items mitenthalten.

Im Folgenden wurden Testmerkmale von KiSS.2 anhand der KiSS.2-Validierungsstudie und HSP-Daten von Tomasik et al. (2020) verglichen. Deskriptive Angaben zum Schwierigkeitsindex und zur Eigentrennschärfe von KiSS.2 befinden sich im Elektronischen Supplement 1 (Tab. E3) und Elektronischen Supplement 2 (Tab. E6). Unterschiede im Schwierigkeitsindex waren nicht statistisch signifikant ($Z = -0,45$, $p = 0,651$, $\hat{p} = 0,45$). Die Eigentrennschärfe sowohl im Sinne der Korrelationen Item vs. Untertestscores ($Z = -5,43$, $p < 0,001$, $\hat{p} = 0,16$) als auch im Sinne der Korrelationen Item vs. Gesamtscore ($Z = -3,80$, $p < 0,001$, $\hat{p} = 0,24$) fiel in der KiSS.2-Validierungsstudie im Vergleich mit dem HSP signifikant niedriger aus. Laut Cocron-Berechnungsverfahren lag das Cronbachs Alpha in der KiSS.2-Validierungsstudie (0,89) signifikant niedriger als im HSP (0,90): $\chi^2_{(1)} = 16,60$, $p < 0,001$.

Die Spearman-Korrelationen zwischen dem Schwierigkeitsindex und der Eigentrennschärfe von KiSS-Items wurde für alle KiSS-Stichproben (siehe Tabelle 1 oben) außer die von Tomasik et al. (2020) zusammen berechnet. Bei MO war der Zusammenhang nicht signifikant ($\rho = 0,070$,

$p = 0,700$), bei ZM korrelierte ein höherer Anteil der richtigen Antworten mit der Verschlechterung der Eigentrennschärfe ($\rho = -0,456$, $p = 0,008$).

Aus der linearen Regression (siehe Tabelle 5) geht hervor, dass die höhere Eigentrennschärfe von KiSS in erster Linie mit niedrigerem Schwierigkeitsindex einherging ($p < 0,001$), zusätzlich bei ZM verglichen mit MO ($p = 0,008$) und abhängig von KiSS-Untertests ($p = 0,028$).

	Regressionskoeffizient Beta nicht standardisiert	Regressionskoeffizient Beta standardisiert	T
Schwierigkeitsindex	-0,003	-0,452	-10,30*
Lingualität (MO vs. ZM)	0,032	0,110	2,65*
Einzelne Stichproben	0,001	0,006	0,22
KiSS.1- vs. KiSS.2- Stichproben	-0,017	-0,058	-0,76
KiSS-Untertests	0,010	0,086	2,20*
Konstante	0,476	—	11,06*
Gesamtmodell (F)		44,04*	
Korrigiertes R^2		0,286	

Tabelle 5: Mögliche Einflussfaktoren auf die Eigentrennschärfe (Item vs. Gesamtscore) in allen KiSS.1- und KiSS.2-Stichproben^a: Lineare Regression, Methode „Einschluss“

Anmerkungen: MO = monolingual deutschsprachige Kinder, ZM = zwei-/mehrsprachige Kinder.

^a Ohne KiSS.2-Daten von Tomasik et al. (2020), da in diesen zwischen monolingual deutschsprachigen und zwei-/mehrsprachigen Kindern nicht unterschieden wird.

* $p < 0,05$

Beim Vergleich der SiZM und SuZM wurde als Erstes geprüft, ob unter SuZM sprachpädagogisch förderbedürftige Kinder signifikant häufiger vorkamen. Laut Chi-Quadrat-Test war es tatsächlich der Fall: Während in der SiZM-Gruppe 50,0% (108/216) nach KiSS.2-Auffälligkeitskriterien sprachpädagogisch förderbedürftig waren, lag dieser Anteil in der SuZM-Gruppe bei 93,8% (345/393); $\chi^2_{(1)} = 106,57$, $p < 0,001$ (cf. 17,9% bei MO, 64/357; 37,7% bei SiZM, die Deutsch seit Geburt erwerben, 52/138). Auch beim klinischen Abklärungsbedarf bestand ein höchst signifikanter Unterschied mit den Auffälligkeitsraten von 10,2% (22/216) bei SiZM und 28,6% (65/227) bei SuZM; $\chi^2_{(1)} = 23,87$, $p < 0,001$ (cf. 14,8% bei MO, 53/357; 10,1% bei SiZM, die Deutsch seit Geburt erwerben, 14/138). Daher konnten wesentliche Unterschiede in den KiSS.2-Güteindizes zwischen SiZM und SuZM vermutet werden. Im nächsten Schritt wurden KiSS.2-Güteindizes für SiZM und SuZM verglichen. Für SiZM waren KiSS.2-Aufgaben signifikant einfacher als für SuZM: $Z = -3,60$, $p < 0,001$, $\hat{p} = 0,28$ (deskriptive Angaben zum Schwierigkeitsindex: siehe Tabelle E7 im Elektronischen Supplement 4). Die Schwierigkeitsgrade der SuZM bewegten sich viel häufiger in extremen Bereichen und fast nie im mittleren. Unterschiede in den Eigentrennschärfen „Item vs. Untertestscore“ ($Z = -0,72$, $p = 0,474$, $\hat{p} = 0,45$) und „Item vs. Gesamtscore“ ($Z = -1,23$, $p = 0,218$, $\hat{p} = 0,42$) zwischen SiZM und SuZM waren zwar nicht signifikant laut Mann-Whitney U-Tests, aber entsprechende Korrelationskoeffizienten lagen numerisch höher eher bei SiZM als bei SuZM (siehe Tabelle E8 im Elektronischen Supplement 4). Cronbachs Alpha und damit die Reliabilität vom Gesamttest KiSS.2

lag in der SiZM-Gruppe bei $\alpha = 0,88$, in der SuZM-Gruppe bei $\alpha = 0,85$; der Unterschied zwischen diesen Alpha-Werten war laut Cocron-Verfahren nicht signifikant ($p > 0,05$).

Im letzten Schritt wurde durch Kreuztabellierung geprüft, ob Anteile sprachauffälliger Kinder unter ehemaligen Teilnehmern und Teilnehmerinnen bzw. Nicht-Teilnehmern und Nicht-Teilnehmerinnen der KiSS.2-Validierungsstudie in der Schuleingangsuntersuchung vergleichbar waren. Sprachpädagogischer Förderbedarf wurde bei 44/151 (29,1%) der KiSS-Kinder und 72/216 (33,3%) anderer Kinder festgestellt; der Unterschied war nicht signifikant ($\chi^2_{(1)} = 0,72$, $p = 0,395$). Der Anteil klinisch Abklärungsbedürftiger lag dagegen bei KiSS-Kindern deutlich höher als bei anderen Kindern: 22/151 (14,6%) vs. 17/216 (7,9%), $\chi^2_{(1)} = 4,20$, $p = 0,040$.

4 Diskussion

In der aktuellen Studie wurde untersucht, inwiefern die Eigentrennschärfe, der Schwierigkeitsindex und die interne Konsistenz der Sprachscreenings KiSS.1 und KiSS.2 in den Validierungsstudien und im HSP variierten. Beim Übergang von Validierungsstudien zum HSP kam es zu einer signifikanten Verbesserung der Eigentrennschärfe und internen Konsistenz bei MO. Diese Verbesserung ging mit der Erhöhung des Schwierigkeitsindex (Anteil der richtigen Antworten) einher. Bei ZM wurden dagegen nur minimale, nicht signifikante Fluktuationen der Testkennwerte festgestellt.

Um Änderungen in den Testkennwerten unter KiSS-Stichproben besser interpretieren zu können, soll zunächst auf den Zusammenhang zwischen dem Schwierigkeitsindex, der Trennschärfe und internen Konsistenz eingegangen werden (cf. zum Folgenden Bühner 2021). Mittlere Schwierigkeitsgrade sollten in Sprachtests gehäuft vorkommen, weil ein mathematischer Zusammenhang zwischen Schwierigkeitsindex und Trennschärfe besteht: Je extremer (sehr leicht oder sehr schwierig) der Schwierigkeitsindex ist, desto geringer wird die Trennschärfe (cf. Moosbrugger/Kelava 2012; Neumann/Euler 2009). Mittlere Schwierigkeitsgrade resultieren daher in besseren Trennschärfen und zeugen zudem von einer guten internen Konsistenz. Auch in den KiSS-Validierungsstudien wurden ursprünglich mittlere Schwierigkeitsgrade angestrebt. Letztendlich wurden aber bewusst eher leichte Items aufgenommen, um einerseits die Motivation der Kinder in der Testsituation zu stützen, i. e. die Compliance möglichst hoch zu halten, und andererseits eher im unteren Bereich der Sprachkompetenz besser differenzieren zu können (cf. Euler et al. 2010). ZM liegen mit Schwierigkeitsindex-Mittelwerten unter 0,60 oberhalb des optimalen (mittleren) Bereichs, bei MO ist dies mit Werten unter 0,80 in allen geprüften Stichproben noch stärker der Fall. Die Aufnahme von relativ leichten Items in KiSS erfolgte damit auf Kosten der Eigentrennschärfe, vor allem bei MO.

Dementsprechend war zu erwarten, dass die Erhöhung des Schwierigkeitsindex mit der Reduzierung der Trennschärfe und der internen Konsistenz einhergeht. Dies war aber in der aktuellen Studie nur bei ZM der Fall. Bei MO wurde kein statistisch signifikanter Zusammenhang zwischen dem Schwierigkeitsindex und der Eigentrennschärfe beobachtet.

Der fehlende Zusammenhang zwischen dem Schwierigkeitsindex und der Eigentrennschärfe von KiSS bei MO, der in einer Spearman-Korrelation demonstriert wurde, ist testkonstruktionsbedingt. Da KiSS eher für den Einsatz bei ZM optimiert war, wurden bei MO Deckeneffekte

auf Kosten der Trennschärfe zugelassen. Hohe Schwierigkeitsindex-Werte, i. e. ein hoher Prozentsatz der richtigen Antworten, waren sowohl mit besseren als auch mit schlechteren Eigentrennschärfen assoziiert. Fluktuationen der Schwierigkeitsgrade führten daher bei MO mal zur leichten Verbesserung, mal zur leichten Verschlechterung der Eigentrennschärfe. Bei ZM resultierte dagegen ein höherer Prozentsatz der richtigen Antworten in der Verschlechterung der Eigentrennschärfe, weil der Schwierigkeitsindex in solchen Fällen den optimalen Bereich häufiger verließ. Damit liegt kein Widerspruch in den Daten vor, sondern ein Artefakt aus der Testentwicklungsphase, der darin begründet war, dass die Testautoren und Testautorinnen versuchten, statt zwei separate KiSS-Versionen für MO und ZM eine gemeinsame Version zu entwickeln.

Die Tatsache, dass es bei MO zur (leichten) Verbesserung der Eigentrennschärfe von KiSS im HSP kam, muss also nicht zwingend mit der Erhöhung des Schwierigkeitsindex zusammenhängen, zumal diese Erhöhung bei durchschnittlich gerade mal 2% in der Gegenüberstellung KiSS.1-Stichprobe vs. HSP-Stichprobe und 1% in der Gegenüberstellung KiSS.XL-Stichprobe vs. HSP-Stichprobe lag. Sollte aber der (minimale, durch die o. g. Spearman-Korrelation nicht erkannte) Zusammenhang trotzdem bestehen, kann dieser folgendermaßen erklärt werden. Da die Verteilung des KiSS-Gesamtscores bei MO in allen KiSS-Validierungsstudien linksschief war (cf. z. B. Neumann/Euler 2009, 2010; Zaretsky/Hey 2021), wurden bei der KiSS-Entwicklung bewusst Items ausgewählt, die in der MO-Untergruppe trotz eines hohen Anteils an richtigen Antworten relativ trennscharf bleiben. Dabei muss berücksichtigt werden, dass der Zusammenhang zwischen dem Schwierigkeitsindex und der Eigentrennschärfe bei MO nicht ganz linear, sondern eher kurvilinear ist (vielleicht deswegen konnte die Spearman-Korrelation kein signifikantes Ergebnis demonstrieren). Aus der Abbildung E3 im Elektronischen Supplement 2 geht hervor, dass bei MO (a) der Schwierigkeitsindex um 0,40 mit einer niedrigen Eigentrennschärfe assoziiert ist und (b) ein klarer Rückgang der Trennschärfe erst ab dem Schwierigkeitsindex von ca. 0,90 erkennbar ist. In den Stichproben mit einem relativ hohen Schwierigkeitsindex (z. B. HSP) war der problematische Bereich mit den Schwierigkeitsindex-Werten um 0,40 nicht vorhanden und fast alle Untertestmittelwerte lagen trotzdem unterhalb des problematischen Bereichs 0,90+. Durch das Fehlen dieser Schwierigkeitsgrade, die die Eigentrennschärfe und folglich auch die interne Konsistenz reduzieren, fielen die Testkennwerte bei MO in entsprechenden Stichproben etwas besser aus. Es handelt sich allerdings um minimal signifikante Unterschiede, die die Qualitätsmerkmale von KiSS nur schwach beeinflussen.

Bei ZM waren dagegen keine signifikanten Änderungen der Testkennwerte von KiSS nachweisbar. Generell lagen die mittleren Schwierigkeitsindex-Werte der ZM in allen KiSS-Stichproben knapp über 0,50 und kamen damit den optimalen, anzustrebenden Werten nahe, die sich positiv auf die Eigentrennschärfe auswirken sollten. Auch der Anteil der Extremwerte, der – wenn hoch – in der Regel die Eigentrennschärfe reduziert, war bei ZM durchgehend niedriger als bei MO (siehe Tabelle E1 im Elektronischen Supplement 1). Daher fielen Eigentrennschärfen in der ZM-Untergruppe in allen Stichproben höher aus als die in der MO-Untergruppe (auf die Unterscheidung SiZM vs. SuZM wird weiter unten eingegangen).

Die meisten geprüften Einflussfaktoren auf die Eigentrennschärfe erwiesen sich in der linearen Regression als signifikant. Vor allem besteht auch in der aktuellen Studie der zu erwartende

negative Zusammenhang zwischen hohem Schwierigkeitsindex und niedriger Trennschärfe. Relativ einfache Aufgaben waren also mit einer niedrigen Trennschärfe assoziiert, relativ komplizierte mit einer hohen. Höhere Eigentrennschärfen wurden zudem eher bei ZM als bei MO festgestellt, was auf die eingeschränkte Varianz in den Antworten der MO zurückzuführen ist (siehe oben). Bei ZM lässt sich dagegen ein breites Spektrum der Schwierigkeitsgrade finden, darunter Werte im hohen und niedrigen Bereich, was sich auf die Stärke des Zusammenhangs positiv auswirkte. Auch je nach KiSS-Untertest variierte die Eigentrennschärfe beträchtlich. Einzelne KiSS-Untertests lieferten unterschiedliche Ergebnisse abhängig von ihrem Umfang, wobei eine kleine Item-Zahl eher in einer niedrigen Trennschärfe resultierte (weil in größeren Untertests mehr Spielraum für unterschiedliche Ausprägungen der Schwierigkeitsgrade besteht). Obwohl auch Schwierigkeitsgrade einzelner Untertests nicht gleichmäßig verteilt waren (siehe Tabelle E2), war dieser Faktor in der Regression durch die Variable „Schwierigkeitsindex“ abgedeckt, und nicht durch „KiSS-Untertests“, daher ist der letztgenannte Faktor eher mit der Item-Anzahl pro Untertest gleichzusetzen. Diese Einflussfaktoren waren gleichermaßen bedeutsam in allen geprüften Stichproben und in beiden KiSS-Versionen (KiSS.1 und KiSS.2). Die Tatsache, dass in der Regression nur ein Drittel der Varianz erklärt wurde, deutet darauf hin, dass weitere wichtige Einflussfaktoren existieren. Die Verschlechterung der Deutschkenntnisse sowohl bei MO als auch bei ZM in den letzten Jahren (cf. Zaretsky/Lange/Hey 2023) führte z. B. dazu, dass in der KiSS.3-Validierungsstichprobe der Schwierigkeitsindex von KiSS.2 absank und sich auch andere psychometrische Kennwerte entsprechend veränderten. Als möglicher Einflussfaktor kommt daher u. a. das Jahr der KiSS-Testung in Frage.

Hohe Eigentrennschärfen begünstigen auch eine hohe interne Konsistenz. Die Letztere lag, ohne Ausnahmen, in den Stichproben höher, in denen auch die Eigentrennschärfe höher war. Da die Trennschärfe bei ZM durchgehend höher ausfiel als bei MO, war es auch bei der internen Konsistenz der Fall. Damit erklären sich Fluktuationen der internen Konsistenz aus Änderungen in der Eigentrennschärfe, ohne dass weitere Einflussfaktoren herangezogen werden müssen.

Die numerische Erhöhung des Schwierigkeitsindex (Anteils der richtigen Antworten) im HSP, verglichen mit den KiSS-Validierungsstudien, ist höchstwahrscheinlich darauf zurückzuführen, dass im Ersteren die Altersgruppe von KiSS um einen Monat erweitert wurde. In sämtlichen KiSS-Validierungsstudien wurden Kinder im Alter von 4;0–4;5 Jahren getestet, im HSP dagegen, auf Empfehlung der KiSS-Autoren und KiSS-Autorinnen (cf. Neumann/Zaretsky/Euler 2011), mit 4;0–4;6 Jahren. Auch wenn Kinder im Alter von 4;6 Jahren im HSP stark unterrepräsentiert waren (cf. Tomasik et al. 2020), lieferten sie häufiger richtige Antworten als jüngere Kinder, was den mittleren Schwierigkeitsindex nach oben korrigierte und bei MO andere Testkennwerte mitbeeinflusste. Bei ZM ist dagegen weniger das chronologische Alter als die Dauer des Deutschkontakts für die Deutschkenntnisse ausschlaggebend (cf. Zaretsky/Lange 2023), daher erfolgten im HSP keine signifikanten Änderungen in den Güteindizes.

Die Tatsache, dass die angestrebte Auffälligkeitsrate – klinischer Abklärungsbedarf bei 10% der Kinder – im HSP weit übertroffen wurde, lässt sich durch mehrere Faktoren erklären. Erstens kam es in den letzten Jahren in Deutschland nach Angaben der Krankenkasse BARMER

zur Vergrößerung des Anteils therapiebedürftiger Kinder, gemessen an der Diagnose „Umschriebene Entwicklungsstörungen des Sprechens und der Sprache“, von 10% im Jahr 2010 auf 13% im Jahr 2021 (cf. Grobe/Szecsényi 2023). Zweitens wurde in allen KiSS-Validierungsstudien in den Nachsprechaufgaben, die eines der Auffälligkeitskriterien für den klinischen Abklärungsbedarf bei MO darstellen, streng zwischen Aussprachefehlern und reinen Nachsprechfehlern unterschieden (cf. Zaretsky/Hey 2021), was aber aus unbekanntem Gründen keinen Eingang ins KiSS-Handbuch fand (cf. Neumann/Zaretsky/Euler 2011). Damit kann nicht ausgeschlossen werden, dass der Anteil klinisch abklärungsbedürftiger Kinder im HSP unnötigerweise durch solche, die aufgrund vom Schetismus bzw. Sigmatismus fehlerhaft Kunstwörter nachsprachen, erhöht wurde. Dabei stellen Schetismus und Sigmatismus in KiSS kein Auffälligkeitskriterium für klinischen Abklärungsbedarf dar. Drittens kann die Stichprobe des HSP nicht als ganz repräsentativ betrachtet werden, weil nicht alle Kinder daran teilnehmen. Damit ist ein Rekrutierungsbias zu vermuten, mit einem erhöhten Anteil an sprachauffälligen Kindern.

Die Repräsentativität der HSP-Stichproben kann damit bezweifelt werden. Auch die eingeschränkte Repräsentativität der KiSS.2-Validierungsstichprobe wurde in der aktuellen Studie durch den Vergleich der Ergebnisse der Schuleingangsuntersuchung zwischen Kindern, die etwa anderthalb Jahre davor an KiSS.2 teilgenommen oder nicht teilgenommen hatten, bestätigt. Der Anteil klinisch abklärungsbedürftiger war unter KiSS-Kindern fast doppelt so hoch wie unter anderen Kindern, die an denselben Tagen zur Schuleingangsuntersuchung kamen. Damit bestand zumindest in dieser Validierungsstudie ein Rekrutierungsbias.

Die geprüften Testkennwerte von KiSS waren nicht in allen Fällen zufriedenstellend. Beispielsweise liegt die interne Konsistenz in einigen Untertests unter 0,7 und damit in einem zu niedrigen Bereich. Dagegen zeigten KiSS.1 und KiSS.2 als Gesamttests in allen Stichproben und Untergruppen (MO, ZM) eine interne Konsistenz oberhalb von 0,8 und damit in einem mittleren Bereich, teilweise mit exzellenten Werten oberhalb von 0,9 (cf. Bühner 2021). Auch die Trennschärfen waren teils optimierungsbedürftig, mit Durchschnittswerten fast ausschließlich unterhalb von 0,5 und teilweise sogar unterhalb von 0,4. Bei der Testentwicklung werden in der Regel Trennschärfen oberhalb von 0,5 angestrebt (cf. Döring/Bortz 2016). Die neue KiSS-Version, die im HSP noch nicht eingeführt wurde, nämlich KiSS.3 (cf. Zaretsky/Hey 2021), erreicht dagegen durchgehend die Trennschärfen von mind. 0,5.

Sowohl in der ersten (2007–2012) als auch in der zweiten (2017–2021) Entwicklungsphase von KiSS blieb das Verhältnis der SiZM und SuZM stabil bei ungefähr 1:1. Dies soll aber nicht darüber hinwegtäuschen, dass der Anteil der ZM, die Deutsch seit Geburt erwerben, rückläufig ist. Während in der ersten KiSS-Entwicklungsphase solche Kinder 38% aller ZM ausmachten (143/381), waren es in der zweiten Entwicklungsphase 31% (138/443). Damit stellt sich die Frage, inwiefern KiSS-Ergebnisse durch die Item-Konstellation, die eher für sprachlich stärkere Kinder konzipiert war, beim wachsenden Anteil der Kinder mit späterem Deutscherwerb verfälscht werden (können). Die aktuelle Studie zeigt, dass bei SuZM, trotz eines hohen Anteils an Extremwerten unter den Schwierigkeitsgraden, die Eigentrennschärfe und Reliabilität nur numerisch beeinträchtigt wurden, ohne dass die Grenze der statistischen Signifikanz erreicht wurde. Trotzdem wäre es empfehlenswert, vor allem angesichts einer steigenden Zahl der Kin-

der mit minimalen bzw. fehlenden Deutschkenntnissen, zusätzlich zur eigentlichen KiSS-Testung Fragebogen-basierte Prädiktoren für den Deutscherwerb bis zur Schuleingangsuntersuchung zu entwickeln. Solche Prädiktoren wurden tatsächlich 2017–2021 definiert und sind nun ein Bestandteil der noch nicht eingeführten Version KiSS.3 (cf. Zaretsky/Hey 2021). Dabei ist KiSS.3 nicht mit der von den KiSS-Testentwicklern nicht autorisierten Zwischenversion „KiSS dritte Generation“ (Tomasik et al. 2020) zu verwechseln, die in der Fachliteratur kritisiert wurde (cf. van Minnen/Lange 2022).

Bemerkenswert ist der hohe Anteil sprachpädagogisch förderbedürftiger Kinder unter SuZM (94%, cf. 50% bei SiZM, 18% bei MO). Damit wäre eine flächendeckende Sprachförderung der SuZM auch ohne Teilnahme am HSP bzw. ohne Sprachtestung durchaus berechtigt, zumal auch die 6%, die altersgemäße Deutschkenntnisse zeigten, meist nur die niedrigsten als unauffällig geltende Punktzahlen erreichten. Die „verdeckten Sprachschwierigkeiten“, die noch von Knapp (1999) thematisiert wurden und die nicht immer für die Eltern der ZM bzw. Kita-Erzieher und -Erzieherinnen ersichtlich sind, mögen den Anteil von 94% auf den ersten Blick als zu hoch erscheinen lassen, aber beim strikten algorithmischen Einsatz der KiSS-Auffälligkeitskriterien wäre dieser Prozentsatz noch höher. Die Auswertung durch Sprachexperten bzw. -expertinnen weicht erfahrungsgemäß immer wieder von den algorithmischen Kriterien ab im Sinne vom „Ausländerbonus“, der ebenfalls im o. g. Artikel von Knapp (1999) angesprochen wurde (z. B. beim altersgemäßen Sprachstand in der nicht-deutschen Muttersprache in Kombination mit relativ kurzer Dauer des Deutscherwerbs und fehlenden Hinweisen auf sprachbezogene medizinische Auffälligkeiten). Ein hoher Anteil an sprachpädagogisch förderbedürftigen Kindern unter SiZM, verglichen mit MO (sogar unter solchen, die Deutsch seit Geburt erwerben), zeigt, dass ihre Eingliederung in ZM durchaus berechtigt ist.

Hervorzuheben ist zudem ein unerwartet hoher Anteil klinisch abklärungsbedürftiger Kinder unter SuZM (29%, cf. 10% bei SiZM, 15% bei MO). Als mögliche Erklärung kommt in diesem Fall ein hoher Prozentsatz neu eingewanderter Kinder mit noch nicht abgeklärten bzw. therapierten sprachbezogenen medizinischen Auffälligkeiten (z. B. Hörstörungen) unter SuZM in Frage. Der Elternbrief, der zu allen KiSS-Versionen gehört, informiert Eltern über den klinischen Abklärungsbedarf und ermöglicht damit eine schnellere Einleitung der medizinischen Diagnostik, darunter auch bei SuZM.

Zusammenfassend lässt sich festhalten, dass die psychometrischen Kennwerte von KiSS je nach der Zusammensetzung der Stichprobe variieren können. Bei MO fielen die Eigentrennschärfe und interne Konsistenz, aufgrund einer speziellen Item-Selektion, in den Stichproben mit eher fortgeschrittenen Deutschkenntnissen etwas besser aus als in den Stichproben mit schwächeren Deutschkenntnissen. Da bei der Übertragung von KiSS aus Validierungsstudien ins HSP die KiSS-Altersgruppe um einen Monat erweitert wurde, führte es bei MO zu einem höheren Schwierigkeitsindex und zu einer Verbesserung anderer psychometrischer Kennwerte. Bei ZM blieben dagegen geprüfte Kennwerte unverändert, weil der Sprachstand in dieser Untergruppe weniger mit dem chronologischen Alter, sondern eher mit der Länge des Deutschkontakts zusammenhängt. Da KiSS von Anfang an eher für ZM optimiert war, waren seine psychometrischen Kennwerte durchgehend besser in ZM- als in MO-Untergruppen. Wegen einer nicht obligatorischen Teilnahme sowohl an KiSS-Validierungsstudien als auch am HSP ist

in beiden Fällen ein Selektionsbias zu vermuten, der für die Validierungsstudien in der aktuellen Arbeit zumindest indirekt nachgewiesen wurde. Die eingangs erwähnten zu hohen Auffälligkeitsraten im HSP sind damit vermutlich auf eine nicht ganz repräsentative Stichprobe zurückzuführen, in der klinisch abklärungsbedürftige Kinder überrepräsentiert waren. Eine repräsentative Stichprobe kann nur durch eine flächendeckende Teilnahme am HSP erreicht werden, wie es u. a. im Neugeborenen-Hörscreening der Fall ist. Aktuell muss damit gerechnet werden, dass infolge des Selektionsbias die Stichprobe der Kinder aus dem HSP für die Errechnung der neuen KiSS-Normwerte (die für alle hessischen Kinder gelten sollen) nur eingeschränkt geeignet ist. Aus demselben Grund sind die Deutschkenntnisse der Kinder, die im HSP getestet wurden, trotz einer hohen Fallzahl nicht mit den Deutschkenntnissen aller hessischen Kinder gleichzusetzen.

Die gezeigten Fluktuationen in den Güteindizes von KiSS sollen die Wichtigkeit und Notwendigkeit dieses Testverfahrens nicht in Frage stellen. Erstens verbesserten sich KiSS-Güteindizes bei der Übertragung dieses Screenings in das Sprachstandserfassungsprogramm. Die Fluktuationen sind damit nicht mit der Verschlechterung der Güteindizes gleichzusetzen. Zweitens ist KiSS nach wie vor einer der wenigen deutschen Tests mit separaten Normwerten für MO und ZM (ein anderes Beispiel wäre *Linguistische Sprachstandserhebung – Deutsch als Zweitsprache* (LiSe-DaZ) von Schulz/Tracy 2011). Drittens deckt KiSS im Gegensatz zu den meisten anderen Sprachtests bzw. -screenings alle linguistischen Ebenen ab und liefert darüber hinaus Hinweise auf Redefluss- und Stimmstörungen. Viertens kann KiSS im Gegensatz zur üblichen Dichotomisierung der Sprachtestergebnisse in „pass/fail“ zusätzlich zwischen sprachpädagogischem Förderbedarf und klinischem Abklärungsbedarf unterscheiden. Laut Mercator-Studie, in der mehrere in Deutschland verfügbare Sprachscreenings bewertet wurden (cf. Neugebauer/Becker-Mrotzek 2013), teilte KiSS Platz 1 mit dem *Dortmunder Entwicklungsscreening für den Kindergarten* (DESK 3–6 von Tröster/Flender/Reineke 2004) bzgl. seiner Anwendbarkeit und Güteindizes. Vor allem wurden die Berücksichtigung sprachlicher Basisqualifikationen, die Validität, die Objektivität, die Reliabilität und die Normierung des Verfahrens positiv hervorgehoben. Insgesamt erfüllte KiSS, wie auch DESK 3–6, 25 von 32 Qualitätskriterien. Die (noch nicht eingeführte) neue Version von KiSS – KiSS.3 – hat die Güteindizes dieses Screenings nochmal verbessert.

Literaturverzeichnis

- Armon-Lotem, Sharon (2012): “Specific language impairment in bilingual children”. *Bilingualism: Language and Cognition* 15/1: 1–4.
- BBMFI (Die Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration) (2019): *Deutschland kann Integration: Potenziale fördern, Integration fordern, Zusammenhalt stärken*. Berlin: BBMFI.
- Bühner, Markus (2021): *Einführung in die Test- und Fragebogenkonstruktion*. 4. Aufl. München: Pearson Studium.
- De Jong, Jan/Çavus, Nazife/Baker, Anne (2010): “Language impairment in Turkish-Dutch bilingual children”. In: Topbaş, Seyhun/Yavas, Mehmet (eds.): *Communication disorders in Turkish in monolingual and multilingual settings*. Clevedon, Multilingual Matters: 288–300.

- Diedenhofen, Birk/Musch, Jochen (2016): “Cocron: A web interface and R package for the statistical comparison of Cronbach’s alpha coefficients”. *International Journal of Internet Science* 11: 51–60. comparingcronbachalphas.org [24.02.2025].
- Döpfner, Manfred et al. (2005): *S-ENS. Screening des Entwicklungsstandes bei Einschulungsuntersuchungen: theoretische und statistische Grundlagen. Manual Screening des Entwicklungsstandes bei Einschulungsuntersuchungen*. Göttingen: Hogrefe.
- Döring, Nicola/Bortz, Jürgen (2016): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 5. Aufl. Berlin: Springer.
- Euler, Harald A. et al. (2010): „Psychometrische Gütekriterien eines Kurztests zur Erfassung des Sprachstands 4-jähriger Kinder“. *HNO* 58: 1116–1123.
- Genesee, Fred/Paradis, Johanne/Crago, Martha B. (2004): *Dual language development and disorders: A handbook on bilingualism and second language learning*. Baltimore: Brookes.
- Gogolin, Ingrid (2002): “Sprachlich-kulturelle Differenz und Chancengleichheit – (un)versöhnlich in staatlichen Bildungssystemen?” In: Lohmann, Ingrid/Rilling, Rainer (eds.): *Die verkaufte Bildung. Kritik und Kontroversen zur Kommerzialisierung von Schule, Weiterbildung, Erziehung und Wissenschaft*. Opladen, Dietz-Verlag: 153–168.
- Grimm, Angela/Schulz, Petra (2012): “Das Sprachverstehen bei frühen Zweitsprachlernern: Erste Ergebnisse der kombinierten Längs- und Querschnittstudie MILA”. In: Ahrenholz, Bernt (ed.): *Einblicke in die Zweitspracherwerbsforschung und ihre methodischen Verfahren. Reihe DaZ-Forschung. Deutsch als Zweitsprache, Mehrsprachigkeit und Migration*. Berlin, de Gruyter: 195–218.
- Grimm, Angela/Schulz, Petra (2014): “Specific Language Impairment and early second language acquisition: The risk of over- and underdiagnosis”. *Child Indicators Research* 7: 821–841.
- Grimm, Angela/Schulz, Petra (2016): „Warum man bei mehrsprachigen Kindern dreimal nach dem Alter fragen sollte: Sprachfähigkeiten simultan-bilingualer Lerner im Vergleich mit monolingualen und frühen Zweitsprachlernern“. *Diskurs Kindheits- und Jugendforschung* 11(1) 27–42. doi.org/10.3224/diskurs.v11i1.22247.
- Grimm, Hannelore (2001): *SETK 3–5. Sprachentwicklungstest für drei- bis fünfjährige Kinder*. Göttingen: Hogrefe.
- Grissom, Robert J./Kim, John J. (2012): *Effect sizes for research: Univariate and multivariate applications*. 2nd ed. New York: Routledge.
- Grobe, Thomas G./Szecsenyi, Joachim (2023): *Barmer-Arztreport 2023*. Berlin: Barmer.
- Hanack, Peter (2019): „Kiss im Kindergarten“. *Frankfurter Rundschau*. fr.de/rhein-main/kiss-kindergarten-11594817.html [06.06.2024]
- Holler-Zittlau, Inge/Dux, Winfried/Berger, Roswitha (2003): *Marburger Sprach-Screening für 4- bis 6-jährige Kinder (MSS)*. Hamburg: Horneburg Verlag.
- Holler-Zittlau, Inge/Euler, Harald A./Neumann, Katrin (2011): „Kindersprachscreening (KiSS) – das hessische Verfahren zur Sprachstandserfassung“. *Sprachheilarbeit* 5: 263–268.
- IQWiG (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen) (2009): *Früherkennungsuntersuchung auf umschriebene Entwicklungsstörungen des Sprechens und der Spra-*

- che* (Version 1.0). *Abschlussbericht S06-01*. Köln: IQWiG. qwig.de/download/s06-01abschlussbericht_frueherkennung_umschriebener_stoerungen_des_sprechens_und_der_sprache.pdf [31.01.2025].
- Jeuk, Stefan (2009): „Probleme der Sprachstandserhebung bei mehrsprachigen Kindern“. *Zeitschrift für Soziologie der Erziehung und Sozialisation* 29/2: 141–156.
- Kiese-Himmel, Christiane (2005): *AWST-R. Aktiver Wortschatztest für 3- bis 5-jährige Kinder – Revision*. Göttingen: Hogrefe.
- Knapp, Werner (1999): „Verdeckte Sprachschwierigkeiten“. *Grundschule* 31/5: 30–33.
- Meisel, Jürgen M. (2009): „Second language acquisition in early childhood“. *Zeitschrift für Sprachwissenschaft* 28/1: 5–34. doi.org/10.1515/ZFSW.2009.002.
- Moosbrugger, Helfried/Kelava, Augustin (2012): *Testtheorie und Fragebogenkonstruktion*. 2. Aufl. Berlin: Springer.
- 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.
- Neumann, Katrin/Euler, Harald A. (2009): *Einführung einer flächendeckenden Sprachstandserfassung in Hessen. Forschungsbericht 2010*. Frankfurt am Main: Goethe-Universität Frankfurt am Main.
- Neumann, Katrin/Euler, Harald A. (2010): *Einführung einer flächendeckenden Sprachstandserfassung in Hessen. Forschungsbericht 2010*. Frankfurt am Main: Goethe-Universität Frankfurt am Main.
- Neumann, Katrin/Euler, Harald A./Schneider, Peter (2014): *Identifikation von Stottern im Vorschulalter. Projektabschlussbericht*. Köln: Demosthenes.
- Neumann, Katrin/Zaretsky, Yevgen/Euler, Harald A. (2011): *Einführung einer flächendeckenden Sprachstandserfassung in Hessen. Forschungsbericht*. Frankfurt am Main: Goethe-Universität Frankfurt am Main.
- Paradis, Johanne (2010): „The interface between bilingual development and specific language impairment“. *Applied Psycholinguistics* 31/2: 227–252.
- Paradis, Johanne/Genesee, Fred/Crago, Martha (2011): *Dual language development and disorders: A handbook on bilingualism and second language learning*. 2. Aufl. Baltimore: Brookes Publishing.
- Paradis, Johanne et al. (2003): “French-English bilingual children with SLI: How do they compare with their monolingual peers?” *Journal of Speech, Language, and Hearing Research* 46/1: 113–127.
- Presse- und Informationsamt der Bundesregierung. (2007): *Der Nationale Integrationsplan: Neue Wege – Neue Chancen*. Drucksache 16/6281 des Deutschen Bundestags. dserver.bundestag.de/btd/16/062/1606281.pdf [06.06.2024].
- Schneider, Hansjakob et al. (2013): *Expertise. Wirksamkeit von Sprachförderung*. Köln: Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache.
- Schulz, Petra/Tracy, Rosemarie (2011): *Linguistische Sprachstandserhebung – Deutsch als Zweitsprache (LiSe-DaZ)*. Göttingen: Hogrefe.

- Schulz, Petra/Tracy, Rosemarie/Wenzel, Ramona (2008): „Linguistische Sprachstandserhebung – Deutsch als Zweitsprache (LiSe-DaZ): Theoretische Grundlagen und erste Ergebnisse“. In: Ahrenholz, Bernt (ed.): *Zweitspracherwerb. Diagnosen, Verläufe, Voraussetzungen*. Fillibach, Freiburg im Breisgau: 17–42.
- Tomasik, Martin J. et al. (2020): „Landesweite Normierung des Hessischen Kindersprachscreenings“. *Diagnostica* 66/3: 178–189. doi.org/10.1026/0012-1924/a000246.
- Tröster, Heinrich/Flender, Judith/Reineke, Dirk (2004): *DESK 3–6. Dortmunder Entwicklungsscreening für den Kindergarten*. Göttingen: Hogrefe.
- Van Minnen, Susanne/Lange, Benjamin P. (2022): „Ein Kommentar zu Tomasik et al. (2020), ‚Landesweite Normierung des Hessischen Kindersprachscreenings‘“. *Praxis Sprache* 67/1: 61–62.
- Weiland, Marina et al. (2019): „Sprachstandserfassung im Elementarbereich für ein- und mehrsprachige Kinder am Beispiel des hessischen Kindersprachscreenings KiSS“. *Sprachförderung und Sprachtherapie* 2: 84–88.
- Yağmur, Kutlay/Nap-Kolhoff, Elma (2010): „Aspects of acquisition and disorders in Turkish-Dutch bilingual children“. In: Topbaş, Seyhun/Yavas, Mehmet (eds.): *Communication disorders in Turkish in monolingual and multilingual settings*. Clevedon, Multilingual Matters: 269–288.
- Zaretsky, Eugen/Hey, Christiane (2021): *Förderung der nachhaltigen Integration von Vorschulkindern mit minimalen bzw. fehlenden Deutschkenntnissen durch frühzeitige Prädiktion einer Störung des Zweitspracherwerbs. Entwicklung von KiSS.3. Abschlussbericht 2017–2021*. Marburg: Philipps-Universität Marburg.
- Zaretsky, Eugen/Lange, Benjamin P. (2023): „Sprachkontaktdauer: Auswirkungen auf den Sprachstand im Deutschen abhängig vom Alter des Zweitspracherwerbs und Sprachgebrauch im familiären Umfeld.“ *Linguistik online* 122/4: 115–137. doi.org/10.13092/lo.122.10201.
- Zaretsky, Eugen/Lange, Benjamin P./Hey, Christiane (2022): „Fit für die Schule? Der Sprachstand 5-jähriger Kinder in Hessen und seine soziodemografischen Korrelate“. *Monatsschrift Kinderheilkunde* 171: 340–347. doi.org/10.1007/s00112-022-01568-8.
- Zaretsky, Eugen/Lange, Benjamin P./Hey, Christiane (2023): „Änderungen im Sprachstand vierjähriger Kitakinder zwischen 2008 und 2019: medizinische und soziolinguistische Einflussfaktoren“. *Forschung Sprache* 11/2: 196–204.
- Zaretsky, Eugen et al. (2021): „Sprachtherapeutische Versorgung 4-jähriger Kinder in Hessen: ein Überblick“. *Monatsschrift Kinderheilkunde*. doi.org/10.1007/s00112-021-01376-6.
- Zieleniewicz, Eva D. (2014): *Optimierung der Sprachstandserfassung bei mehrsprachig aufwachsenden Kindern mit dem Kindersprachscreening (KiSS.2)*. Dissertation, Goethe-Universität Frankfurt am Main.

Elektronische Supplemente (ESM)**Elektronisches Supplement 1: Schwierigkeitsgrad von KiSS.1- und KiSS.2-Items**

KiSS.1-Items	KiSS.1-Validierungsstudie		KiSS.XL-Validierungsstudie		HSP	
	MO	ZM	MO	ZM	MO	ZM
<i>n</i>	143	79	94	87	4280	1864
Spontansprache bzw. Sprachproduktion						
1: Kind spricht spontan über das Bild?	121 (84,6%)	62 (78,5%)	94 (100,0%)	84 (96,6%)	4132 (96,5%)	1719 (92,2%)
8: Sag mir, was ich Dir zeigen soll.	116 (81,1%)	66 (83,5%)	49 (52,1%)	33 (37,9%)	3338 (78,0%)	1237 (66,4%)
Sprachverständnis						
4: Zeige mir einen Jungen mit einer blauen Hose.	126 (88,1%)	58 (73,4%)	93 (98,9%)	74 (85,1%)	4095 (95,7%)	1684 (88,4%)
5: Zeige mir den Jungen neben dem Klettergerüst.	120 (83,9%)	43 (54,4%)	70 (74,5%)	35 (40,2%)	3616 (84,5%)	905 (48,6%)
6: Zeige mir die grüne Kiste unter dem Brett.	95 (66,4%)	22 (27,8%)	70 (74,5%)	35 (40,2%)	3633 (84,9%)	1059 (56,8%)
7: Zeige mir das Mädchen, das den Sand in den Anhänger schüttet.	116 (81,1%)	24 (30,4%)	76 (80,9%)	40 (46,0%)	4063 (94,9%)	1319 (70,8%)
Wortschatz						
9: Dach	126 (88,1%)	36 (45,6%)	83 (88,3%)	28 (32,2%)	3493 (81,6%)	922 (50,5%)
10: Fenster	137 (95,8%)	62 (78,5%)	93 (98,9%)	62 (71,3%)	4129 (96,5%)	1449 (77,7%)
11: Klettergerüst	120 (83,9%)	37 (46,8%)	75 (79,8%)	31 (35,6%)	3751 (87,6%)	905 (48,6%)
12: Rad	134 (93,7%)	50 (63,3%)	54 (57,4%)	32 (36,8%)	3920 (91,6%)	1337 (71,7%)
13: Trommel	133 (93,0%)	49 (62,0%)	91 (96,8%)	52 (59,8%)	3950 (92,3%)	1186 (63,6%)
14: Farbe: der Pulli (gelb)	131 (91,6%)	75 (94,9%)	92 (97,9%)	73 (83,9%)	3974 (92,9%)	1557 (83,5%)
15: Farbe: die Hose des Kindes (blau)	132 (92,3%)	70 (88,6%)	90 (95,7%)	74 (85,1%)	3992 (93,3%)	1614 (86,6%)

16: Eigenschaft: der Teddy (weich/kuschelig)	114 (79,7%)	43 (54,4%)	67 (71,3%)	23 (26,4%)	3244 (75,8%)	831 (44,6%)
17: Eigenschaft: die Rut- sche (hart, glatt, kalt)	96 (67,1%)	33 (41,8%)	60 (63,8%)	22 (25,3%)	2926 (68,4%)	716 (38,4%)
18: Form: der Ball (rund)	91 (63,6%)	31 (39,2%)	49 (52,1%)	18 (20,7%)	2263 (52,9%)	451 (24,2%)
19: Form: das Fenster ((vier)eckig)	62 (43,4%)	14 (17,7%)	30 (31,9%)	13 (14,9%)	1821 (42,5%)	321 (17,2%)
Aussprache						
20: Dach (Ziellaut: /h/)	133 (93,0%)	68 (86,1%)	91 (96,8%)	79 (90,8%)	4065 (95,0%)	1717 (92,1%)
21: Bücher (Ziellaut: /č/)	123 (86,0%)	52 (65,8%)	81 (86,2%)	71 (81,6%)	3282 (76,7%)	1342 (72,0%)
22: Sand (Ziellaut: /z/)	128 (89,5%)	65 (82,3%)	64 (68,1%)	63 (72,4%)	3749 (87,6%)	1672 (89,7%)
23: Korb (Ziellaut: /k/)	122 (87,8%)	65 (82,3%)	88 (93,6%)	82 (94,3%)	3875 (90,5%)	1719 (92,2%)
24: Schaufel/Schippe (Ziellaut: /ʃ/)	116 (81,1%)	68 (86,1%)	77 (81,9%)	60 (69,0%)	3358 (78,5%)	1414 (75,9%)
25: Schmetterling (Ziel- laute: /ʃm/)	111 (77,6%)	60 (75,9%)	72 (76,6%)	58 (66,7%)	3211 (75,0%)	1300 (69,7%)
26: Kiste (Ziellaute: /k/ und /st/)	104 (72,7%)	48 (60,8%)	55 (58,5%)	60 (69,0%)	3553 (83,0%)	1615 (86,6%)
27: Brille (Ziellaute: /br/)	123 (86,0%)	59 (74,7%)	86 (91,5%)	70 (80,5%)	3829 (89,5%)	1501 (80,5%)
28: Klettergerüst (Ziel- laute: /kl/ und /st/)	113 (79,0%)	49 (62,0%)	58 (61,7%)	56 (64,4%)	3262 (76,2%)	1341 (71,9%)
29: Rad (Ziellaut: /r/)	127 (88,8%)	65 (82,3%)	88 (93,6%)	72 (82,8%)	3978 (92,9%)	1634 (87,7%)
30: Trommel (Ziellaute: /tr/)	112 (78,3%)	58 (73,4%)	80 (85,1%)	68 (78,2%)	3539 (82,7%)	1495 (80,2%)
31: Krone (Ziellaute: /kr/)	107 (74,8%)	59 (74,7%)	78 (83,0%)	65 (74,7%)	3546 (82,9%)	1489 (79,9%)
Grammatik						
32: Plural: Apfel, Äpfel	85 (59,4%)	21 (26,6%)	68 (72,3%)	25 (28,7%)	2928 (68,4%)	558 (29,9%)
33: Plural: Ball, Bälle	115 (80,4%)	34 (43,0%)	83 (88,3%)	33 (37,9%)	3580 (83,6%)	783 (42,0%)

34: Plural: Auto, Autos	122 (85,3%)	53 (67,1%)	90 (95,7%)	46 (52,9%)	3995 (93,3%)	1236 (66,3%)
35: Partizipbildung: gekrabbelt/gekrochen/gelaufen	101 (70,6%)	35 (44,3%)	65 (69,1%)	30 (34,5%)	3037 (71,0%)	864 (46,4%)
36: Partizipbildung: hingefallen/gefallen/gestürzt	114 (79,7%)	52 (65,8%)	79 (84,0%)	55 (63,2%)	3406 (79,6%)	1151 (61,7%)
37: Partizipbildung: (vor-)gelesen	104 (72,7%)	32 (40,5%)	67 (71,3%)	33 (37,9%)	3281 (76,7%)	887 (47,6%)
38: 2. Person Singular: sollst, darfst	69 (48,3%)	22 (27,8%)	50 (53,2%)	34 (39,1%)	2199 (51,4%)	692 (37,1%)
39: 2. Person Singular: riechst/duftest	77 (53,8%)	32 (40,5%)	61 (64,9%)	30 (34,5%)	2578 (60,2%)	803 (43,1%)
40: Präposition im Akkusativkontext: ...in den Sandkasten	76 (53,1%)	25 (31,6%)	79 (84,0%)	36 (41,4%)	2830 (66,1%)	695 (37,3%)
41: Präposition im Akkusativkontext: ...auf das Dach	94 (65,7%)	18 (22,8%)	54 (57,4%)	15 (17,2%)	2393 (55,9%)	360 (19,3%)
42: Präposition im Akkusativkontext: durch das Rohr	69 (48,3%)	8 (10,1%)	56 (59,6%)	14 (16,1%)	2489 (58,2%)	315 (16,9%)
43: Präposition im Dativkontext: aus dem Sandkasten	70 (49,0%)	12 (15,2%)	53 (56,4%)	14 (16,1%)	2119 (49,5%)	352 (18,9%)
44: Präposition im Dativkontext: auf der Wippe	102 (71,3%)	22 (27,8%)	79 (84,0%)	25 (28,7%)	3238 (75,7%)	529 (28,4%)
45: Nebensatzbildung mit Konjunktion: ...die Kinder reinkommen können	103 (72,0%)	29 (36,7%)	71 (75,5%)	41 (47,1%)	2993 (69,9%)	754 (40,5%)
46: Nebensatzbildung mit Konjunktion: ...die Kinder bauen können	109 (76,2%)	31 (39,2%)	67 (71,3%)	29 (33,3%)	2861 (66,8%)	591 (31,7%)
Extremwerte (Schwierigkeit 0–10, 90–100%)	6	2	11	3	12	3
Mittlere Werte (Schwierigkeit 45–55%)	5	6	3	3	3	5

KiSS.2	KiSS.2-Validierungsstudie		BASS-Validierungsstudie		KiSS.3-Validierungsstudie	
	MO	ZM	MO	ZM	MO	ZM
<i>n</i>	131	180	279	352	357	528
Spontansprache bzw. Sprachproduktion						
1: Spontansprache	123 (93,9%)	162 (90,0%)	264 (94,6%)	332 (94,3%)	344 (96,4%)	455 (86,2%)
16: Sag mir, was ich Dir zeigen soll.	116 (88,5%)	150 (83,3%)	235 (84,2%)	243 (69,0%)	319 (89,4%)	350 (66,3%)
Sprachverständnis						
4: Junge mit einer blauen Hose	125 (95,4%)	160 (88,9%)	267 (95,7%)	302 (85,8%)	340 (95,2%)	415 (78,6%)
5: Junge neben dem Klettergerüst	96 (73,3%)	85 (47,2%)	220 (78,9%)	165 (46,9%)	249 (69,7%)	186 (35,2%)
6: die grüne Kiste unter dem Brett	99 (75,6%)	98 (54,4%)	207 (74,2%)	174 (49,4%)	244 (68,3%)	171 (32,4%)
Wortschatz						
7: Dach	93 (71,0%)	74 (41,1%)	234 (83,9%)	164 (46,6%)	278 (77,9%)	174 (33,0%)
8: Fenster	127 (96,9%)	139 (77,2%)	263 (94,3%)	262 (74,4%)	342 (95,8%)	302 (57,2%)
9: Klettergerüst	111 (84,7%)	76 (42,2%)	231 (82,8%)	135 (38,4%)	319 (89,4%)	192 (36,4%)
10: Fahrrad	129 (98,5%)	161 (89,4%)	274 (98,2%)	329 (93,5%)	281 (78,7%)	259 (49,1%)
11: Trommel	117 (89,3%)	111 (61,7%)	261 (93,5%)	225 (63,9%)	333 (93,3%)	269 (50,9%)
12: Eigenschaft: der Teddy	92 (70,2%)	69 (38,3%)	180 (64,5%)	89 (25,3%)	210 (58,8%)	90 (17,0%)
13: Eigenschaft: die Rutsche	91 (69,5%)	64 (35,6%)	176 (63,1%)	92 (26,1%)	185 (51,8%)	89 (16,9%)
14: Form: der Ball (rund)	70 (53,4%)	51 (28,3%)	133 (47,7%)	90 (25,6%)	172 (48,2%)	89 (16,9%)
15: Form: das Fenster ((vier)eckig)	59 (45,0%)	34 (18,9%)	97 (34,8%)	60 (17,0%)	104 (29,1%)	41 (7,8%)

Aussprache						
17: Dach (Ziellaut: /h/)	121 (92,4%)	160 (88,9%)	271 (97,1%)	325 (92,3%)	335 (93,8%)	483 (91,5%)
18: Bücher (Ziellaut: /č/)	102 (77,9%)	129 (71,7%)	220 (78,9%)	245 (69,6%)	252 (70,6%)	363 (68,8%)
19: Sand (Ziellaut: /z/)	117 (89,3%)	158 (87,8%)	253 (90,7%)	317 (90,1%)	328 (91,9%)	496 (93,9%)
20: Korb (Ziellaut: /k/)	117 (89,3%)	161 (89,4%)	264 (94,6%)	316 (89,8%)	322 (90,2%)	496 (93,9%)
21: Schaufel (Ziellaut: /f/)	117 (89,3%)	128 (71,1%)	218 (78,1%)	256 (72,7%)	272 (76,2%)	383 (72,5%)
22: Schmetterling (Ziellaute: /ʃm/)	109 (83,2%)	123 (68,3%)	217 (77,8%)	245 (69,6%)	268 (75,1%)	355 (67,2%)
23: Brille (Ziellaute: /br/)	116 (88,5%)	138 (76,7%)	249 (89,2%)	286 (81,3%)	319 (89,4%)	387 (73,3%)
24: Klettergerüst (Ziellaute: /kl/)	112 (85,5%)	157 (87,2%)	235 (84,2%)	265 (75,3%)	328 (91,9%)	456 (86,4%)
25: Klettergerüst (Ziellaute: /st/)	112 (85,5%)	146 (81,1%)	235 (84,2%)	265 (75,3%)	330 (92,4%)	419 (79,4%)
26: Flugzeug (Ziellaute: /fl/)	117 (89,3%)	147 (81,7%)	252 (90,3%)	297 (84,4%)	332 (93,0%)	447 (84,7%)
27: Trommel (Ziellaute: /tr/)	108 (82,4%)	144 (80,0%)	229 (82,1%)	276 (78,4%)	291 (81,5%)	392 (74,2%)
28: Krone (Ziellaute: /kr/)	109 (83,2%)	150 (83,3%)	243 (87,1%)	286 (81,3%)	290 (81,2%)	422 (79,9%)
29: Krone (Ziellaut: /r/)	118 (90,1%)	153 (85,0%)	260 (93,2%)	305 (86,6%)	333 (93,3%)	460 (87,1%)
Grammatik						
30: Plural: Apfel, Äpfel	92 (70,2%)	51 (28,3%)	170 (60,9%)	95 (27,0%)	229 (64,1%)	86 (16,3%)
31: Plural: Ball, Bälle	100 (76,3%)	68 (37,8%)	222 (79,6%)	160 (45,5%)	294 (82,4%)	146 (27,7%)
32: Plural: Auto, Autos	117 (89,3%)	96 (53,3%)	211 (75,6%)	177 (50,3%)	318 (89,1%)	252 (47,7%)
33: Partizipbildung: gekrabbelt	94 (71,8%)	80 (44,4%)	202 (72,4%)	168 (47,7%)	227 (63,6%)	127 (24,1%)
34: Partizipbildung: (vor-)gelesen	91 (69,5%)	71 (39,4%)	201 (72,0%)	167 (47,4%)	255 (71,4%)	171 (32,4%)

35: 2. Person Singular: sollst, darfst	64 (48,9%)	71 (39,4%)	134 (48,0%)	141 (40,1%)	160 (44,8%)	160 (30,3%)
36: 2. Person Singular: riechst	78 (59,5%)	70 (38,9%)	180 (64,5%)	157 (44,6%)	196 (54,9%)	196 (37,1%)
37: auf das Dach	83 (63,4%)	28 (15,6%)	175 (62,7%)	72 (20,5%)	164 (45,9%)	48 (9,1%)
38: durch das Rohr	78 (59,5%)	31 (17,2%)	165 (59,1%)	57 (16,2%)	183 (51,3%)	48 (9,1%)
39: aus dem Sandkasten	75 (57,3%)	34 (18,9%)	157 (56,3%)	75 (21,3%)	170 (47,6%)	55 (10,4%)
40: auf der Wippe	94 (71,8%)	50 (27,8%)	228 (81,7%)	103 (30,7%)	265 (74,2%)	103 (19,5%)
41: ...die Kinder reinkom- men können	112 (86,2%)	98 (54,4%)	219 (78,5%)	166 (47,2%)	313 (87,7%)	206 (39,0%)
42: ...die Kinder bauen können	93 (71,0%)	70 (38,9%)	188 (67,4%)	123 (34,9%)	254 (71,1%)	146 (27,7%)
Nachsprechen von Kunstwörtern						
43: Triser	88 (67,2%)	101 (56,1%)	236 (84,6%)	285 (81,0%)	288 (80,7%)	382 (72,3%)
44: Misküranok	61 (46,6%)	82 (45,6%)	128 (45,9%)	154 (43,8%)	143 (40,1%)	180 (34,1%)
45: Nabolira	79 (60,3%)	94 (52,2%)	158 (56,6%)	167 (47,4%)	169 (47,3%)	220 (41,7%)
46: Verklasaft	54 (41,2%)	66 (36,7%)	144 (51,6%)	151 (42,9%)	151 (42,3%)	197 (37,3%)
Nachsprechen von Sätzen (maximale Punktzahl pro Item)						
47: Das braune Pony rennt über...	74 (56,5%)	26 (14,4%)	139 (49,8%)	65 (18,5%)	195 (54,6%)	70 (13,3%)
48: Bevor Lena abgeholt wird, muss...	60 (45,8%)	16 (8,9%)	130 (46,6%)	60 (17,0%)	130 (36,4%)	44 (8,3%)
Extremwerte (Schwierig- keit 0–10, 90–100%)	6	2	10	4	11	8
Mittlere Werte (Schwierig- keit 45–55%)	5	6	6	9	8	3

Tabelle E1: Schwierigkeitsgrade einzelner KiSS.1- und KiSS.2-Items (% der richtigen Antworten)^a

Anmerkungen: MO = monolingual deutschsprachige Kinder, ZM = zwei-/mehrsprachige Kinder; HSP = hessisches Sprachstandserfassungsprogramm.

^a KiSS-Items 2 und 3 (Redefluss- und Stimmstörungen) wurden in der aktuellen Arbeit nicht berücksichtigt und sind weder in Tab. E1 noch in anderen Tabellen enthalten.

KiSS.1	KiSS.1-Validierungsstudie		KiSS.XL-Validierungsstudie		HSP	
	MO	ZM	MO	ZM	MO	ZM
Untertest "Sprachverständnis"	0,80 ±0,09	0,47 ±0,22	0,82 ±0,12	0,53 ±0,21	0,90 ±0,06	0,66 ±0,17
Untertest "Wortschatz"	0,81 ±0,17	0,58 ±0,23	0,76 ±0,22	0,45 ±0,26	0,80 ±0,18	0,55 ±0,24
Untertest "Aussprache"	0,83 ±0,07	0,76 ±0,09	0,81 ±0,13	0,77 ±0,09	0,84 ±0,07	0,82 ±0,08
Untertest "Grammatik"	0,66 ±0,13	0,36 ±0,16	0,73 ±0,13	0,35 ±0,13	0,68 ±0,12	0,38 ±0,15
Insg. (Mittelwert)	0,76 ±0,14	0,55 ±0,23	0,77 ±0,16	0,52 ±0,23	0,78 ±0,14	0,58 ±0,24
KiSS.2	KiSS.2-Validierungsstudie		BASS-Validierungsstudie		KiSS.3-Validierungsstudie	
Untertest "Sprachverständnis"	0,81 ±0,13	0,64 ±0,22	0,83 ±0,11	0,61 ±0,22	0,78 ±0,15	0,49 ±0,30
Untertest "Wortschatz"	0,75 ±0,19	0,48 ±0,23	0,74 ±0,22	0,46 ±0,26	0,69 ±0,23	0,32 ±0,18
Untertest "Aussprache"	0,87 ±0,04	0,81 ±0,07	0,87 ±0,06	0,81 ±0,08	0,86 ±0,08	0,81 ±0,09
Untertest "Grammatik"	0,69 ±0,11	0,35 ±0,13	0,68 ±0,10	0,36 ±0,12	0,65 ±0,16	0,25 ±0,12
Untertest "Nachsprechen" ^a	0,53 ±0,10	0,36 ±0,20	0,56 ±0,15	0,42 ±0,23	0,50 ±0,16	0,35 ±0,23
Insg. (Mittelwert)	0,75 ±0,16	0,55 ±0,25	0,75 ±0,17	0,55 ±0,25	0,72 ±0,19	0,47 ±0,28

Tabelle E2: Schwierigkeitsindex des „Kindersprachscreenings“ (KiSS.1 und KiSS.2): Mittelwerte und Standardabweichungen

Anmerkungen: MO = monolingual deutschsprachige Kinder, ZM = zwei-/mehrsprachige Kinder; HSP = hessisches Sprachstandserfassungsprogramm, BASS = „Bochum-Aachener Stotterscreening“.

^a Untertest „Nachsprechen“ war in KiSS.1 nicht enthalten.

	KiSS.2-V	HSP
Untertest „Sprachverständnis“	0,77 ±0,15	0,82 ±0,10
Untertest „Wortschatz“	0,69 ±0,19	0,70 ±0,24
Untertest „Aussprache“	0,85 ±0,04	0,84 ±0,06
Untertest „Grammatik“	0,60 ±0,11	0,61 ±0,13
Untertest „Nachsprechen“	0,49 ±0,11	0,62 ±0,13
Insg. (Mittelwert)	0,70 ±0,17	0,71 ±0,17

Tabelle E3: Gewichtete Angaben zum Schwierigkeitsindex von KiSS.2-Items in der KiSS.2-Validierungsstudie (KiSS.2-V) und im hessischen Sprachstandserfassungsprogramm (HSP, laut Tomasik et al. (2020)^a), Verhältnis 1:3 von monolingual deutschsprachigen und zwei-/mehrsprachigen Kindern: Mittelwerte und Standardabweichungen

^a Als Grundlage für Tab. E3 (Spalte „HSP“) wurden eigene Berechnungen (Mittelung der Werte) mit den Daten aus der Elektronischen Anlage 1 des Artikels von Tomasik et al. (2020) herangezogen.

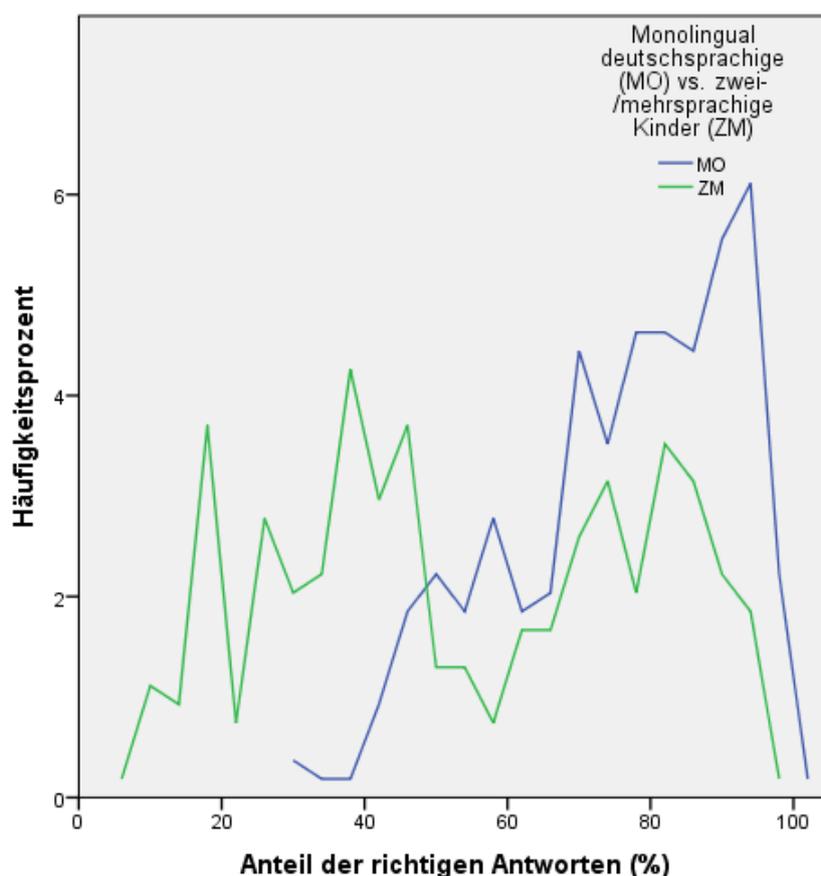


Abbildung E1: Häufigkeit der richtigen Antworten in allen KiSS.1- und KiSS.2-Stichproben^a, getrennt für monolingual deutschsprachige (MO) und zwei-/mehrsprachige Kinder (ZM)

^a Ohne KiSS.2-Daten von Tomasik et al. (2020), da in diesen zwischen monolingual deutschsprachigen und zwei-/mehrsprachigen Kindern nicht unterschieden wird.

Elektronisches Supplement 2: Eigentrennschärfe von KiSS.1- und KiSS.2-Items

KiSS.1: Item-Num- mern ^a	KiSS.1-Validierungs- studie		KiSS.XL-Validierungs- studie		HSP	
	MO	ZM	MO	ZM	MO	ZM
<i>n</i>	143	79	94	87	4280	1864

Punktbiseriale Korrelationen einzelner Items mit dem jeweiligen KiSS.1-Untertest-score ohne dieses Item („Item vs. Untertestscore“)

Sprachverständnis						
4	0,278*	0,401*	0,266*	0,266*	0,217*	0,368*
5	0,447*	0,554*	0,272*	0,358*	0,291*	0,393*
6	0,370*	0,466*	0,358*	0,392*	0,314*	0,451*
7	0,291*	0,567*	0,362*	0,284*	0,209*	0,404*
Wortschatz						
9	0,108	0,589*	0,284*	0,582*	0,408*	0,574*
10	0,216*	0,614*	0,340*	0,560*	0,314*	0,544*
11	0,150	0,627*	0,408*	0,523*	0,398*	0,562*
12	0,189*	0,539*	0,044	0,346*	0,278*	0,451*
13	0,138	0,525*	0,231*	0,545*	0,317*	0,570*
14	0,253*	0,339*	0,118	0,224*	0,298*	0,441*
15	0,240*	0,368*	0,097	0,348*	0,318*	0,400*
16	0,250*	0,461*	0,461*	0,546*	0,481*	0,582*
17	0,342*	0,544*	0,499*	0,566*	0,482*	0,560*
18	0,540*	0,442*	0,462*	0,558*	0,500*	0,458*
19	0,434*	0,313*	0,274*	0,376*	0,455*	0,436*
Aussprache						
20	0,414*	0,399*	0,197	0,473*	0,367*	0,349*
21	0,205*	0,271*	0,248*	0,143	0,342*	0,307*
22	0,480*	0,219	0,369*	0,296*	0,361*	0,363*
23	0,564*	0,400*	0,330*	-0,012	0,512*	0,432*
24	0,456*	0,411*	0,483*	0,588*	0,495*	0,438*
25	0,582*	0,283*	0,503*	0,594*	0,540*	0,482*
26	0,527*	0,250*	0,432*	0,430*	0,569*	0,484*
27	0,532*	0,268*	0,411*	0,536*	0,592*	0,572*

28	0,546*	0,384*	0,369*	0,342*	0,601*	0,548*
29	0,582*	0,499*	0,396*	0,634*	0,520*	0,528*
30	0,485*	0,479*	0,479*	0,352*	0,483*	0,561*
31	0,606*	0,490*	0,474*	0,442*	0,577*	0,573*

Grammatik

32	0,427*	0,553*	0,387*	0,587*	0,381*	0,485*
33	0,439*	0,611*	0,432*	0,782*	0,467*	0,646*
34	0,373*	0,458*	0,192	0,703*	0,368*	0,562*
35	0,173*	0,378*	0,284*	0,523*	0,415*	0,537*
36	0,228*	0,473*	0,359*	0,372*	0,406*	0,491*
37	0,329*	0,209	0,438*	0,638*	0,450*	0,539*
38	0,208*	0,384*	0,245*	0,573*	0,323*	0,452*
39	0,337*	0,417*	0,075	0,483*	0,387*	0,479*
40	0,269*	0,561*	0,179	0,453*	0,420*	0,491*
41	0,443*	0,330*	0,450*	0,431*	0,521*	0,504*
42	0,375*	0,283*	0,462*	0,368*	0,525*	0,547*
43	0,374*	0,519*	0,433*	0,481*	0,432*	0,509*
44	0,644*	0,661*	0,469*	0,675*	0,552*	0,594*
45	0,377*	0,495*	0,316*	0,550*	0,448*	0,567*
46	0,383*	0,433*	0,525*	0,508*	0,503*	0,582*

Punktbiserialen Korrelationen einzelner Items mit dem jeweiligen KiSS.1-Gesamtscore ohne dieses Item („Item vs. Gesamtscore“)

Spontansprache bzw. Sprachproduktion^b

1	0,080	0,211	– ^c	0,120	0,149*	0,234*
8	0,090	0,380*	0,310*	0,217	0,259*	0,340*

Sprachverständnis

4	0,142	0,347*	0,238*	0,395*	0,260*	0,336*
5	0,220*	0,543*	0,308*	0,474*	0,328*	0,425*
6	0,125	0,325*	0,486*	0,557*	0,356*	0,495*
7	0,248*	0,552*	0,170	0,414*	0,226*	0,366*

Wortschatz						
9	0,196*	0,603*	0,280*	0,630*	0,397*	0,561*
10	0,319*	0,552*	0,348*	0,597*	0,311*	0,534*
11	0,359*	0,674*	0,307*	0,560*	0,446*	0,583*
12	0,249*	0,472*	0,024	0,300*	0,281*	0,439*
13	0,274*	0,537*	0,255*	0,455*	0,309*	0,550*
14	0,144	0,407*	0,078	0,137	0,321*	0,438*
15	0,137	0,338*	0,188	0,303*	0,325*	0,414*
16	0,183*	0,452*	0,520*	0,618*	0,431*	0,567*
17	0,172	0,523*	0,526*	0,603*	0,432*	0,538*
18	0,415*	0,490*	0,481*	0,435*	0,458*	0,441*
19	0,336*	0,287*	0,398*	0,382*	0,426*	0,429*
Aussprache						
20	0,332*	0,177	0,120	0,385*	0,335*	0,287*
21	0,253*	0,200	0,059	0,249*	0,307*	0,217*
22	0,412*	0,003	0,198	0,165	0,230*	0,183*
23	0,348*	0,197	0,185	-0,061	0,327*	0,255*
24	0,139	0,285*	0,171	0,358*	0,326*	0,272*
25	0,284*	0,160	0,254*	0,398*	0,396*	0,345*
26	0,436*	0,032	0,131	0,273*	0,347*	0,262*
27	0,357*	0,398*	0,117	0,440*	0,429*	0,410*
28	0,441*	0,334*	0,063	0,224*	0,427*	0,406*
29	0,505*	0,405*	0,227*	0,437*	0,380*	0,325*
30	0,487*	0,308*	0,148	0,251*	0,395*	0,362*
31	0,475*	0,286*	0,175	0,358*	0,417*	0,358*
Grammatik						
32	0,452*	0,615*	0,329*	0,620*	0,389*	0,483*
33	0,353*	0,603*	0,411*	0,782*	0,486*	0,638*
34	0,332*	0,441*	0,267*	0,723*	0,394*	0,594*
35	0,250*	0,454*	0,381*	0,515*	0,396*	0,536*
36	0,183*	0,524*	0,353*	0,338*	0,410*	0,514*
37	0,327*	0,262*	0,444*	0,657*	0,460*	0,535*
38	0,223*	0,325*	0,245*	0,508*	0,295*	0,421*

39	0,325*	0,366*	0,071	0,398*	0,378*	0,480*
40	0,266*	0,478*	0,166	0,476*	0,398*	0,476*
41	0,380*	0,259*	0,485*	0,354*	0,477*	0,453*
42	0,350*	0,282*	0,446*	0,442*	0,491*	0,486*
43	0,380*	0,551*	0,378*	0,454*	0,421*	0,468*
44	0,502*	0,559*	0,470*	0,671*	0,537*	0,556*
45	0,255*	0,430*	0,344*	0,539*	0,427*	0,546*
46	0,344*	0,321*	0,522*	0,595*	0,464*	0,536*

KiSS.2: Item-Num- mern	KiSS.2-Validierungs- studie		BASS-Validierungs- studie		KiSS.3-Validierungs- studie	
	MO	ZM	MO	ZM	MO	ZM
<i>n</i>	131	180	279	352	357	528

Punktbiseriale Korrelationen einzelner Items mit dem jeweiligen KiSS.2-Untertest-score ohne dieses Item („Item vs. Untertestscore“)

Sprachverständnis

4	0,187*	0,237*	0,219*	0,370*	0,149*	0,332*
5	0,274*	0,260*	0,260*	0,395*	0,315*	0,405*
6	0,182*	0,366*	0,245*	0,419*	0,325*	0,408*

Wortschatz

7	0,392*	0,572*	0,332*	0,516*	0,157*	0,609*
8	0,168	0,562*	0,305*	0,520*	0,178*	0,580*
9	0,438*	0,468*	0,327*	0,561*	0,220*	0,666*
10	0,259*	0,341*	0,196*	0,272*	0,163*	0,387*
11	0,311*	0,551*	0,256*	0,571*	0,184*	0,602*
12	0,501*	0,505*	0,499*	0,564*	0,414*	0,520*
13	0,580*	0,498*	0,459*	0,463*	0,417*	0,521*
14	0,474*	0,431*	0,434*	0,444*	0,456*	0,500*
15	0,496*	0,407*	0,413*	0,411*	0,326*	0,382*

Aussprache

17	0,332*	0,424*	0,250*	0,230*	0,361*	0,369*
18	0,386*	0,268*	0,315*	0,323*	0,249*	0,150*
19	0,211*	0,278*	0,303*	0,288*	0,205*	0,188*

20	0,443*	0,524*	0,176*	0,387*	0,359*	0,205*
21	0,509*	0,367*	0,591*	0,426*	0,369*	0,358*
22	0,644*	0,474*	0,554*	0,491*	0,499*	0,515*
23	0,808*	0,547*	0,579*	0,513*	0,446*	0,472*
24	0,607*	0,405*	0,505*	0,438*	0,511*	0,446*
25	0,526*	0,357*	0,505*	0,438*	0,389*	0,331*
26	0,643*	0,445*	0,524*	0,507*	0,510*	0,501*
27	0,686*	0,457*	0,375*	0,476*	0,582*	0,532*
28	0,477*	0,507*	0,504*	0,472*	0,488*	0,489*
29	0,527*	0,464*	0,422*	0,474*	0,475*	0,457*

Grammatik

30	0,489*	0,639*	0,335*	0,545*	0,382*	0,497*
31	0,560*	0,723*	0,366*	0,640*	0,308*	0,674*
32	0,532*	0,635*	0,319*	0,546*	0,272*	0,511*
33	0,493*	0,415*	0,406*	0,536*	0,355*	0,531*
34	0,541*	0,688*	0,449*	0,562*	0,304*	0,477*
35	0,353*	0,373*	0,195*	0,302*	0,328*	0,339*
36	0,392*	0,477*	0,243*	0,496*	0,312*	0,404*
37	0,512*	0,440*	0,475*	0,526*	0,296*	0,462*
38	0,439*	0,473*	0,498*	0,500*	0,396*	0,448*
39	0,441*	0,538*	0,431*	0,530*	0,360*	0,467*
40	0,616*	0,615*	0,531*	0,598*	0,395*	0,620*
41	0,508*	0,611*	0,288*	0,571*	0,317*	0,629*
42	0,396*	0,532*	0,347*	0,535*	0,367*	0,605*

Nachsprechen (phonologisches Kurzzeitgedächtnis)

43	0,197*	0,311*	0,428*	0,489*	0,122*	0,232*
44	0,254*	0,432*	0,453*	0,443*	0,321*	0,312*
45	0,394*	0,456*	0,361*	0,456*	0,335*	0,283*
46	0,281*	0,359*	0,399*	0,427*	0,241*	0,332*
47	0,458*	0,351*	0,425*	0,374*	0,293*	0,206*
48	0,430*	0,328*	0,353*	0,401*	0,391*	0,332*

**Punktbiserial Korrelationen einzelner Items mit dem jeweiligen KiSS.2-Gesamtscore
ohne dieses Item („Item vs. Gesamtscore“)**

Spontansprache bzw. Sprachproduktion

1	0,135	0,247*	0,000	-0,060	0,079	0,236*
16	0,343*	0,218*	0,138	0,247*	-0,006	0,224*

Sprachverständnis

4	0,223*	0,284*	0,237*	0,425*	0,146*	0,390*
5	0,411*	0,398*	0,252*	0,505*	0,388*	0,442*
6	0,331*	0,547*	0,242*	0,487*	0,331*	0,475*

Wortschatz

7	0,397*	0,571*	0,358*	0,521*	0,234*	0,627*
8	0,181*	0,505*	0,270*	0,542*	0,185*	0,597*
9	0,566*	0,600*	0,388*	0,578*	0,214*	0,672*
10	0,207*	0,356*	0,144*	0,291*	0,166*	0,424*
11	0,377*	0,465*	0,107	0,552*	0,114*	0,601*
12	0,461*	0,424*	0,351*	0,541*	0,369*	0,492*
13	0,549*	0,377*	0,296*	0,467*	0,348*	0,473*
14	0,522*	0,434*	0,396*	0,369*	0,434*	0,494*
15	0,454*	0,406*	0,321*	0,348*	0,348*	0,370*

Aussprache

17	0,148	0,203*	0,232*	0,215*	0,254*	0,318*
18	0,328*	0,186*	0,236*	0,268*	0,166*	0,165*
19	0,042	0,095	0,181*	0,135*	0,088	0,187*
20	0,182*	0,264*	0,168*	0,234*	0,169*	0,106*
21	0,268*	0,214*	0,347*	0,264*	0,230*	0,183*
22	0,500*	0,254*	0,287*	0,379*	0,341*	0,414*
23	0,359*	0,415*	0,467*	0,396*	0,278*	0,389*
24	0,489*	0,237*	0,298*	0,483*	0,397*	0,374*
25	0,415*	0,248*	0,298*	0,483*	0,197*	0,406*
26	0,476*	0,277*	0,532*	0,473*	0,441*	0,381*
27	0,423*	0,313*	0,333*	0,340*	0,329*	0,321*
28	0,188*	0,181*	0,302*	0,276*	0,319*	0,290*
29	0,224*	0,260*	0,263*	0,289*	0,330*	0,311*

Grammatik						
30	0,509*	0,604*	0,319*	0,506*	0,382*	0,427*
31	0,542*	0,688*	0,367*	0,639*	0,254*	0,652*
32	0,493*	0,585*	0,240*	0,551*	0,234*	0,523*
33	0,496*	0,445*	0,352*	0,575*	0,370*	0,530*
34	0,531*	0,390*	0,129	0,296*	0,257*	0,341*
35	0,243*	0,488*	0,244*	0,478*	0,373*	0,428*
36	0,307*	0,677*	0,433*	0,563*	0,261*	0,498*
37	0,508*	0,411*	0,436*	0,476*	0,304*	0,407*
38	0,464*	0,492*	0,454*	0,473*	0,416*	0,439*
39	0,439*	0,524*	0,472*	0,498*	0,319*	0,430*
40	0,629*	0,620*	0,515*	0,618*	0,414*	0,615*
41	0,544*	0,663*	0,301*	0,548*	0,294*	0,637*
42	0,430*	0,529*	0,425*	0,495*	0,347*	0,591*
Nachsprechen (phonologisches Kurzzeitgedächtnis)						
43	0,322*	0,148	0,295*	0,295*	0,156*	0,207*
44	0,348*	0,411*	0,324*	0,315*	0,232*	0,266*
45	0,373*	0,356*	0,274*	0,439*	0,238*	0,163*
46	0,279*	0,403*	0,299*	0,371*	0,217*	0,358*
47	0,497*	0,467*	0,451*	0,447*	0,354*	0,374*
48	0,451*	0,430*	0,433*	0,508*	0,452*	0,533*

Tabelle E4: Eigentrennschärfe einzelner KiSS.1- und KiSS.2-Items

Anmerkungen: MO = monolingual deutschsprachige Kinder, ZM = zwei-/mehrsprachige Kinder; HSP = hessisches Sprachstandserfassungsprogramm.

^a Nähere Angaben zu einzelnen KiSS-Items befinden sich in Tab. E1 im Elektronischen Supplement 1. In Tab. E4 werden nur Item-Nummern angegeben.

^b Bei der Berechnung der Eigentrennschärfe „Item vs. Untertestscore“ wurden Items zur Spontansprache bzw. Sprachproduktion nicht berücksichtigt, da diese keinem Untertest angehören. Bei Korrelationen mit dem KiSS-Gesamtscore, d. h. bei der Bestimmung der Eigentrennschärfe „Item vs. Gesamtscore“, waren diese zwei Items hingegen mitenthalten.

^c Die Korrelation konnte nicht berechnet werden, weil alle Kinder die Frage richtig beantwortet hatten.

* $p < 0,05$ (ohne Korrekturen der p-Werte)

KiSS.1-Stichproben	KiSS.1-V		KiSS.XL-V		HSP	
	MO	ZM	MO	ZM	MO	ZM
Item vs. Untertestscore: „Sprachverständnis“	0,35 ±0,08	0,50 ±0,08	0,32 ±0,05	0,33 ±0,06	0,26 ±0,05	0,40 ±0,03
Item vs. Untertestscore: „Wort- schatz“	0,26 ±0,13	0,49 ±0,11	0,29 ±0,16	0,47 ±0,12	0,39 ±0,08	0,51 ±0,07
Item vs. Untertestscore: „Aus- sprache“	0,50 ±0,11	0,36 ±0,10	0,39 ±0,10	0,40 ±0,19	0,50 ±0,09	0,47 ±0,09
Item vs. Untertestscore: „Gram- matik“	0,36 ±0,12	0,45 ±0,12	0,35 ±0,13	0,54 ±0,12	0,44 ±0,07	0,53 ±0,05
Item vs. Gesamtscore	0,29 ±0,12	0,39 ±0,16	0,28 ±0,15	0,44 ±0,17	0,37 ±0,08	0,43 ±0,11
KiSS.2-Stichproben	KiSS.2-V		BASS-V		KiSS.3-V	
	MO	ZM	MO	ZM	MO	ZM
Item vs. Untertestscore: „Sprachverständnis“	0,21 ±0,05	0,29 ±0,07	0,24 ±0,02	0,40 ±0,03	0,26 ±0,10	0,38 ±0,04
Item vs. Untertestscore: „Wort- schatz“	0,40 ±0,13	0,48 ±0,07	0,36 ±0,10	0,48 ±0,10	0,28 ±0,12	0,53 ±0,10
Item vs. Untertestscore: „Aus- sprache“	0,52 ±0,16	0,42 ±0,09	0,43 ±0,14	0,42 ±0,09	0,42 ±0,11	0,39 ±0,13
Item vs. Untertestscore: „Gram- matik“	0,48 ±0,08	0,55 ±0,11	0,38 ±0,10	0,53 ±0,08	0,34 ±0,04	0,51 ±0,10
Item vs. Untertestscore: „Nach- sprechen“	0,34 ±0,11	0,37 ±0,06	0,40 ±0,04	0,43 ±0,04	0,28 ±0,09	0,28 ±0,05
Item vs. Gesamtscore	0,38 ±0,14	0,40 ±0,15	0,31 ±0,11	0,42 ±0,14	0,28 ±0,11	0,41 ±0,15

Tabelle E5: Eigentrennschärfe (part-whole-korrigierte punktbiseriale Korrelationen „Item vs. Untertestscore“ und „Item vs. Gesamtscore“) des „Kindersprachscreenings“ (KiSS.1 und KiSS.2): Mittelwerte und Standardabweichungen

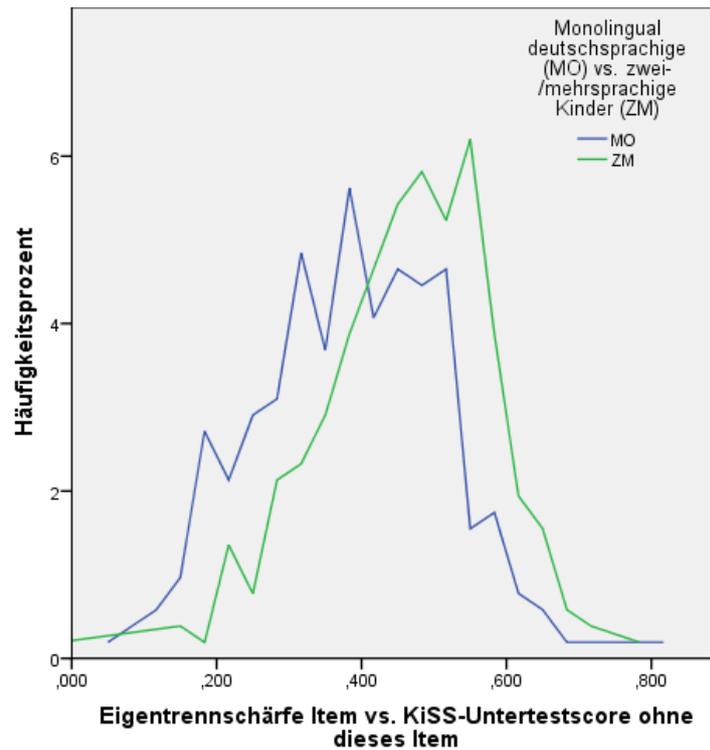
Anmerkungen: MO = monolingual deutschsprachige Kinder, ZM = zwei-/mehrsprachige Kinder; KiSS.1-V = KiSS.1-Validierungsstudie, KiSS.XL-V = KiSS.XL-Validierungsstudie, HSP = hessisches Sprachstandserfassungsprogramm, KiSS.2-V = KiSS.2-Validierungsstudie, BASS-V = BASS-Validierungsstudie (BASS = „Bochum-Aachener Stotterscreening“), KiSS.3-V = KiSS.3-Validierungsstudie.

	KiSS.2-V	HSP
Item vs. Untertestscore: „Sprachverständnis“	0,23 ±0,04	0,38 ±0,04
Item vs. Untertestscore: „Wortschatz“	0,42 ±0,10	0,50 ±0,07
Item vs. Untertestscore: „Aussprache“	0,50 ±0,13	0,47 ±0,09
Item vs. Untertestscore: „Grammatik“	0,50 ±0,08	0,52 ±0,10
Item vs. Untertestscore: „Nachsprechen“	0,35 ±0,08	0,54 ±0,17
Item vs. Gesamtscore	0,39 ±0,13	0,44 ±0,15

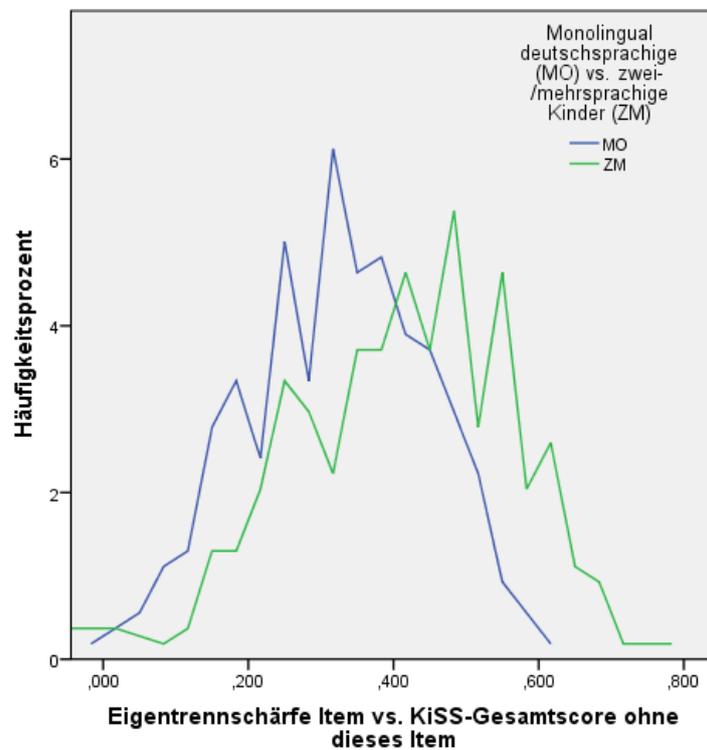
Tabelle E6: Gewichtete Angaben zur Eigentrennschärfe (part-whole-korrigierte punktbiseriale Korrelationen „Item vs. Untertestscore“ und „Item vs. Gesamtscore“) von KiSS.2-Items in der KiSS.2-Validierungsstudie (KiSS.2-V) und im hessischen Sprachstandserfassungsprogramm (HSP, laut Tomasik et al. (2020)^a), Verhältnis 1:3 von monolingual deutschsprachigen und zwei-/mehrsprachigen Kindern: Mittelwerte und Standardabweichungen

^a Als Grundlage für Tab. E6 (Spalte „HSP“) wurden eigene Berechnungen (Mittelung der Werte) mit den Daten aus der Elektronischen Anlage 1 des Artikels von Tomasik et al. (2020) herangezogen.

(a) Item vs. KiSS-Untertestscore

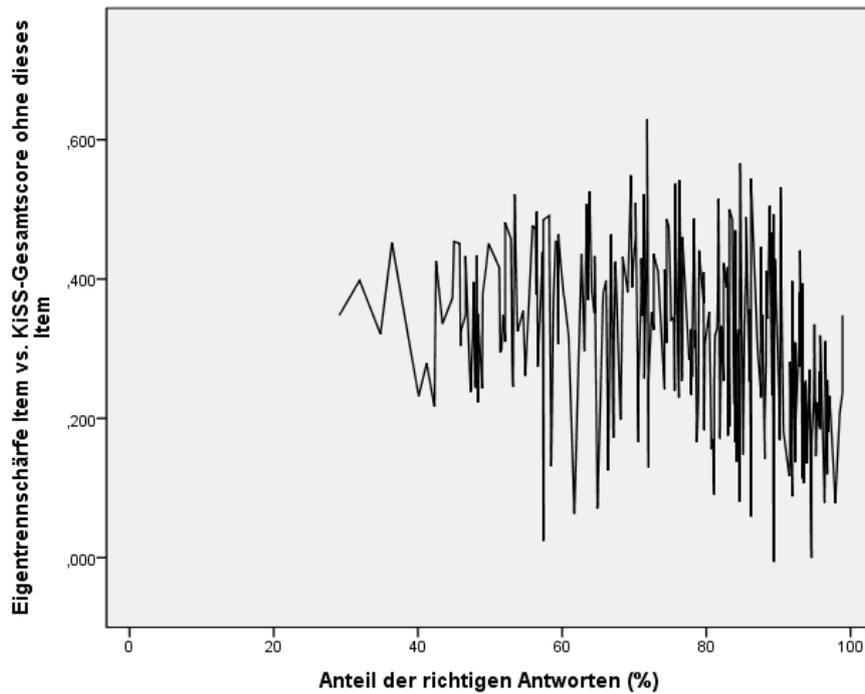


(b) Item vs. KiSS-Gesamtscore

Abbildung E2: Eigentrennschärfe von KiSS in allen KiSS.1- und KiSS.2-Stichproben^a

^a Ohne KiSS.2-Daten von Tomasik et al. (2020), da in diesen zwischen monolingual deutschsprachigen und zwei-/mehrsprachigen Kindern nicht unterschieden wird.

(a) Monolingual deutschsprachige Kinder



(b) Zwei-/mehrsprachige Kinder

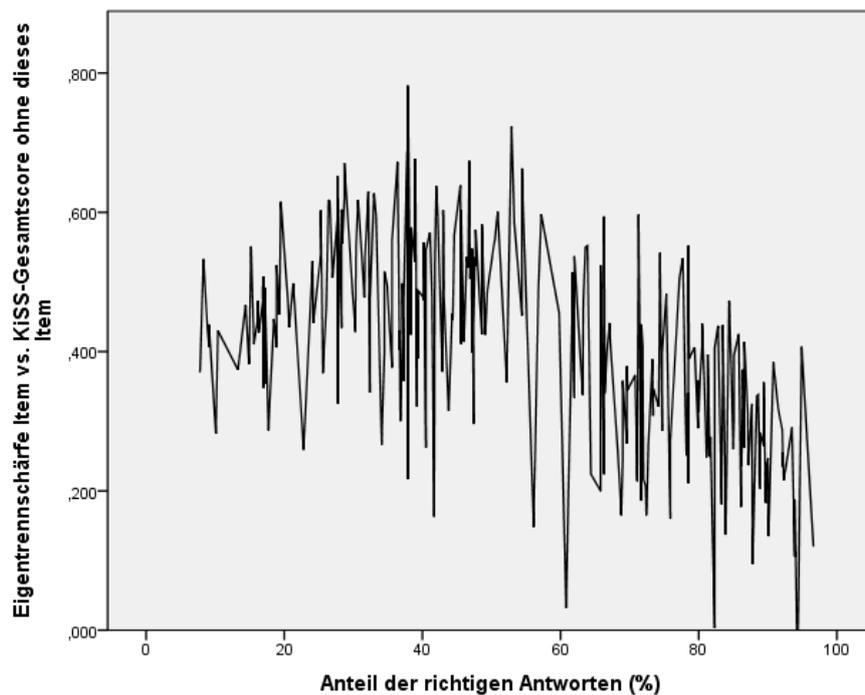


Abbildung E3: Zusammenhang zwischen der Eigentrennschärfe (Item vs. KiSS-Gesamtscore) und dem Schwierigkeitsindex in allen KiSS.1- und KiSS.2-Stichproben^a

^a Ohne KiSS.2-Daten von Tomasik et al. (2020), da in diesen zwischen monolingual deutschsprachigen und zwei-/mehrsprachigen Kindern nicht unterschieden wird.

Elektronisches Supplement 3: Interne Konsistenz von KiSS.1 und KiSS.2

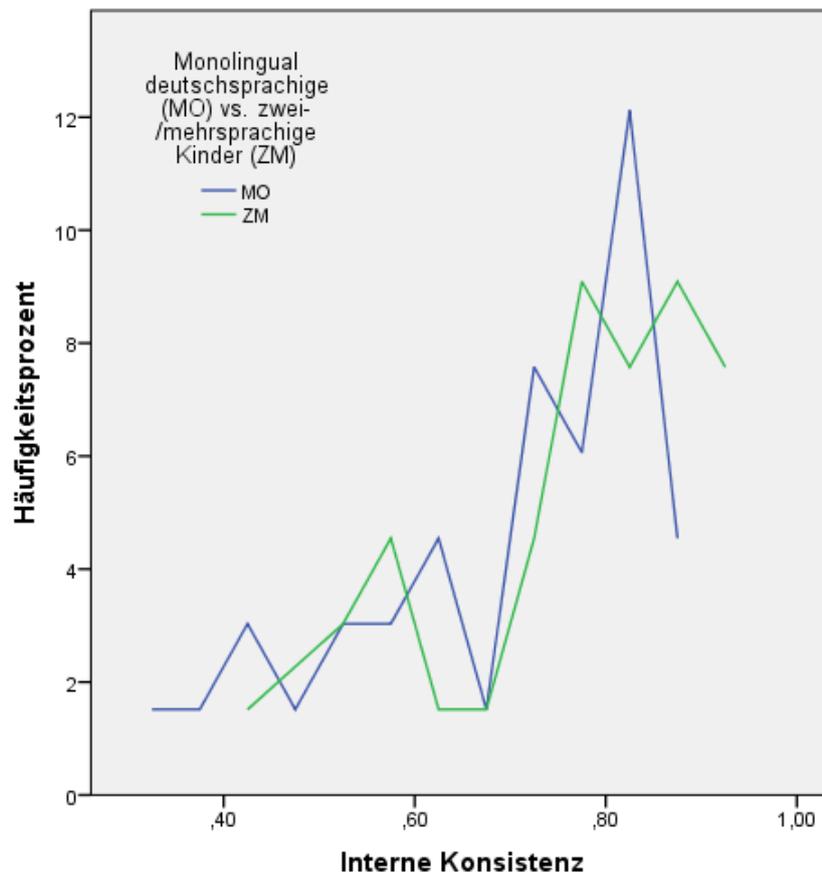


Abbildung E4: Interne Konsistenz (Cronbachs Alpha) von KiSS in allen KiSS.1- und KiSS.2-Stichproben^a

^a Ohne KiSS.2-Daten von Tomasik et al. (2020), da in diesen zwischen monolingual deutschsprachigen und zwei-/mehrsprachigen Kindern nicht unterschieden wird.

Elektronisches Supplement 4: KiSS.2-Ergebnisse der simultan und sukzessiv zwei-/mehrsprachigen Kinder (KiSS.3-Stichprobe)

KiSS.2-Items	SiZM	SuZM
<i>n</i>	216	227
Spontansprache bzw. Sprachproduktion		
1: Spontansprache	197 (91,2%)	180 (79,3%)
16: Sag mir, was ich Dir zeigen soll.	156 (72,2%)	137 (60,4%)
Sprachverständnis		
4: Junge mit einer blauen Hose	119 (88,4%)	161 (70,9%)
5: Junge neben dem Klettergerüst	99 (45,8%)	57 (25,1%)
6: die grüne Kiste unter dem Brett	99 (45,8%)	51 (22,5%)
Wortschatz		
7: Dach	120 (55,6%)	24 (10,6%)
8: Fenster	178 (82,4%)	73 (32,2%)
9: Klettergerüst	134 (62,0%)	28 (12,3%)
10: Fahrrad	140 (64,8%)	81 (35,7%)
11: Trommel	156 (72,2%)	65 (28,6%)
12: Eigenschaft: der Teddy	63 (29,2%)	11 (4,8%)
13: Eigenschaft: die Rutsche	64 (29,6%)	7 (3,1%)
14: Form: der Ball (rund)	55 (25,5%)	20 (8,8%)
15: Form: das Fenster ((vier)eckig)	27 (12,5%)	7 (3,1%)
Aussprache		
17: Dach (Ziellaut: /h/)	211 (97,7%)	195 (85,9%)
18: Bücher (Ziellaut: /č/)	157 (72,7%)	148 (65,2%)
19: Sand (Ziellaut: /z/)	206 (95,4%)	208 (91,6%)
20: Korb (Ziellaut: /k/)	207 (95,8%)	207 (91,2%)
21: Schaufel (Ziellaut: /f/)	169 (78,2%)	152 (67,0%)
22: Schmetterling (Ziellaute: /jm/)	175 (81,0%)	124 (54,6%)
23: Brille (Ziellaute: /br/)	182 (84,3%)	138 (60,8%)
24: Klettergerüst (Ziellaute: /kl/)	199 (92,1%)	184 (81,1%)
25: Klettergerüst (Ziellaute: /st/)	188 (87,0%)	163 (71,8%)
26: Flugzeuge (Ziellaute: /fl/)	196 (90,7%)	176 (77,5%)

27: Trommel (Ziellaute: /tr/)	174 (80,6%)	156 (68,7%)
28: Krone (Ziellaute: /kr/)	189 (87,5%)	167 (73,6%)
29: Krone (Ziellaut: /r/)	203 (94,0%)	182 (80,2%)
Grammatik		
30: Plural: Apfel, Äpfel	66 (30,6%)	9 (4,0%)
31: Plural: Ball, Bälle	106 (49,1%)	16 (7,0%)
32: Plural: Auto, Autos	155 (71,8%)	60 (26,4%)
33: Partizipbildung: gekrabbelt	86 (39,8%)	17 (7,5%)
34: Partizipbildung: (vor-)gelesen	106 (49,1%)	40 (17,6%)
35: 2. Person Singular: sollst, darfst	86 (39,8%)	52 (22,9%)
36: 2. Person Singular: riechst	116 (53,7%)	53 (23,3%)
37: auf das Dach	37 (17,1%)	4 (1,8%)
38: durch das Rohr	34 (15,7%)	4 (1,8%)
39: aus dem Sandkasten	39 (18,1%)	9 (4,0%)
40: auf der Wippe	73 (33,8%)	10 (4,4%)
41: ...die Kinder reinkommen können	133 (61,6%)	43 (18,9%)
42: ...die Kinder bauen können	98 (45,4%)	21 (9,3%)
Nachsprechen von Kunstwörtern		
43: Triser	159 (73,6%)	153 (67,4%)
44: Misküranok	88 (40,7%)	62 (27,3%)
45: Nabolira	91 (42,1%)	90 (39,6%)
46: Verklasaft	100 (46,3%)	69 (30,4%)
Nachsprechen von Sätzen (maximale Punktzahl pro Item)		
47: Das braune Pony rennt über...	48 (22,2%)	6 (2,6%)
48: Bevor Lena abgeholt wird, muss...	33 (15,3%)	1 (0,4%)
Extremwerte (Schwierigkeit 0–10, 90–100%)	7	16
Mittlere Werte (Schwierigkeit 45–55%)	7	1

**Tabelle E7: Schwierigkeitsgrade einzelner KiSS.2-Items in der KiSS.3-Studie
(% der richtigen Antworten)**

Anmerkungen: SiZM = simultan zwei-/mehrsprachige Kinder; SuZM = sukzessiv zwei-/mehrsprachige Kinder.

KiSS.2-Items	Eigentrennschärfe „Item vs. Untertestscore“		Eigentrennschärfe „Item vs. Gesamtscore“	
	SiZM	SuZM	SiZM	SuZM
<i>n</i>	216	227	216	227
Spontansprache bzw. Sprachproduktion^a				
1: Spontansprache	—	—	0,089	0,288*
16: Sag mir, was ich Dir zeigen soll.	—	—	0,243*	0,211*
Sprachverständnis				
4: Junge mit einer blauen Hose	0,297*	0,271*	0,311*	0,280*
5: Junge neben dem Klettergerüst	0,417*	0,287*	0,422*	0,314*
6: die grüne Kiste unter dem Brett	0,380*	0,336*	0,452*	0,348*
Wortschatz				
7: Dach	0,481*	0,377*	0,548*	0,347*
8: Fenster	0,443*	0,432*	0,442*	0,399*
9: Klettergerüst	0,543*	0,536*	0,605*	0,406*
10: Fahrrad	0,247*	0,306*	0,286*	0,329*
11: Trommel	0,518*	0,480*	0,507*	0,493*
12: Eigenschaft: der Teddy	0,470*	0,327*	0,411*	0,253*
13: Eigenschaft: die Rutsche	0,416*	0,327*	0,347*	0,163*
14: Form: der Ball (rund)	0,447*	0,439*	0,439*	0,359*
15: Form: das Fenster ((vier)eckig)	0,353*	0,277*	0,414*	0,197*
Aussprache				
17: Dach (Ziellaut: /h/)	0,093	0,401*	0,127	0,300*
18: Bücher (Ziellaut: /č/)	0,149*	0,093	0,139*	0,169*
19: Sand (Ziellaut: /z/)	0,194*	0,173*	0,162*	0,138*
20: Korb (Ziellaut: /k/)	0,162*	0,207*	0,039	0,109
21: Schaufel (Ziellaut: /f/)	0,349*	0,317*	0,026	0,158*
22: Schmetterling (Ziellaute: /jm/)	0,491*	0,443*	0,122	0,439*
23: Brille (Ziellaute: /br/)	0,472*	0,421*	0,243*	0,359*
24: Klettergerüst (Ziellaute: /kl/)	0,332*	0,452*	0,308*	0,313*
25: Klettergerüst (Ziellaute: /st/)	0,341*	0,259*	0,367*	0,402*
26: Flugzeug (Ziellaute: /fl/)	0,377*	0,467*	0,284*	0,331*
27: Trommel (Ziellaute: /tr/)	0,463*	0,463*	0,256*	0,285*

28: Krone (Ziellaute: /kr/)	0,471*	0,502*	0,150*	0,285*
29: Krone (Ziellaut: /r/)	0,539*	0,540*	0,148*	0,307*
Grammatik				
30: Plural: Apfel, Äpfel	0,439*	0,247*	0,376*	0,090*
31: Plural: Ball, Bälle	0,574*	0,543*	0,607*	0,369*
32: Plural: Auto, Autos	0,388*	0,354*	0,378*	0,279*
33: Partizipbildung: gekrabbelt	0,389*	0,342*	0,423*	0,285*
34: Partizipbildung: (vor-)gelesen	0,379*	0,479*	0,401*	0,449*
35: 2. Person Singular: sollst, darfst	0,230*	0,405*	0,204*	0,326*
36: 2. Person Singular: riechst	0,366*	0,301*	0,334*	0,327*
37: auf das Dach	0,403*	0,430*	0,366*	0,221*
38: durch das Rohr	0,408*	0,288*	0,441*	0,119
39: aus dem Sandkasten	0,433*	0,393*	0,459*	0,244*
40: auf der Wippe	0,560*	0,440*	0,649*	0,310*
41: ...die Kinder reinkommen können	0,535*	0,519*	0,547*	0,546*
42: ...die Kinder bauen können	0,537*	0,463*	0,537*	0,376*
Nachsprechen von Kunstwörtern				
43: Triser	0,225*	0,324*	0,152*	0,228*
44: Misküranok	0,288*	0,302*	0,263*	0,243*
45: Nabolira	0,277*	0,284*	0,220*	0,104
46: Verklasaft	0,269*	0,344*	0,250*	0,409*

Tabelle E8: Eigentrennschärfe einzelner KiSS.2-Items aus der KiSS.3-Studie

Anmerkungen: SiZM = simultan zwei-/mehrsprachige Kinder; SuZM = sukzessiv zwei-/mehrsprachige Kinder.

^a Bei der Berechnung der Eigentrennschärfe „Item vs. Untertestscore“ wurden Items zur Spontansprache bzw. Sprachproduktion nicht berücksichtigt, da diese keinem Untertest angehören. Bei Korrelationen mit dem KiSS.2-Gesamtscore, d. h. bei der Bestimmung der Eigentrennschärfe „Item vs. Gesamtscore“, waren diese zwei Items hingegen mitenthalten.

* $p < 0,05$ (ohne Korrekturen der p-Werte)