

Messung von Wortschatztiefe und -breite bei mono- und bilingualen Vorschulkindern

Dr. Astrid Jurecka

Goethe-Universität Frankfurt
Fachbereich Erziehungswissenschaften
Institut für Pädagogik der Elementar- und Primarstufe und Empirische Bildungsforschung
E-Mail: jurecka@em.uni-frankfurt.de

Melihan Cinar

Goethe-Universität Frankfurt
Fachbereich Erziehungswissenschaften
Institut für Pädagogik der Elementar- und Primarstufe und Empirische Bildungsforschung

Prof. Dr. Ilonca Hardy

Goethe-Universität Frankfurt
Fachbereich Erziehungswissenschaften
Institut für Pädagogik der Elementar- und Primarstufe und Empirische Bildungsforschung

Abstract: Obgleich bildungssprachlicher Wortschatz bereits in der Grundschule eine wichtige Voraussetzung für Bildungserfolg darstellt, existieren diesbezüglich bislang keine vorschulspezifischen Instrumente. Ziel der Studie war die Messung eines an die Bildungssprache anschlussfähigen Wortschatzes im Vorschulalter. Unter der Annahme, dass Wortschatz aus den Dimensionen Wortschatzbreite (Anzahl bekannter Wörter) und Wortschatztiefe (Wortwissen) besteht, erfolgte die Entwicklung eines passiven Wortschatztests. Erste Ergebnisse (n=150 mono- und bilingual deutsche Vorschulkinder) weisen auf gute psychometrische Eigenschaften des Instruments hin. Die Dimensionalität des Tests wurde anhand exploratorischer Faktorenanalysen überprüft, Fit-Indices weisen auf eine zwei- als auch eine dreifaktorielle Lösung hin, so dass von einem mehrdimensionalen Wortschatz-Konstrukt ausgegangen werden kann. Die Ergebnisse werden insbesondere bezüglich der Wortschatz-Tiefe-Skala kritisch diskutiert.

Although academic language is a prerequisite for academic success already in primary school, no preschool-specific tests focusing academic lexical proficiency in German have been developed so far. The main objective of the study was therefore to assess German preschoolers' passive academic vocabulary knowledge. Based on theoretical considerations on the dimensions of vocabulary breadth (known words) and vocabulary depth (knowledge about words), a test assessing preschoolers' German vocabulary knowledge was developed. First results (n=150 mono- and bilingual German preschoolers) show good psychometric properties. The dimensionality of the test was examined using exploratory factor analyses. Fit-indices suggest two-factor and three-factor solutions pointing to a multidimensional vocabulary construct. The results on the dimension of vocabulary depth are critically discussed.

Schlüsselwörter: Wortschatztest; Wortschatztiefe; Wortschatzbreite; Vorschulkinder; Bildungssprache; Vocabulary test; vocabulary depth; vocabulary breadth; preschoolers; academic language proficiency

Jurecka, Astrid; Cinar, Melihan & Hardy, Ilonca (2019), Messung von Wortschatztiefe und -breite bei mono- und bilingualen Vorschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 24: 1, 145–163.
Abrufbar unter <http://tjournals.ulb.tu-darmstadt.de/index.php/zif/>.

Sprachkompetenzen in der Unterrichtssprache Deutsch stellen eine wichtige Voraussetzung für Schulerfolg und Chancengleichheit dar (vgl. z.B. Baumert & Schümer 2001; Bos, Valtin, Hußmann, Wendt & Goy, 2017; Söhn 2005). Jedoch zeigen sich deutliche Unterschiede zwischen dem sprachlichen Entwicklungsstand von bilingual deutsch (BL) und monolingual deutsch (ML) aufgewachsenen Kindern zugunsten der monolingualen (vgl. z.B. Dubowy, Ebert, von Maurice & Weinert 2008), die für festgestellte Leistungsunterschiede und Nachteile bzgl. der Bildungsteilhabe wesentlich verantwortlich gemacht werden (vgl. z.B. Baumert & Schümer 2001; Ramm, Prenzel, Heidemeier & Walter 2004). Insbesondere wird dabei die sogenannte Bildungssprache (vgl. z.B. Cummins 2000; Heppt, Stanat, Dragon, Berendes & Weinert 2014) als bedeutsam für den Schulerfolg erachtet. Bei dieser handelt es sich um die typische Form der Sprachverwendung in Schul- und Bildungskontexten, welche insbesondere für das Verstehen schulischer Inhalte relevant ist (vgl. zusammenfassend Morek & Heller 2012). Sie zeichnet sich durch Merkmale wie Komplexität, Abstraktheit, Explizitheit, Kohärenz (vgl. Schmölzer-Eibinger, Dorner, Langer & Helten-Pacher 2013) sowie eine im Vergleich zur Alltagssprache geringere soziale und situativ kontextuelle Einbettung aus (vgl. z.B. Heppt et al. 2014); auch auf Lernprozesse bezogene Sprachkompetenzen, wie etwas erklären oder erörtern zu können, werden der Bildungssprache zugerechnet (vgl. z.B. Feilke 2012). Des Weiteren sind vor allem lexikale Merkmale wie Fremdwörter, Nominalisierungen, lexikalische Dichte (Bailey, Butler, Stevens & Lord 2007; Eckhardt 2008; Gogolin & Lange 2011), das Kennen von Fachvokabular und schwierigen Alltagswörtern (vgl. Heppt et al. 2014) sowie das semantische Wortwissen (Weinert 2004) bezüglich schul- und fachbezogener Inhalte zentral. Um die Chancengleichheit zwischen mono- und bilingual deutsch aufwachsenden Kindern zu erhöhen, wird daher empfohlen, sprachliche Kompetenzen, und hier insbesondere den Wortschatz, bereits vor dem Schuleintritt zu erfassen und ggf. angemessen zu fördern (vgl. z.B. Niedersächsisches KM 2007; Neugebauer & Becker-Mrotzek 2013). Inwieweit jedoch bereits im Vorschulalter auch bildungssprachlich und schulkontextuell relevanter Wortschatz über grundlegende Wortschatzkenntnisse hinaus differenziert erfasst und gefördert werden sollte, ist bislang unklar. Bezüglich der Definition bildungssprachlichen Wortschatzes für das Vorschulalter sowie dessen Differenzierung von alltagssprachlichem Wortschatz existieren dabei unterschiedliche Herangehensweisen. So differenzieren etwa Bialystok, Luk, Peets und Yang (2010: 5) Wörter in primär „school-relevant contexts“ und „home contexts“, und betrachten schulkontextuell relevantes Vokabular als bildungssprachliche Vorläuferfähigkeit. Van Kleeck (2014: 725) wiederum unterteilt für das Vorschulalter in die beiden Sprachregister „casual talk“ und (der Entwicklungsstufe angepassten) „academic talk“; Letzteres sieht sie dabei als Grundlage für Literalität in der Schule. Hinsichtlich spezifisch lexikaler Eigenschaften definiert sie dabei akademischen Wortschatz u.a. als literarischer, unbekannter, länger und fachspezifischer. In beiden Arbeiten wird dabei für das Vorschulalter nicht von einer klaren Differenzierbarkeit von alltagssprachlichem versus bildungssprachlichem Wortschatz, sondern eher von Nutzungsschwerpunkten ausgegangen. Angelehnt an die beschriebenen Vorgehensweisen wird im Rahmen der vorliegenden Arbeit bildungssprachlicher Wortschatz im Vorschulalter definiert als bildungssprachliche Vorläuferfähigkeit bezüglich der Kenntnis von und des Wissens über Wörter, die in schulisch-fachlichen Kontexten in der Schuleingangsphase relevant ist.

Bezüglich der Erklärung von Leistungsunterschieden in Sekundar- und Primarstufe weisen erste Studien auf inkrementelle, jedoch eher geringe Effekte von bildungssprachlichem Wortschatz hin (vgl. Heppt, Henschel & Haag 2016; Townsend, Filippini, Collins & Biancarosa 2012; Uccelli, Barr, Dobbs, Galloway, Meneses & Sanchez 2015). Bislang existieren jedoch keine vorschulspezifischen Studien hinsichtlich des Einflusses vorschulischen bildungssprachlichen Wortschatzes auf schulische Leistungen. Dies ist nicht zuletzt auf einen Mangel an validen Testinstrumenten zur Messung bildungssprachlichen Wortschatzes in dieser Altersgruppe zurückzuführen.

Das Konstrukt Wortschatz wurde zunächst primär über die *Anzahl bekannter Wörter* (vgl. z.B. Meara & Buxton 1987), vermehrt jedoch auch über das *Ausmaß des Wissens über einzelne Wörter* definiert (vgl. z.B. Nation 2001; Read 1998, 2004). Dabei wird der erste, eher quantitativ geprägte Aspekt des Wortschatzes häufig als *Wortschatzbreite* (WS-Breite) bezeichnet, der zweite als *Wortschatztiefe* (WS-Tiefe; vgl. Anderson & Freebody 1981). Messinstrumente zur Erfassung des Wortschatzes beschränken sich im Vorschulalter jedoch zumeist auf die Erfassung von WS-Breite (z.B. CITO-Wortschatztest: Konak, Duindam & Kamphuis 2005; PPVT: Dunn & Dunn 2007, deutsch: Lenhard, Lenhard, Segerer & Suggate 2015).

Bezüglich der vorschulspezifischen Wortschatzdiagnostik lassen sich daraus folgende Probleme ableiten: Erstens führt die oft eindimensionale Erfassung des Wortschatzes durch die weitgehende Vernachlässigung der Wortschatztiefe-Dimension sowie des bildungssprachlichen Wortschatzes möglicherweise zu einer eingeschränkten Konstruktvalidität solcher Verfahren bezogen auf schulkontextuell relevantes Wortwissen. Zweitens folgen daraus eine eingeschränkte Diagnostizierbarkeit und Konsequenzen für die Förderung solchen Wortschatzes. Und drittens können bildungserfolgsrelevante Forschungsfragen, wie etwa nach der Differenzierbarkeit unterschiedlicher Wortschatzaspekte und -formen sowie deren möglicherweise differenziellem Einfluss auf schulische Leistungen, bislang nicht beantwortet werden. Ziele der Studie sind daher die Entwicklung und erste Validierung eines Tests zur Erfassung von WS-Breite und WS-Tiefe von bildungssprachlich anschlussfähigem Wortschatz bei Vorschulkindern sowie eine erste Untersuchung der Dimensionalität des damit erfassten Konstrukts.

1. Zum Konstrukt des Wortschatzes

Ogleich bezüglich des Gesamtkonstrukts Wortschatz hinsichtlich der exakten Anzahl und Beschreibung konstruktrelevanter Dimensionen bislang keine einheitliche Definition existiert (z.B. zweidimensional: Anderson & Freebody 1981; dreidimensional: Henriksen 1999; Daller, Milton & Treffers-Daller 2007; Nation 2001; vierdimensional: Qian 2002), wird über die Modelle hinweg doch mindestens von einer Zweidimensionalität des Konstrukts ausgegangen. Dabei inkludieren die Modelle passive sowie aktive (auch als rezeptive und produktive bezeichnet) Wortschatzkenntnisse. Ferner wird angenommen, dass neben der bloßen Anzahl von Wörtern im mentalen Lexikon (WS-Breite) auch der Erwerb tiefergehenden Wissens über Wörter (WS-Tiefe) relevant für den Wortschatzerwerb ist.

Während WS-Breite einheitlich als die Anzahl bekannter Wörter definiert wird, existiert bislang keine einheitliche Definition von WS-Tiefe. Read (1998: 41) beschreibt diese recht allgemein als „[...] the extent to which learners are familiar with the range of meaning and uses of a set of target words“, bei Nation (2001) hingegen sind 12 der 18 Aspekte seines Wortschatz-Modells (z.B. semantische Bezüge zu anderen Wörtern, Assoziationen, Kollokationen) der Dimension WS-Tiefe zuzuordnen (vgl. Gyllstad 2013). Nach Qian (2002) wiederum beinhaltet diese Dimension sämtliche Charakteristika (z.B. phonemische, graphemische, morphemische, syntaktische, kollokationale, semantische, phraseologische) von Wörtern und Juska-Bacher und Jakob (2014) schlagen eine Unterteilung in die Bereiche semantisches, formales und relationales Wortwissen vor.

Ein wiederholt genanntes Merkmal von WS-Tiefe stellt die *kontextuelle Einbettung von Wörtern* dar. So bezeichnen etwa Schoonen und Verhallen (2008) WS-Tiefe als das *dekontextualisierte Wissen* über Wörter. Lerner werden demnach beim Spracherwerb mit dem Erlernen gänzlich neuer Wörter sowie mit neuen Bedeutungen bereits bekannter Wörter konfrontiert. Durch anhaltende Generalisierungs-, Kategorisierungs- und Abstraktionsprozesse werde die Bedeutung eines Wortes stetig vertieft und unterliege dabei einer zunehmenden Dekontextualisierung (213). Von unterschiedlichen Qualitäten von Wortwissen und kontextueller Einbettung als Quelle für das Wissen über Wörter geht auch Ruhl (1989) aus. Dabei beschreibt er die erste Wissensquelle als das Wissen um die inhärente lexikale Bedeutung eines Wortes; die zweite Quelle stellen die kontextuellen Hinweise eines Wortes dar, indem sich aus dem direkten semantischen Kontext von Wörtern deren Bedeutungen ableiten lassen. Auch Sternberg (1987) nimmt an, dass das Erlernen von Wörtern durch semantische Kontexthinweise, d.h. die semantische Einbettung von Wörtern, erfolgt. Zusammenfassend zeigt sich, dass die kontextuell-semantische Einbettung von Wörtern für den lexikalen Wissenserwerb und -abruf von großer Relevanz zu sein scheint. Diese wird jedoch in gängigen Instrumenten bislang wenig berücksichtigt.

Die Darstellung von WS-Tiefe und WS-Breite als unterschiedliche Dimensionen eines Konstrukts wird jedoch verschiedentlich auch kontrovers diskutiert (vgl. Read 2004; Schoonen & Verhallen 2008). So wird etwa, da sich WS-Breite auf das gesamte Vokabular bezieht, WS-Tiefe jedoch eine Charakteristik individueller Wörter darstellt, die Zugehörigkeit beider Dimensionen zum gleichen Konstrukt durchaus kritisch hinterfragt (vgl. Gyllstad 2013; Meara & Wolter 2004). Andere Autoren wiederum stellen in Frage, ob sich Wortschatzaspekte wie WS-Tiefe und WS-Breite überhaupt voneinander differenzieren lassen (Qian 2002; Vermeer 2001). Trotz Uneindeutigkeit der

Forschungslage zeigt sich jedoch zusammenfassend, dass der größere Teil der empirischen Befunde auf die Mehrdimensionalität des Konstrukts hinweist, wobei jedoch nicht von einer Unabhängigkeit der Dimensionen auszugehen ist. So fanden Tannenbaum, Torgesen und Wagner (2006), dass es sich bei WS-Breite und -Tiefe oder -Flüssigkeit um zwei voneinander differenzierbare, wenn auch latent hoch korrelierte Faktoren ($r=.87$) handelt. Vermeer (2001) fand bei Kindergartenkindern ähnlich hohe Korrelationen, schlussfolgerte jedoch daraus, dass beide Dimensionen nicht trennbar sind. Oullette (2006) hingegen konnte bei Viertklässlern differenzielle Zusammenhänge zwischen unterschiedlichen Aspekten der Lesekompetenz mit Aspekten der WS-Breite und -Tiefe aufzeigen, bei Tannenbaum (2008) wiederum weist für Kinder in der dritten Klasse ein einfaktorielles Wortschatz-Modell den besten Modellfit auf, für Siebtklässler hingegen ein zweifaktorielles Modell (WS-Flüssigkeit und WS-Tiefe/-Breite), und in einer Studie von Qian (1999) konnten mithilfe der Erfassung von WS-Tiefe 11 % der Varianz des Leseverständnisses von Studierenden zusätzlich zu der durch WS-Breite erklärten Varianz aufgeklärt werden.

2. Die Erfassung von Wortschatzbreite und -tiefe

Um mögliche sprachliche Defizite von Kindern bereits vor Schuleintritt diagnostizieren und so ggf. die Sprachentwicklung möglichst früh fördern zu können, wurden in vielen deutschen Bundesländern in den letzten Jahren flächendeckend Verfahren zur Erfassung des Sprachstandes von Vorschulkindern eingeführt (z.B. Lisker 2013). Jedoch wird kein einheitliches Instrumentarium verwendet, die verwendeten Verfahren weisen u.a. in Hinblick auf Qualität und Methode (z.B. Tests, Einschätzungs- und Beobachtungsverfahren) eine große Varianz auf. So definieren etwa Becker-Mrotzek et al. (2013) im Rahmen einer Studie des Mercator-Instituts Qualitätsstandards für Sprachstanderhebungsverfahren im Vorschulalter in insgesamt zehn Handlungsfeldern. Diese beinhalten etwa die Testgütekriterien Normierung, Reliabilität, Objektivität, Validität, zeitliche Anforderungen an die Kinder, Mehrsprachigkeit, die Berücksichtigung sprachlicher Basisqualifikationen, Spezifität der Diagnostik, Fehlerquote sowie Qualifikation der durchführenden pädagogischen Fachkräfte. Neugebauer und Becker-Mrotzek (2013: 4) überprüften die verwendeten Verfahren systematisch hinsichtlich der Erfüllung der Qualitätskriterien und konstatierten, dass – obgleich eine Verbesserung hinsichtlich der Erfüllung der Kriterien zu beobachten sei – an dieser Stelle noch immer großer Handlungsbedarf bestehe, unter anderem in den Bereichen Objektivität, Validität und Mehrsprachigkeit.

Neben standardisierten Verfahren wie dem CITO-Sprachtest (Konak et al. 2005) oder dem HSET (Grimm & Schöler 1991) existieren noch weitere standardisierte und normierte Verfahren, die den in der vorliegenden Studie fokussierten Wortschatz berücksichtigen (z.B. SET 5-10, Petermann 2012; PPVT, Dunn & Dunn 2007); bezüglich des passiven Wortschatzes wird jedoch meist nur die WS-Breite, d.h. das Kennen oder Erkennen von Wörtern, berücksichtigt. Diesbezüglich erfolgt die Operationalisierung üblicherweise über das Zuordnen von Wörtern zu vorgegebenen Wortdefinitionen oder über das Erkennen eines verbal vorgegebenen Wortes unter mehreren Bildern (vgl. z.B. PPVT; CITO-Sprachtest).

Der WS-Tiefe zuzuordnenden Aspekte werden zwar in einigen Instrumenten teilweise mit einbezogen, dann jedoch meist bezogen auf den aktiven Wortschatz, wie z.B. bei SET 5-10: Kategorienbildung (Wissen über semantische Relationen), AWST-R (Kiese-Himmel 2005): Wortbedeutungswissen oder HSET: Kategorienbildung, Korrektur semantischer Inkonsistenzen. Der WS-Tiefe-Aspekt der semantischen Kontextualisierung von Zielwörtern wurde bei der Erfassung passiven Wortschatzes bislang noch nicht gezielt berücksichtigt. In der Erwachsenenentstung (teilweise auch in der Primarstufe; z.B. Schoonen & Verhallen 2008) wird Wortschatztiefe fast ausschließlich anhand des Word Association Tests (WAT, Read 1998, 2004, bzw. ähnlicher Eigenentwicklungen, z.B. Oullette 2006) und damit über das Erkennen von Synonymen und Bilden von Assoziationen operationalisiert. Diese setzen jedoch Lesekompetenz voraus, weshalb deren Einsatz bei Vorschulkindern in dieser Form nicht möglich ist.

Des Weiteren zeigt sich, dass gängige standardisierte Verfahren für das Vorschulalter (z.B. CITO-Sprachtest, PPVT) – auch wenn zwischen diesen und dem neuen Verfahren aufgrund der fehlenden Trennschärfe zwischen alltags- und bildungssprachlichem Wortschatz im Vorschulalter Überschneidungen bezüglich einiger weniger

Zielwörter bestehen mögen – neben der fehlenden Berücksichtigung der Wortschatztiefe zumeist auf alltags-sprachlichen Wortschatz abzielen und keine gezielt auf den schulisch-fachlichen Kontext fokussierende Wortauswahl aufweisen.

Die vorliegende Studie soll daher einen ersten Beitrag dazu leisten, den Erwerb von passivem bildungssprachlichem Vokabular unter Einbezug der Wortschatztiefe bereits bei Vorschulkindern zu untersuchen. Dies soll durch die semantisch-kontextuelle Einbettung von Wörtern (vgl. Ruhl 1989; Schoonen & Verhallen 2008; Sternberg 1987), genauer durch die Einbettung von Zielwörtern in unterschiedliche, semantisch anomale bzw. nicht-anomale Kontexte, sowie die anschließende Beurteilung dieser Kontexte hinsichtlich semantischer Kohärenz durch Vorschulkinder operationalisiert werden (siehe Abschnitt 4.2).

3. Fragestellungen und Hypothesen

Aus dem derzeitigen Forschungsstand lassen sich mehrere Forschungsdesiderate ableiten: Zum einen ist eine Überprüfung der Dimensionalität des Wortschatz-Konstruktes speziell bei Kindern im Elementarbereich unter Verwendung von Maßen zur Erfassung von WS-Breite und WS-Tiefe anzustreben. Diesbezüglich interessieren sowohl der Zusammenhang zwischen den beiden angenommenen Wortschatz-Dimensionen Tiefe und Breite als auch eine explorative Betrachtung der WS-Tiefe-Dimension im Hinblick auf Existenz und Anzahl psychometrischer Subdimensionen. Zum anderen existieren bislang keine Erkenntnisse im Hinblick auf inkrementelle Effekte schulkontextuell-sachfachlich relevanten Wortschatzes von Vorschulkindern auf deren schulische Leistung in der Primarstufe. Eine dazu notwendige Voraussetzung – und damit das primäre Ziel der aktuellen Studie – ist jedoch die Entwicklung eines im Elementarbereich einsetzbaren Instruments zur Erfassung von WS-Breite und -Tiefe bei bildungssprachlich anschlussfähigem Vokabular. Dabei soll, entsprechend den gängigen Empfehlungen für pädagogisch-psychologische Testverfahren (z.B. APA 2014) sowie für Verfahren zur Erfassung des Sprachstandes im Vorschulalter (Becker-Mrotzek et al. 2013), das Instrument auf die Erfüllung der Testgütekriterien Objektivität, Reliabilität sowie verschiedene Aspekte der Validität hin überprüft werden. In Hinblick auf die Konstruktvalidität erfolgt dies dabei – dem ersten der o.g. Forschungsdesiderate entsprechend – anhand einer Untersuchung der Testdimensionalität. Da der größere Teil existierender empirischer Studien auf mindestens zwei voneinander abgrenzbare, jedoch mittelhoch bis hoch korrelierte Dimensionen hinweist, wird diesbezüglich die Annahme getroffen, dass sich der Test als mehrdimensional erweist, mindestens jedoch aus den beiden voneinander differenzierbaren, jedoch korrelierten Konstruktbereichen WS-Tiefe und WS-Breite besteht. Konkret wird angenommen, dass sich die Fit-Indices eines zweidimensionalen passiven Wortschatz-Modells im Rahmen einer explorativen Faktorenanalyse als besser erweisen als die eines eindimensionalen. Die Untersuchung externer Validität erfolgt mithilfe der Betrachtung von korrelativen Zusammenhängen zu weiteren Sprachmaßen (konvergente Validität) sowie allgemeiner, nonverbaler kognitiver Fähigkeiten (diskriminante Validität). Bezüglich der WS-Breite wird dabei ein hoher, jedoch nicht perfekter Zusammenhang zu Maßen des allgemeinen Wortschatzes erwartet, da das neue Testinstrument vor allem auf schwierigeren sowie bildungssprachlichen Wortschatz abzielt. Somit wird auch erwartet, dass sich das neue Instrument als schwieriger als existierende Maße erweist. Niedrigere Zusammenhänge werden hingegen mit anderen Sprachmaßen wie der phonologischen Bewusstheit sowie allgemeinen nonverbalen kognitiven Fähigkeiten erwartet. Bezüglich der WS-Tiefe können aufgrund der uneindeutigen theoretisch-empirischen Basis hinsichtlich erwartbarer Zusammenhänge zu externen Maßen keine Hypothesen formuliert werden. Diese Betrachtung erfolgt daher explorativ. Basierend auf empirischen Studien zu sprachlichen Kompetenzunterschieden (vgl. z.B. Schoonen & Verhallen 2008) wird des Weiteren erwartet, dass sich mono- und bilingual deutsch aufgewachsene Kinder bezüglich beider Maße zugunsten der monolingualen Gruppe unterscheiden (Known Group Validity); dies wird anhand von Kovarianzanalysen überprüft. Zusammenfassend lassen sich folgende Forschungsfragen ableiten: Besteht ein den Annahmen entsprechender Zusammenhang zwischen den mithilfe des Instruments erfassten Wortschatz-Dimensionen Tiefe und Breite? Erfüllt das entwickelte Instrument zur Messung der passiven Wortschatzkenntnisse in Tiefe und Breite die Testgütekriterien Objektivität, Reliabilität sowie interne und externe Validität?

4. Methode

4.1. Studie, Stichprobe und Durchführung

Die Studie erfolgte an N=150 monolingual deutsch (Sprachgruppe ML, n=57) und bilingual türkisch-deutsch (Sprachgruppe BL, n=93) aufgewachsenen 4- bis 7-jährigen Vorschulkindern (Tab. 1) in 18 Kindergärten des Rhein-Main-Gebiets. Die bilingual türkisch-deutschen Kinder wurden nach Angaben der pädagogischen Fachkräfte der Kindertageseinrichtungen sowie anhand eines Elternfragebogens zur familiären Sprachpraxis identifiziert. Hierbei wurde über die Indikatoren der Familiensprache (deutsch/türkisch) sowie des Geburtslands der Kinder und der Eltern der Migrationshintergrund festgestellt (zur Stichprobenbeschreibung vgl. auch Cinar & Hardy 2017).

Tab. 1: Altersverteilung der Kinder

	4-Jährige	5-Jährige	6-Jährige	7-Jährige	M/SD (Jahre)	M/SD (Monate)
Gesamtstichprobe (N=150)	17	73	46	14	5,38 (0,81)	68,11 (7,07)
Teilstichprobe BL (n=93)	13	46	32	2	5,23 (0,90)	68,25 (7,41)
Teilstichprobe ML (n=57)	4	27	14	12	5,24 (0,71)	67,85 (6,42)

Anmerkung: ML=monolingual Deutsch; BL=bilingual deutsch-türkisch; ML=44 % männlich; BL=44 % männlich).

Die Stichprobe entstammt einer Interventionsstudie¹, deren Ziel die Entwicklung und Überprüfung eines Wortschatztrainings (20 Interventionssitzungen) für bildungssprachlich anschlussfähiges Vokabular für bilingual türkisch-deutsche Vorschulkinder war. Dabei wurden Zielwörter in auditiv vorgegebene Geschichten eingebettet, welche im Rahmen eines 5-wöchigen Trainings entweder in Deutsch, Türkisch oder beiden Sprachen in drei Untersuchungsbedingungen präsentiert wurden (drei BL-Experimentalgruppen, trainierte BL-Kontrollgruppe, untrainierte ML-Kontrollgruppe; zum Design siehe Cinar & Hardy 2017). Ein Motiv für die Entwicklung des dargestellten Messinstruments war dessen Einsatz zur Erfassung von WS-Breite sowie von WS-Tiefe im Rahmen der Intervention.

WS-Breite wurde sowohl im Prä- als auch im Posttest erhoben, WS-Tiefe aus ökonomisch-zeitlichen Gründen sowie aus Gründen der Zumutbarkeit und Testlänge lediglich im Posttest, jeweils von ausführlich geschulten studentischen Testleiterinnen (Grundschullehramt, Psychologie, Erziehungswissenschaften) und in Einzeltestungen. Obgleich sich varianzanalytisch bzgl. des Vergleichs des WS-Zuwachses (von Prä- zu Posttest) von vier Untersuchungsgruppen ein tendenziell signifikanter, spezifischer Effekt für eine rein deutschsprachige Intervention im Vergleich zu einer trainierten BL-Kontrollgruppe zeigt, findet sich regressionsanalytisch unter Kontrolle nonverbaler Intelligenz und des Alters insgesamt kein signifikanter Effekt für die Interventionsgruppenzugehörigkeit der bilingual türkisch-deutschen Kinder ($p=.905-.157$) (Hardy & Cinar eingereicht), weshalb für die vorliegenden Analysen ein Globalvergleich zwischen der BL- und ML-Gruppe durchgeführt werden konnte.

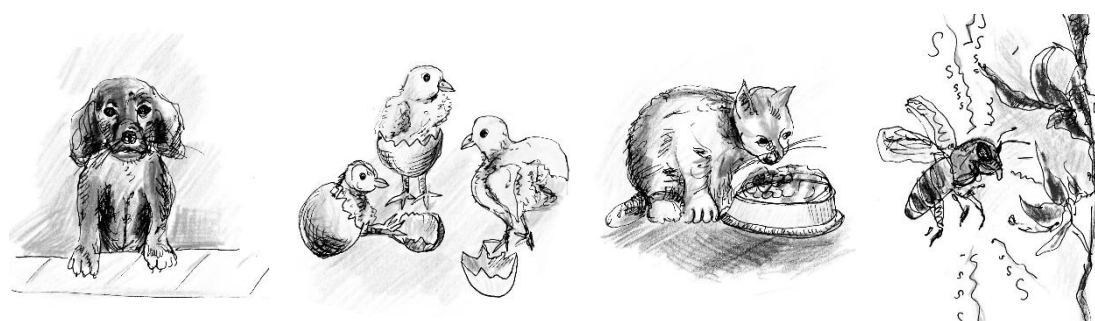
Auch ändern sich die Ergebnisse hinsichtlich der Dimensionalität der WS-Breite nicht relevant von Prä- zu Posttest. Daher wurden aufgrund der zeitlichen Nähe die Posttest-Daten für die Wortschatzmaße WS-Breite und WS-Tiefe verwendet. Ferner wurden im Prätest standardisierte Maße des allgemeinen passiven Wortschatzes in Deutsch und Türkisch, der phonologischen Bewusstheit in Deutsch (beides Skalen des CITO-Sprachtests; Konak et al. 2005), nonverbaler kognitiver Fähigkeiten (K-ABC; Melchers & Preuß 2009), und per Elternfragebogen Sprache (BL/ML) und Alter erhoben (vgl. auch Cinar & Hardy 2017). Bei den hier eingesetzten Verfahren handelt sich um gut überprüfte und normierte Instrumente für das Vorschulalter, die der Mehrsprachigkeit der Stichprobe Berücksichtigung schenken, da der K-ABC ein nonverbaler Intelligenztest und der CITO-Sprachtest ein deutsch-türkisches Sprachdiagnostikinstrument ist.

4.2. Test- und Skalenkonstruktion zur Messung von Wortschatztiefe und -breite

Auswahl des Zielwortschatzes: Es wurden insgesamt zwei Wortschatz-Skalen zur Messung der WS-Breite, d.h. das Kennen/Erkennen von Wörtern, sowie der Wortschatztiefe, d.h. Wortwissen, mit jeweils 60 Items konstruiert. Die verwendeten Zielwörter wurden dabei wie folgt ausgewählt: Da die Testkonstruktion vor dem Hintergrund der Rahmenstudie (Cinar & Hardy 2017) auch auf die Einsatzfähigkeit des Instruments zur Überprüfung der Effektivität eines Wortschatztrainings zur Förderung des Wortschatzerwerbs von bilingual türkisch-deutschen Vorschulkindern abzielte, erfolgte eine erste gröbere Vorauswahl zu verwendender Zielwörter an die Inhalte des Interventionsmaterials sowie die dort verwendeten Hörtexte angelehnt. Dabei stand im Vordergrund, bereits bei Vorschulkindern passiven, bildungssprachlich anschlussfähigen Wortschatz abzubilden, welcher insbesondere an sachunterrichtliche sowie schulkontextuell relevante Themen des ersten Schuljahres anschlussfähig sein sollte. Entsprechende Begriffe aus den verschiedenen Themenfeldern wurden gemäß folgenden Aspekten selektiert und pilotiert:

Basierend auf dem „Perspektivrahmen Sachunterricht“ (GDSU 2012: 5) wurden aus unterschiedlichen gesellschafts- und natur-/umweltbezogenen Themenfeldern (z.B. Tiere/Pflanzen/Natur, Technik/Arbeit, Räume/Naturgrundlagen) Wörter aus schulischer und vorschulischer Literatur ausgewählt. Dies führte zu einer Auswahl von insgesamt 153 Wörtern, die basierend auf von der GDSU (2012) benannten Themenfeldern, sowie – in Anlehnung an das von Bialystok et al. (2010: 5) beschriebene Vorgehen bei der Unterteilung von Wörtern des PPVT in primär „schoolrelevant contexts“ und „home contexts“ – in schulkontextlich relevante Kategorien eingeordnet wurden (siehe Anhang A). Die Einordnung in die Kategorien erfolgte dabei durch zwei Rater (Konsensurteil; Studienleitung sowie Projektmitarbeiterin). Da kein Korpus bezüglich eines bildungssprachlich anschlussfähigen Wortschatzes im Vorschulalter vorliegt, wurden die ausgewählten Wörter anhand von Wörterbüchern zum Grundwortschatz für das Kindergarten- und Schulalter abgeglichen und sichergestellt, dass sie über den Grundwortschatz für das Kindergartenalter hinausgehen. Die Wörter wurden in einer Pilotierungsstudie an 35 bilingual türkisch-deutschen und 21 monolingual deutschen Vorschulkindern hinsichtlich ihres Unbekanntheits- und Schwierigkeitsgrads getestet. Vor dem Hintergrund der Überschneidung mit Alltagskontexten wurden die schwierigsten 60 Wörter (49 Nomen, 8 Verben, 3 Adjektive; Anhang A) ausgewählt.

Wortschatzbreite: Die passive WS-Breite wurde anhand der Anzahl korrekt erkannter Zielwörter operationalisiert. Im Rahmen einer Einzeltestung (ca. 20 Min.) wurden dazu den Kindern die insgesamt 60 Zielwörter einzeln vorgelesen. Jedes Zielwort wurde dabei in Form eines Bildes präsentiert und musste unter vier Antwortmöglichkeiten (Multiple-Choice-Format) korrekt erkannt werden. Die Distraktoren wurden anhand verschiedener visueller, phonemischer und semantischer Ähnlichkeitskriterien entwickelt (Abb. 1), die Anordnung sowie Zusammensetzung der Distraktoren wurde nach vorhergehender Überprüfung im Rahmen der Pilotierungsstudie variiert.



Zielwort Welp; Ähnlichkeiten Distraktoren: Küken=semantisch, Katze=visuell, Wespe=phonemisch

Abb. 1: Beispielitem Wortschatzbreite

Wortschatztiefe: Als Maß für WS-Tiefe wurde der oben beschriebene WS-Tiefe-Aspekt „semantisch-kontextuelle Einbettung von Wörtern“ fokussiert. Aus Gründen der Zumutbarkeit wurden zunächst aus den insgesamt 60 Zielwörtern des WS-Breite-Tests – möglichst gleichverteilt über die oben dargestellten schulkontextlichen Kategorien sowie inhaltlich angelehnt an die in der Intervention verwendeten Hörspiele – jeweils ein Zielwort pro Interventionsinhalt ausgewählt. Dies ergab insgesamt einen Umfang von 20 Zielwörtern zur Erfassung der WS-Tiefe (siehe Anhang A). Im nächsten Schritt wurde dann der Grad der semantisch-kontextuellen Einbettung der Zielwörter durch deren jeweilige Einbettung in sowohl semantisch anomale als auch semantisch nicht-anomale Satzkontexte operationalisiert (siehe Beispielsätze: Tab. 2).

Semantisch anomale Kontexte zeichneten sich dabei dadurch aus, dass ein Satz zwar grammatikalisch korrekt gebildet wurde, jedoch der semantische Kontext keinerlei Passung zum jeweiligen Zielwort aufwies („Quatsch-Satz“). Ein semantisch nicht-anomaler Kontext zeichnete sich dabei entsprechend dadurch aus, dass es sich um eine für das jeweilige Zielwort inhaltlich-semantisch sinnvolle Einbettung handelt und der Satz folglich semantisch korrekt ist. Aufgabe der Kinder bei der Testbearbeitung sollte es sein, zu erkennen, ob es sich jeweils um einen semantisch nicht anomalen oder semantisch anomalen Satz handelte. Basierend auf den oben dargestellten theoretischen Annahmen wurde dabei davon ausgegangen, dass das Erkennen von semantisch anomalen bzw. nicht-anomalen Sätzen lexikalisches Wissen über die zentralen Wörter eines Satzes erfordert, dass also ein gewisses inhärentes Wissen über mögliche semantische Verknüpfungen und semantische Kontexte eines Wortes, d.h. WS-Tiefe, Voraussetzung für eine korrekte Beantwortung dieser Aufgabe sind. Obgleich den Kindern die jeweiligen Zielwörter nicht explizit genannt wurden, wurden diese jeweils an inhaltlich so zentralen Stellen des Satzes untergebracht, dass – so die Annahme – die Frage nach semantischer Korrektheit eines Satzes ohne tiefergehendes Wortwissen hinsichtlich möglicher semantischer Verknüpfungen und Relationen des jeweiligen Zielwortes nicht korrekt beantwortet werden konnte. Dabei wurde angenommen, dass semantisch anomale Sätze aufgrund fehlender semantisch-kontextueller Hinweise auf die Bedeutung eines Wortes einen geringeren Grad an kontextueller Einbettung aufweisen und somit für das Erkennen eines Zielwortes ein tiefergehendes, inhärentes Wissen um die Bedeutung eines Wortes, also ein höherer Grad an WS-Tiefe, notwendig sein sollte als bei semantisch nicht-anomalen Sätzen.

Neben der Unterteilung in semantisch anomale und semantisch nicht-anomale Kontexte wurden zusätzlich, um mögliche interventionsspezifische inhaltliche Kontexteffekte auf das Wortwissen zu prüfen, für einen Teil der Sätze jeweils 1-2 weitere Wörter aus Interventionsinhalten verwendet.

Folgende 60 Items wurden konstruiert: Da aus Belastungsgründen im Rahmen der Studie nicht die Gesamtheit des Testmaterials entsprechend angereichert werden konnte, wurden zunächst für 10 zufällig ausgewählte Wörter jeweils 4 Sätze (=40 Items) formuliert (jeweils 2 pro semantischer Bedingung, davon jeweils einmal inhaltlich interventionsgebunden/frei). So wurden beispielsweise für die beiden Zielwörter *Welpen* und *Verwandte* jeweils folgende 4 Sätze formuliert:

Tab. 2: Semantisch anomale/nicht-anomale Sätze in unterschiedlichen Kontexten

Semantisch anomal:

Bsp. I (*Zielwort Welpen*): *Welpen* gehen morgens in die Schule; Die *Welpen* erzählen dem Tierarzt eine Geschichte (zusätzlicher Interventionskontext: Tierarzt)

Bsp. II (*Zielwort Verwandte*): Morgens ziehe ich mir meine *Verwandten* an; Hugos *Verwandte* sind aus Schokoladeneis gemacht (zusätzlicher Interventionskontext: Schokoladeneis, Hugo)

Semantisch nicht-anomal:

Bsp. I: Die Kinder eines Hundes werden *Welpen* genannt; Der Tierarzt impft die *Welpen* (zusätzlicher Interventionskontext: Tierarzt, impfen)

Bsp. II: Mama, Papa, Omi und Opi sind meine *Verwandten*; Hugos *Verwandte* leben in einer Burg (zusätzlicher Interventionskontext: Burg, Hugo)

Für die restlichen 10 Wörter wurden jeweils 2 Sätze (für je 5: anomal / nicht-anomal, für je 5: inhaltlich interventionsgebunden / frei) gebildet. Den Kindern wurden in Einzeltestungen (ca. 20 Minuten) die insgesamt 60 Sätze (davon 30 semantisch anomal / 30 nicht-anomal) vorgelesen. Nach jedem Satz wurden die Kinder gefragt, ob es sich ihrer Ansicht nach um einen „Quatschsatz“ oder um einen „richtigen“ Satz handele. Die Aufgabe wurde zu Beginn anhand eines Beispielitems verdeutlicht. Um das Verständnis der Kinder zu Testbeginn sicherzustellen, wurde, falls notwendig, das Beispiel seitens der Testleitung mehrfach erneut erläutert. Zur Beantwortung der Items wurden den Kindern dann verschiedenfarbige, an das vertraute Ampelsystem angelehnte Smileys vorgelegt (grün/lächelnd= „richtiger“ (semantisch korrekter) Satz, rot/traurig= (semantisch anomaler) „Quatschsatz“). Im Rahmen einer qualitativen Pilotierung (N=10) wurde das Aufgabensetting (Verständlichkeit der Aufgabenstellung, Testlänge) überprüft, indem der Test durchgeführt und die Kinder bezüglich der Verständlichkeit der Items von der Testleitung befragt wurden. Ferner wurden die Eindrücke der Testleitung einbezogen. Basierend auf diesen Ergebnissen wurde eine Überarbeitung der Items zu diesem Zeitpunkt als nicht notwendig erachtet. Als Maß für die WS-Tiefe diente die Anzahl korrekt gelöster Items.

5. Ergebnisse

5.1. Dimensionalität des Tests und interne Validität

Zur Beantwortung der ersten Fragestellung bezüglich der Dimensionalität sowie der Konstruktvalidität des Tests wurden explorative Faktorenanalysen (EFA, oblique) mit MPlus 7 (Muthén & Muthén 1998-2012) durchgeführt²; auf die Durchführung konfirmatorischer Faktorenanalysen wurde aufgrund der Stichprobengröße verzichtet. Bezüglich des Gesamttests „Wortschatz“ (120 Items) zeigt sich (Tab. 3) den Annahmen entsprechend ein deutlich schlechterer Modellfit des einfaktoriellen gegenüber dem zweifaktoriellen Modell ($r(\text{Faktor 1(F1)/Faktor 2 (F2)})=.37, p<.05$; Items Faktor 1: 28 semantisch anomale sowie 9 semantisch nicht-anomale WS-Tiefe-Items wiesen die höchste Itemladungen auf Faktor 1 auf; signifikante Nebenladungen von 9 weiteren semantisch nicht-anomalen WS-Tiefe-Items; Items Faktor 2: höchste Itemladungen von 54 WS-Breite-Items sowie 21 semantisch nicht-anomalen WS-Tiefe-Items). Zwar weisen die Fit-Indices für ein dreifaktorielles Modell minimal bessere Werte auf ($r(\text{F1/F2})=.26, r(\text{F2/F2})=.591; r(\text{F1/F3})=.12$; höchste Itemladungen F1: 21 semantisch anomale WS-Tiefe-Items, 16 semantisch nicht-anomale WS-Tiefe-Items; F2: 12 semantisch nicht-anomale WS-Tiefe-Items, 24 WS-Breite-Items, 6 semantisch anomale WS-Tiefe-Items; F3: 35 WS-Breite Items, 3 Items WS-Tiefe semantisch anomal), dieses ist jedoch bezüglich der inhaltlichen Interpretierbarkeit weniger eindeutig als das zweifaktorielle Modell, in dem jeweils ein Faktor von WS-Breite und WS-Tiefe-Items dominiert wird. Auffällig bei beiden Lösungen ist, dass die semantisch nicht-anomalen WS-Tiefe-Items auf dem ersten Faktor negativ (Haupt- oder Nebenladungen), die semantisch anomalen Items jedoch positiv laden. Mögliche Gründe werden in der Diskussion aufgegriffen. Vor diesem Hintergrund sowie der nicht eindeutig interpretierbaren Dimensionalität der semantisch nicht-anomalen WS-Tiefe-Items werden in den folgenden Analysen beide semantischen Bedingungen sowohl gemeinsam als auch separat betrachtet.

Tab. 3: Fit-Indices EFA Gesamttest „Wortschatz“

Modell	RMSEA	CFI	χ^2/df	χ^2	p	df	SRMR
1-faktoriell	0.022	0.871	1,07	7536.100	.000	7020	0.143
2-faktoriell	0.013	0.956	1,03	7077.552	.067	6901	0.119
3-faktoriell	0.012	0.962	1,02	6934.873	.097	6783	0.114
4-faktoriell	0.012	0.964	1,02	6808.538	.109	6666	0.111
5-faktoriell	0.012	0.967	1,02	6682.975	.123	6550	0.107

5.2. Objektivität, Reliabilität und Schwierigkeit

Durchführungs- sowie Auswertungsobjektivität waren aufgrund der ausführlichen Schulung der Testleiterinnen sowie des standardisierten Materials und des standardisierten Vorgehens bei der Testung gegeben. Die Auswertungsobjektivität wurde durch die einfache und eindeutige Zuordnung der Antworten der Kinder in „richtig“ und „falsch“ durch die Testleitung gegeben. Für jede korrekte Antwort wurde pro Item ein Punkt vergeben. Zur Überprüfung möglicher Reihenfolgeeffekte wurde die Reihenfolge der WS-Tiefe-Testitems anhand von vier Itemblöcken in vier Testversionen systematisch variiert. Eine univariate Varianzanalyse zeigte diesbezüglich keinen Haupteffekt „Testversion“ ($df=3$, $F=0,688$, $p=.561$, $\eta^2=.016$).

Tab. 4: Deskriptive Statistiken WS-Breite /-Tiefe und CITO-Sprachtest

	WS-Breite			WS-Tiefe			CITO PW	CITO PW (Norm)	CITO PB	CITO PB (Norm)
	<i>M</i>	(<i>SD</i>)	min-max	<i>M</i>	(<i>SD</i>)	min-max	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>
Gesamt	24.08	(12.12)	8-50	37.94	(7.68)	23-54	45.70	42.47	24.28	19.83
ML	35.87	(10.32)	11-50	44.07	(7.20)	28-54	53.82	49.10	27.56	21.81
BL	17.11	(6.55)	8-39	34.23	(5.24)	23-47	40.67	31.26	22.25	16.32

Anmerkungen: WS-Breite=Wortschatzbreite (60 Items); WS-Tiefe=Wortschatztiefe (60 Items); CITO-Test: Norm=Mittlere Werte CITO-Normstichproben (4-6 J.): passiver Wortschatz (PW)=60 Items, Phonologische Bewusstheit (PB)=30. ML=monolingual deutsch, BL=bilingual deutsch-türkisch; WS-Tiefe semantisch anomal: $M(ML)=24,3$ ($SD=4,16$), $M(BL)=18,57$ ($SD=5,07$); semantisch nicht-anomal: $M(ML)=18,67$ ($SD=5,07$), $M(BL)=15,81$ ($SD=4,52$).

Im Mittel wurden von der Gesamtstichprobe $M=24.08$ der 60 WS-Breite-Items sowie $M=37.94$ der 60 WS-Tiefe-Items korrekt beantwortet. Dabei zeigte sich deskriptiv bezüglich beider Maße wie erwartet ein deutlicher Mittelwertunterschied zugunsten der ML-Gruppe (Tab. 4). Entsprechend der Fragestellung wurde ferner die Reliabilität der Skalen (interne Konsistenz für WS-Tiefe und WS-Breite, zusätzlich die Retest-Reliabilität für WS-Breite aufgrund von zwei Messzeitpunkten) für die Gesamtstichprobe sowie für die beiden Substichproben berechnet. Die 60 Items des WS-Breite-Tests wiesen dabei im Prätest vor der Intervention eine innere Konsistenz (Cronbach's Alpha) von $\alpha=.86$, im Posttest von $\alpha=.93$ (ML: $\alpha=.93$; BL: $\alpha=.75$), sowie eine Retest-Reliabilität von $r=.90$ auf. Die innere Konsistenz des WS-Tiefe-Tests erwies sich größtenteils als befriedigend ($\alpha=.76$; (ML: $\alpha=.83$; BL: $\alpha=.5$ nicht ausreichend)), die Reliabilitäten der WS-Tiefe-Subskalen (semantisch nicht-anomal: $\alpha=.79$ (ML: $\alpha=.86$; BL: $\alpha=.67$; semantisch anomal: $\alpha=.83$ (ML: $\alpha=.89$; BL: $\alpha=.76$)) lagen jedoch darüber. Die Reliabilität der Testskalen ist damit größtenteils als befriedigend bis sehr gut zu interpretieren. Die Itemschwierigkeiten wiesen nicht auf extreme Boden- oder Deckeneffekte hin (0=falsch, 1=richtig; Anzahl korrekter Antworten/Gesamte Antworten; WS-Breite: 0.11-0.74; WS-Tiefe: 0.41-0.8), was als erster Hinweis auf eine großen Teils gute Differenzierungsfähigkeit der Items gewertet werden kann.

Des Weiteren zeigte sich, dass im WS-Tiefe-Test von beiden Sprachgruppen (ML/BL) durchschnittlich eine größere Anzahl Items korrekt beantwortet wurde als im WS-Breite-Test. Dies kann darauf zurückzuführen sein, dass – wie im Rahmen der Itemkonstruktion beschrieben wurde – im WS-Tiefe-Testteil wiederholt die gleichen Zielwörter – wenn auch in unterschiedlichen Kontexten – vorgegeben wurden. Im Vergleich zum allgemeinen deutschen CITO-Wortschatztest zeigte sich, dass sowohl die mono- als auch die bilingual aufgewachsenen Kinder in der vorliegenden Studie deutlich über den im CITO-Manual angegebenen Normwerten lagen (Tab. 4). Ferner weist der neu entwickelte Test den Erwartungen entsprechend insgesamt eine größere Schwierigkeit auf als der allgemeine deutsche CITO-Wortschatztest, d.h. es wurde eine geringere Anzahl Items korrekt beantwortet (Tab. 4).

Ob des dichotomen Formats der Items wurde zunächst betrachtet, inwieweit sich für jedes Kind die tatsächlich erfolgte Anzahl korrekter Antworten pro Skala von der bei zufälliger Beantwortung zu erwartenden Häufigkeit

(15 korrekte Antworten pro Skala) unterscheiden. Diese wurde von einem Großteil der Kinder hinsichtlich beider Skalen deutlich überschritten: Bezüglich der semantisch nicht-anomalen Items lagen 66,7 % der Kinder darüber, bezüglich der semantisch anomalen Items sogar 77,1 %.

Hinsichtlich der Schwierigkeit (mittlere Anzahl korrekt gelöster Items) der beiden semantischen Bedingungen wurde dann überprüft, inwieweit sich die beiden WS-Tiefe-Skalen (semantisch anomal/ semantisch nicht-anomal) diesbezüglich unterscheiden. Hier zeigten sich rein deskriptiv zunächst Mittelwertunterschiede ($M(\text{semantisch nicht-anomal})=17.23$; $M(\text{semantisch anomal})=20.72$). Im Rahmen einer multivariaten Varianzanalyse mit Messwiederholung (Innersubjektfaktor semantischer Kontext; Zwischensubjektfaktor Sprache; Kontrollvariablen: Alter, nonverbale Intelligenz) zeigte sich entgegen der Annahmen allerdings kein Haupteffekt des Messwiederholungsfaktors semantischer Kontext ($df=1$; $F=1,056$; $p=.308$; $\eta^2=.018$). Ferner existierten keine signifikanten Wechselwirkungen des Messwiederholungsfaktors mit den Variablen Sprache (BL/ML), Alter oder nonverbale Intelligenz. Ein erwartbarer Haupteffekt zeigte sich hingegen für den Zwischensubjektfaktor Sprache zugunsten der ML-Gruppe ($df=1$; $F=15,742$; $p<.01$; $\eta^2=.216$). Auch zeigte sich für die Gesamtstichprobe kein Mittelwertunterschied bezüglich der freien und der interventionsgebundenen Kontexte ($t=-1.213$; $p=.23$). Dies deutet darauf hin, dass die Art des zur Einbettung eines Zielwortes verwendeten Kontextes entgegen der Annahme kaum Einfluss auf die Schwierigkeit zu haben scheint.

5.3. Externe Validität: Zusammenhänge zu sprachlich-kognitiven Maßen

Aufgrund des großen Altersranges (siehe Tab. 1) der Stichprobe werden altersbereinigte Partialkorrelationen berichtet. Wie erwartet, weist in der Gesamtstichprobe das WS-Breite-Maß eine hohe, jedoch keine perfekte Korrelation zu einem Maß des allgemeinen deutschen passiven Wortschatzes (CITO-Wortschatztest) auf (Tab. 5). Gleichfalls den Annahmen entsprechend finden sich signifikante, jedoch niedrigere Korrelationen zwischen WS-Breite und der phonologischen Bewusstheit sowie nonverbalen kognitiven Fähigkeiten. Das explorativ untersuchte gesamte WS-Tiefe-Maß weist ähnliche Korrelationsmuster mit den externen Maßen wie die WS-Breite-Skala auf. Diese Korrelationen sind bei beiden einzelnen WS-Tiefe-Skalen etwas niedriger, jedoch ähnlich, die niedrigsten Zusammenhänge weist die semantisch nicht-anomale Skala auf.

Tab. 5: Altersbereinigte Korrelationen WS-Tiefe und -Breite mit Sprach- und kognitiven Maßen

	WS-Breite	WS-Tiefe Gesamt	WS-Tiefe S-A	WS-Tiefe S-NA	Deutsch CITO-PW	Deutsch CITO-PB
WS-Tiefe Gesamt	.80**					
BL	.519**					
ML	.714**					
WS-Tiefe S-A	.589**	.674**				
BL	.222(*)	.566**				
ML	.601**	.529**				
WS-Tiefe S-NA	.486**	.669**	-.098			
BL	.324**	.479**	-.45**			
ML	.378*	.766**	-.14			
Deutsch CITO-PW	.688**	.665**	.659**	.233*		
BL	.261*	.410**	.469**	-.056		
ML	.668**	.650**	.606**	.300 (*)		
Deutsch CITO-PB	.405**	.322**	.370**	.061	.493**	
BL	.134	.247*	.315**	-.068	.358**	
ML	.223	.003	.034	-.022	.226	
K-ABC-NV	.399**	.425**	.347**	.224*	.449**	.346**
BL	.205(*)	.246*	.216(*)	.037	.293*	.210(*)
ML	.362*	.474**	.198	.403*	.334*	.287(*)

Anmerkungen: Pearson Produkt-Moment-Korrelationen; (*) $p<.1$; * $p<.05$; ** $p<.01$; BL=bilingual türkisch-deutsch; ML=monolingual deutsch; KV=Alter; $r(\text{Alter})$ mit: WS-Breite: $r=.251$ ($p=.003$); WS-Tiefe: $r=.274$ ($p=.001$); CITO PW: $.391$ ($p<.001$); CITO PB: $r=.82$ ($p=.032$); K-ABC: $.047$ ($p=.68$). SA=semantisch anomal, S-NA=semantisch nicht-anomal.

Die BL-Gruppe weist insgesamt zwar ein ähnliches Korrelationsmuster, jedoch durchgehend niedrigere Korrelationen als die ML-Gruppe auf; auffällig ist ferner die signifikant negative Korrelation zwischen den beiden einzelnen WS-Tiefe-Skalen bei den bilingualen Kindern. Eine detaillierte Analyse der Rohwertdifferenzen der beiden semantischen Bedingungen zeigt zwar, dass sich beide Gruppen hinsichtlich des mittleren Betrags der Differenz zwischen den beiden WS-Tiefe-Skalen nicht unterscheiden ($ML=6,667$, $BL=6,53$; $df = 1$; $F=.021$; $p=.89$, $\eta^2 = < 0,001$), die Kinder in der bilinguale Gruppe jedoch *entweder* einen besseren Wert auf der semantisch nicht-anomalen WS-Tiefe-Skala (40 %) *oder* der semantisch anomalen WS-Tiefe-Skala (54,44 %, 5,5 % gleiche Werte) aufweisen, während bei der monolingualen Gruppe 72,78 % einen besseren Wert bezüglich der semantisch anomalen Sätze aufzeigen. Dies könnte möglicherweise auf unterschiedliche Itemlösungsstrategien hinweisen.

5.4. Validität: Known Group Validity

Anhand einer univariaten Kovarianzanalyse wurde überprüft, ob sich die erwarteten und deskriptiv gefundenen Mittelwertsunterschiede zwischen der ML- und BL-Gruppe bestätigen lassen (Tab. 6). Bezüglich aller Wortschatzmaße zeigte sich, wie erwartet, ein Haupteffekt für den Faktor Sprache (BL/ML) (näherungsweise signifikant für WS-Tiefe semantisch nicht-anomal), außerdem fanden sich signifikante Effekte für die beiden Kontrollvariablen (nonverbale Intelligenz; Alter) sowie eine Wechselwirkung zwischen Alter und Sprache (BL/ML), die auf eine unterschiedliche Verteilung der 4- ($n=17$ ($n(ML)=4$, $n(BL)=13$) bzw. 7-Jährigen ($n=13$ ($n(ML)=11$, $n(BL)=2$) zurückzuführen ist. Der Faktor Sprache (BL/ML) klärt für die WS-Breite- sowie die semantisch anomale WS-Tiefe-Skala deutlich den größten Varianzanteil auf, für die Gesamtskala WS-Tiefe sind die Effekte von Sprache und Alter etwa gleich groß. Bezüglich der semantisch nicht-anomalen Skala weist hingegen das Alter einen größeren Effekt auf.

Tab. 6: Kovarianzanalysen WS-Breite und WS-Tiefe

	WS-Breite			WS-Tiefe								
	<i>F</i>	<i>p</i>	η^2	Gesamt			Sem. anomal			Sem. nicht-anomal		
	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2
Sprache	64.41	.00	.35	27.12	.00	.19	18.5	.00	.14	2.92	.09	.02
Alter	13.4	.00	.25	10.09	.00	.21	3.4	.02	.08	3.04	.03	.07
K-ABC NV	8.69	.00	.05	11.18	.00	.09	4.16	.04	.04	1.85	.18	.02
Sprache*Alter	4.41	.01	.10	3.12	.03	.08	.37	.78	.00	4.24	.01	.10

Anmerkungen: R^2 Wortschatzbreite=.71; R^2 Wortschatztiefe=.57; R^2 WS-Tiefe semantisch anomal=.38; R^2 WS-Tiefe semantisch nicht-anomal=.26; $n(ML)=42$; $n(BL)=74$.

Diese Ergebnisse sprechen bezüglich der meisten Skalen für die Validität des Instruments, da die theoretisch erwartbaren Unterschiede aufgezeigt werden und so zwischen unterschiedlichen Gruppen differenziert werden kann. Bezüglich der semantisch nicht-anomalen WS-Tiefe-Skala sind die Ergebnisse jedoch nicht eindeutig interpretierbar, da der Faktor Sprache hier lediglich näherungsweise signifikant wird, und das Alter einen größeren Effekt aufzuweisen scheint.

6. Diskussion

Zur Konstruktion des Tests zur Messung der WS-Tiefe und WS-Breite bei Vorschulkindern wurden 60 für den Schul- und Sachunterrichtskontext relevante Wörter mit einer für Vorschulkinder hohen Itemschwierigkeit ausgewählt sowie Skalen zur Erfassung von sowohl WS-Breite als auch WS-Tiefe entwickelt. Erste Ergebnisse weisen insgesamt auf die Objektivität, Reliabilität sowie teilweise auf interne und externe Validität des neu entwickelten Testinstruments hin; dies erfüllt somit in Teilen die von Becker-Mrotzek et al. (2013) definierten Qualitätskrite-

rien. Dies gilt auch für die zeitlichen Anforderungen an die Kinder sowie die Berücksichtigung von Mehrsprachigkeit. Bezüglich prognostischer Validität, Normierung und diagnostischer Spezifität können anhand der vorliegenden Daten keine Aussagen getroffen werden.

Hinsichtlich verschiedener Ergebnisse bezüglich der WS-Tiefe-Skala ist das neu konstruierte Instrument jedoch kritisch zu diskutieren. Bezüglich der eingangs formulierten Forschungsfrage nach der Dimensionalität des gesamten Wortschatztests zeigen sich zwar erwartungsgemäß bessere Fit-Indices für sowohl eine zwei- als auch eine dreidimensionale Lösung, was hypothesenkonform darauf hinweist, dass es sich – bezüglich der hier untersuchten 4- bis 7-Jährigen – um ein mehrdimensionales WS-Konstrukt handelt, und als erster Hinweis auf Konstruktvalidität gedeutet werden kann. Ferner deuten beide Faktorenlösungen auf jeweils von WS-Breite- bzw. WS-Tiefe-Items dominierte, wenn auch nicht ganz eindeutig interpretierbare Faktoren hin. Jedoch zeigt sich, dass lediglich ein Teil der semantisch anomalen WS-Tiefe-Items (positiv) und der semantisch nicht-anomalen WS-Tiefe-Items (negativ) auf einem gemeinsamen Faktor laden, was – gemeinsam mit den deutlich besseren Reliabilitäten für getrennte WS-Tiefe-Skalen sowie den unerwarteterweise nicht beobachtbaren Schwierigkeitsunterschieden zwischen semantisch anomalen und semantisch nicht-anomalen Items – auf eine Mehrdimensionalität der beiden WS-Tiefe-Skalen hinweisen könnte; dies müsste in Folgestudien anhand konfirmatorischer Faktorenanalysen (CFA) an einer größeren Stichprobe geprüft werden; auch wird eine Überarbeitung von Items angestrebt.

Die Korrelationen zwischen WS-Breite und den verschiedenen WS-Tiefe-Skalen weisen den Annahmen entsprechend für die Gesamtstichprobe sowie für beide Sprachgruppen (obgleich inkonsistenter für die BL-Gruppe) die erwarteten mittleren bis hohen Zusammenhänge auf, was darauf hinweist, dass es sich erwartungsgemäß um ein Konstrukt mit voneinander differenzierbaren Dimensionen handelt. Obgleich beide WS-Tiefe-Skalen mit der WS-Breite korrelieren, was erwartungsgemäß auf die Zugehörigkeit zum gleichen Konstrukt hindeutet, müssen jedoch die nicht signifikante, im Falle der bilingualen Gruppe sogar signifikant negative Korrelation zwischen den beiden WS-Tiefe-Skalen sowie die in der EFA beobachtbaren unterschiedlichen Ladungsrichtungen von semantisch anomalen und nicht-anomalen Items auf einem gemeinsamen Faktor diskutiert werden. Diesbezüglich sind folgende Erklärungen denkbar: Eine Betrachtung der *Richtungen der Rohwertdifferenzen* zwischen den beiden WS-Tiefe-Skalen weist darauf hin, dass bei der monolingualen Gruppe häufiger (oft deutlich) höhere Testscores für die semantisch anomale Skala erreicht wurden, während die Kinder der bilingualen Gruppe *entweder* in der semantisch anomalen *oder* in der semantisch nicht-anomalen Skala bessere Werte aufwiesen. Dies deutet auf einen Gruppeneffekt und möglicherweise unterschiedliche Itemlösungsstrategien der Gruppen hin. Denkbar wäre hier, dass dies bei der bilingualen Gruppe durch einen auf das Wortschatztraining rückführbaren Effekt des verwendeten Itemformats begründbar ist. So zielte die Intervention auf das Erlernen von Wörtern aus einem semantisch korrekten Kontext ab, was bei einer größeren Anzahl der trainierten Kinder zu einer im Vergleich stärkeren Fokussierung auf grammatisch korrekte Kontexte führte, während die untrainierten Kinder dazu tendierten, verstärkt auf semantisch nicht korrekte „Quatschsätze“ zu achten. Möglicherweise weisen die unterschiedlichen manifesten Korrelationen für die mono- und bilingualen Kinder zwischen den WS-Tiefe-Skalen jedoch auch auf eine unterschiedliche Struktur des Tests oder sogar des Konstrukts WS-Tiefe an sich für die unterschiedlichen Sprachgruppen hin; dies muss anhand von Mehrgruppenanalysen an einer größeren Stichprobe und ohne begleitende Intervention geprüft werden.

Basierend auf den Ergebnissen der manifesten Korrelationen zwischen den beiden WS-Tiefe-Subskalen ist es ferner naheliegend zu vermuten, dass dort teilweise unterschiedliche Aspekte von Wortwissen abgebildet werden, was drei Fragen aufwirft: Erstens nach der inhaltlichen Interpretation der beiden WS-Tiefe-Skalen, zweitens nach den bei der Lösung von semantisch anomalen bzw. nicht-anomalen Items zugrundeliegenden kognitiv-sprachlichen Prozessen sowie drittens bezüglich des Testscores und der Bedeutung für die praktische Anwendbarkeit und Weiterentwicklung des Verfahrens.

Zunächst erfordern beide WS-Tiefe-Itemtypen für deren korrekte Beantwortung Wissen über das Zielwort, über die weiteren Inhaltswörter des Satzes, sowie über die semantische Verknüpfung zwischen dem Ziel- und mindestens einem weiteren Inhaltswort, im Folgenden bezeichnet als Referenzwort. Möglich wäre, dass sich die Art des

erforderlichen Wissens über Wörter und semantische Verknüpfungen bei beiden Itemtypen unterscheidet. Denkbar ist, dass zum Lösen eines semantisch anomalen Items nun primär Wissen darüber notwendig ist, was *nicht* zu einem Konzept gehört, dass also *keine* sinnvolle semantische Verknüpfung zwischen einem Ziel- und einem Referenzwort besteht. So ist bei dem Satz „Hugos Verwandte sind aus Schokoladeneis gemacht“ das Zielwort „Verwandte“, Referenzwort ist „Schokoladeneis“, anzunehmen, dass Kindern im Vorschulalter das Konzept „Schokoladeneis“ sowie dessen Eigenschaften (süß, kalt, schmilzt etc.) bekannt sind. Bezüglich des abstrakteren und als schwieriger einzuschätzenden Zielworts „Verwandte“ ist hingegen möglich, dass zwar noch nicht sämtliche semantischen Verknüpfungen zu anderen Wörtern, jedoch ein grundlegendes, konkretes Wissen wie „Verwandte sind Menschen, Familie“ vorhanden ist. Auch im Falle eines unvollständigen Konzepts hinsichtlich des Worts „Verwandte“ wäre es für ein Kind nun erschließbar, dass zwischen Verwandten und Schokoladeneis keine direkte semantische Verknüpfung existiert, da Menschen nicht aus Schokoladeneis bestehen. Hier ist zur Lösung des Items also lediglich ein grundlegendes Wissen über das Zielwort notwendig (vorausgesetzt, das notwendige Wissen über das leichtere, konkretere Referenzwort ist vorhanden), um eine bestimmte semantische Verknüpfung ausschließen zu können.

Zur Lösung des entsprechenden semantisch nicht-anomalen Satzes „Omi, Opi, Mama, Papa sind meine Verwandten“ wird hingegen vor allem das Wissen über *zum Konzept zugehörige* semantische Verknüpfungen benötigt. Hier würde dies das Wissen darüber bedeuten, welche Personen unter die Bezeichnung „Verwandte“ fallen. Bezüglich der beiden WS-Tiefe-Dimensionen bedeutet dies zusammenfassend, dass jeweils unterschiedliches Wissen über semantische Verknüpfungen relevant sein könnte: im Falle semantisch anomaler Kontexte um deren *Ausschluss* bzw. Nicht-Zugehörigkeit, in nicht-anomalen Kontexten um deren *Einschluss* bzw. Zugehörigkeit zum Konzept. Inwieweit sich diese beiden Bereiche des Wissens über die Art von semantischen Verknüpfungen tatsächlich von Vorschulkindern bereits voneinander differenzieren lassen, muss durch weitere Studien überprüft werden. Geplant ist hier der Einsatz der Methode des „lauten Denkens“, um mehr über die während der Lösung der Items stattfindenden sprachlich-kognitiven Prozesse und möglicherweise unterschiedlich eingesetzten Antwortstrategien der Sprachgruppen zu erfahren; erste qualitative Studien werden diesbezüglich derzeit ausgewertet.

In der Praxis sollten ferner getrennte Testscores für die beiden WS-Tiefe-Skalen vergeben werden, da eine differenzierte Betrachtung und Interpretation der semantischen Bedingungen in Praxis, Diagnostik und Forschung – auch bezüglich der möglicherweise unterschiedlichen prognostischen Validität der Dimensionen – notwendig zu sein scheint.

Kritisch zu diskutieren ist ferner, dass dichotome Items, wie sie zur Erfassung der WS-Tiefe verwendet wurden, eine hohe Ratewahrscheinlichkeit aufweisen. Auch wenn diesbezüglich erste Analysen durchgeführt wurden, wäre doch zu überlegen, ob möglicherweise eine Vorgabe der Items als *Forced-Choice-Items*, in denen sich die Kinder jeweils zwischen einer semantisch anomalen und einer semantisch nicht-anomalen Variante entscheiden müssen, die Ratewahrscheinlichkeit senken und die Testvalidität erhöhen könnte.

Gleichfalls kritisch hinterfragt werden sollte, ob anhand der Einbettung der Zielwörter in unterschiedliche semantische Kontexte tatsächlich WS-Tiefe, oder doch etwa Wissen über semantische Inkonsistenzen (vgl. Bialystok 1986) erfasst wird, bzw. ob und inwieweit diese beiden Arten von Wissen überhaupt differenzierbar sind. Dies lässt sich anhand der vorliegenden Studie wegen der geringen Anzahl von Items nicht zuverlässig überprüfen. Für weitere Studien wären jedoch eine Formulierung weiterer Items mit jeweils gleichen Zielwörtern in gleichen Bedingungen zur Überprüfung der Wortwissenskonsistenz sowie auch hier ein Einsatz der Methode „lautes Denken“ vorstellbar.

Insgesamt haben sich während Konstruktion und Validierung des Instruments relevante Fragen hinsichtlich des Testkonstrukts ergeben, die in weiteren Forschungsvorhaben näher betrachtet und zur weitergehenden Entwicklung und Validierung des Instruments genutzt werden sollten. Zunächst sollten dazu die oben diskutierten Annahmen bezüglich der Dimensionalität sowie der exakten inhaltlichen Interpretierbarkeit der WS-Tiefe-Dimensionen anhand von CFA überprüft werden. Dies erfordert die Konstruktion weiterer bzw. veränderter Items. Ferner wurde lediglich ein passives WS-Tiefe-Merkmal fokussiert. Für eine weitere Untersuchung des Konstrukts WS-Tiefe ist

es jedoch unabdingbar, weitere, auch aktive WS-Aspekte wie etwa Flüssigkeit einzubeziehen. Auch bleibt bislang unklar, inwieweit sich die Ergebnisse replizieren oder auf andere Alters-, Kompetenzniveau- und Sprachgruppen generalisieren lassen; hierzu sind weitere Studien mit größeren Stichproben notwendig, die Mehrgruppen-Analysen und damit einen Vergleich von Strukturen bei unterschiedlichen Sprach- sowie Altersgruppen zulassen. Ferner wurden aus Gründen der Zumutbarkeit zunächst nicht für jedes Zielwort Sätze jeder WS-Tiefe-Bedingung formuliert; diesbezüglich sollte der Test noch ergänzt werden.

Des Weiteren ist die verwendete Stichprobe hinsichtlich des Alters recht heterogen und es zeigen sich deutliche Alterseffekte in den Analysen. Hier sollte in Folgestudien gezielt ein Augenmerk auf die unterschiedlichen Altersgruppen, auch im Hinblick auf mögliche Altersnormen, gelegt werden.

Trotz dieser kritisch zu diskutierenden Punkte zeigt sich jedoch zusammenfassend, dass die vorliegende Studie einen Beitrag zu der Diskussion um die Messung und Struktur des Konstrukts Wortschatz im Vorschulalter leisten kann. Die Berücksichtigung von WS-Tiefe und -Breite bildungssprachlich anschlussfähigen Wortschatzes kann wiederum die Grundlage zur Analyse differenzieller Effekte in Bezug auf die Nutzung von Erwerbs- und Fördergelegenheiten bieten. Einsatzmöglichkeiten des Instruments liegen für die Praxis beispielsweise im formativen Assessment gezielt bildungssprachlicher lexikaler Vorläuferfähigkeiten der Kinder seitens der Fachkräfte, für die Forschung im Bereich der theoretischen Konstrukt- und Kompetenzmodellierung, in Zusammenhangsanalysen der gewählten Wörter mit bildungssprachlichen Fähigkeiten in der Grundschule, in der Überprüfung inkrementeller und differenzieller Effekte bildungssprachlicher Vorläuferfähigkeiten sowie in der Diagnose und Überprüfung der Wirksamkeit von Wortschatzförderung im Vorschulalter.

Literatur

- American Educational Research Association; American Psychological Association; National Council on Measurement in Education (Hrsg.) (2014), *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Assoc.
- Anderson, Richard C. & Freebody, Peter (1981), Vocabulary knowledge. In: Guthrie, John T. (Hrsg.), *Comprehension and Teaching: Research Reviews*. Newark, DE: International Reading Association, 77-117.
- Bailey, Alison L.; Butler, Frances; Stevens, Robin & Lord, Carol (2007), Further specifying the language demands of schooling. In: Bailey, Alison L. (Hrsg.), *The Language Demands of School. Putting Academic English to the Test*. New Haven, CT: Yale University Press, 103-156.
- Baumert, Jürgen & Schümer, Gundel (2001), Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In: Deutsches PISA-Konsortium (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Wiesbaden: Verlag VS für Sozialwissenschaften, 323-410.
- Bialystok, Ellen (1986), Factors in the growth of linguistic awareness. *Child Development* 57, 498-510.
- Bialystok, Ellen; Luk, Gigi; Peets, Kathleen F. & Yang, Sujin (2010), Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition* 13, 525-531.
- Becker-Mrotzek, Michael et al. (2013), Qualitätsmerkmale für Sprachstandsverfahren im Elementarbereich. Ein Bewertungsrahmen für fundierte Sprachdiagnostik in der Kita. Köln: Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache.
- Bos, Wilfried; Valtin, Renate; Hußmann, Anke; Wendt, Heike & Goy, Martin (2017), IGLU 2016: Wichtige Ergebnisse im Überblick. In: Hußmann, Anke; Wendt, Heike; Bos, Wilfried; Bremerich-Vos, Albert; Kasper, Daniel; Lankes, Eva-Maria; McElvany, Nele; Stubbe, Tobias C. & Valtin, Renate (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann, 13-28.
- Cinar, Melihan & Hardy, Ilonca (2017), Impliziter Wortschatzerwerb in der Bildungssprache von Kindern mit Migrationshintergrund im Vorschulalter. *Empirische Pädagogik* 31: 4, 411-431.
- Cummins, Jim (2000), *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon, Multilingual Matters.
- Daller, Helmut; Milton, James & Treffers-Daller, Jeanine (2007), Editors' introduction: Conventions, terminology and an overview of the book. In: Daller, Helmut; Milton, James & Treffers-Daller, Jeanine (Hrsg.), *Modelling and assessing vocabulary knowledge*. Cambridge, UK: Cambridge University Press, 1-32.
- Dubowy, Minja; Ebert, Susanne; von Maurice, Jutta & Weinert, Sabine (2008), Sprachlich-kognitive Kompetenzen beim Eintritt in den Kindergarten. Ein Vergleich von Kindern mit und ohne Migrationshintergrund. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 40: 3, 124-134.
- Dunn, Lloyd M. & Dunn, Douglas M. (2007), *Peabody Picture Vocabulary Test-4*. Deutsche Adaption: Lenhard, Alexandra; Lenhard, Wolfgang; Segerer, Robin & Suggate, Sebastian (2015), Pearson Assessment.
- Eckhardt, Andrea G. (2008), *Sprache als Barriere für den schulischen Erfolg. Potentielle Schwierigkeiten beim Erwerb schulbezogener Sprache für Kinder mit Migrationshintergrund*. Münster: Waxmann.
- Feilke, Helmut (2012), Bildungssprachliche Kompetenzen - fördern und entwickeln. *Praxis Deutsch* 233, 5-13.
- Gesellschaft für Didaktik des Sachunterrichts (Hrsg.) (2012), *Perspektivrahmen Sachunterricht*. Bad Heilbrunn: Klinkhardt.
- Gogolin, Ingrid & Lange, Imke (2011), Bildungssprache und Durchgängige Sprachbildung. In: Fürstenau, Sara & Gomolla, Mechthild (Hrsg.), *Migration und schulischer Wandel: Mehrsprachigkeit*. Wiesbaden: VS Verlag für Sozialwissenschaften, 107-128.
- Grimm, Hannelore & Schöler, Hermann (1991), *Heidelberger Sprachentwicklungstest*. Göttingen: Hogrefe.

- Gyllstad, Henrik (2013), Looking at L2 Vocabulary Knowledge Dimensions from an Assessment Perspective – Challenges and Potential Solutions. In: Bardel, Camilla; Laufer, Batia & Lindqvist, Christina (Hrsg.), *L2 Vocabulary Acquisition, Knowledge and Use: New Perspectives on Assessment and Corpus Analysis*. Eurosla Monographs Series, Eurosla, 11-28.
- Hardy, Ilonca & Cinar, Melihan (eingereicht), Vocabulary Acquisition of Bilingual Preschoolers in Different Training Conditions: The Role of Children's First Language.
- Henriksen, Birgit (1999), Three dimensions of vocabulary development. *Studies in Second Language Acquisition* 21: 2, 303-317.
- Heppt, Birgit; Henschel, Sofie & Haag, Nicole (2016), Everyday and academic language comprehension: Investigating their relationships with school success and challenges for language minority learners. *Learning and Individual Differences* 47, 244-251.
- Heppt, Birgit; Stanat, Petra; Dragon, Nina; Berendes, Karin & Weinert, Sabine (2014), Bildungssprachliche Anforderungen und Hörverstehen bei Kindern mit deutscher und nicht-deutscher Familiensprache. *Zeitschrift für Pädagogische Psychologie* 28: 3, 139-149.
- Juska-Bacher, Britta & Jakob, Sabrina (2014), Wortschatzumfang und Wortschatzqualität und ihre Bedeutung im fortgesetzten Spracherwerb. *Zeitschrift für angewandte Linguistik* 61: 1, 49-75.
- Kiese-Himmel, Christiane (2005), *Aktiver Wortschatztest für 3-5jährige Kinder*. Göttingen: Beltz.
- Konak, Ömer; Duindam, Tom & Kamphuis, Frans (2005), *CITO-Sprachtest. Wissenschaftlicher Bericht*. Arnhem: CITO.
- Lisker, Andrea (2013), *Sprachstandserhebung und Sprachförderung vor der Einschulung – Eine Bestandsaufnahme in den Bundesländern*. Expertise im Auftrag des Deutschen Jugendinstituts. München: Verlag Deutsches Jugendinstitut.
- Meara, Paul & Buxton, Brian (1987), An alternative to multiple choice vocabulary tests. *Language Testing* 4: 2, 142-154.
- Meara, Paul & Wolter, Brent (2004), Beyond vocabulary depth. In: Albrechtsen, Dorte; Haastrup, Kirsten & Henriksen, Birgit (Hrsg.), *Angles on the English-Speaking World*. Copenhagen: Museum Tusulanum Press, 85-96.
- Melchers, Peter & Preuss, Ulrich (2009), *Kaufman assessment battery for children*. Deutschsprachige Fassung. Frankfurt a. Main: Swets und Zeitlinger.
- Morek, Melanie & Heller, Vivien (2012), Bildungssprache – Kommunikative, epistemische, soziale und interaktive Aspekte ihres Gebrauchs. *Zeitschrift für angewandte Linguistik* 57, 67-101.
- Muthén, Linda K. & Muthén, Bengt O. (1998-2012), *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.
- Nation, Paul (1983), Testing and teaching vocabulary. *Guidelines* 5, 12-25.
- Nation, Paul (2001), *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.
- 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.
- Niedersächsisches Kultusministerium (2007), *Sprachförderung in Kindergarten und Schule*. Hannover: Niedersächsisches Kultusministerium.
- Oullette, Gene P. (2006), What's Meaning Got to Do With It: The Role of Vocabulary in Word Reading and Reading Comprehension. *Journal of Educational Psychology* 98: 3, 554-566.
- Petermann, Franz (2012), *SET 5–10. Sprachstandserhebungstest für Kinder im Alter zwischen 5 und 10 Jahren*. Göttingen: Hogrefe.
- Qian, Davis (1999), Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review* 56: 2, 282-308.
- Qian, David (2002), Investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective. *Language Learning* 52: 3, 513-536.

- Ramm, Gesa; Prenzel, Manfred; Heidemeier Heike & Walter, Oliver (2004), Soziokulturelle Herkunft. Migration. In: Prenzel, Manfred; Baumert, Jürgen; Blum, Werner; Lehmann, Rainer; Leutner, Detlev & Neubrand, Michael (Hrsg.), *Pisa 2003. Der Bildungsstand der Jugendlichen in Deutschland. Ergebnisse des zweiten Internationalen Vergleichs*. Münster: Waxmann, 254-272.
- Read, John (1998), Validating a test to measure depth of vocabulary knowledge. In: Kunnan, Antony (Hrsg.), *Validation in language assessment*. Mahwah, NJ: Lawrence Erlbaum, 41-60.
- Read, John (2004), Plumbing the depths: how should the construct of vocabulary knowledge be defined. In: Bogaards, Paul & Laufer, Batia (Hrsg.), *Vocabulary in a Second Language*. Amsterdam: John Benjamins, 209-227.
- Ruhl, Charles (1989), *On Monosemy: a Study in Linguistic Semantics*. Albany: State University of New York Press.
- Schmölzer-Eibinger, Sabine; Dorner, Magdalena; Langer, Elisabeth & Helten-Pacher, Maria R. (2013), *Sprachförderung im Fachunterricht in sprachlich heterogenen Klassen*. Stuttgart: Fillibach/Klett.
- Schoonen, Rob & Verhallen, Marianne (2008), The assessment of deep word knowledge in young first and second language learners. *Language Testing* 25: 2, 211-236.
- Söhn, Janina (2005), *Zweisprachiger Schulunterricht für Migrantenkinder. Ergebnisse der Evaluationsforschung zu seinen Auswirkungen auf Zweitspracherwerb und Schulerfolg. AKI-Forschungsbilanz 2*. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB)/ Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI).
- Sternberg, Robert J. (1987), Most vocabulary is learned from context. In: McKeown, Margaret & Curtis, Mary E. (Hrsg.), *The Nature of Vocabulary Acquisition*. Hillsdale, N.J.: Erlbaum, 89-105.
- Tannenbaum, Kendra R.; Torgesen, Joseph K. & Wagner, Richard K. (2006), Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading* 10: 4, 381-398.
- Tannenbaum, Kendra R. (2008), *Relationships between Measures of Word Knowledge and Reading Comprehension in Third- and Seventh-Grade Children*. Electronic Theses, Treatises and Dissertations. Paper 1703. [Online unter <http://diginole.lib.fsu.edu/etd/1703/>. 10.08.2015].
- Townsend, Diana; Filippini, Alexis; Collins, Penelope & Biancarosa, Gina (2012), Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students. *The Elementary School Journal* 112: 3, 497-518.
- Uccelli, Paola; Barr, Christopher D.; Dobbs, Christina L.; Galloway, Emily P.; Meneses, Alejandra & Sanchez, E. (2015), Core academic language skills: An expanded operational construct and a novel instrument to chart school-relevant language proficiency in pre-adolescent and adolescent learners. *Applied Psycholinguistics* 36, 1077-1109.
- Van Kleeck, Anne (2014), Distinguishing between casual talk and academic talk beginning in the preschool years: An important consideration for speech-language pathologists. *American Journal of Speech-Language Pathology* 23, 724-741.
- Vermeer, Anne (2001), Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics* 22, 217-234.
- Weinert, Sabine (2004), Wortschatzerwerb und kognitive Entwicklung. *Sprache – Stimme – Gehör* 28: 1, 20-28.

Anhang A: Zielwörter WS-Breite und WS-Tiefe / schulkontextuell-sachfachliche Kategorien

Berufsbezogene Wörter/Kontexte	Tiere/Pflanzen, Natur/Landschaft	Technik/Infrastruktur	Schulbezogene Wörter/Aktivitäten
Absperrung <i>Architekt</i> Asphalt Bau Fließband Grube <i>Herberge</i> Kajüte Kittel <i>Kombüse</i> Patient Pflug Rohr Spurensicherung <i>Stethoskop</i> Streife_fahren Tresen <i>Wetterbericht</i> Wirt <i>Nahrung</i> Schlemmen	Bach <i>Beet</i> Beschnüffeln Fell Fohlen <i>Fliegen</i> Getreide Koppel Kralle Lichtung Niederschlag Oase Prasseln Quelle Rinde <i>Stall</i> <i>Stängel</i> Trog Ufer <i>Verwandte</i> <i>Welp</i> Züchten (auf etwas) starren	Mikrofon Horchen <i>Bildschirm</i> <i>Scheinwerfer</i> <i>Wolkenkratzer</i> Veranda Gasse Pforte	<i>Globus</i> Lexikon Rangelei <i>Einzelgänger</i> Erschöpft <i>Vereint</i> Erwartungsvoll Betrachten

Anmerkung: kursiv=WS-Tiefe-Zielwörter

Anmerkungen

¹ Das Verbund-Projekt „Potenzial der Muttersprache zur Verringerung der Bildungsgleichheit – Wortschatzerwerb von Kindern nicht-deutscher Familiensprache vor zentralen Übergängen des Bildungssystem (InterMut)“ wurde vom Bundesministerium für Bildung und Forschung im Forschungsschwerpunkt „Chancengerechtigkeit und Teilhabe. Sozialer Wandel und Strategien der Förderung“ gefördert.

² Kategoriale Indikatoren; Faktorladungen der 120 Items werden aus Platzgründen nicht im Einzelnen dargestellt.