

Erfassung der Erzählfähigkeit mehrsprachig aufwachsender Vorschulkinder

Überprüfung der Testgütekriterien des Verfahrens MAIN (Multilingual Assessment Instrument for Narratives)

Assessing narrative abilities of bilingual children – Psychometric evaluation of MAIN (Multilingual Assessment Instrument for Narratives)

Schlüsselwörter: Erzählfähigkeit, Mehrsprachigkeit, MAIN, Testgütekriterien

Keywords: Narrative abilities, bilingualism, MAIN, psychometric properties

Zusammenfassung: Erzählfähigkeit wird als vielversprechender Untersuchungsbereich für die Erfassung sprachlicher Fähigkeiten bei mehrsprachigen Kindern betrachtet (Paradis et al., 2010). Speziell für Teilfähigkeiten auf der sog. „Makrostrukturebene“ werden eine geringere Abhängigkeit vom Kontakt zur Einzelsprache und Übertragungseffekte aus der stärker entwickelten Sprache angenommen. Vor dem Hintergrund dieser Annahmen wurde das Verfahren MAIN (Multilingual Assessment Instrument for Narratives; Gagarina et al., 2012) entwickelt, um die Erzählfähigkeit drei- bis zehnjähriger mehrsprachiger Kinder zu erfassen. Das Verfahren und dessen Entwicklung wurden in dem Impulsbeitrag „Multilingual Assessment Instrument for Narratives (MAIN) – Ein Verfahren zur Erfassung der Erzählkompetenz von Kindern“ von Gagarina (in der letzten Ausgabe (29/1)) näher beschrieben. Als Erzählstimulus dienen bei MAIN

verschiedene Bildergeschichten mit parallelisiertem Handlungsverlauf, anhand derer mehrsprachig aufwachsende Kinder in jeweils beiden Sprachen getestet werden können. Da für MAIN bislang noch keine umfassende Überprüfung der Testgüte vorliegt, wurden anhand von Erzählraten aus dem Forschungsprojekt IMKi (Effekte einer aktiven Integration von Mehrsprachigkeit in Kindertageseinrichtungen) erste Untersuchungen zur Auswertungsobjektivität, Retest- und Paralleltest-Reliabilität sowie Konstruktvalidität durchgeführt. Die Ergebnisse weisen darauf hin, dass das Verfahren nach intensiver Einarbeitung in das Auswertungsvorgehen als ausreichend objektiv und reliabel eingeschätzt werden kann. Bei der Überprüfung der Reliabilität der Paralleltestversionen zeigte sich allerdings ein deutlicher Reliabilitätsverlust. Bezüglich der Validität ergeben sich Hinweise darauf, dass MAIN eine Fähigkeit erfasst, die einen deutlichen

Zusammenhang zu expressiven Wortschatzleistungen aufweist, jedoch von diesen abgrenzbar ist. Zusammenfassend weisen die Ergebnisse darauf hin, dass sich MAIN für den Einsatz im Forschungskontext eignet, wobei bei Anwendung der beiden Paralleltestformen die geringe Paralleltest-Reliabilität beachtet werden muss. Als zukünftiges Forschungsvorgehen im Projekt ist geplant, Erzählfähigkeit auf ihr Potenzial zur Vorhersage der sprachlichen Entwicklung drei- bis sechsjähriger mehrsprachig aufwachsender Kinder zu untersuchen.

Abstract: The assessment of narrative abilities is discussed to be an adequate method to gain insights into bilingual children's language abilities (Paradis et al., 2010). Especially abilities on the macro structure level are assumed to be less dependent on the amount of input a bilingual child has received in each language and might be transferred from

the child's more advanced language to the less advanced one. Based on these assumptions, the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012) was developed to assess the narrative abilities of three- to ten-year-old bilingual children. MAIN has recently been used in a number of studies focusing on bilingual children's narrative abilities. However, the psychometric properties of the instrument have not been examined yet. Therefore, a psychometric evaluation analysing the objectivity, reliability and validity of MAIN was conducted as part of the research project IMKi (The effects of actively integrating multilingualism into daycare facilities). Results indicate that the instrument is applicable for research purposes as it shows adequate objectivity and reliability. However, when using the two parallel test forms a considerable drop in reliability was observed and must therefore be considered. To ensure objectivity of analysis, intensive training of the testers has to be taken into account. MAIN seems to assess an ability which is closely associated with but not identical to expressive vocabulary. For future research it is planned to examine narrative abilities assessed by MAIN regarding their predictive value for bilingual language development.

KURZBIOGRAFIE

Tamara Lautenschläger studierte Sonderpädagogik mit erster sonderpädagogischer Fachrichtung Sprache an der Pädagogischen Hochschule Heidelberg. Seit Juni 2018 arbeitet sie dort als wissenschaftliche Mitarbeiterin im Forschungsprojekt „IMKi“ (Effekte einer aktiven Integration von Mehrsprachigkeit in Kindertagesstätten). Im Rahmen ihrer Projektstelle promoviert sie zur sprachlichen Entwicklung mehrsprachig aufwachsender Kinder im Alter von drei bis sechs Jahren. Als Elternzeitvertretung lehrt sie seit dem Sommersemester 2019 im Bereich der Sprachbehindertenpädagogik.

Einleitung

Aktuell hat in Deutschland etwa jedes dritte Kind unter sechs Jahren einen Migrationshintergrund (38%; Autorengruppe Bildungsberichterstattung, 2018). Viele dieser Kinder erleben in ihren ersten Lebensjahren alltäglich Kommunikationssituationen, in denen mehrere Sprachen Verwendung finden, und wachsen somit mehrsprachig auf (Reich, 2010). Die Sprachentwicklungsbedingungen und -verläufe dieser Kinder sind sehr heterogen, wodurch die zuverlässige Einschätzung ihrer sprachlichen Fähigkeiten erschwert wird. Geeignete diagnostische Möglichkeiten zu identifizieren ist jedoch entscheidend, um Aussagen dahingehend treffen zu können, wie eine Vorhersage der weiteren (mehr-)sprachigen Entwicklung gelingen kann oder wie Sprachentwicklungsstörungen von typischen Phänomenen des mehrsprachigen Erwerbs unterschieden werden können. Erzählfähigkeit wird diesbezüglich als ein vielversprechender Untersuchungsbereich diskutiert, da kindliche Erzählungen die Analyse unterschiedlicher sprachlicher Fähigkeiten ermöglichen und zudem für einzelne Analyseebenen eine geringe Abhängigkeit von individuellen Inputbedingungen angenommen wird (Gagarina et al., 2012). Um Erzählfähigkeit erfassen zu können, sind qualitativ hochwertige Testverfahren erforderlich. Die Qualität eines Verfahrens bemisst sich neben der Darlegung der jeweiligen theoretischen Fundierung des Messgegenstandes auch an der AnwenderUnabhängigkeit, Stabilität und Gültigkeit der erfolgenden Messung und damit daran, inwieweit das jeweilige Testverfahren die psychometrischen Hauptgütekriterien Objektivität, Reliabilität und Validität erfüllt. Für das Verfahren MAIN (Multilingual Assessment Instrument for Narratives; Gagarina et al., 2012), welches speziell für die Erfassung der Erzählfähigkeit bei mehrsprachigen Kindern entwickelt wurde, liegt nach unserem Kenntnisstand bislang keine umfassende Untersuchung der Hauptgütekriterien vor. Es sollen hier daher Ergebnisse aus einer Untersuchung zur Objektivität, Reliabilität und Validität des Verfahrens vorgestellt werden.

Theoretischer Hintergrund

Der vorliegende Artikel bezieht sich auf die Erfassung von Erzählfähigkeit über das monologische Erzählen zu einer fiktiven Bildergeschichte. Neben sprachlichen Fähigkeiten auf lexikalischer, morpho-syntaktischer und pragmatischer Ebene sind hierfür auch nichtsprachliche kognitive Fähigkeiten, wie das Erkennen zeitlicher, räumlicher und kausaler Zusammenhänge zwischen einzelnen Ereignissen sowie Fähigkeiten aus dem Bereich der „Theory of mind“ erforderlich (Kapica et al., 2014; Pfeffer, 2015). Das Konzept der „Theory of mind“ bezieht sich auf die kognitive Fähigkeit, sich selbst und Anderen mentale Zustände wie Gedanken, Emotionen oder Intentionen zuschreiben zu können (Goldman, 2012). Dies beinhaltet etwa das gedankliche Hineinversetzen in die Situation von Anderen, welches beim Erzählen zu einer Bildvorlage bspw. erforderlich ist, um die Intentionen von ProtagonistInnen nachvollziehen und versprachlichen zu können (Gagarina et al., 2015).

Die Analyse kindlicher Erzählungen ermöglicht folglich die Untersuchung einer Reihe unterschiedlicher Fähigkeiten anhand einer einzelnen Aufgabe. Dabei besteht jedoch keine Einigkeit dahingehend, welche Teilfähigkeiten das Konstrukt „Erzählfähigkeit“ konstituieren, sodass die jeweils analysierten Bereiche je nach Verfahren variieren. Einen verbreiteten Analyserahmen stellt die Differenzierung zwischen Makro- und Mikrostrukturebene dar. Während auf Mikrostrukturebene die verwendeten lexikalischen und grammatischen Mittel, z.B. die Anzahl der Wörter oder die syntaktische Komplexität, untersucht werden, zielt die Analyse der Makrostrukturebene auf die Abbildung von Fähigkeiten, die der gelingenden Strukturierung der Geschichte zugrunde liegen (Licandro, 2016; Ringmann, 2013). Die Basis der Makrostrukturanalyse bildet häufig der „Story-Grammar-Ansatz“ (Gagarina et al., 2015). Dessen Ausgangspunkt besteht in der Annahme, dass Erzählungen unabhängig von individuellen inhaltlichen Details eine identische Grundstruktur aufweisen, auf die Erzählende beim Erzählen implizit zurückgreifen (Stein & Glenn, 1975). Insbesondere für Problemlösegeschichten

wird hierbei angenommen, dass sie aus einem Setting und einer oder mehreren Episoden aufgebaut sind (Licandro, 2016). Eine Episode entspricht in etwa einem Handlungsstrang und ist aus einer festen Abfolge zentraler Informationskategorien, sogenannter Episodenstrukturelemente, aufgebaut (z. B. Ziel des Protagonisten – Lösungsversuch – Resultat; Mandler & Johnson, 1977). Innerhalb des Story-Grammar-Ansatzes existieren verschiedene Modelle solcher Grundstrukturen für den Aufbau einer Episode, die sich in der Anzahl der angenommenen Strukturelemente unterscheiden. Gemeinsam ist den Ansätzen jedoch die Annahme, dass sich über die Analyse der Struktur der durch das Kind versprachlichten Erzählung Rückschlüsse auf dessen implizites Wissen zur Geschichtenstruktur ableiten lassen (Schröder, 2009).

Erzählfähigkeit bei Mehrsprachigkeit

Die Erfassung von Erzählfähigkeit wird als vielversprechender Untersuchungsbereich in der Sprachentwicklungsdiagnostik bei mehrsprachigen Kindern eingeschätzt. Speziell Teilfähigkeiten auf Makrostrukturebene könnten für die Einschätzung des sprachlichen Entwick-

lungsstandes bzw. -potenzials besonders geeignet sein, da angenommen wird, dass diese weniger stark von der Inputmenge in der Einzelsprache abhängig sind als die Größe des Wortschatzes oder der Erwerb des grammatischen Regelsystems. Zudem werden Übertragungseffekte von der stärker auf die schwächer entwickelte Sprache diskutiert (Boerma et al., 2016). Einige Untersuchungen weisen in Übereinstimmung mit dieser Annahme darauf hin, dass mehrsprachige Kinder auf Makrostrukturebene vergleichbare Leistungen zeigen wie monolinguale Kinder (z. B. Boerma et al., 2016; Kunnari et al., 2016) bzw. in ihren beiden Sprachen ähnliche Ergebnisse erzielen, auch wenn eine der beiden Sprachen stärker entwickelt ist (Rodina, 2017; Kunnari et al., 2016). Diese Tendenzen konnten jedoch nicht in allen Untersuchungen, in denen solche Vergleiche angestellt wurden, gezeigt werden (z. B. Rezzonico et al., 2016). Untersucht wird Erzählfähigkeit zudem hinsichtlich ihrer Eignung zur Differenzierung zwischen mehrsprachigen Kindern mit typischem Spracherwerbsverlauf und -entwicklungsstörungen. Erste Studienergebnisse weisen darauf hin, dass über die Untersuchung der Erzählfähigkeit auf Makrostrukturebene Hinweise auf Sprachentwicklungsstörungen bei mehrsprachigen Kindern gewonnen werden könnten (Boerma et al., 2016).

Ein Verfahren zu Erfassung der Erzählfähigkeit – MAIN

Das „Multilingual Assessment Instrument for Narratives“ (MAIN; Gagarina et al., 2012) wurde speziell zur Erfassung der Erzählfähigkeit drei- bis zehnjähriger mehrsprachig aufwachsender Kinder konzipiert. Das Verfahren wurde sowohl als Forschungsinstrument als auch für die Nutzung im klinischen Kontext entwickelt.

Anhand von MAIN können Erzählungen in den Erzählmodi „Erzählen“, „Nach-erzählen“ oder „Erzählen nach Hören einer Modellgeschichte“ elizitiert und hinsichtlich mehrerer Teilaspekte auf Makro- und Mikrostrukturebene analysiert werden. Neben produktiven Erzählleistungen kann das Verständnis für die Geschichtenstruktur erfasst werden (Skerra et al., 2012). Kern des Verfahrens

bilden vier jeweils sechsteilige Bilder-geschichten (Download unter <https://www.leibniz-zas.de/de/publications/schriftenreihe/zaspil/zaspil-56/main-start/>). Für die Erhebung der Erzählleistung ohne vorheriges Modell sind die Geschichten-vorlagen „Baby Birds“ und „Baby Goats“ vorgesehen. In beiden Geschichten wird ein Tierjunges (Vogelküken bzw. Ziegenjunges) durch ein Raubtier (Katze bzw. Fuchs) angegriffen, jedoch von einem weiteren Tier (Hund bzw. Rabe) vor dem Angriff gerettet. Durch diese inhaltliche Parallelisierung der Bildvorlagen soll es ermöglicht werden, mehrsprachige Kinder in ihren beiden Sprachen testen und die jeweiligen Ergebnisse miteinander vergleichen zu können (Gagarina et al., 2012).

Das Bildmaterial bei MAIN ist so konstruiert, dass alle Geschichtenvorlagen drei ineinander verwobene Episoden enthalten. Wie sich aus der unten aufgeführten Beispielerzählung bereits erkennen lässt, sind diese in der Vorlage „Baby birds“ bspw.:

1. Mutter holt Futter für ihre hungrigen Küken.
2. Eine hungrige Katze klettert auf den Baum und schnappt eines der Küken.
3. Ein Hund zieht die Katze vom Baum und rettet damit die Küken vor der Katze.

Beispielerzählung „Baby Birds“ (Junge, 6;1 J.)

„Die Babyvögel haben Hunger. Da sehen die Babyvögel plötzlich eine Katze. Und die Mama fliegt weg. Die Katze klettert auf den ... hm der Vogel kam zurück und gibt den Babyvögeln zum Essen. Die Katze kletterte nach oben. Die Katze war oben und wollte den Vogel und wollt ... hm ... und die Mamavogel hat den andere Vogel gefüttert. Und der ... und die Katze ist nach oben geklettert und wollte den anderen Vogel essen. Der Hund hat gesehen, dass dort die Katze ist. Die Katze ... hm der Hund ... hat die unten ... die der Hund hat die Katze an Schwanz gebissen und hat sie nach unten gezogen. Hat sie dann der Hund hat sie dann weggejagt.“

Der Schwerpunkt der Auswertung liegt auf der Makrostruktur, da für diese die Vergleichbarkeit über verschiedene Sprachen hinweg angenommen wird (ebd.).

KURZBIOGRAFIE

Katja Schneller studierte Logopädie (B.Sc.) an der Hochschule Fresenius und Experimental Clinical Linguistics (M.Sc.) an der Universität Potsdam. Sie arbeitet an der Pädagogischen Hochschule Heidelberg als Dozentin am Institut für Sonderpädagogik im Bereich Sprachbehindertenpädagogik sowie als wissenschaftliche Mitarbeiterin im Forschungsprojekt „IMKi“ (Effekte einer aktiven Integration von Mehrsprachigkeit in Kindertagesstätten) am Institut für Psychologie. Seit 2009 ist sie als akademische Sprachtherapeutin tätig. Im Rahmen ihrer Promotion interessiert sie sich für Sprachentwicklung und deren Störungen im Kontext von Mehrsprachigkeit.

Bewertet werden die Anzahl der Episodenstrukturelemente, die Episodenkomplexität sowie die Gesamtzahl der „Internal State Terms“ (IST). Letztere bezeichnen Ausdrücke, die innere Zustände von ProtagonistInnen (z.B. Gedanken oder Emotionen) wiedergeben. Basis für die Erfassung der Episodenstrukturelemente bildet eine adaptierte Version des „Story-Grammar-Modells“ nach Glenn und Stein (1975). Gemäß diesem adaptierten Modell besteht jede Episode aus den fünf Strukturelementen „initialer innerer Zustand eines Protagonisten“, „Ziel“, „Lösungsversuch“, „Resultat des Lösungsversuchs“ und „innerer Zustand, der auf das Resultat des Lösungsversuchs folgt“. Jedes dieser fünf Strukturelemente kann in jeder der drei Episoden jeweils einmal kodiert werden, sodass insgesamt maximal 15 Punkte für das Vorhandensein der relevanten Episodenstrukturelemente vergeben werden können. Zusätzlich kann jeweils ein Punkt für die Nennung des Ortes bzw. der Zeit des Geschehens vergeben werden (Setting). Insgesamt ergibt sich dadurch eine Maximalpunktzahl von 17 (Gagarina et al., 2012). Für jedes Strukturelement sind bei MAIN konkrete inhaltliche Vorgaben aufgeführt, anhand derer beurteilt werden soll, ob ein Kind in seiner Erzählung jeweils die entsprechende bzw. eine als äquivalent einzustufende Äußerung produziert hat. In Tabelle 1 sind die erfassten Episodenstrukturelemente mit entsprechenden Äußerungsbeispielen für die Episode „Katze“ aus der Geschichtenvorlage „Baby Birds“ aufgeführt.

MAIN wurde im Zeitraum von 2011 bis 2012 an insgesamt 306 bilingualen Kindern zwischen drei und neun Jahren für 17 Sprachen und 14 Sprachenkombinationen pilotiert (Gagarina et al., 2015). Das Verfahren wurde seitdem in mehreren internationalen Studien als Forschungsinstrument für die Untersuchung der Erzählfähigkeit bei mehrsprachigen Kindern mit verschiedenen Sprachenkombinationen genutzt (z. B. Bohnacker, 2016; Kunnari et al., 2016; Maviş et al., 2016). Eine Normierung sowie eine offizielle Überprüfung der Testgütekriterien liegen bislang noch nicht vor.

Element	Beschreibung	Beispiele aus MAIN („Baby birds“, Episode Katze)
IST initial	innerer Zustand, der eine Handlung initiiert	Die Katze war hungrig.
GOAL	Ziel der ProtagonistInnen in Form einer Idee, wie das entstandene Problem gelöst werden könnte	Die Katze wollte die Vögel fressen.
ATTEMPT	Handlung, um das Ziel zu erreichen	Die Katze kletterte den Baum hoch.
OUTCOME	Ereignisse, die kausal auf den Lösungsversuch folgen	Die Katze fing (fast) einen Vogel.
IST Reaktion	Gefühle der ProtagonistInnen in Bezug auf das Ergebnis des Lösungsversuchs	Die Katze war glücklich.

Tabelle 1 **Durch MAIN erfasste Episodenstrukturelemente**

Zentrale Testgütekriterien

Testgütekriterien dienen als Qualitätsmaße für psychologische Testverfahren. Die Hauptgütekriterien ermöglichen Aussagen hinsichtlich der Anwenderunabhängigkeit (Objektivität), Messpräzision (Reliabilität) und Gültigkeit (Validität) von Testverfahren und stellen somit einen zentralen Aspekt für deren Beurteilung dar.

(Auswertungs-)Objektivität

Die Objektivität eines Verfahrens gibt an, inwieweit das zu messende Merkmal unabhängig von den jeweiligen Untersuchenden gemessen wird. Unterschieden wird zwischen Durchführungs-, Auswertungs-, und Interpretationsobjektivität (Bühner, 2006).

Um die Objektivität der Auswertung eines Verfahrens zu gewährleisten, muss sichergestellt sein, dass sich die Vergabe von Punkten oder Leistungswerten für eine erfasste Testleistung nicht zwischen verschiedenen Auswertenden unterscheidet, indem etwa genaue Regeln festgelegt sind, wann eine Antwort als richtig gilt und wie viele Punkte für welche Antwort vergeben werden (Moosburger & Kelava, 2012).

(Retest- und Paralleltest-) Reliabilität

Die Reliabilität gibt die Messgenauigkeit eines Tests an und erlaubt somit Aussagen über die Größe des Einflusses von Messfehlern (Moosburger & Kelava, 2012). Die Reliabilität kann u. a. darüber bestimmt werden, inwieweit der- bzw.

dieselbe ProbandIn zu zwei unterschiedlichen Zeitpunkten (Retest-Reliabilität) oder in zwei parallelen Versionen desselben Testes (Paralleltest-Reliabilität) jeweils vergleichbare Werte erzielt. Da verschiedene Merkmale unterschiedlich stabil erfasst werden können, sind pauschale Aussagen hinsichtlich der erforderlichen Höhe des Reliabilitätskoeffizienten allerdings nicht möglich. Bei einem Reliabilitätskoeffizienten unter 0.7 kann jedoch nicht mehr von einem reliablen Test ausgegangen werden (Moosburger & Kelava, 2012; Schermelleh-Engel & Werner, 2012).

(Konstrukt-)Validität

Die Validität gibt das Ausmaß an, mit dem ein Test das Merkmal, das er zu messen vorgibt, tatsächlich misst (Bühner, 2006). Da der direkte Vergleich zwischen erfassten Testwerten und der realen Ausprägung des Merkmals anwendungspraktisch nicht möglich ist, kann u. a. auf die Bestimmung der Konstruktvalidität zurückgegriffen werden. Hierbei wird überprüft, inwiefern die Ergebnisse des Testverfahrens Zusammenhänge mit Ergebnissen aus solchen mit ähnlichem Messanspruch aufweisen (konvergente Validität) und inwieweit sich Unterschiede zu Verfahren mit anderem Messanspruch abbilden lassen (divergente Validität). Allerdings bestehen keine Richtwerte dahingehend, welche Höhe die ermittelten Korrelationen jeweils haben müssen bzw. maximal haben dürfen, um von einer zufriedenstellenden Konstruktvalidität sprechen zu können (Bühner, 2006).

Testgütekriterien bei MAIN

Eine umfassende Überprüfung der Testgütekriterien liegt nach unserem Kenntnisstand für MAIN aktuell noch nicht vor. Aus Studien können jedoch erste Hinweise zur Testgüte gewonnen werden. Für einen ersten Eindruck bezüglich der Objektivität kann die für einige der relevanten Studien ermittelte Beurteilungsübereinstimmung betrachtet werden. Diese wird in den Studien zwar als Maß für die Messgenauigkeit genutzt, gleichzeitig weisen die Ergebnisse der Überprüfungen jedoch auch darauf hin, in welchem Ausmaß jeweils eine objektive Auswertung möglich war. Die für die Überprüfung herangezogenen methodischen Vorgehensweisen sind sehr unterschiedlich. Alle berichteten Übereinstimmungen werden aber von den jeweiligen AutorInnen als zufriedenstellend angegeben (Boerma et al., 2016; Bohnacker, 2016; Kunnari et al., 2016; Maviş et al., 2016).

Eine Überprüfung der Retest- oder Paralleltest-Reliabilität erscheint für MAIN bisher noch nicht erfolgt zu sein. Für zwei Studien wurde die Vergleichbarkeit der beiden Parallelgeschichten in der gleichen Sprache überprüft, wobei allerdings der Vergleich zwischen zwei verschiedenen Kindergruppen stattfand. Hier ergaben sich zwischen den Untersuchungsgruppen keine Mittelwertunterschiede in Abhängigkeit von der erzählten Geschichtenversion (Bohnacker, 2016; Boerma et al., 2016). Bislang scheint allerdings nicht überprüft worden zu sein, inwieweit sich innerhalb derselben Kindergruppe in derselben Sprache vergleichbare Leistungen für die beiden Geschichtenversionen erfassen lassen.

Als ein Indikator für die Validität von MAIN wurde in einer Studie von Boerma et al. (2016) die diagnostische Validität des Verfahrens überprüft. Die Ergebnisse weisen auf eine unzureichende Sensitivität bei der Identifikation bilingualer Kinder mit Sprachentwicklungsstörungen hin (79%). Durch eine von den AutorInnen vorgenommene Adaption der Bewertung (höhere Berücksichtigung der Internal State Terms) ließ sich eine verbesserte diagnostische Validität erreichen (Boerma et al., 2016).

Forschungsprojekt IMKi

Die vorliegende Untersuchung wurde im Rahmen des Forschungsprojektes IMKi (Effekte einer aktiven Integration von Mehrsprachigkeit in Kindertageseinrichtungen) durchgeführt. Innerhalb des Projektes sollen anhand eines längsschnittlichen Designs die Bedingungen für eine gelingende sprachliche Entwicklung mehrsprachig aufwachsender Vorschulkinder im Kontext der Kindertagesstätte untersucht werden. Hierfür werden in jährlich stattfindenden Sprachstandserhebungen alle teilnehmenden Kinder hinsichtlich ihrer Sprachfähigkeiten im Deutschen getestet. Kinder mit türkischer bzw. russischer Herkunftssprache werden zusätzlich im Türkischen bzw. Russischen überprüft. Für alle untersuchten Sprachen werden der rezeptive und expressive Wortschatz, die expressiven Grammatikleistungen sowie die Erzählfähigkeit erfasst. Die Testungen werden in einer Einzelsituation innerhalb der jeweiligen Kindertagesstätte durch vorab geschulte TestleiterInnen durchgeführt und per Audio- und Videoaufnahme aufgezeichnet. Für alle teilnehmenden Kinder liegen vonseiten der Erziehungsberechtigten entsprechende Einverständniserklärungen zur Erfassung und Auswertung der Sprachdaten vor. Das Studiendesign wurde durch ein positives Votum der Ethikkommission der zuständigen Einrichtung bestätigt.

Verwendung von MAIN

MAIN wurde für den Einsatz im Rahmen des Projektes gewählt, da das Verfahren explizit für die Erfassung und den Vergleich der Erzählfähigkeit in beiden Sprachen mehrsprachiger Kinder entwickelt wurde. Erhoben wurde das freie Erzählen zu einer Bildergeschichte ohne vorangehendes Modell. Die Elizitation der Erzählung erfolgte für die Erhebung im Deutschen für alle Kinder über die Bildergeschichte „Baby Birds“. Für die Erfassung der Erzählfähigkeit im Türkischen bzw. Russischen wurde die parallele Geschichtenversion „Baby Goats“ als Erzählstimulus verwendet.

Die Auswertung der Erzählungen beschränkt sich aktuell auf die Ermittlung der Gesamtzahl der versprochenen Episodenstrukturelemente. Für die Auswer-

tung der Erzähldaten wurde zunächst das Auswertungsraster aus dem Manual von MAIN (Version von 2012; Skerra et al., 2012) herangezogen. Hier ist explizit angegeben, dass das bestehende Auswertungsraster nicht exhaustiv ist, sondern jede als adäquat beurteilte Äußerung als entsprechendes Strukturelement gewertet werden soll. Einheitliche Entscheidungen darüber, welche Äußerungen als adäquat für eine Bewertung als Strukturelement eingestuft werden sollten, wurden in der vorliegenden Untersuchung jedoch durch verschiedene Aspekte erschwert. Zum einen nutzten die untersuchten Kinder viele Äußerungen, die nicht im Manual aufgeführt waren, um einzelne Strukturelemente zu versprachlichen. Darüber hinaus wird bei der Bewertung der Strukturelemente zwar das Vorhandensein einzelner Äußerungen gewertet (z.B. die Vogelmutter füttert, der Hund verjagt die Katze etc.). Diese werden von den Kindern jedoch nicht als isoliert, sondern im Rahmen einer Geschichte produziert und stehen somit in Bezug zu vorangegangenen und nachfolgenden Äußerungen. Daher muss für Fehler bzw. Ungenauigkeiten bspw. bei den für die Referenz auf ProtagonistInnen gewählten sprachlichen Mitteln oder in der Reihenfolge der versprachlichten Strukturelemente der Erzählung jeweils vor dem Hintergrund des Kontextes der einzelnen Äußerung entschieden werden, wie diese zu bewerten ist.

Insgesamt führten die aufgeführten Aspekte im Rahmen der vorliegenden Untersuchung bereits bei der Auswertung der Erzähldaten des ersten Messzeitpunktes (2015) zu erheblichen Inkongruenzen hinsichtlich der Bewertungsentscheidungen, sodass eine einheitliche Bewertung über mehrere AuswerterInnen hinweg nicht erreicht wurde. Daher wurde im Rahmen des Projektes zunächst eine Erweiterung des im Manual von MAIN aufgeführten Auswertungsrasters vorgenommen. Hierfür wurden zusätzliche, als adäquat eingestufte Äußerungen in das Raster aufgenommen, um in möglichst vielen Fällen eine eindeutige Bewertungsentscheidung zu ermöglichen. Im Unterschied zu dem bei MAIN vorgegebenen Auswertungsraster wurden auch Beispiele für nicht zu wertende Äußerungen aufgeführt, um die Ein-

deutigkeit bei einzelnen Bewertungsentscheidungen weiter zu erhöhen. Bspw. sind im MAIN-Manual für das Strukturelement „Attempt“ der Episode „Katze“, „klettert den Baum hoch“, „springt hoch/auf den Baum“ sowie „versucht die Vögel zu erreichen/zu bekommen“ angegeben. Im Rahmen der vorliegenden Studie wurden zusätzlich „geht/krabbelt/kommt hoch/auf den Baum“ als richtig zu wertend aufgenommen. Als nicht zu werten wurde „ist hoch“ aufgeführt, da dies nicht die Handlung der Katze ausdrückt. Außerdem nicht als Attempt gewertet wurden Äußerungen mit „will“ (z. B. „die Katze will hoch“), um eine Konfundierung mit der Kategorie „GOAL“ zu vermeiden.

Darüber hinaus erschien es für eine einheitliche Bewertung äußerungsübergreifender Aspekte, wie etwa den Umgang mit der häufig unklaren oder inkorrekten Referenz auf ProtagonistInnen, zusätzlich erforderlich, Bewertungskriterien zu formulieren. Hinsichtlich der Referenz wird im Manual von MAIN explizit darauf verwiesen, dass sich Fehlbenennungen nicht auf die Punktevergabe auswirken. Dies wurde innerhalb der vorliegenden Untersuchungen entsprechend gehandhabt. Allerdings zeigte sich bei der Auswertung, dass viele der untersuchten Kinder auf einzelne ProtagonistInnen lediglich mit Pronomen referierten (z. B. „Die Katze will die essen“, wenn ‚die Vögel‘ vorher durch das Kind noch nicht eingeführt waren) oder bereits eingeführte ProtagonistInnen nicht rechtzeitig wieder benannten, um eine eindeutige Zuordnung durch den/die ZuhörerIn zu gewährleisten. Des Weiteren musste bezüglich der Referenz geklärt werden, wie mit grammatischen Fehlern (z. B. „Der Hund beißt den in den Schwanz“ für „Der Hund beißt die Katze in den Schwanz“) oder potenziell irreführender Referenz (z. B. weil sowohl ‚die Katze‘ als auch ‚der Hund‘ in der Geschichte als „Hund“ bezeichnet wurde) umgegangen werden sollte. Als Kriterien im Hinblick auf fehlerhafte oder unklare Referenz wurde daher festgehalten, dass Äußerungen mit nicht eindeutiger Benennung des Protagonisten oder der Protagonistin als Strukturelement gewertet werden, sofern unter Kenntnis der Bildergeschichte eindeutig erschlossen werden kann, auf wen

referiert wird. Diese Formulierung war entscheidend, um eine eindeutige Orientierung für Auswertende herzustellen, von welchem Vorwissen des Zuhörers/der ZuhörerIn an dieser Stelle ausgegangen werden sollte. Im unten aufgeführten Beispiel wird entsprechend „und da ist sie hochgeklettert“ als Attempt für die Episode „Katze“ gewertet. Die Äußerung „dann wollte sie noch essen“ wird dagegen nicht als Strukturelement gewertet, da hier nicht eindeutig erkennbar ist, ob das Kind auf die Katze oder auf das Vogelkücken referiert.

„(...) und da ist sie hochgeklettert. Und dann ist die Mama wieder zurückgekommen. Und dann wollte sie noch essen.“

Für das formulierte Kriterium wurde entsprechend expliziert, dass Äußerungen, die inhaltlich eindeutig einem Protagonisten/einer Protagonistin zugeordnet werden können, auch dann als Strukturelemente zählen, wenn der/die entsprechende ProtagonistIn zuvor nicht eingeführt wurde oder wenn Fehler bei der Referenz auftreten (z. B. durch falsch verwendete Pronomen wie in „und dann hat er mit seine Krallen angefasst den Vogel“ für „die Katze fasst den Vogel mit ihren Krallen an“).

Fragestellungen

Im Rahmen der vorliegenden Untersuchung sollen Hinweise darauf gewonnen werden, inwieweit das Verfahren MAIN die Hauptgütekriterien Objektivität, Reliabilität und Validität erfüllt. Die durchgeführte Objektivitätsprüfung richtet sich auf die Auswertungsobjektivität des Verfahrens. Untersucht werden soll, inwieweit eine von der/dem jeweiligen BeurteilerIn unabhängige Auswertung gewährleistet werden kann.

Ziel der Reliabilitätsprüfung ist die Untersuchung der Messgenauigkeit des Verfahrens über die Ermittlung der Retest- sowie der Paralleltest-Reliabilität. Für die Überprüfung der Retest-Reliabilität wird untersucht, ob die erfassten Testergebnisse bei wiederholter Vorlage der gleichen Bildergeschichte („Baby birds“) über zwei Messzeitpunkte hinweg stabil bleiben. Bezüglich der Paralleltest-Reliabilität wird untersucht, inwiefern die

KURZBIOGRAFIE

Jens Kaiser-Kratzmann ist Professor für Pädagogik mit Schwerpunkt frühe Kindheit an der Kath. Universität Eichstätt-Ingolstadt. Seine Arbeitsschwerpunkte sind (mehr-)sprachige Bildung, Professionalisierung pädagogischer Fachkräfte, Qualitätsentwicklung in Kindertageseinrichtungen, Übergänge sowie soziale und migrationsgekoppelte Ungleichheit in der Kindheit.

untersuchten Kinder in den beiden parallelisierten Geschichtenversionen „Baby Birds“ und „Baby Goats“ vergleichbare Ergebnisse erzielen und ob Leistungsunterschiede zwischen den Kindern durch beide Testversionen in gleicher Weise abgebildet werden.

Mit der Überprüfung der Validität sollen Hinweise dahingehend gewonnen werden, inwieweit die Messintention des Verfahrens erfüllt wird. Die Erfassung der Episodenstrukturelemente zielt gemäß des Story-Grammar-Ansatzes darauf ab, Rückschlüsse auf das implizite Wissen zum schematischen Aufbau von Erzählungen zu ermöglichen und somit zu erfassen, inwieweit es dem Kind gelingt, die für die Geschichte wesentlichen Informationen darzustellen. Da für die Elizitation eine Bildergeschichte als Vorlage dient, ist der hierfür erforderliche Wortschatz zu einem gewissen Grad bereits vordefiniert. Vor diesem Hintergrund erscheint es erforderlich zu untersuchen, ob sich das durch MAIN gemessene Merkmal von rein expressiven Wortschatzleistungen abgrenzen lässt. Ergebnisse aus einer Untersuchung von Uccelli und Pérez (2007) weisen diesbezüglich darauf hin, dass das Erzählen zu einer Bildvorlage keine von den expressiven Wortschatzleistungen des Kindes unabhängige Fähigkeit abbildet. Für die vorliegende Untersuchung wird daher erwartet, dass die zu ermittelnden Zusammenhänge darauf hinweisen, dass über MAIN eine mit dem expressiven Wortschatz assoziierte, jedoch von diesem abgrenzbare Fähigkeit erfasst wird.

Überprüfung der Gütekriterien

Objektivität

Die Erzählenden, die zur Überprüfung der Auswertungsobjektivität herangezogen wurden, sind Teil der deutschen Sprachdaten, die innerhalb des Projektes zum zweiten Messzeitpunkt im Jahr 2016 erhoben wurden. Für die geplante Untersuchung wurde aus den getesteten Kindern ($n=278$) eine Zufallsstichprobe von 8% ($n=22$) gezogen. Innerhalb der Zufallsstichprobe waren 50% weiblich und 50% männlich ($n=11$). Das Alter lag zwischen 4;01 und 6;06 Jahren, der Altersdurchschnitt betrug 5;01 Jahre (61 Monate, $SD=8,26$).

Für die Überprüfung der Auswertungsobjektivität wurden die Transkripte der Zufallsstichprobe von zwei unabhängigen Beurteilerinnen ausgewertet. Alle Transkriptionen waren im Vorfeld durch dieselbe Person vorgenommen worden, um eine Konfundierung von möglichen

Transkriptions- und Auswertungsunterschieden zu vermeiden. Für die Bewertung wurden neben den Vorgaben des Manuals von MAIN das erweiterte Auswertungsraster und die zusätzlich formulierten Kriterien verwendet.

Die Beurteilungsübereinstimmung wurde über drei unterschiedliche Methoden ermittelt. Hierbei wurde zunächst eine Korrelation zwischen den jeweils vergebenen Gesamtpunktwerten berechnet. Entsprechend der Untersuchung von Maviş et al. (2016), bei der eine Korrelation der Gesamtwerte von $r=.82$ angegeben wird, wurde dabei ein Wert von mindestens $r=.80$ als zufriedenstellende Übereinstimmung festgelegt. Da bei abschließlicher Ermittlung einer Korrelation systematische Bewertungsunterschiede zwischen den beiden Bewertenden, bspw. eine durchgehend höhere Bewertung durch eine der Beurteilerinnen bei gleichbleibender Reihung der Kinder, nicht erkannt werden könnte, wurde darüber hinaus ein Mittelwertvergleich berechnet. Zusätzlich wurde überprüft, zu welchem prozentualen Anteil die Kodierung einzelner Episodenstrukturelemente als vorhanden bzw. nicht vorhanden über alle 22 analysierten Erzählungen hinweg (insgesamt 374 Entscheidungen) übereinstimmte. Hinsichtlich des prozentualen Anteils gleicher Kodierungsentscheidungen wurde der bei Kunnari et al. (2016) ermittelte Wert von 92,7% Übereinstimmung zur Festlegung eines kritischen Wertes herangezogen. In der vorliegenden Untersuchung wird darauf bezugnehmend eine Übereinstimmung für mindestens 90% aller Kodierungen als zufriedenstellend betrachtet.

Die Gesamtpunktzahlen der durch die beiden Beurteilerinnen jeweils erfassten Episodenstrukturelemente korrelieren zu $r=.985$ ($p<.001$), was einem sehr hohen Zusammenhang entspricht. Der Mittelwert der Gesamtpunktzahlen liegt bei der Bewertung durch Rater 1 bei 3,59 ($SD=2,69$) und bei der Bewertung durch Rater 2 bei 3,45 ($SD=2,65$). Der Unterschied zwischen den beiden Mittelwerten ist statistisch nicht signifikant (vgl. Tab. 2). Bei der Kodierung aller Episodenstrukturelemente als jeweils vorhanden bzw. nicht vorhanden erfolgte in 98,13% der Fälle eine übereinstimmende Bewertung.

Reliabilität

Für die Überprüfung der Reliabilität und damit der Messpräzision des Verfahrens wurden sowohl die Retest- als auch die Paralleltest-Reliabilität bestimmt.

Die Stichprobe der Kinder, die für die Überprüfung der Retest-Reliabilität herangezogen wurde, stammt nicht aus der Stichprobe des Projektes, sondern wurde im Frühjahr 2017 zusätzlich rekrutiert. Hierfür wurden in zwei Kindertagesstätten in Hessen Einverständniserklärungen an Eltern mehrsprachiger Kinder ausgegeben. Insgesamt konnten die Daten von zehn Kindern für den Vergleich der Erzählenden herangezogen werden. Das Alter der untersuchten Kinder lag zwischen 3;08 und 6;06 Jahren (44 bis 78 Monate) bei einem Mittelwert von 5;01 Jahre (61 Monate).

Die untersuchten Kinder wurden zu jeweils zwei verschiedenen Testzeitpunkten mit der Geschichte „Baby Birds“ innerhalb ihrer jeweiligen Kindertagesstätte getestet. Alle Testungen fanden in einem zeitlichen Abstand von 14 Tagen statt, um Übungseffekte gering zu halten, gleichzeitig jedoch einen zu langen Abstand zwischen den Testungen und damit möglicherweise einhergehende Merkmalsveränderungen zu vermeiden. Für die Ermittlung der Retest-Reliabilität wurden jeweils Korrelationen zwischen der Gesamtzahl der Episodenstrukturelemente zu den beiden Testzeitpunkten berechnet. Für die Höhe des als zufriedenstellend erachteten Reliabilitätskoeffizienten wurde gemäß Moosburger und Kelava (2012) ein Wert von mindestens $r=.70$ festgelegt. Zusätzlich sollte überprüft werden, ob sich die erreichten Punktzahlen über die beiden Messzeitpunkte hinweg in überzufälligem Ausmaß unterscheiden. Für beide Auswertungsebenen wurde zur Überprüfung der festgestellten Unterschiede daher jeweils ein T-Test für verbundene Stichproben durchgeführt.

Die ermittelte Korrelation für die Gesamtzahl der erfassten Episodenstrukturelemente liegt bei $r=.978$ ($p=.000$) und somit deutlich über dem festgesetzten Mindestwert von $r=.70$. Zum ersten Messzeitpunkt wurden durchschnittlich $M=4,6$ ($SD=2,41$), zum zweiten Messzeitpunkt $M=5,00$ ($SD=2,49$) Episoden-

KURZBIOGRAFIE

Steffi Sachse studierte Psychologie an der Humboldt-Universität in Berlin. In der Abteilung für Entwicklungsstörungen der Klinik für Kinder- und Jugendpsychiatrie der LMU München arbeitete und promovierte sie mit einer Längsschnittstudie zu Kindern mit verspätetem Sprechbeginn (Late Talker). An der Universität Ulm leitete sie am ZNL-Transfer Zentrum für Neurowissenschaften und Lernen eine Arbeitsgruppe zum Thema „Sprachentwicklung und -förderung in Kindertagesstätten“. Seit Oktober 2013 ist sie Professorin für Entwicklungspsychologie mit dem Schwerpunkt Sprachentwicklung an der Pädagogischen Hochschule Heidelberg. Ihre Forschungsschwerpunkte liegen im Bereich der Früherkennung und Sprachdiagnostik, der sprachlichen Bildung und Förderung im Rahmen von Kindertagesstätten, der diagnostischen Kompetenzen von Lehrkräften sowie in entwicklungspsychologischen Aspekten von kindlicher Mehrsprachigkeit.

strukturelemente produziert. Die untersuchten Kinder erreichten damit zum zweiten Messzeitpunkt im Durchschnitt signifikant höhere Punktzahlen ($t(9)=-2,449$; $p=.037$).

Paralleltest-Reliabilität

Die Daten für die Analyse der Paralleltest-Reliabilität entstammen zwei unterschiedlichen Stichproben. Dies war erforderlich, da aufgrund von potenziellen Übungseffekten die Ausbalancierung der Durchführungsreihenfolge gewährleistet werden musste. Insgesamt basieren die Analysen auf den Erzählenden von 28 Kindern. Alle erzählten jeweils sowohl die Geschichtenversion „Baby Birds“ als auch die Geschichtenversion „Baby Goats“, wobei eine gleichmäßige Aufteilung nach Reihenfolge der bearbeiteten Testversionen gegeben war (1. Bedingung „Baby Birds“ – „Baby Goats“ $n=14$; 2. Bedingung „Baby Goats“ – „Baby Birds“ $n=14$).

Für die erste Testbedingung (Testreihenfolge T1: „Baby Birds“, T2: „Baby Goats“) wurden Kinder aus der Projektstichprobe des zweiten Messzeitpunktes (2016) herangezogen. Die Auswahl erfolgte in Abhängigkeit von der besuchten Kindertagesstätte. In insgesamt drei der teilnehmenden Kindertagesstätten war eine zweite Testung mit MAIN möglich. Da die Geschichtenversion „Baby Goats“ zur Elizitation der Erzählung im Rahmen der Erstspracherhebung eingesetzt wird, wurden Kinder mit türkischer und russischer Herkunftssprache nicht in die Untersuchungsgruppe eingeschlossen, um Übungs- und Übertragungseffekte aus der Herkunftsspracherhebung auszuschließen. Die unter der zweiten Bedingung (T1: „Baby Goats“ – T2: „Baby Birds“) getesteten Kinder stammen aus einer zusätzlich rekrutierten Stichprobe, um innerhalb der Sprachstandserhebungen des Projektes gleichbleibende Bedingungen für das Erzählen der Geschichte „Baby Birds“ und damit die Vergleichbarkeit der Erzählleistungen zu gewährleisten. Daher wurden im Frühjahr 2017 in zwei Kindertagesstätten in Baden-Württemberg zusätzlich mehrsprachige Kinder rekrutiert. Insgesamt konnten 14 Kinder aus den beiden Einrichtungen getestet werden.

Die untersuchten Kinder waren im

Durchschnitt 5;04 Jahre alt (64 Monate, SD: 8,92). Die Altersspanne lag zwischen 3;07 und 6;06 Jahren (43 bis 78 Monate). 46,6% ($n=13$) der Kinder waren weiblich und 53,6% ($n=15$) männlich.

Zwischen den beiden Gruppen für die unterschiedlichen Testbedingungen ergaben sich dabei keine signifikanten Altersunterschiede ($t(27)=-.886$; $p=.384$). Die Kinder wurden an jeweils zwei verschiedenen Terminen in einem ruhigen Raum ihrer Kindertagesstätte untersucht. Alle Untersuchungen wurden von derselben Person durchgeführt.

Kindertagesstätten war es innerhalb der zweiten Testbedingung nicht möglich, den bei MAIN empfohlenen Zeitabstand von vier bis sieben Tagen für die Testung mit den beiden parallelisierten Geschichten einzuhalten. Bei drei Kindern musste die Erhebung unmittelbar am selben Tag erfolgen, da sie zum ersten Testzeitpunkt nicht in der Einrichtung waren und kein weiterer Untersuchungstermin möglich war. Es ergab sich hierbei kein signifikanter Zusammenhang zwischen dem Leistungsunterschied in den beiden Geschichtenversionen und dem Testabstand in Tagen ($r=.002$; $p=.994$). Grundlage für die Beurteilung bildeten das Auswertungsraster für die Geschichte „Baby Birds“ sowie das entsprechend übertragene Raster für die Geschichte „Baby Goats“. Um zu ermitteln, inwieweit die einzelnen Kinder in den beiden Geschichten vergleichbare Ergebnisse erzielen, wurden Korrelationen zwischen den jeweils erzielten Gesamtpunktzahlen der Episodenstrukturelemente berechnet. Die Höhe der als zufriedenstellend einzustufenden Korrelation wurde dabei auf mindestens $r=.70$ festgelegt. Darüber hinaus wurde ein Vergleich mit dem Ergebnis aus der Überprüfung der Retest-Reliabilität herangezogen, da sich bei der Annahme der beiden Geschichten als parallele Testversionen keine bedeutsamen Unterschiede zwischen den jeweils ermittelten Reliabilitätskoeffizienten ergeben sollten. Da zudem nicht ausgeschlossen werden konnte, dass eine der beiden Geschichtenvorlagen höhere Anforderungen an die Erzählenden stellt, wurde zusätzlich über einen Mittelwertvergleich untersucht, ob sich die Erzählleistungen der Kinder in den beiden Geschichten systematisch unterscheiden.

Vergleich	Ergebnis
Korrelation der Gesamtpunktwerte	$r=.985$ ($p=.000$)
Mittelwertvergleich der Gesamtpunktwerte	Rater 1: $M=3,59$ ($SD=2,67$) Rater 2: $M=3,45$ ($SD=2,65$) $t(21)=1,368$; $p=.186$
Prozentzahl übereinstimmender Kodierungen	98,13% (367 von 374 Kodierungen)

Tabelle 2 Ergebnisse der Objektivitätsprüfung

Die Korrelation für die Anzahl der erfassten Episodenstrukturelemente für die beiden Geschichtenversionen „Baby Birds“ und „Baby Goats“ liegt bei $r=.767$ ($p=.000$) und damit über dem festgelegten Mindestwert von $r=.70$. Wie in Tabelle 3 dargestellt wird, liegt der ermittelte Koeffizient jedoch deutlich unter dem Ergebnis aus der Überprüfung für die Retest-Reliabilität. Die erfassten Leistungen weisen somit eine wesentlich geringere Stabilität auf als bei erneuter Vorlage derselben Geschichte. Für die Mittelwerte der versprochenen Episodenstrukturelemente ergibt sich kein signifikanter Unterschied zwischen den beiden Geschichtenversionen ($t(27)=1.563$; $p=.13$).

Validität

Erste Hinweise zur Validität des Verfahrens sollten über die Untersuchung der Konstruktvalidität ermittelt werden. Als Indikator für diese wurde die divergente Validität des Verfahrens ermittelt. Hierfür wurde untersucht, inwieweit sich Zusammenhänge zwischen der Anzahl der produzierten Episodenstrukturelemente und den expressiven Wortschatzleistungen der Kinder ergeben.

Die Stichprobe für die Überprüfung der Validität entspricht der Stichprobe, die zur Überprüfung der Objektivität herangezogen wurde und umfasst somit 22 Kinder (8%) aus der Gesamtstichprobe des Projektes für den zweiten Messzeitpunkt.

Zur Ermittlung der Konstruktvalidität wurde die Anzahl der Episodenstrukturelemente mit den expressiven Wortschatzleistungen aus einem Bildbenennungsverfahren (Aktiver Wortschatztest für 3- bis 5-jährige Kinder-Revision (AWST-R); Kiese-Himmel, 2005) kor-

reliert. Der AWST-R besteht aus insgesamt 75 Test-Items und ist für Kinder im Alter von 3;0 bis 5;5 Jahren normiert. Da ein Teil der untersuchten Kinder zum Zeitpunkt der Testung bereits älter als 5;5 Jahre war, wurden für die Berechnung der Korrelationen zu den Testwerten aus MAIN die erfassten Rohwerte herangezogen.

Die Korrelation zwischen den erfassten Episodenstrukturelementen und den Ergebnissen des AWST-R liegt bei $r=.648$ ($p=.001$). Es zeigt sich somit ein hoher Zusammenhang zwischen den beiden erfassten Leistungen.

Diskussion und Limitationen

Die Ergebnisse der hier vorgestellten Überprüfung der Gütekriterien für das Verfahren MAIN sind insgesamt zufriedenstellend.

Die Ergebnisse der Objektivitätsprüfung weisen darauf hin, dass unter Verwendung des ausdifferenzierten Auswertungsrasters sowie zusätzlich formulierter Kriterien eine hohe Beurteilungsübereinstimmung und damit eine zufriedenstellende Auswertungsobjektivität erreicht werden kann. Bei der Interpretation der Ergebnisse muss jedoch berücksichtigt werden, dass die hier untersuchte Stichprobe im Durchschnitt lediglich knapp vier von insgesamt 17 möglichen Gesamtpunkten erzielte. Der Grad der Übereinstimmung könnte daher auch durch die insgesamt geringen Punktzahlen beeinflusst sein. Die Ergebnisse der vorliegenden Untersuchung sollten daher nicht ungeprüft für die Anwendung des Instrumentes bei anderen Untersuchungsgruppen, etwa bei älteren Kindern, angenommen werden. Beachtet werden muss zudem, dass für eine einheitliche Auswertung innerhalb von Forschungsprojekten eine intensive Einarbeitung aller Beteiligten sowie die regelmäßige Abstimmung zu Auswertungsentscheidungen im Team erforderlich sind.

Die Ergebnisse der Reliabilitätsprüfung sind nur teilweise zufriedenstellend, da sich deutliche Unterschiede zwischen

Analyse	Retest	Paralleltest
Korrelation	.978 ($p=.000$)	.767 ($p=.000$)
Mittelwertvergleich	T1: $M=4,6$ ($SD=2,41$) T2: $M=5,00$ ($SD=2,49$) $t(9)=-2,449$, $p=.037$	Baby birds: $M=3,89$ ($SD=2,23$) Baby goats: $M=3,46$ ($SD=1,93$) $t(27)=1,563$, $p=.13$

Tabelle 3 Ergebnisse der Reliabilitätsprüfung

den beiden überprüften Reliabilitätsformen ergeben. Die Untersuchung der Retest-Reliabilität weist auf eine hohe Zuverlässigkeit des Verfahrens bei Verwendung der gleichen Testversion hin. Obgleich sich Hinweise auf mögliche Übungseffekte beim erneuten Erzählen der Geschichte ergeben, weist die Korrelation von $r=.978$ darauf hin, dass die im Forschungskontext angestrebte Abbildung von Leistungsunterschieden zwischen einzelnen Kindern stabil gewährleistet werden kann. Die erneute Testung mit derselben Geschichtenvorlage gestaltete sich überdies für alle getesteten Kinder als unproblematisch. Auch Kinder, welche die Geschichtenvorlage wiedererkannten, erzählten diese bereitwillig erneut.

Im Gegensatz dazu sind die Ergebnisse aus der Überprüfung der Paralleltest-Reliabilität wenig zufriedenstellend. Bei der Verwendung der beiden parallelisierten Geschichtenversionen zeigt sich eine wesentlich geringere Reliabilität als bei Verwendung derselben Geschichtenversion zu zwei Messzeitpunkten. Da innerhalb des Projektes in Abhängigkeit von der überprüften Sprache (Deutsch bzw. Herkunftssprache) unterschiedliche Geschichtenversionen verwendet wurden, muss beim Vergleich der Testleistungen in den beiden Sprachen eines Kindes der deutliche Reliabilitätsverlust mitberücksichtigt werden.

Bei der Überprüfung der Konstruktvalidität ergab sich der erwartete Zusammenhang zwischen der Gesamtzahl der Episodenstrukturelemente und den expressiven Wortschatzleistungen der untersuchten Kinder. Die ermittelte Korrelation von $r=.648$ ($p=.001$) liegt dabei über den bei Uccelli und Páez (2007) berichteten Werten von $r=.55$, $p<.01$

(für Englisch) bzw. $r=.56$, $p<.01$ (für Spanisch). Der Unterschied in der Höhe des Zusammenhanges könnte hierbei auf Unterschiede in der Bewertung der Geschichtenstruktur zurückführbar sein. Im Gegensatz zur Auswertung bei MAIN wird in dieser Untersuchung auf ein Verfahren zurückgegriffen, innerhalb dessen für einzelne Episodenstrukturelemente nur über deren Vorhandensein entschieden

werden muss, ohne dass bereits konkrete inhaltliche Vorgaben bestehen (z. B. „offers resolution“) (vgl. Uccelli & Páez, 2007, S. 16). Möglicherweise ergibt sich der für MAIN ermittelte höhere Zusammenhang zu den Wortschatzleistungen aufgrund der bereits stärker vordefinierten inhaltlichen Angaben bei der Auswertung. Die Höhe der hier ermittelten Korrelation erscheint jedoch dahingehend interpretierbar, dass MAIN zwar eine mit dem expressiven Wortschatz zusammenhängende, jedoch von diesem unterscheidbare Fähigkeit abbildet. Einschränkend ist allerdings anzumerken, dass dieses Ergebnis lediglich als ein erster grober Indikator auf die Konstruktvalidität des Verfahrens interpretiert werden darf, da für eine umfassende Überprüfung neben der divergenten auch die konvergente Validität ermittelt werden müsste.

Insgesamt muss vor dem Hintergrund der geringen Stichprobengröße für alle Untersuchungen zudem beachtet werden, dass die berichteten Ergebnisse nur einen ersten Hinweis auf die grundsätzliche Anwendbarkeit sowie mögliche Schwierigkeiten bei der Anwendung des Verfahrens geben können.

Gesamtfazit und Ausblick

Zusammenfassend scheint das Verfahren MAIN für die Nutzung als Forschungsinstrument geeignet, da es als hinreichend objektiv und reliabel eingestuft werden kann, sowie erste Hinweise darauf gegeben sind, dass die Messintention des Verfahrens erfüllt wird. Hinsichtlich der Verwendung der beiden unterschiedlichen Testversionen muss jedoch der ermittelte Reliabilitätsverlust berücksichtigt werden.

Über die Eignung des Verfahrens zur

klinischen Diagnostik können auf Grundlage der ermittelten Daten aktuell noch keine Aussagen getroffen werden. Für weitere Forschungsvorhaben innerhalb des Projektes ist jedoch geplant, den prädiktiven Aussagewert der durch MAIN ermittelten Erzählfähigkeit auf die weitere mehrsprachige Entwicklung zu untersuchen. Sofern sich das Verfahren als prognostisch hinreichend valide erweist, um Sprachentwicklungsverläufe vorherzusagen und jene mehrsprachigen Kinder mit Risiko für eine ungünstige sprachliche Entwicklung zu identifizieren, könnte in einem nächsten Schritt geprüft werden, wie der Einsatz des Verfahrens in der klinischen Praxis gelingen kann und welche Voraussetzungen hierfür gegeben sein müssen.

Interessenkonflikt

Die AutorInnen geben an, dass keine Interessenkonflikte vorliegen.
Das Förderkennzeichen lautet 01 JM 1402B.

Literatur

- Autorengruppe Bildungsberichterstattung. (Hrsg.) (2018). *Bildung in Deutschland 2018. Ein indikatorengestützter Bericht mit einer Analyse zu Wirkungen und Erträgen von Bildung*. <https://www.bildungsbericht.de/de/bildungsberichte-seit-2006/bildungsbericht-2018/pdf-bildungsbericht-2018/bildungsbericht-2018.pdf>.
- Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: implications for clinical practice. *International journal of language & communication disorders*, 51(6), 626–638.
- Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37(1), 19–48.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. Pearson Studium.
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 243–276). Multilingual Matters.
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual assessment instrument for narratives. <https://www.leibniz-zas.de/de/forschung/publikationen/details/publications/3832-main-multilingual-assessment-instr/>.
- Goldman, A. I. (2012). *Theory of mind. The Oxford handbook of philosophy of cognitive science*. http://diabeto.enseiht.fr/download/perception/Goldman_2012.pdf.
- Kapica, N., Klages, H., & Pagonis, G. (2014). Narrative Kompetenz: Anforderungen an ein Verfahren der Sprachstandsfeststellung im Elementarbereich. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 19(2), 5–19.
- Kiese-Himmel, C. (2005). *Aktiver Wortschatztest für 3- bis 5-jährige Kinder: AWST-R: Revision*. Beltz Test.
- Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish-Swedish children. *Applied Psycholinguistics*, 37(1), 123–144.
- Licandro, U. (2016). *Narrative Skills of Dual Language Learners: Acquisition and Peer-Assisted Support in Early Childhood Education and Care*. Springer.
- Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive psychology*, 9(1), 111–151.
- Maviş, İ., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish-German bilingual children. *Applied Psycholinguistics*, 37(1), 69–89.
- Moosburger, H., & Kelava, A. (2012). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosburger, & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 7–26). Springer.
- Paradis, J., Genesee, F., & Crago, M. B. (2010). *Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning*. Brookes Publishing Company.
- Pfeffer, K. (2015). *Selektieren, verknüpfen, sprachlich umsetzen: zu viel für Kinder mit einer Sprachentwicklungsstörung? Narrative Fähigkeiten bei Kindern mit und ohne Sprachentwicklungsstörung im Grundschulalter*. Schulz-Kirchner.
- Reich, H. H. (2010). *Frühe Mehrsprachigkeit aus linguistischer Perspektive*. DJI.
- Rezzonico, S., Goldberg, A., Ka-Yan Mak, K., Yap, S., Milburn, T., Beletti, A., & Girolametto, L. (2016). Narratives in Two Languages: Storytelling of Bilingual Cantonese-English Preschoolers. *Journal of Speech, Language, and Hearing Research*, 59(3), 521–532.
- Ringmann, S. (2013). Therapie der Erzählfähigkeit. In S. Ringmann, & J. Siegmüller (Hrsg.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen. Schuleingangsphase* (S. 163–188). Elsevier.
- Rodina, Y. (2017). Narrative abilities of preschool bilingual Norwegian-Russian children. *International Journal of Bilingualism*, 21(5), 617–635.
- Schermelleh-Engel, K., & Werner, C. S. (2012). Methoden der Reliabilitätsbestimmung. In H. Moosburger, & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 118–141). Springer.
- Schröder, A. (2009). *Vergleichende Analyse interaktiver Erzählfähigkeiten bei sechsjährigen Kindern mit einer sogenannten spezifischen Sprachentwicklungsstörung und Kindern mit unauffälligem Spracherwerb* [Dissertation]. <https://eldorado.tu-dortmund.de/bitstream/2003/27128/1/Dissertation.pdf>.
- Skerra, A., Reichenbach, K., Reichardt, V., & Gagarina, N. (2012). Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics*, 56. <https://www.leibniz-zas.de/de/publications/schriftenreihe/zaspil/zaspil-56/main-start/>.
- Stein, N. L., & Glenn, C. G. (1975). *An Analysis of Story Comprehension in Elementary School Children: A Test of a Schema*. <http://files.eric.ed.gov/fulltext/ED121474.pdf>.
- Uccelli, P., & Páez, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools*, 38(3), 225–236.



AutorInnen

Tamara Lautenschläger
Pädagogische Hochschule Heidelberg
Institut für Psychologie/
Institut für Sonderpädagogik
Keplerstr. 87, D-69120 Heidelberg
lautenschlaeg@ph-heidelberg.de

Katja Schneller (M.Sc.)
Pädagogische Hochschule Heidelberg
Institut für Psychologie/
Institut für Sonderpädagogik
Keplerstraße 87, D-69120 Heidelberg
schneller@ph-heidelberg.de

Prof. Dr. Jens Kaiser-Kratzmann
Kath. Universität Eichstätt-Ingolstadt
Fakultät für Soziale Arbeit
Kapuzinergasse 2 (Raum 217),
D-85072 Eichstätt
Jens.Kratzmann@ku.de

Prof. Dr. Steffi Sachse
Pädagogische Hochschule Heidelberg
Institut für Psychologie
Keplerstr. 87, D-69120 Heidelberg
sachse@ph-heidelberg.de



DOI dieses Beitrags
(www.doi.org)
10.7345/prolog-2021164