Zeitschrift für Interkulturellen Fremdsprachenunterricht
Didaktik und Methodik im Bereich Deutsch als Fremdsprache
ISSN 1205-6545 Jahrgang 19, Nummer 2 (Oktober 2014)

Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern

Dr. Astrid Jurecka
Goethe-Universität Frankfurt am Main
Fachbereich Erziehungswissenschaften/
Institut für Pädagogik der Elementar- und Primarstufe
Grüneburgplatz 1
D-60629 Frankfurt am Main
E-Mail: jurecka@em.uni-frankfurt.de, Tel.: 069-798-36262

Dr. Jules Bündgens-Kosten
Goethe-Universität Frankfurt am Main
Akademie für Bildungsforschung und Lehrerbildung
Senckenberganlage 31
Juridicum Raum 1007
D-60486 Frankfurt am Main
E-Mail: Buendgens-kosten@em.uni-frankfurt.de, Tel.: 069-798-23293

Prof. Dr. Daniela Elsner
Goethe-Universität Frankfurt am Main
Institut für England und Amerikastudien
Sprachlehrforschung und Didaktik
Grüneburgplatz 1
D-60629 Frankfurt am Main
E-Mail: elsner@em.uni-frankfurt.de
Tel.: 069-798-32518, Fax: 069-798-32-509

Abstract: Fremdsprachliche (englischsprachige) Interaktion in der Primarstufe wird im Rahmen existierender Testverfahren häufig primär anhand stark strukturierter Settings (beispielsweise Interview-Settings) erfasst. Dies resultiert jedoch möglicherweise in einer Einschränkung der Konstruktvalidität von Tests, da anhand solch gesteuerter, häufig asymmetrischer Test-Settings nicht die gesamte Bandbreite interaktiver Sprachhandlungen erfasst werden kann. Die vorliegende Studie untersucht die Frage, ob der zusätzliche Einsatz von in geringerem Maße gesteuerten, symmetrischen Testformaten die Konstruktvalidität von Tests zur Messung interaktiver Sprachhandlungen auch in der Primarstufe erhöhen könnte. Dazu wurde ein Vergleich von Interview- und Peer-to-Peer-Settings durchgeführt mit dem Ziel, deren Wirkung auf die Sprachproduktion sowie auf die Art interaktiver Sprachhandlungen zu untersuchen. Es zeigt sich, dass die SchülerInnen (n=38; 4. Klasse) insgesamt häufig auf die deutsche Sprache zurückgriffen, in Interview-Settings signifikant mehr englische Wörter produzierten als in Peer-to-Peer-Settings, letztere jedoch sprachübergreifend (Deutsch/Englisch) insgesamt eine größere Bandbreite interaktiver Sprachhandlungen hervorriefen.

Existing test formats for the assessment of primary school pupils´ interaction in a foreign language (English) commonly make use of highly structured interview settings, for instance interviews. However, such a setting may have a negative impact on the construct validity of the tests because such controlled, frequently asymmetrical test settings do not permit the assessment of the whole range of interactive language uses. This study looks at the question whether the added use of less controlled peer-to-peer settings could increase the construct validity of tests even on the elementary level. For this purpose, interview and peer-to-peer settings are compared in order to analyze their effect on young learners´ (English) language production and the type of peer interactions generated. It will be shown that while primary school pupils (n=38; 4th grade) repeatedly resorted to using German

they produced significantly more English words in the interview situations than in peer-to-peer settings, but the latter setting gave rise to a broader range of interactive language functions.

Schlagwörter: Messung fremdsprachlicher Interaktion im Primarbereich, Eignung von Testsettings, Konstruktvalidität fremdsprachlicher Testverfahren

1. Einleitung

Mit dem Ziel, europaweit die Mobilität und damit auch die sprachenübergreifende Interaktion zwischen BürgerInnen zu erleichtern, fordert die Europäische Kommission dazu auf, den Erwerb fremdsprachlicher Kompetenzen bereits im frühen Kindesalter zu unterstützen (vgl. z.B. Europarat 1998; 2002). Um dieser Forderung nachzukommen, wurde in Deutschland im Schuljahr 2004/2005 flächendeckend obligatorischer fremdsprachlicher Unterricht spätestens ab der dritten Klasse eingeführt (siehe dazu Elsner 2010: 20-21). In der Folge wurden in vielen deutschen Bundesländern Bildungsstandards für fremdsprachliche Kompetenzen in der Grundschule entwickelt (für eine Übersicht siehe Dausend 2014), welche sich, ebenso wie die Bildungsstandards für die Sekundarstufe, an den Kompetenzskalen des Gemeinsamen Europäischen Referenzrahmens für Sprachen (GER; Europarat 2001; siehe z.B. KM Hessen 2010) sowie dem dort zugrundeliegenden Ansatz kommunikativer Kompetenz (vgl. z.B. Bachman & Palmer 1996; Canale & Swain 1980; Hymes 1966; Savignon 2002) orientieren. Dieser grenzt sich von einem rein grammatisch orientierten Ansatz ab, indem er das Zusammenspiel strategischer, soziokultureller und diskursiver Funktionen von Sprachen beim Lernen (und Testen) von Sprachen als essentiell für die Entwicklung einer kommunikativen Kompetenz erkennt: „Although the relative importance of the various components depends on the overall level of communicative competence, each is essential. (…) Rather, when an increase occurs in one area, that component interacts with other components to produce a corresponding increase in overall communicative competence" (Savignon 2002: 8). In der Primarstufe wird diese Verknüpfung von Kompetenzbereichen in den Bildungsplänen ebenfalls propagiert, jedoch fokussiert man sich hier – ob der starken zeitlichen Limitierung des Unterrichts – vorwiegend auf die Ausbildung mündlicher Sprachkompetenzen (Hörverstehen und Sprechen) (vgl. Engel 2009; Roos 2006: 27; Wolff 2009).

Der GER hat sich in den letzten Jahren nicht nur als eine Leitschrift für die Entwicklung von Lehr- und Lernzielen im Bereich des Fremdsprachenunterrichts (z.B. KMK 2003) etabliert, sondern ist auch maßgebend für die Entwicklung von Testinstrumenten zur sprachlichen Leistungs- und Kompetenzmessung (z.B. Preliminary English Test (PET)/Key English Test (KET); UCLES 2011a, b). Der GER folgt dabei einem handlungs- und kriterienorientierten Ansatz. Mit Hilfe sogenannter *‚Can Do' Statements* beschreibt er für sechs aufeinander aufbauende Kompetenzniveaus (A1/A2: elementare Sprachverwendung; B1/B2: selbständige Sprachverwendung; C1/C1: kompetente Sprachverwendung), welche sprachlichen Handlungen Fremdsprachenlernende beispielsweise ausführen können sollten, um ein spezifisches Kompetenzniveau zu erreichen. In der deutschen Primarstufe werden dabei die Niveaus der elementaren Sprachverwendung anvisiert (vgl. hierzu Dausend 2014: 45-46), wenngleich der Referenzrahmen ursprünglich für erwachsene Lernende konzipiert wurde.

Mit der Verwendung von GER-Skalen zur Erfassung fremdsprachlicher Kompetenzen speziell von jungen Lernenden setzte sich beispielsweise das norwegische „Bergen ‚Can do' project" (Hasselgren 2003, 2005) auseinander. Dort wurde untersucht, wie GER-basierte Kompetenzdeskriptoren und Kompetenzbeschreibungen, unter anderem auch für die fremdsprachlichen Fähigkeiten „Sprechen" und „Interaktion", speziell für junge Lernende thematisch und inhaltlich adaptiert werden können. Die entwickelten Skalen erwiesen sich insgesamt als gut geeignet, allerdings für eine etwas ältere Zielgruppe (13-15 Jahre; s. Hasselgreen 2003) als die im vorliegenden Beitrag fokussierte (6-10 Jahre).

Kommunikative Sprachaktivitäten werden im GER unterteilt in Rezeption, Produktion und Interaktion, wobei diese Aktivitäten jeweils sowohl mündlich als auch schriftlich ausgeführt werden können (vgl. Europarat 2001: 25). Mündliche Kompetenzen umfassen dabei die rezeptive Teilkompetenz „Hören" sowie die produktiven Teilkompetenzen „zusammenhängendes Sprechen" (z.B. Vorträge) und „mündliche fremdsprachliche Interaktion" (z.B. Gespräche; ebd.: 63; 80). Während das Hören und das zusammenhängende Sprechen als „primäre" Prozesse verstanden

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

werden (25), welche nicht zwingend direkte GesprächspartnerInnen erfordern, kann die Interaktion nur dann zustande kommen, wenn ein oder mehrere Personen miteinander kommunizieren. Dabei können Produktion und Rezeption von Sprache ständig abwechselnd, teilweise überlappend, oder sogar gleichzeitig stattfinden (26), mit dem Ziel, „durch das Aushandeln von Bedeutung auf der Basis des Prinzips der Kooperation das Gespräch gemeinsam entstehen zu lassen“ (78). Dabei wird angenommen, dass es sich bei der Interaktion um eine eigenständige Teilkompetenz mündlicher fremdsprachlicher Kommunikation über die reine mündliche Produktion und die reine Sprachrezeption hinaus handelt, da während eines Gesprächs zeitweise produktive und rezeptive Prozesse synchron stattfinden (26). Aufgrund der Komplexität interaktionaler Prozesse werden im Rahmen des GER Sprachhandlungen und Sprachstrategien, die der *mündlichen fremdsprachlichen Interaktion* zuzuordnen sind, gesondert anhand eigener, spezifischer Kompetenzskalen erfasst und beschrieben. Neben den während einer Interaktion ständig verwendeten Rezeptions- und Produktionsstrategien werden dabei zusätzlich *kognitive und kooperative Strategien* (z.B. Sprecherwechsel, Kooperation oder Bitten um Klärung; ebd.: 88-89) definiert, die von einem Sprachverwender zur Steuerung von Kooperation und Interaktion angewendet werden. Diesen Strategien wiederum werden jeweils bestimmte interaktive Sprachaktivitäten bzw. Sprachhandlungen als zugehörig zugeordnet, wie etwa das Vorschlagen und Evaluieren von Lösungen, das Zusammenfassen des Gesprächstands oder das Beantworten von Fragen (ebd.). Ferner werden spezifische *Interaktionskontexte* definiert (z.B. Interviews, zwanglose Gespräche), in deren Rahmen diese interaktiven *Aktivitäten* und Sprachhandlungen durchgeführt werden. Des Weiteren betreffen auch Teile der zur pragmatischen Kompetenz gehörenden Sprachhandlungen die mündliche Interaktion, wie etwa die Diskurskompetenz (Flexibilität, Turn-Taking, Thematische Entwicklung sowie Kohärenz und Kohäsion) oder die funktionale Kompetenz (angemessene Wahl der Sprechakte, richtige Interpretation des Gegenübers; adäquate Reaktionsweise; ebd.: 123-130).

2. Messung mündlich-interaktiver fremdsprachlicher Kompetenzen in der Primarstufe

Obwohl Interaktion insgesamt eine hohe Relevanz für das Sprachenlernen zugeschrieben wird (auch explizit in Bildungsstandards für den Fremdsprachenerwerb in der Primarstufe; z.B. KM Baden-Württemberg 2004), wurde die mündliche fremdsprachliche Interaktion speziell von jungen Lernenden in der Primarstufe, also von Kindern im Alter zwischen 6 und 10 Jahren, im deutschsprachigen Raum bislang wenig empirisch erforscht (z.B. Diehr & Frisch 2008; Diehr & Polte 2009; Haudeck & Schwab 2011; Keßler 2009). Obgleich für die Testung von Erwachsenen (teilweise auch von älteren Kindern oder Jugendlichen) mittlerweile eine Reihe von Testverfahren existieren, die auch die Messung von Interaktion im Rahmen unterschiedlicher Test-Settings und Testsituationen beinhalten (z.B. Fit in Deutsch 1/2, Goethe-Institut 2013[1]; KET/PET, UCLES 2011a,b), wird sich im Folgenden auf die Darstellung speziell für die Primarstufe konzipierter Testverfahren beschränkt, nicht zuletzt da Kinder andere thematische Inhalte und Aufgabenformate benötigen und als motivierend empfinden als Erwachsene (siehe dazu beispielsweise Hasselgren 2003/2005; Diehr & Frisch 2008).

Im Rahmen der Bildungsstandards der Länder für den Fremdsprachenerwerb in der Primarstufe wird üblicherweise erwartet, dass SchülerInnen, je nach Anzahl absolvierter Unterrichtsjahre, am Ende der Grundschule hinsichtlich mündlicher fremdsprachlicher Kompetenzen mindestens das GER-Niveau A1 bzw. A2 erreichen (vgl. Dausend 2014). Um das Erreichen dieser Lernziele valide überprüfen zu können, ist es gerade für Lehrkräfte von hoher Relevanz, die Kompetenzen ihrer SchülerInnen in Konformität mit den jeweiligen Bildungsstandards richtig einschätzen zu können. Während zur Erfassung rezeptiver Fähigkeiten im Primarbereich bereits eine Reihe von Instrumenten entwickelt wurden (z.B. KESS - Kompetenzen und Einstellungen von Schülerinnen und Schülern: Bos & Pietsch 2006; EVENING - Evaluation Englisch in der Grundschule: z.B. Groot-Wilken & Paulick 2009; YLE - Young Learners English: UCLES 2013), existieren zur Erfassung der in der Grundschule fokussierten produktiven mündlichen Fähigkeiten bislang lediglich einzelne Instrumente, die zumeist auf die Erfassung der Bereiche monologisches Sprechen (Produktion) oder dialogisches Sprechen in klassischer Interviewform abzielen (z.B. Keßler 2009; TOEFL Primary (ab 8 Jahre, computerbasiert; ETS 2013); Kötter 2009; YLE-Tests/UCLES 2013; Zangl 2000).

Eine Ausnahme stellen dabei die kriterienorientierten TaPS-Skalen (Testing and Assessing Spoken English in Primary School; Diehr & Frisch 2008) dar, die innovative Test-Settings wie erlernte Mini-Dialoge, Vorführungen oder kleine erlernte Rollenspiele beinhalten. Allerdings wird hier bezüglich der interaktiven Aktivitäten ausschließlich die Diskurskompetenz bewertet (Flexibilität, Gestik/Mimik, Zuschauerbezug, Themenentwicklung). Im Rahmen der

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

TaPs-Skalen untersuchten Diehr und Polte (2009) dabei in einer Studie diskursivspezifische Merkmale (Kohäsion) anhand der von 20 Grundschulkindern (allerdings monologisch) produzierten englischen Texte (Nacherzählen einer Geschichte). Sie stellen dabei eine große Varianz hinsichtlich der Verwendung kohäsiver Sprachmittel fest, kommen jedoch insgesamt zu dem Schluss, dass die untersuchten L2-Lernenden „beim Nacherzählen einer Geschichte (…) produktiv (vorgehen) und (…) die curricularen Anforderungen, die an sie in einem deutschen Grundschulkontext gestellt werden (übertreffen)" (171). Der Young Learners English-Test (YLE; UCLES 2013) beinhaltet gleichfalls einen Testteil zur Erfassung von mündlichen Kompetenzen (Interview/Bildergeschichte). Dieser ist jedoch, gerade bezüglich der hier interessierenden untersten Kompetenzniveaus A1 und A2, sehr kurz (Test-Review: siehe Bailey 2005).

Auch in der EVENING-Studie (Groot-Wilkens, Engel & Thürmann 2007) wurde im Rahmen der Untersuchung mündlichen Sprachgebrauchs im Englischen bei ViertklässlerInnen neben einer monologischen eine dialogische Komponente an einer Substichprobe (n=120) erhoben (vgl. Kötter 2009). Die Testung war jedoch auch hier mit lediglich 6 Minuten sehr kurz und beinhaltete im Hinblick auf die Messung von Interaktion ausschließlich eine kurze Interview-Situation zwischen SchülerIn und Testleiter. Auch wurden zwar interaktive Sprachelemente bei der Codierung der Schülerantworten mit berücksichtigt, allerdings anhand relativ grober Indikatoren wie die Reaktion oder die Komplexität der Antwort (überarbeitete Version: Zeitpunkt/Spontaneität von Gegenfragen; Sprachqualität von Äußerungen; vgl. Keßler 2009: 149). Die Ergebnisse diesbezüglich zeigen, dass jede der gestellten sieben Fragen von mindestens 50 % der SchülerInnen korrekt beantwortet werden konnte. Keßler (2009) kommt zu dem Schluss, dass die Kinder „in der Lage sind, in dialogischen Kommunikationssituationen in der Zielsprache Englisch zu kommunizieren" (168), obgleich es ihnen leichter zu fallen scheint, auf Interviewfragen zu reagieren als eigene Fragen an den Interviewer zu formulieren.

Auch im Rahmen einer Schweizer Studie mit Fokus auf mündlichen und interaktiven Kompetenzen im Primarbereich (Haenni Hoti & Werlen 2007), in der unter anderem ein erlerntes Rollenspiel zwischen zwei SchülerInnen als Test-Setting diente, kommen die Autorinnen zu dem Schluss, dass die Kinder bereits nach einem Jahr Englischunterricht eine hohe Fähigkeit besitzen, sich auszudrücken, und auch komplexere sprachliche Äußerungen in Bezug auf bestimmte Themen erfolgen (vgl. Husfeldt & Bader-Lehmann 2009).

In der europäischen ELLiE-Studie (Early Language Learning in Europe; Enever 2011), einer Längsschnittstudie, die fremdsprachlichen Unterricht und fremdsprachliche Kompetenzen in der Grundschule in sieben europäischen Ländern erforschte, wurde sowohl monologisches (Beschreibung) als auch dialogisches Sprechen (*role-play*) untersucht, ebenso mussten die Kinder in einer kurzen spielerischen Übung (*guessing game*) nach Orten oder Dingen fragen und dabei mit dem Interviewpartner in Interaktion treten (17, 127). Die über vier Jahre durchgeführte longitudinal angelegte Studie kann aufzeigen, dass Lernende sich in allen Kompetenzbereichen, so auch im Sprechen weiterentwickeln, dabei jedoch individuell unterschiedliche Kompetenzniveaus in den einzelnen Kompetenzbereichen aufzeigen. Zudem kommt die Autorin zu dem Schluss, dass der GER keine geeignete Grundlage für die Messung fremdsprachlicher Kompetenzen von Grundschulkindern zur Verfügung stelle (5, 7).

Zusammenfassend zeigt sich, dass PrimarschülerInnen insgesamt bereits in der Lage sind, nach 2-4 Jahren (teilweise sogar bereits nach einem Jahr) Instruktion in einem gewissen Maße interaktive Sprachelemente zu produzieren. Testinstrumente zur Messung mündlicher Kompetenzen im Primarstufenalter berücksichtigen teilweise zwar dialogische und interaktive Kompetenzen, jedoch zielen sie häufig entweder auf ältere Zielgruppen, oder die interaktiven Sprachhandlungen werden nur partiell erfasst. Dies ist insoweit problematisch, als dass die Interaktion als Teilkompetenz der mündlichen Kompetenz definiert ist (siehe z.B. Europarat 2001). Ein lediglich partieller Einbezug dieser Teilkompetenz bei der Testung junger Fremdsprachenlernender führt damit möglicherweise zu einer Einschränkung der Konstruktvalidität der existierenden Verfahren.[2]

Jedoch existieren für diesen Alters- und Kompetenzbereich, nicht zuletzt da der GER für erwachsene Lernende konzipiert wurde, bislang nur wenige bzw. lediglich ausschnittsweise adäquate Kompetenzskalen, die hier eine Orientierung darstellen (z.B. Enever 2011).

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

2.1. Test-Settings zur Messung interaktiver mündlicher Kompetenzen

Bei der Betrachtung existierender Testinstrumente zur Messung interaktiver mündlicher Kompetenzen von PrimarschülerInnen wird neben einer oft wenig differenzierten Erfassung von Interaktionskompetenz deutlich, dass bislang in den Verfahren meist Interviewsituationen als Test-Settings verwendet werden (Ausnahmen stellen beispielsweise die TaPS-Skalen sowie die ELLiE-Studie dar). Dabei beantworten die jungen SprachanwenderInnen Fragen einer meist älteren, sprachlich kompetenteren Person (TestleiterIn/LehrerIn). Die Fragen und Antworten bestehen dabei, gerade auf niedrigen Kompetenzniveaus, zumeist aus eingeübten Phrasen wie „how are you" und „I'm fine". Obgleich dieses Format insgesamt durchaus als Test-Setting für die jungen SchülerInnen gut geeignet scheint (z.B. Enever 2011; Kessler 2009; Kötter 2009), besteht jedoch die Gefahr einer Einschränkung der Konstruktvalidität, da bereits erworbene Sprachkompetenzen auf diese Weise möglicherweise nicht in ihrer kompletten Breite erfasst werden können. So existieren etwa aus der Erwachsenenforschung Hinweise, dass symmetrische, weniger gesteuerte Test-Settings, wie beispielsweise sogenannte Peer-to-Peer-Settings, es den Getesteten erlauben, eine größere Bandbreite interaktionaler Kompetenzen aufzuzeigen (z.B. Ducasse & Brown 2009: 423). Gerade bei Lernenden auf einem sehr niedrigen Kompetenzniveau mit eher inselhaften mündlichen fremdsprachlichen Kompetenzen (z.B. Enever 2011: 7), wie es meist bei PrimarschülerInnen der Fall ist, könnte ein Nichteinbeziehen solcher Peer-to-Peer-Settings möglicherweise zu einer Unterschätzung der Sprachkompetenz führen. Dies ist zum Teil der Tatsache geschuldet, dass sich das jeweils Gesprochene thematisch auf die durch den Interviewer gestellten Fragen beschränkt und in einem sehr gesteuerten und engen zeitlichen Rahmen durchgeführt wird, so dass die Themen durch die interviewte Person nicht frei wählbar sind. Dies erhöht die Gefahr von fragenbedingten Kontexteffekten und somit der Einführung konstruktirrelevanter Schwierigkeitsvarianz. In freieren Settings hingegen haben Sprachenlernende im jeweiligen kommunikativen Rahmen die Möglichkeit, ein ihnen geläufigeres Thema einzubeziehen.

Im Bereich der Leistungsmessung von Erwachsenen wird daher mittlerweile häufig auf die simultane Testung zweier oder mehrerer Personen in Gruppen-Settings und zwischen fremdsprachlich „gleichrangigen" Personen („Peers"; symmetrische Settings) zurückgegriffen („Peer-to-Peer"-Settings; dazu z.B. Brooks 2009; Ducasse & Brown 2009; Taylor & Wigglesworth 2009). So werden beispielsweise im Rahmen des KET (Key English Test (A2); UCLES 2011a) auch bereits auf niedrigen Kompetenzniveaus solche symmetrischen Peer-to-Peer-Settings angewandt. Üblicherweise wird dabei durch die Testleitung ein Konversationsthema vorgegeben (siehe z.B. „Fit in Deutsch" 1/2; Goethe-Institut 2013). Aufgabe der Getesteten ist es dann, frei über den vorgegebenen Inhalt zu kommunizieren. Eine Bewertung der Sprechleistung erfolgt dabei üblicherweise durch geschulte Testleiter anhand von Rating-Skalen (z.B. ebd.). Für erwachsene Getestete hat sich dabei in empirischen Studien beispielsweise gezeigt, dass Peer-to-Peer-Test-Settings – im Gegensatz zu den asymmetrischen, eher traditionellen Test-Settings in Interviewform – zu insgesamt besseren Testergebnissen sowie auch zu komplexeren Interaktionen und Bedeutungsaushandlungsprozessen führten (z.B. Brooks 2009: 353; Taylor & Wigglesworth 2009: 329).

Es stellt sich daher die Frage, ob die Einbindung solcher offenerer, symmetrischer Test-Settings auch für junge Lernende sinnvoll sein könnte mit dem Ziel, eine mögliche Kompetenzunterschätzung durch zu starke thematische Festlegungen zu vermeiden und somit die Test- und Konstruktvalidität von eingesetzten Verfahren zu erhöhen. Jedoch ist durchaus fraglich, ob und inwieweit offene Settings für junge Sprachlernende geeignet sind. So existieren einerseits empirische Studien, die darauf hinweisen, dass junge Sprachlernende auf niedrigen Kompetenzniveaus vor allem auf gelernte Phrasen und Sätze beim Sprechen zurückgreifen (z.B. Roos 2009), was möglicherweise mehr für eine Verwendung von klar strukturierten und thematisch begrenzten Interview-Settings spricht; andererseits existieren Hinweise, dass Kinder in der Fremdsprache bereits auf einem niedrigen Kompetenzniveau Komponenten interaktiver Sprache produzieren können (z.B. Diehr & Frisch 2008; Diehr & Polte 2009; Enever 2011). In den oben beschriebenen Testverfahren werden zwar teilweise auch Test-Settings wie erlernte Mini-Dialoge zwischen Peers eingesetzt; diese sind jedoch durch die starke thematische Vorgabe und das vorherige Erlernen des Dialogs ebenso thematisch eingeschränkt und gesteuert wie es bei Interview-Settings der Fall ist. Insgesamt bleibt unklar, inwieweit Test-Settings mit einer freieren sprachlichen Gestaltbarkeit seitens der Getesteten geeignet sind für eine junge Gruppe von Lernenden und ob durch das Ermöglichen zusätzlicher interaktiver Sprachhandlungen auch bei dieser Gruppe die Konstruktvalidität erhöht werden kann. Bezüglich des Primarbereichs wurde unseres Wissens bislang kein systematischer Vergleich gesteuerter vs. offener Testsituationen realisiert. Somit wurde auch nicht gezielt unter-

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

sucht, *wie genau* die Art des Test-Settings (z.B. Interview- oder freie Peer-to-Peer-Settings) auf junge SprachlernanfängerInnen und deren durch das jeweilige Test-Setting induzierte interaktive Sprachhandlungen – und damit auch auf die Testvalidität bezüglich des Konstrukts „mündliche Interaktion" – wirkt.

3. Fragestellung

Ziel dieser Studie ist es, für eine konstruktvalide Erfassung mündlicher fremdsprachlicher Interaktionskompetenzen unterschiedliche (gesteuert/weniger gesteuerte bzw. symmetrische/asymmetrische) Test-Settings am Beispiel von Englisch als Fremdsprache in der Primarstufe zu untersuchen und zu vergleichen. Dabei soll hier explizit betont werden, dass es nicht Ziel der Studie ist, Aussagen über das Kompetenzniveau von SchülerInnen bzw. das Schwierigkeitsniveau von Sprachhandlungen zu machen.

Folgende drei Hauptziele werden verfolgt: Erstens soll untersucht werden, inwieweit offenere (d.h. in geringerem Maße gesteuerte), symmetrische Peer-to-Peer-Settings für Primarschulkinder im Vergleich zu den gut erprobten (gesteuerten und meist asymmetrischen) Interview-Settings hinsichtlich der Aktivierung von Sprachproduktion und Interaktion geeignet sind. Zweitens soll untersucht werden, inwieweit die Sprachproduktion in den beiden Settings in englischer Sprache stattfindet beziehungsweise in welchem Ausmaß in den unterschiedlichen Settings von den hier fokussierten jungen SprachlernanfängerInnen zur Bearbeitung der jeweiligen kommunikativen Aufgabe auf die Verkehrssprache Deutsch zurückgegriffen wird. Damit beziehen sich die beiden erstgenannten Punkte auf einen eher quantitativen Aspekt von Sprache im Sinne des Ausmaßes gesprochener Wörter und stattfindender Interaktionen (z.B. *turn taking* oder Wort ergreifen). Drittens interessiert, welche qualitativ unterschiedlichen interaktiven Sprachhandlungen *genau* in den verschiedenen Test-Settings ausgeführt werden, und inwieweit sich somit möglicherweise unterschiedliche Bereiche des Konstrukts „mündliche fremdsprachliche Interaktion" mithilfe der Verwendung unterschiedlicher Test-Settings abbilden lassen. Dies zielt auch auf die Frage ab, ob ein kombinierter Einsatz unterschiedlicher interaktiver Test-Settings die Konstruktvalidität erhöht oder ob der alleinige Einsatz von Interview-Settings – wie es bisher oft der Fall ist – gleiche Sprachaktivitäten abbildet und der Einsatz weiterer freier Settings aus Gründen der Testökonomie bei PrimarschülerInnen eher nicht notwendig ist.

Basierend auf oben genannten Ergebnissen aus der Erwachsenenforschung (z.B. Ducasse & Brown 2009) wird zusammenfassend die Hypothese aufgestellt, dass auch junge SprachlernanfängerInnen qualitativ und quantitativ unterschiedliche Sprachhandlungen in den unterschiedlichen Test-Settings produzieren. Folgende spezifische Forschungsfragen lassen sich aus den genannten Zielen ableiten:

3.1. Quantitativer Sprachhandlungsaspekt

Inwieweit unterscheiden sich die beiden Test-Settings „Interview" und „Peer-to-Peer" im Hinblick auf die Anzahl gesprochener Wörter in der Fremdsprache Englisch und in der Verkehrssprache Deutsch sowie hinsichtlich der Anzahl von zwischen den GesprächspartnerInnen stattfindenden Interaktionen?

Aufgrund der Tatsache, dass die an der Studie teilnehmenden SchülerInnen zu Beginn der vierten Klasse verhältnismäßig wenige Stunden fremdsprachlichen Unterrichts genossen haben (1 Jahr/2 Stunden pro Woche in Hessen), wird angenommen, dass insgesamt die englische Sprache eher bruchstückhaft und in formelhaften Ausdrücken (z.B. Roos 2009; vgl. auch GER-Deskriptoren für Niveau A1/A2) produziert wird. Es wird daher erwartet, dass die SchülerInnen in beiden mündlichen Settings, für die interaktiven und produktiven mündlichen Äußerungen, auch auf die deutsche Sprache (L1 bzw. L2) zurückgreifen, um ein Kommunikationsziel zu erreichen. Dabei soll hier betont werden, dass ein Rückgriff auf die deutsche Sprache nicht ausschließlich aufgrund eines niedrigen Kompetenzniveaus seitens der SprecherInnen erfolgen muss, sondern dass die Ursachen auch sozialer Natur sein können. So erfolgt die Kommunikation zwischen Peers in der Schule üblicherweise auf Deutsch, und auch die Gruppenarbeit im Englischunterricht findet häufig in deutscher Sprache statt (vgl. Elsner, Buendgens-Kosten & Hardy im Druck). Die fremdsprachliche Konversation im Englischunterricht findet dagegen häufig mit der Lehrperson statt, das Antworten auf Englisch auf eine von einer Lehrperson gleichfalls in englischer Sprache gestellte Frage ist eine häufige Unterrichtshandlung (z.B. Wolff 2009). Auch ist anzunehmen, dass von den SchülerInnen in den strukturierteren Inter-

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

view-Settings insgesamt mehr unterrichtlich erlernte fremdsprachliche Dialogabläufe, Redemittel und *chunks* angewendet werden können. Es wird daher ferner erwartet, dass im Rahmen von Interview-Settings insgesamt seltener auf die deutsche Sprache zurückgegriffen wird als in Peer-to-Peer-Settings.

Wie aus der Betrachtung existierender Instrumente (z.B. Enever 2011; Keßler 2009; Kötter 2009; UCLES 2013) ersichtlich wird, werden für Interview-Settings zumeist Fragen ausgewählt, die auf niedrigem Sprachkompetenzniveau beantwortet werden können (z.B. „How are you?“, „What's your name?“) und die üblicherweise lediglich eine relativ kurze Antwort erfordern. Bezüglich der freieren Peer-to-Peer-Settings hingegen wird angenommen, dass (im vorgegebenen Rahmen) eine stärkere sprachlich-thematische Lenkung seitens der Getesteten möglich ist, was zu vermehrter Sprachproduktion führt. Daher wird zusammenfassend erwartet, dass in Peer-to-Peer-Settings insgesamt sprachübergreifend mehr Sprache produziert wird, dazu jedoch seitens der SchülerInnen häufiger auf die deutsche Sprache zurückgegriffen wird als in Interview-Settings. Bezüglich der Häufigkeit von Interaktionen in den beiden Settings wird keine gerichtete Hypothese aufgestellt: Einerseits kann eine freiere thematische Lenkbarkeit in den Peer-to-Peer-Settings zu mehr Diskussion und damit zu mehr Interaktion zwischen den SchülerInnen führen. Andererseits existiert eine vergleichbare Anzahl von standardisierten Sprachanlässen in den beiden Settings (vgl. 4.2), woraus möglicherweise auch eine vergleichbare Anzahl von Interaktionen erfolgt. Die Betrachtung von Unterschieden in den beiden Settings im Hinblick auf die Anzahl stattfindender Interaktionen erfolgt daher explorativ.

3.2. Qualitativer Sprachhandlungsaspekt

Inwieweit unterscheiden sich die beiden Test-Settings im Hinblick auf Art und Anzahl interaktiver Sprachhandlungen?

Basierend auf den im GER beschriebenen interaktiven Sprachhandlungen in Interview-Kontexten (vgl. Europarat 2001: 85) sowie der Analyse empirischer Studien und existierender Testverfahren wird erwartet, dass im Rahmen von Interview-Settings vor allem das Beantworten von Fragen die vorwiegend vorkommende Sprachaktivität sein wird (siehe z.B. Keßler 2009; vgl. auch GER-Interaktions-Skala „Interview“: z.B. „Fragen beantworten“, „nachfragen“). In den offeneren Peer-to-Peer-Settings hingegen ist aufgrund der freieren thematischen Gestaltbarkeit seitens der getesteten Personen insgesamt eine größere Bandbreite unterschiedlicher Interaktions- und Sprachhandlungen zu erwarten. Auch bezieht sich ein größerer Teil der im GER beschriebenen interaktiven Sprachhandlungen auf Sprachaktivitäten, die vermutlich tendenziell (wenn auch nicht ausschließlich) eher in Peer-to-Peer-Settings als in klassischen Interview-Settings von Relevanz sind (etwa Skala „Zielorientierte Kooperation“: z.B. „Diskutieren, was als nächstes getan werden soll“, „Begründen und Erklären“; „Informelle Diskussion“: „Informationen austauschen“, „Vorschläge machen“; Europarat 2001: 81-83). Es wird daher erwartet, dass in Peer-to-Peer-Settings schwerpunktmäßig qualitativ andere Sprachhandlungen (bzw. eine größere Bandbreite von Sprachhandlungen) als in Interview-Settings stattfinden und somit das Konstrukt, so wie es im Rahmen des GER anhand der dort definierten interaktiven Sprachhandlungen beschrieben ist, dort insgesamt breiter erfasst wird. Dies sollte sich darin äußern, dass zum einen der Anteil der Aktivität „Fragen beantworten“ im Peer-to-Peer-Setting geringer sein sollte, als es im Interview-Setting der Fall ist, zum anderen die Anteile anderer interaktiver Sprachhandlungen, wie den oben beschriebenen, entsprechend größer sind und damit häufiger im Peer-to-Peer-Setting als im Interview-Setting getätigt werden.

4. Methode

4.1. Stichprobe

Die Daten wurden im Herbst 2012 an n=38 SchülerInnen (M(Alter)=9,4 Jahre; männlich=18; weiblich=20; Deutsch als Muttersprache=26; Türkisch als Muttersprache=12) zu Beginn der vierten Klasse an drei Schulen im Raum Frankfurt/Main erhoben[3]. Die SchülerInnen hatten zu dem Zeitpunkt etwa ein Jahr fremdsprachlichen Unterricht absolviert. Die Erhebung wurde dabei paarweise durchgeführt. Die Zusammensetzung der Paare erfolgte basierend auf der Muttersprache der SchülerInnen. Dies resultierte in insgesamt n=19 Paaren (13=DaM, 6=DaZ).

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

4.2. Durchführung

Hintergrund des Settings bildete die computerbasierte Geschichte „Ruben and the magic stones“ (MuViT, „Multilingual Virtual Talking Book“; Elsner 2011). Wir gehen dabei davon aus, dass durch die Bearbeitung einer Geschichte die Kinder besonders zum Sprechen und Mitdenken angeregt wurden (vgl. hierzu Cameron 2001; Ellis & Brewster 2002). Die Geschichte besteht aus insgesamt 16 Bildschirmseiten und wurde auf unterschiedlichen Input-Kanälen (visuell (Bild/Schrift); auditiv) vorgegeben: Die Kinder konnten vor- und zurückblättern, die Geschichte seitenweise durch die Software vorlesen lassen, den dazugehörigen Text an- und ausblenden. Die paarweise Bearbeitung (ca. 25 Min.) wurde videografiert. Die Kinder wurden zu Beginn sowie wiederholt auch während der Testung explizit instruiert, miteinander über die Geschichte zu kommunizieren sowie beim Sprechen über die Geschichte (Peer-to-Peer-Setting) und beim Beantworten der durch die Testleiterin auf Englisch gestellten Fragen (Interview-Setting) möglichst die englische Sprache zu verwenden. Da nach nur einem Jahr fremdsprachlichen Unterrichts jedoch ein eher niedriges Kompetenzniveau der SchülerInnen angenommen wurde, hatten diese außerdem die Möglichkeit, bei Nichtverstehen des Inhalts die Sprache, in der die Geschichte vorgegeben wurde, zu ändern bzw. bei fehlendem englischem Vokabular zur Bearbeitung der Kommunikationsaufgabe auf ihre jeweilige L1 oder L2 zurückzugreifen (Türkisch oder Deutsch). Des Weiteren wurde das Textverstehen der Kinder nach Abschluss der Kommunikationsaufgabe anhand von 18 Items verschiedenen Formats (Multiple Choice, Ergänzungsaufgaben, Richtig-Falsch, Zuordnungsaufgaben) schriftlich überprüft. Die Kinder wiesen dabei insgesamt ein hohes Textverstehen auf (m=15,3; s=3,11).

Die Bearbeitung der Geschichte erfolgte in mehreren sequentiell abwechselnd aufeinanderfolgenden Phasen, die jeweils eines der beiden fokussierten Test-Settings „Interview“ oder „Peer-to-Peer“ abbildeten. So erfolgten etwa zu Beginn der Testung als erster Testteil ein standardisiertes Eingangsinterview auf Englisch sowie die auf die Geschichte bezogene Instruktion inklusive Darstellung relevanten Schlüsselvokabulars (siehe Tabelle 1).

Tabelle 1: Sprachanlässe im Interview- und Peer-to-Peer-Setting

Sprachanlässe Interview	**Sprachanlässe Peer-to-Peer**
What’s your name?	Seiten 1-16 der Geschichte (Instruktion: Beim gemeinsamen Lesen und Bearbeiten der Geschichte miteinander mündlich über die Geschichte kommunizieren, soweit möglich in englischer Sprache)
How are you?	
How old are you?	
Do you like English?	
Can you repeat what I said?	
Do you know what (a book/flag/which flag this/xy) is?*	
Have you understood/got any more Questions?	
Prompts/Interview-Sprachanlässe während der Geschichte:	
What can you see in the picture?	
Ruben wants to be at school. Do you know why?	
What happens next? What do you think?	
What does Ruben say the next morning?	
What would you like to be? Ruben wants to be a builder, a pilot and a policeman; what about you?	

(variiert teilweise situativ (abhängig v. Schülerverständnis); je Video 5-6 Verständnisfragen)

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

Zu Beginn der eigentlichen Geschichte erfolgte dann die Überleitung in das erste Peer-to-Peer-Setting, indem die Kinder von den Testleiterinnen (geschulte Englisch-Lehramtsstudierende) instruiert wurden, sich von dort an – möglichst in englischer Sprache – frei über die Geschichte zu unterhalten (z.B. Fragen stellen, gegenseitig Inhalte erklären). Das Peer-to-Peer-Setting wurde dabei an insgesamt fünf Stellen von kürzeren, auf die Geschichte bezogenen und standardisierten Interview-Phasen (*Prompts*) abgewechselt, was – trotz der damit einhergehenden zwischenzeitlichen Strukturierung der Kommunikation – den Zweck hatte, die Interview-Phasen an die Thematik der Geschichte zu knüpfen. Es wird davon ausgegangen, dass so weitgehend eine inhaltliche Vergleichbarkeit der unterschiedlichen Test-Settings geschaffen wurde. Wir gehen ferner davon aus, dass trotz der Unterschiedlichkeit der Test-Settings deren Vergleichbarkeit außerdem durch eine vergleichbare Anzahl standardisierter Sprachanlässe gegeben ist: Im Peer-to-Peer-Setting durch das Bearbeiten einer jeweils neuen Seite des virtuellen Buches und somit eines neuen Aspektes der Geschichte, in den Interview-Settings durch vier standardisierte Interview-Fragen zu Beginn, meist fünf bis sechs darauf folgende Instruktions- und verständnisbezogene Fragen sowie fünf standardisierte Prompts während der Geschichte. So ist der Anteil an Interview- (6) und Peer-to-Peer-Sequenzen (5) sowie die Anzahl der jeweils pro Setting vorhandenen Sprachanlässe (16 in Peer-to-Peer-, ca. 15-16 in Interview-Settings) weitestgehend ausbalanciert. Um eine zuverlässige Codierung der Sprachhandlungen gewährleisten zu können, wurden die Videos im Anschluss von geschulten studentischen Hilfskräften transkribiert (nach Krummheuer & Fetzer 2009).

4.3. Auswertung: Codierung der quantitativen und qualitativen Sprachhandlungsaspekte

Die Codierung des *quantitativen Sprachhandlungsaspekts* (*Menge* der Sprachhandlungen) als Maß für die interaktive Sprachaktivierung in den beiden Test-Settings erfolgte anhand zweier Indikatoren. Zum einen wurden die während der Testung stattfindenden Interaktionen codiert und anschließend ausgezählt. Interaktionen wurden dabei definiert als Äußerungen, die entweder explizit und erkennbar an eine Person gerichtet wurden (beispielsweise Fragen stellen oder Inhalt übersetzen/erklären), oder die eine direkte Reaktion auf die Äußerung einer anderen Person darstellten (z.B. Fragen beantworten, Aussage ergänzen). Zum anderen erfolgte eine Auszählung der jeweils in englischer und deutscher Sprache gesprochenen Wörter (*tokens*) in den einzelnen Test-Settings als ein globales Maß für Sprachproduktion in den unterschiedlichen Settings.

Aufgrund der sehr geringen Nutzung der englischen Sprache vor allem im Rahmen des Peer-to-Peer-Settings (vgl. Tabelle 4) wurden für die Analyse des *qualitativen Sprachhandlungsaspekts* Äußerungen der getesteten SchülerInnen sowohl in deutscher als auch in englischer Sprache einbezogen. Zum einen hatte dies das Ziel, sprachübergreifend zu analysieren, welche potentiellen Sprachhandlungen die Settings jeweils hervorrufen. Dabei ist zu betonen, dass den Autorinnen bewusst ist, dass eine mögliche Wechselwirkung zwischen Sprache und Art der Äußerung hierbei nicht berücksichtigt wird. Da wir jedoch davon ausgehen, dass eine Sprachsituation bzw. ein Sprachsetting bestimmte Sprachhandlungen sprachübergreifend hervorrufen sollte und die Wahl der Sprachhandlung damit eher situations- und weniger sprachabhängig ist, haben wir das beschriebene Vorgehen gewählt. Dennoch soll auch betont werden, dass durchaus die Notwendigkeit gesehen wird, die Frage einer möglichen Wechselwirkung von verwendeter Sprache und Test-Setting im Rahmen weiterer Studien systematisch zu untersuchen. Zunächst soll hier jedoch primär die Frage beantwortet werden, ob anhand der unterschiedlichen Test-Settings überhaupt systematisch und gezielt unterschiedliche Teile des Konstrukts „fremdsprachliche Interaktion“ bei PrimarschülerInnen abgebildet werden *können.*

Die Codierung hinsichtlich der *Art* der Sprachhandlungen (qualitativer Sprachhandlungsaspekt) erfolgte auf Basis des GER sowie den dort beschriebenen Sprachhandlungen (Skalen: interaktive Aktivitäten und Strategien). Da die GER-Skalen jedoch häufig als teilweise unvollständig (vgl. z.B. Alderson, Figueras, Kuijper, Nold, Takala & Tardieu 2006; Figueras, North, Takala, Verhelst & van Avermaet 2005), bzw. als für PrimarschülerInnen nicht geeignet (vgl. Enever 2011) kritisiert werden, war es zunächst notwendig zu untersuchen, ob diese sich als Basis für die Codierung der interaktiven Sprachhandlungen eigneten. Da im Rahmen der GER-Skalen interaktive Sprachhandlungen größtenteils für spezifische Kontexte beschrieben werden, die zudem häufig eher für Erwachsene als für Kinder geeignet sind (siehe dazu z.B. Diehr & Frisch 2008; Hasselgren 2003, 2005), wurden zunächst alle dort beschriebenen Sprachhandlungen, unabhängig von Kompetenzniveau und inhaltlichem Kontext, extrahiert. Dies diente dazu,

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

beschriebene Sprachhandlungen situationsunabhängig und somit auf andere Test-Settings übertragbar zu machen. Im Rahmen der hier dargestellten Studie wurden zudem nur spezifische, konkrete Sprachhandlungen wie „Frage beantworten“ für die Codierung verwendet, übergreifendere und globalere Sprachhandlungen wie beispielsweise „sich verständigen“ wurden wegen ihres wenig konkreten, eher globalen Charakters nicht mit einbezogen. Auch wurden ähnliche, in unterschiedlichen Deskriptorskalen oder auf unterschiedlichen Kompetenzniveaus beschriebene Sprachhandlungen wie „Frage beantworten“ und „Auf eine Frage reagieren“ zusammengefasst. So wurden zunächst insgesamt 35 Sprachhandlungen identifiziert (vgl. Tabelle 7 im Anhang).

Im nächsten Schritt wurde dann überprüft, welche dieser interaktiven Sprachhandlungen tatsächlich in den Settings getätigt wurden, d.h. in den Transkripten mindestens einmal von mindestens einer Beurteilerin codiert werden konnten. Bis auf zwei im GER nicht explizit aufgeführte Sprachhandlungen („beschließen“ und „Überlegung anstellen“), die zusätzlich in die Liste der codierbaren Sprachhandlung mit aufgenommen wurden, erwiesen sich die im GER beschriebenen mündlich-produktiven Interaktionshandlungen für eine solche Codierung als geeignet. Daraus resultierten insgesamt 22 in den beiden Test-Settings codierte Sprachhandlungen. Hier zeigt sich, dass nicht alle der vormals anhand der GER-Skalen identifizierten interaktiven Sprachhandlungen im Rahmen der beiden untersuchten Settings getätigt wurden, also das im GER beschriebene Konstrukt auch anhand dieser beiden Test-Settings nicht vollständig abgebildet werden konnte. Folgende Sprachhandlungen wurden in das Codierschema (vgl. Tabelle 2) und die anschließenden Analysen mit einbezogen:

Tabelle 2: GER-basiertes Codierschema für Sprachhandlungen

Sprachhandlungen	
Frage stellen	Unwissenheit/ Unentschiedenheit anzeigen
Frage beantworten/auf Frage reagieren	Beschließen
Inhalt/erzählen/Übersetzen/zusammenfassen	Vorschlagen
Verstehen anzeigen	Feststellen/ kommentieren
Vermutung anstellen	Anweisungen geben/auffordern
Rückfrage stellen/klären	Inhalt aushandeln
Einverständnis signalisieren	Begründen
Anweisungen folgen	Vorgehen besprechen
Aufgabe selbständig weiterführen	Information prüfen
Informationen ergänzen/an Aussage anknüpfen	Überlegungen anstellen
Beipflichten/zustimmen	Ablehnen/widersprechen

Im Anschluss daran erfolgte die Codierung aller Sprachhandlungen in allen 19 Transkripten durch zwei unabhängige geschulte Beurteilerinnen. Als Maß für die Beurteilerübereinstimmung[4] wurde der Intra-Klassen-Koeffizient (ICC)[5] berechnet (vgl. Tabelle 6 im Anhang) für alle Sprachhandlungen, die häufiger als 20 Mal (d.h. häufiger als im Mittel einmal pro Video) codiert wurden. Bezüglich des Peer-to-Peer-Settings beträgt dabei die mittlere Übereinstimmung über alle Sprachhandlungen hinweg ICC=.76, bezüglich des Interview-Settings hingegen sind die Übereinstimmungen insgesamt etwas niedriger (mittlere Übereinstimmung: ICC=.65), jedoch noch immer als ausreichend zu bewerten. Des Weiteren zeigt sich, dass die einzelnen Beurteilerübereinstimmungen größtenteils Signifikanz aufweisen und weitestgehend im ausreichenden bis sehr guten Bereich liegen. Für die folgenden Analysen werden die über beide Beobachterinnen gemittelten Ratings verwendet.

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

5. Analysen & Ergebnisse

Da die getesteten Schülerpaare teilweise aus SchülerInnen mit Deutsch als Zweitsprache bzw. Türkisch als Erstsprache bestanden, wurde zunächst anhand von t-Tests für abhängige Stichproben überprüft, ob sich die Testpaare mit unterschiedlichen Herkunftssprachen signifikant hinsichtlich der Häufigkeit bzw. Art der Sprachhandlungen unterschieden. Da dies nicht der Fall war (p=.09-.97; df=17) und auch (bis auf sehr wenige Ausnahmen bezüglich einzelner Wörter) hinsichtlich der Sprachproduktion seitens der Kinder nicht auf die türkische Sprache zurückgegriffen wurde, sondern ebenfalls ggf. auf die Verkehrssprache Deutsch, wird die Herkunftssprache bei den folgenden Auswertungen nicht mit berücksichtigt.

5.1. Quantitativer Sprachhandlungsaspekt

Um die Frage bezüglich der unterschiedlichen Aktivierung von Sprachproduktion in den beiden Test-Settings beantworten zu können, wurden die in den verschiedenen Peer-to-Peer-Sequenzen bzw. Interview-Sequenzen pro Paar getätigten Interaktionen und Wörter jeweils ausgezählt und addiert. Es wurde nicht unterschieden zwischen Interaktionen, die zwischen den Peers bzw. zwischen einem der Kinder und der Testleitung stattfanden, jedoch wurden ausschließlich Schüleräußerungen in die Analysen einbezogen.

Tabelle 3: Deskriptive Statistiken und t-Tests (Anzahl Wörter und Interaktionen)

	Mittelwert*	**sd**	**T**	**(p)**
Wörter Englisch Gesamt	39.42	39.49	-11.12	(.000)
Wörter Deutsch Gesamt	535.84	190.17		
Wörter Englisch Interview	27.84	20.54	3.70	(.002)
Wörter Englisch Peer-to-Peer	11.58	23.26		
Wörter Deutsch Interview	118.16	50.39	-9.43	(.000)
Wörter Deutsch Peer-to-Peer	417.68	158.55		
Interaktion Interview	54.05	12.24	-. 89	(.385)
Interaktion Peer-to-Peer	58.42	21.94		
Wörter Interview Gesamt	146.00	48.53	-8.430	(.000)
Wörter Peer-to-Peer Gesamt	429.26	164.77		
Wörter/Interaktion Interview	2.78	0.95	-7.590	(.000)
Wörter/Interaktion Peer-to-Peer	7.75	2.95		

*t-Tests für abhängige Stichproben; *Mittlere Anzahl gesprochener Wörter/Interaktionen pro Video*

In den Interview-Settings fanden dabei im Mittel 54,05 Interaktionen zwischen den beiden TestpartnerInnen bzw. zwischen einem der getesteten Kinder und der Testleitung statt, die Anzahl der durchschnittlich getätigten Interaktionen in den Peer-to-Peer-Settings lag mit 58,42 nur geringfügig höher (vgl. Tabelle 3). Dieser deskriptiv ermittelte Unterschied erwies sich in einem t-Test für abhängige Stichproben als statistisch nicht bedeutsam. Dies weist darauf hin, dass beide Settings ähnlich gut dazu geeignet scheinen, Interaktionen auszulösen.

Hinsichtlich der Häufigkeit der in den unterschiedlichen Test-Settings gesprochenen Wörter zeigte sich, dass trotz einer vergleichbaren Anzahl geschaffener Sprachanlässe und der vergleichbaren Anzahl stattfindender Interaktionen insgesamt, d.h. sprachübergreifend, im Mittel deutlich mehr Wörter pro Testpaar in den Peer-to-Peer-Settings als in

den Interview-Settings gesprochen wurden. Auch wurden im Mittel signifikant mehr Wörter pro Interaktion im Peer-to-Peer-Setting produziert. Diese deutlichen Mittelwertsunterschiede erweisen sich den Erwartungen gemäß als statistisch signifikant und können als Effekte des Test-Settings interpretiert werden. Des Weiteren zeigt sich, entsprechend der eingangs aufgestellten Hypothesen, dass insgesamt signifikant häufiger die deutsche als die englische Sprache zur Kommunikation herangezogen wurde. Es wurden zudem, gleichfalls den Annahmen entsprechend, signifikant mehr Wörter in englischer Sprache im Rahmen des Interview-Settings produziert. Diese Ergebnisse hinsichtlich der sehr seltenen Nutzung der englischen Sprache (im Mittel 39,42 Wörter pro Video) seitens der ViertklässlerInnen könnten ein Hinweis darauf sein, dass es den Kindern nach einem Jahr Englisch-Unterricht insgesamt noch schwer fällt, die englische Sprache in eher offenen, symmetrischen Interaktions- bzw. Konversationssituationen zu verwenden. Die wie erwartet signifikant häufigere Nutzung englischer Sprache in den Interview-Sequenzen kann dabei zum einen darauf hindeuten, dass für die GrundschülerInnen solche Settings, in denen auch und vor allem vorformulierte und geübte Redewendungen zum Einsatz kommen können, insgesamt leichter zu sein scheinen, andererseits könnte dies auch auf die Relevanz sozialer Faktoren bei der Sprachwahl hindeuten.

5.2. Qualitativer Sprachhandlungsaspekt

Bezüglich der *Art der Sprachhandlungen* zeigt sich, dass zwischen den beiden Settings, wie erwartet, sprachübergreifend deutliche Unterschiede existieren. Für die Analysen wurden dabei analog dem oben gewählten Vorgehen die Sprachhandlungen einbezogen, die in mindestens einem Test-Setting insgesamt häufiger als 20 Mal codiert wurden. Somit wurden insgesamt 13 Sprachhandlungen in die Analysen einbezogen.

Tabelle 4: Deskriptive Statistiken Sprachhandlungen

	Interview		**Peer-to-Peer**	
Sprachhandlung	Mittelwert*	sd	Mittelwert*	sd
Fragen/ Rückfragen stellen/ klären	2.89	3.08	5.76	3.42
Fragen beantworten/ reagieren	19.74	5.59	6.13	4.43
Inhalt erzählen/ übersetzen/ zusammenfassen	0.82	0.73	15.39	8.13
Verstehen anzeigen	1.31	0.95	1.7	1.84
Vermutung anstellen	1.05	0.86	1.36	2.03
Anweisungen folgen	1.45	1.23	8.05	4.44
Aufgabe weiterführen	0.92	0.90	7.47	4.09
Inhalt ergänzen/ anknüpfen	2.42	1.83	14.39	13.93
Beipflichten/ zustimmen	0.84	0.91	2.53	2.60
Beschließen	0.18	0.30	1.53	1.90
Feststellen/kommentieren	1.00	1.07	4.47	3.55
Anweisung geben/ auffordern	0.26	0.48	1.74	2.31
Ablehnen/widersprechen	0.08	0.19	1.53	1.90

Mittlere Häufigkeit pro Interview (Gemittelte Ratings/2 Rater)

Deskriptiv (Tabelle 4) zeigt sich, dass „Fragen stellen“ seitens der Kinder im Mittel im Interview-Setting häufiger vorkam, während die anderen Sprachhandlungen im Mittel häufiger im Peer-to-Peer-Setting getätigt wurden. Diese deskriptiven Unterschiede wurden im nächsten Schritt auf ihre statistische Bedeutsamkeit hin überprüft. Als Analysemethode wurde eine einfaktorielle Varianzanalyse mit Messwiederholung gewählt. Multivariate Tests zeigen

einen signifikanten Haupteffekt des Messwiederholungs-Faktors „Test-Setting“ (F=20,89; df=13; p=.000; eta²= .978). Dies ist so zu interpretieren, dass insgesamt ein statistisch relevanter Unterschied bezüglich der Häufigkeit von verwendeten Sprachhandlungen zwischen den beiden Test-Settings besteht. Im nächsten Schritt wurden die Sprachhandlungen anhand univariater Varianzanalysen im Hinblick auf Mittelwertsunterschiede in den beiden Test-Settings verglichen.

Tabelle 5: Varianzanalyse mit Messwiederholung/Univariate Effekte

Maß	**F**	**Sig.***	**Eta²**
Fragen/Rückfragen stellen/ klären	11.729	.003	.395
Fragen beantworten/ reagieren	90.983	.000	.835
Inhalt erzählen/ übersetzen/ zusammenfassen	58.900	.000	.766
Aufgabe selbst. weiterführen	61.455	.000	.773
Verstehen anzeigen	1.703	.208	.086
Vermutung anstellen	0.486	.494	.026
Anweisungen folgen	45.521	.000	.717
Inhalt ergänzen/ anknüpfen	15.657	.001	.465
Beipflichten/zustimmen	10.419	.005	.367
Beschließen	10.349	.005	.365
Feststellen/ kommentieren	21.411	.000	.543
Anweisung geben/ auffordern	7.966	.011	.307
Ablehnen/widersprechen	6.761	.018	.273

Darstellung univariater Effekte; Gemittelte Ratings/2 Rater
Haupteffekt Messwiederholungsfaktor: F=20,892 (.001), eta²=.978;

Insgesamt zeigen sich für 11 der 13 Sprachhandlungen signifikante Mittelwertsunterschiede (Tabelle 5). Der Annahme entsprechend existiert dabei zugunsten des Interview-Settings ein signifikanter Mittelwertsunterschied bezüglich der Sprachhandlung „Fragen beantworten“, die im Interview-Setting signifikant häufiger durchgeführt wurde als im Peer-to-Peer-Setting. Bezüglich weiterer zehn Variablen zeigt sich hingegen ein signifikanter Mittelwerts–unterschied zugunsten des Peer-to-Peer-Test-Settings. Auch dies geht mit der eingangs aufgestellten Hypothese einher, dass im Rahmen des Peer-to-Peer-Settings insgesamt eine größere Bandbreite von Sprachhandlungen getätigt wird und hier, zumindest bei häufigem Rückgriff auf die deutsche Sprache, somit potentiell ein größerer Teil des Konstrukts „mündliche Interaktion“ erfasst werden kann.

6. Beantwortung der Fragestellungen und Diskussion

Bezüglich des quantitativen Sprachhandlungsaspekts zeigt sich wie erwartet, dass die SchülerInnen bei der Bearbeitung der Aufgaben insgesamt deutlich häufiger auf die deutsche als auf die englische Sprache zurückgreifen. Dies tun sie im Rahmen von Peer-to-Peer-Settings außerdem signifikant häufiger als im Interview-Setting. Dies kann dafür sprechen, dass den SchülerInnen die produktive und interaktive Verwendung der Fremdsprache im Rahmen von offeneren, symmetrischen Settings, wie dem hier verwendeten Peer-to-Peer-Setting, insgesamt schwer fällt. Diese häufigere Nutzung der Fremdsprache spricht daher für eine Verwendung des Interview-Settings zur Erfassung fremdsprachlicher Kompetenzen in der Primarschule. Eine mögliche Ursache könnte hier eine höhere Schwierigkeit

des Peer-to-Peer-Settings sein: So können in dieser Art von Setting, vermutlich auch aufgrund der Unvorhersehbarkeit des Gesprächsverlaufs, kaum auswendig gelernte und oft stark situationsabhängige Redewendungen verwendet werden, wie sie im fremdsprachlichen Unterricht häufig gelehrt werden (vgl. Kolb 2008). Diese machen jedoch mutmaßlich einen großen Teil der auf niedrigen Kompetenzniveaus vorhandenen und abrufbaren bzw. anwendbaren produktiv-interaktiven Sprachhandlungen aus (z.B. Europarat 2001). Ferner sind in den Interview-Settings gestellte Fragen teilweise anhand eines einzelnen Wortes beantwortbar. So lautet die Eingangsfrage einer der Interview-Sequenzen etwa „What do you see in the picture?". Häufig gegeben Antworten sind hier beispielsweise „cat", „mom" oder „breakfast". Diese Aufgabe kann so von den SchülerInnen in englischer Sprache gelöst werden. Im Peer-to-Peer-Setting hingegen wurden die SchülerInnen instruiert, sich frei über die Geschichte zu unterhalten. Eine mögliche Ursache für die geringe Nutzung der englischen Sprache in diesen Settings könnte hier sein, dass der Wortschatz der meisten SchülerInnen nach nur einem Jahr Englischunterricht noch nicht ausreichend und vielfältig genug ist, um diese Aufgabe in englischer Sprache lösen zu können. Andererseits sprechen empirische Studien (z.B. Diehr & Polte 2009; Enever 2011) dafür, dass SchülerInnen in der Grundschule bereits relativ flüssig englische Sprache produzieren können. Auch sollte bedacht werden, dass in der hier dargestellten Studie das für den Inhalt der Geschichte relevante Schlüsselvokabular zu Beginn der Geschichte sogar über mehrere Inputkanäle explizit vorgegeben und so eine Lerngelegenheit geschaffen wurde, so dass die Kinder vermutlich mehr englische Wörter in ihren geschichtsbezogenen Interaktionen hätten verwenden können. Auch zeigt das häufige Vorkommen der Sprachhandlung „Inhalt erzählen/zusammenfassen/übersetzen" in den Peer-to-Peer-Settings, dass der zumeist auf Englisch vorgegebene Inhalt seitens der SchülerInnen häufig im Sinne einer Übersetzung bzw. Zusammenfassung ins Deutsche wiedergegeben wurde. Dies kann als ein Hinweis darauf interpretiert werden, dass die Kinder den Inhalt des fremdsprachlichen Textes durchaus verstehen konnten, was ferner auch durch das Ergebnis hinsichtlich des separat erhobenen Textverständnisses bestätigt wird.

Zusammengenommen könnte dies darauf hinweisen, dass die Kinder möglicherweise durchaus in der Lage wären, auch in dieser Form von Setting - zumindest in Form einzelner eben erlernter Wörter oder *chunks* - häufiger als hier geschehen auf die englische Sprache zurückzugreifen. Die Verwendung englischer Sprache hängt daher möglicherweise nicht einzig vom Sprachkompetenzniveau, sondern auch vom Aufgabentyp bzw. vom jeweils verwendeten Setting ab. Wie oben bereits kurz dargestellt wurde, spielen dabei vermutlich auch soziale Faktoren eine Rolle. So wird zwischen einander meist gut bekannten MitschülerInnen in vielen anderen sozialen Situationen üblicherweise die Mutter- oder Verkehrssprache verwendet, auf Deutsch gestellte Fragen werden auch wiederum auf Deutsch beantwortet, und auch im Englischunterricht findet Gruppenarbeit zum großen Teil auf Deutsch statt (vgl. Elsner et al. im Druck). Die Verwendung englischer Sprache mit anderen Peers ist daher möglicherweise sehr ungewohnt. Das Beantworten einer durch eine Lehrperson gestellten englischsprachigen Frage auf Englisch (soweit es die sprachlichen Mittel zulassen) hingegen kommt im Unterricht häufig vor (z.B. Engel 2009; Wolff 2009), und dieses sprachliche Verhalten wird von den SchülerInnen möglicherweise auch auf die Testsituation übertragen. Diese Annahmen sowie auch möglicherweise diesbezüglich existierende Unterschiede von Kindern mit Deutsch als L1 oder als L2 müssen jedoch anhand weiterer empirischer Studien überprüft werden.

Die sehr geringe Nutzung der englischen Sprache in den Peer-to-Peer-Settings spricht auf den ersten Blick dafür, dass dieses vergleichsweise offene und weniger strukturierte Test-Setting zur Messung der fremdsprachlichen Interaktionskompetenz junger Sprachlernender (zumindest nach nur einem Jahr Englischunterricht) tendenziell eher ungeeignet ist. Betrachtet man jedoch die sprachübergreifende Sprachproduktion („Anzahl gesprochener Wörter insgesamt") in den beiden Test-Settings, dann werden, unabhängig von der verwendeten Sprache, insgesamt deutlich mehr Wörter im Rahmen des Peer-to-Peer-Settings produziert, und es finden insgesamt vielfältigere Interaktionen zwischen den GesprächspartnerInnen statt. Dies spricht wiederum, zumindest unter weitgehend optimalen Bedingungen wie sie hier durch die Möglichkeit zum Rückgriff auf die Verkehrssprache gegeben waren, für eine insgesamt bessere Sprachaktivierung in den Peer-to-Peer-Settings. Hasselgren (2003, 2005) plädiert dafür, gerade bei jungen Lernenden altersgerechte und motivierende Aufgabentypen und Items zu verwenden. Dabei ist insbesondere das Rezipieren und Arbeiten mit Geschichten ein zentraler Bestandteil des frühen Englischunterrichts, nicht zuletzt deshalb, weil es SchülerInnen zum Sprechen, Nachfragen und Mitdenken anregt (vgl. hierzu Cameron 2001; Ellis & Brewster 2002). Da beide Test-Settings geschichtsbezogene Sprachanlässe beinhalteten, sprechen die Ergebnisse der vorliegenden Studie möglicherweise auch dafür, dass die freie Bearbeitung der verwendeten Geschichte für die

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

GrundschülerInnen noch interessanter und motivierender war als das Beantworten der Fragen in den Interview-Settings, was wiederum für einen Einsatz des Peer-to-Peer-Settings sprechen würde.

Auch die Ergebnisse hinsichtlich des qualitativen Sprachhandlungsaspekts sprechen für dessen Verwendung: Zusammenfassend können diese als ein erster Hinweis darauf gewertet werden, dass die verschiedenen Test-Settings, wie eingangs angenommen, tatsächlich schwerpunktmäßig unterschiedliche Teile des Konstrukts „mündliche fremdsprachliche interaktive Kompetenz" abbilden und erfassen. Vor allem die deutlich größere Bandbreite an durchgeführten Sprachhandlungen in Peer-to-Peer-Settings spricht für diese Interpretation. Ferner ist, wie ebenfalls angenommen, die mit Abstand am häufigsten vorkommende Sprachhandlung im Interview-Setting das Beantworten von Fragen. Im Rahmen des Peer-to-Peer-Settings hingegen spielt dies eine signifikant geringere Rolle, hier liegt der Handlungsschwerpunkt auf dem Erzählen und Übersetzen des Geschichtsinhalts. Dafür, dass dies häufig von beiden Mitgliedern eines Testpaars gemeinsam vollzogen wird und somit tatsächlich interaktive Peer-to-Peer-Sprach–handlungen aktiviert wurden, spricht das häufige Vorkommen der Sprachhandlung „ergänzen/anknüpfen" an eine vorausgehende Aussage des jeweiligen Peers. Dies äußert sich häufig darin, dass ein Gesprächspartner bzw. eine Gesprächspartnerin damit beginnt, den Inhalt einer Seite zu übersetzen oder zu erzählen, und das jeweils andere Kind ihn oder sie dann durch weitere Informationen oder Annahmen ergänzt, was in vielen Fällen zu mehrfachen Sprecherwechseln (*turns*) zwischen den SchülerInnen führt. Auch diese Ergebnisse könnten darauf hindeuten, dass diese Art von Test-Setting (bei Betrachtung aller sprachlicher Äußerungen, unabhängig davon, ob in Verkehrssprache Deutsch oder der Fremdsprache Englisch getätigt) insgesamt tendenziell sprachaktivierender wirkt als Interview-Settings, in denen die Kinder häufig vor allem zwar Fragen der Lehrperson beantworten, darüber hinaus jedoch nicht viel mehr Eigenes beitragen. Dafür spricht auch die im Verhältnis zu den Peer-to-Peer-Settings deutlich geringere Anzahl gesprochener Wörter pro Interaktion: Fragen wie „was denkst Du, was am nächsten Tag passiert" oder „was möchtest Du einmal werden" hätte den SchülerInnen durchaus die Möglichkeit offen gelassen, ausführlicher darüber nachzudenken, zu diskutieren, eigene Ideen einzubringen und so auch im Rahmen dieser Settings deutlich mehr Sprache zu produzieren. Dass diese Möglichkeit nicht oder kaum genutzt wurde, spricht möglicherweise dafür, dass solche Interview-Settings Sprachproduktion, zumindest bei jungen Lernenden, tendenziell eher hemmen, zumal die SchülerInnen auch in diesen Test-Settings auf die deutsche Sprache zurückgreifen konnten (und auch zurückgegriffen haben).

In Bezug auf die Erhöhung der Konstruktvalidität von Testverfahren durch einen zusätzlichen Einsatz von Peer-to-Peer-Settings scheint es durch die additive Verwendung unterschiedlicher Settings gelungen, das Konstrukt interaktiver mündlicher Kompetenz insgesamt breiter abzubilden, als es anhand einzelner Test-Settings der Fall wäre. Dafür spricht die Tatsache, dass den Erwartungen entsprechend in den Peer-to-Peer-Settings insgesamt mehr qualitativ unterschiedliche Sprachhandlungen getätigt wurden als in den Interview-Settings, wohingegen das für Konversationen und interaktive Handlungen hoch relevante Beantworten von Fragen einen Schwerpunkt in Interview-Settings darstellt. Somit scheinen sich die beiden Testformate hinsichtlich der jeweils induzierten Sprachhandlungen zu ergänzen. Eine Verwendung unterschiedlicher Test-Settings führt somit sehr wahrscheinlich zu einer Erhöhung der Konstruktvalidität des Tests. Bei der Entwicklung neuer Testverfahren oder der Erweiterung existierender Verfahren sollte eine Einbindung weiterer Settings, wie es ja bei Testverfahren zur Messung fremdsprachlicher Kompetenzen von Erwachsenen bereits üblich ist, daher auch für die Testung von Kindern im Primarschulalter unbedingt stärker berücksichtigt werden.

Dennoch wurde auch deutlich, dass anhand der hier verwendeten Test-Settings längst nicht alle konstruktrelevanten Sprachhandlungen abgebildet werden konnten. So kamen von den 37 ursprünglich identifizierten Sprachhandlungen lediglich 22 mindestens einmal vor, von denen wiederum einige im Mittel nur einmal oder sogar noch deutlich seltener pro Video getätigt wurden. Es stellt sich hier die Frage, ob diese Sprachhandlungen bei so einer geringen Häufigkeit in den verwendeten Settings überhaupt reliabel und valide erfasst werden können. Denkbar wäre hier beispielsweise eine Erweiterung um ein weiteres Test-Setting, in dem gezielt Sprachhandlungen wie „vorschlagen", „beschließen", „aushandeln", „widersprechen" etc. induziert werden, wie etwa das gemeinsame Lösen einer vorgegebenen Aufgabe oder das gemeinsame Treffen von Entscheidungen basierend auf alternativen Handlungsmöglichkeiten.

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

Die hier berichteten Ergebnisse basieren mit n=38 SchülerInnen und somit n=19 getesteten Paaren auf einer relativ kleinen Stichprobe. Die Ergebnisse können daher lediglich als eine erste Tendenz betrachtet werden. Des Weiteren ist es gerade in den Peer-to-Peer-Settings nicht gelungen, die Kinder zu einer vermehrten Nutzung englischer Sprache zu motivieren. Mögliche Gründe dafür wurden oben bereits diskutiert. Denkbar wäre diesbezüglich möglicherweise, dass in der Instruktion die Verwendung der englischen Sprache noch deutlich mehr betont und die SchülerInnen auf diese Weise dazu angehalten werden sollten, ausschließlich die englische Sprache zu verwenden. Ob eine Nutzung des Peer-to-Peer-Formats unter ausschließlicher Nutzung englischer Sprache auf dem vermutlich niedrigen Kompetenzniveau von jungen SprachlernanfängerInnen jedoch überhaupt möglich ist und ob dann auch im Peer-to-Peer-Setting noch die gleiche Anzahl und die gleiche Art qualitativ unterschiedlicher Sprachhandlungen produziert werden (oder eine Wechselwirkung zwischen Sprache und Test-Setting besteht), konnte anhand der vorliegenden Studie nicht beantwortet werden und sollte daher anhand weiterer empirischer Studien überprüft werden.

Um eine stärkere Verwendung der englischen Sprache in Peer-to-Peer-Settings zu fördern, könnte auch eine gezielte Nutzung von Interview-Settings direkt zu Beginn einer Testung möglicherweise gut als Einstieg dienen: Bekommen die SchülerInnen Fragen gestellt, die ihnen bekannt und geläufig sind, können sie gleich zu Beginn der Testung unter Rückgriff auf bekannte und vertraute englische Ausdrücke, Sätze und Floskeln Fragen korrekt oder zumindest teilweise korrekt in der Fremdsprache beantworten. Neben der Erfassung interaktiver Sprachkompetenzen im Hinblick auf Frage-Antwort-Settings kann das Interview so eventuell zusätzlich als Eisbrecher zum Abbau möglicher Ängste oder sozialer Hemmungen hinsichtlich der Fremdsprachnutzung dienen, bevor das vermutlich eher ungewohnte Peer-to-Peer-Testformat zur Erfassung weiterer interaktiver Sprachhandlungen eingesetzt wird.

Trotz der genannten kritischen Punkte können die Ergebnisse der Studie insgesamt als ein Hinweis darauf gedeutet werden, dass unterschiedliche Test-Settings in der Tat unterschiedliche Sprachhandlungen bei PrimarschülerInnen induzieren und somit die Konstruktvalidität von Testverfahren erhöhen können. Um ein valides Testinstrument zu konstruieren und um letztlich das Ziel einer vollständigen Erfassung des Konstrukts zu erreichen, sollten daher weitere Test-Settings erprobt und additiv eingesetzt werden.

Literatur

Alderson, J. Charles; Figueras, Neus; Kuijper, Henk; Nold, Guenter; Takala, Sauli & Tardieu, Claire (2006), Analysing tests of reading and listening in relation to the common european framework of reference: the experience of the Dutch CEFR construct project. *Language Assessment Quarterly 3: 1*, 3–30.

Bachman, Lyle & Palmer, Adrian (Hrsg.) (1996), *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford etc.: OUP.

Bailey, Alison (2005), Test review: Cambridge young learners English (YLE) tests. *Language Testing 22*: 2, 1-11.

Bos, Wilfried & Pietsch, Marcus (Hrsg.) (2006), *KESS 4 - Kompetenzen und Einstellungen von Schülerinnen und Schülern am Ende der Jahrgangsstufe 4 in Hamburger Grundschulen*. Münster: Waxmann (= HANSE - Hamburger Schriften zur Qualität im Bildungswesen, 1).

Brooks, Lindsay (2009), Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing* 26: 3, 341-366.

Cameron, Lynne (Hrsg.) (2001), *Teaching Languages to Young Learners*. Cambridge: Cambridge University Press.

Canale, Michael & Swain, Merrill (1980), Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1, 1–47.

Dausend, Henriette (Hrsg.) (2014), *Fremdsprachen transcurricular lehren und lernen. Ein methodischer Ansatz für die Grundschule*. Tübingen: Narr.

Diehr, Bärbel & Frisch, Stefanie (Hrsg.) (2008), *Mark their words. Sprechleistung im Englischunterricht der Grundschule fördern und beurteilen*. Braunschweig: Westermann.

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

Diehr, Bärbel & Polte, Linda (2009), Zur Entwicklung diskursiver Fähigkeiten im Englischunterricht der Grundschule. Eine vergleichende Untersuchung von Sprechern des Englischen als Erst- und Fremdsprache. *Zeitschrift für Fremdsprachenforschung* 20: 2, 147-174.

Ducasse, Ana M. & Brown, Annie (2009), Assessing paired orals: Raters' orientation to interaction. *Language Testing* 26: 3, *Special Issue on Paired Interaction*, 423-443.

Ellis, Gail & Brewster, Jean (Hrsg.) (2002), *The Primary English Teacher's Guide*. Essex: Penguin English Guides.

Elsner, Daniela (Hrsg.) (2010), *Englisch in der Grundschule unterrichten. Grundlagen, Methoden, Praxisbeispiele*. München: Oldenbourg.

Elsner, Daniela (2011), Developing Multiliteracies, Plurilingual Awareness & Critical Thinking in the Primary Language Classroom with Multilingual Virtual Talking Books. *Encuentro* 20, 27–38.

Elsner, Daniela; Buendgens-Kosten, Jules & Hardy, Ilonca (im Druck), Affordanzen und Nutzung mehrsprachiger Lernumgebungen – erste Ergebnisse aus der Pilotierung zum Forschungsprojekt LIKE. In: Rymarczyk, Jutta & Kötter, Markus (Hrsg.), *Englischunterricht auf der Primarstufe: neue Forschungen - weitere Entwicklungen*. Frankfurt/Main: Peter Lang.

Enever, Janet (2011), *ELLiE. Early Language Learning in Europe*. London, UK: British Council [Online unter http://www.teachingenglish.org.uk/sites/teacheng/files/B309%20ELLiE%20Book%202011%20FINAL.pdf. 09.05.2014].

Engel, Gaby (2009), EVENING – Konsequenzen für die Weiterentwicklung des Englischunterrichts in der Grundschule. In: Engel, Groot-Wilken & Thürmann (Hrsg.), 197-215.

Engel, Gaby; Groot-Wilken, Bernd & Thürmann, Eike (Hrsg.) (2009), *Englisch in der Primarstufe – Chancen und Herausforderungen. Evaluation und Erfahrungen aus der Praxis*. Berlin: Cornelsen.

[ETS] Educational Testing Service (2013), *TOEFL Primary* [Online unter https://www.ets.org/toefl_primary/. 08.07.2014].

Europarat (1998), *Recommendation 1383 Linguistic Diversification* [Online unter http://assembly.coe.int/Main.asp?link=http://assembly.coe.int/Documents/AdoptedText/ta98/EREC1383.htm. 13.05.2014].

Europarat (2001), *Gemeinsamer Europäischer Referenzrahmen für Sprachen. Lehren, Lernen und Beurteilen.* Berlin: Langenscheidt.

Europarat (2002), *ABl. C 50* vom 23.2.2002 [Online unter http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32002G0223(01)&from=DE. 27.08.2014].

Figueras, Neus; North, Brian; Takala, Sauli; Verhelst, Norman & Van Avermaet, Piet (2005), Relating examinations to the Common European Framework: A manual. *Language Testing* 22, 262–279.

Goethe-Institut (2013), *Fit in Deutsch 1/2. Prüfungsziele Testbeschreibung*. München: Goethe-Institut [Online unter http://www.goethe.de/lrn/prf/pro/hdb/Pruefungsziele_Testbeschreibung_A1_Fit1.pdf. 13.05.2014].

Groot-Wilken, Bernd; Engel, Gaby & Thürmann, Eike (2007), Listening and Reading Comprehension. Erste Ergebnisse einer Studie zu Englisch ab Klasse 3 an nordrhein-westfälischen Grundschulen. *forum schule. Magazin für Lehrerinnen und Lehrer* 1, 35-37.

Groot-Wilken, Bernd & Paulick, Christian (2009), Rezeptive Fähigkeiten und Fertigkeiten am Ende der 4. Klasse unter besonderer Berücksichtigung der sprachlichen Schülerbiografien. In: Engel, Groot-Wilken & Thürmann (Hrsg.), 179-196.

Haenni Hoti, Andrea & Werlen, Erika (2007), Die Zentralschweizer Längsschnittstudie zur Wirksamkeit des Fremdsprachenunterrichts auf der Primarstufe – eine Kurzpräsentation des Forschungsdesigns. In: Werlen, Erika & Weskamp, Ralf (Hrsg.) *Kommunikative Kompetenz und Mehrsprachigkeit*. Baltmannsweiler: Schneider Verlag Hohengehren, 129-137.

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

Hasselgren, Angela (2003), *The Bergen 'Can do' project*. Strasbourg: Council of Europe Publishing.

Hasselgren, Angela (2005), Assessing young language learners. *Language Testing* 22: 3, 337-354.

Haudeck, Helga & Schwab, Götz (2011), Merkmale bedeutungsvoller Interaktion im frühen Fremdsprachenunterricht. In: Kötter, Markus & Rymarczyk, Jutta (Hrsg.), *Fremdsprachenunterricht in der Grundschule. Forschungsergebnisse und Vorschläge zu seiner weiteren Entwicklung.* Frankfurt a. M.: Lang, 135-152.

Husfeldt, Vera & Bader-Lehmann, Ursula (2009), Englisch an der Primarschule: Erfahrungen aus der Schweiz. In: Engel, Groot-Wilken & Thürmann (Hrsg.), 111- 123.

Hymes, Dell Hathaway (1966), Two types of linguistic relativity. In: Bright, William (Hrsg.), *Sociolinguistics*. The Hague: Mouton, 114–158.

Keßler, Jörg-Uwe (2009), Zum mündlichen Sprachgebrauch von Grundschulkindern in Nordrhein-Westfalen am Ende des vierten Schuljahres. In: Engel, Groot-Wilken & Thürmann (Hrsg.), 140-157.

[KMK] Kultusministerkonferenz (2003), *Beschlüsse der Kultusministerkonferenz. Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss*. Beschluss vom 4.12.2003 [Online unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2003/2003_12_04-BS-erste-Fremdsprache.pdf. 04.05.2014].

[KM Baden-Württemberg] Kultusministerium Baden-Württemberg (2004), *Bildungsstandards für Englisch Grundschule-Klassen 2, 4* [Online unter Zur Verfügung unter: http://www.bildung-staerkt-menschen.de/service/downloads/Bildungsplaene/Grundschule. 04.05.2014].

[KM Hessen] Kultusministerium Hessen (2010), *Bildungsstandards und Inhaltsfelder. Das neue Kerncurriculum für Hessen. Primarstufe. Neue Fremdsprachen*. Wiesbaden: Hessisches Kultusministerium [Online unter http://verwaltung.hessen.de/irj/servlet/prt/portal/prtroot/slimp.CMReader/HKM_15/HKM_Internet/med/d9d/d9d1d584-b546-821f-012f-31e2389e4818,22222222-2222-2222-2222-222222222222. 11.05.2014].

Kolb, Elisabeth (2008), Authentische Sprechanlässe schaffen – aber wie? *HotSpot* 9, 2-9.

Kötter, Markus (2009), Entwicklung und Erprobung eines Instruments zur Erfassung produktiver mündlicher Fertigkeiten von Fremdsprachenlernern in der Grundschule. In: Engel, Groot-Wilken & Thürmann (Hrsg.), 140-157.

Krummheuer, Götz & Fetzer, Marei (2009), *Der Alltag im Mathematikunterricht. Beobachten – Verstehen – Gestalten*. München: Spektrum Akademischer Verlag.

McNamara, Tim (2006), Validity in language testing. The challenge of Sam Messick's legacy. *Language Assessment Quarterly* 3: 1, 31-51.

Messick, Samuel (1989), Validity. In: Linn, Robert (Hrsg.), *Educational measurement*. 3rd ed. Washington, DC: American Council on Education / Macmillan, 13–103.

Roos, Jana (2006), Frühes Fremdsprachenlernen – eine Standortbestimmung. In: Pienemann, Manfred; Keßler, Jörg-Uwe & Roos, Eckhard (Hrsg.), *Englischerwerb in der Grundschule. Ein Studien- und Arbeitsbuch*. Paderborn: Ferdinand Schöningh, 24-32.

Roos, Jana (2009), Formelhaftigkeit im frühen Fremdsprachenerwerb. *Zeitschrift für Fremdsprachenforschung* 20: 1, 37-63.

Savignon, Sandra J. (2002), Communicative Language Teaching: Linguistic theory and Classroom Practice. In: Savignon, Sandra J. (Hrsg.), *Interpreting Communicative Language Teaching*. New Haven, London: Yale University Press, 1-27.

Taylor, Lynda & Wigglesworth, Gillian (2009), Are two heads better than one? Pair work in L2 assessment contexts. *Language Testing* July 26: 3, 325-339.

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

[UCLES] University of Cambridge Local Examinations Syndicate (2011a), *Key English Test* [Online unter http://www.cambridgeenglish.org/de/Images/107937-ket-leaflet.pdf. 29.08.2014].

[UCLES] University of Cambridge Local Examinations Syndicate (2011b), *Preliminary English Test* [Online unter http://www.cambridgeenglish.org/de/Images/24943-cambridge-english-pre-document.pdf. 29.08.2014].

[UCLES] University of Cambridge Local Examinations Syndicate (2013), *Young Learners English Test*. Cambridge [Online unter http://www.cambridgeenglish.org/de. 12.05.2014].

Wirtz, Markus & Caspar, Franz (2002), *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.

Wolff, Dieter (2009), Zur Ausbildung von Lehrkräften für den Englischunterricht an Grundschulen – einige Überlegungen vor dem Hintergrund des Projektes EVENING. In: Engel, Groot-Wilken & Thürmann (Hrsg.), 35-46.

Zangl, Renate (2000), Monitoring language skills in Austrian primary (elementary) schools: a case study. *Language Testing* 17: 2, 250-260.

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

Anhang

Tabelle 6: Häufigkeiten Sprachhandlungen und Inter-Rater-Reliabilität

	Interview		**Peer-to-Peer**	
	Häufigkeit gesamt*	ICC (p)	Häufigkeit gesamt*	ICC (p)
Fragen/ Rückfragen stellen/ klären	55	0.83 (.000)	109.5	0.78 (.002)
Fragen beantworten/ reagieren	375	0.87 (.000)	116.5	0.90 (.000)
Inhalt erzählen/ übersetzen/ zusammenfassen	15.5	------	283	0.89 (.000)
Verstehen anzeigen	23.5	0.29 (.247)	32.5	0.74 (.002)
Vermutung anstellen	20	------	26	0.89 (.000)
Einverständnis signalisieren	11	------	13.5	------
Anweisungen folgen	27.5	0.57 (.015)	153	0.76 (.000)
Aufgabe selbst weiterführen	17.5	------	142	0.47 (.015)
Inhalt ergänzen/anknüpfen	46	0.70 (.008)	273.5	0.96 (.000)
Beipflichten/zustimmen	16	------	48	0.67 (.011)
Unwissenheit/Unentschiedenheit	17.5	------	9.5	------
Beschließen	3.5	------	29	0.81 (.001)
Vorschlagen	0	------	6.5	------
Feststellen/kommentieren	19	------	85	0.64 (.012)
Anweisung geben/auffordern	5	------	33	0.88 (.000)
Inhalte aushandeln	0	------	0	------
Begründen	1	------	1	------
Vorgehen besprechen	2.5	------	1.5	------
Information prüfen	1.5	------	2	------
Überlegungen anstellen	5	------	5.5	------
Ablehnen/widersprechen	1.5	------	22.5	0.622 (.025)

(ICC, two-way mixed model, absolute agreement); N=*Häufigkeit der Codierung jeder Sprachhandlung addiert über alle Interview- bzw. alle Peer-to-Peer-Sequenzen*;
** Mittel beider Rater; codiert wurden Sprachhandlungen mit Gesamt-N>20; Die Sprachhandlungen "Fragen" und „Rückfragen/Klären" wurden aufgrund schlechter Differenzierbarkeit zusammengefasst.*

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

Tabelle 7: Codierung interaktiver Sprachhandlungen

Sprachhandlung GER*	**Codiert in Kategorie:**	**Sprachhandlung GER***	**Codiert in Kategorie:**
Fragen stellen	Frage stellen	Verständnis anzeigen	Verstehen anzeigen
Frage beantworten Konkrete Auskünfte geben Informationsfragen beantworten	Frage beantworten/auf Frage reagieren	Zu Meinungsäußerung auffordern; auffordern (mitzumachen/ teilzunehmen)	Auffordern/Anweisen
Nachfragen Um Klärung/ Wiederholung bitten	Klären/nachfragen	Gruß- und Abschiedsformeln gebrauchen	Sprachhandlung wurde nicht getätigt
Anweisungen entgegennehmen (& befolgen) Detaillierte Instruktionen verstehen Fragen/ Anweisungen verstehen	Anweisung befolgen	Präferenzen ausdrücken	
Inhalte diskutieren/ aushandeln	Inhalte aushandeln	Dank und Befinden ausdrücken	
Gespräch selbst in Gang halten	Aufgabe/Gespräch weiterführen	Befragen nach Befinden und Neuigkeiten	
Feststellungen machen/ darauf reagieren Standpunkte klarmachen/ kommentieren	Feststellen/kommentieren	Bedürfnisse ausdrücken	
Informationen zusammenfassen/ weitergeben Interview/Geschichte zusammenfassen Austausch relevanter Informationen/ Sachinformationen austauschen/weitergeben	Inhalt erzählen/ übersetzen/zusammenfassen	Gefühle ausdrücken/darauf reagieren	
Informationen prüfen	Informationen prüfen	Problem erklären	
Anderen beipflichten Informationen bestätigen	Beipflichten/zustimmen/ bestätigen	Alternativen beurteilen	
Einverständnis signalisieren	Einverständnis signalisieren	Lösungsmöglichkeiten diskutieren	
Beschluss fassen**	Beschließen	Jemanden ansprechen	
Anderen widersprechen	Widersprechen/ablehnen	Stand einer Diskussion zusammenfassen	
An Aussagen anknüpfen	Inhalt/Aussage/Information ergänzen	Meinung äußern/Präzise ausdrücken	

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.

Besprechen. was man tun will	Vorgehen besprechen	Kurzes Gespräch beginnen/ in Gang halten/ beenden	
Vorschlagen	Vorschlag machen	Verabredungen treffen	
Unentschiedenheit signalisieren	Unentschiedenheit/ Unwissen signalisieren	Gedanken zu kulturellen Themen äußern	
Standpunkt begründen/ verteidigen/äußern/ Kommentieren Eigene Meinung begründen	Begründen	(Persönliche) Bedeutung von Ereignissen/ Erfahrungen hervorheben	
Hypothesen aufstellen Vermutungen darlegen	Vermutung/ Annahme	Zusammenhänge zwischen Ideen deutlich machen	
Überlegung anstellen **	Überlegungen anstellen überlegen		

**Gemeinsamer Europäischer Referenzrahmen für Sprachen **Ergänzung vorgenommen; keine direkte Entsprechung im Referenzrahmen*

Anmerkungen

[1] Das Mindestalter für die Testung von Niveau A1 beträgt 10, für Niveau A2 12 Jahre. Diese Verfahren sind somit sehr bedingt auch für einen kleinen Teil der PrimarschülerInnen geeignet.

[2] Der Begriff Konstruktvalidität bezieht sich hier dabei auf die inhaltsvalide Abbildung der zu erfassenden Kompetenz anhand eines verwendeten Testverfahrens. Basierend auf Messicks (1989) Validitätsdefinition stellt Konstruktvalidität dabei eine Teilfacette der Validität dar und kann bezeichnet werden als die *„evidential basis"* (20; siehe auch McNamara 2006: 32) für die Interpretation und Verwendung von Tests und Testwerten.

[3] Siehe auch: Creutz, Maya Anastasia (i.Dr.): Effekte von Selbstkonzepten auf die Bearbeitung kooperativer Lernaufgaben im Englischen der Grundschule. In: Elsner, Daniela/ Lohe, Viviane: Multimodalität und Fremdsprachenlernen. Papers of Excellence, Band 5. Aachen: Shaker, 56-86.

Wahl, Lisa (i.Dr.), Bearbeitungsprozesse einer textbasierten Lernaufgabe im Englischen: Ein Vergleich zwischen Kindern mit und ohne Migrationshintergrund. . In: Elsner, Daniela/ Lohe, Viviane: Multimodalität und Fremdsprachenlernen. Papers of Excellence, Band 5. Aachen: Shaker, 88-109.

[4] Der Begriff „Beurteilerübereinstimmung" oder auch „Beurteilerreliabilität" bezieht sich auf die Höhe der Übereinstimmung zweier oder mehr Personen (Beurteilern), die die Schätzung einer Merkmalsausprägung eines nicht beobachtbaren psychologischen Merkmals vornehmen (siehe dazu auch Wirtz & Caspar 2002: 13ff).

[5] Die Intraklassenkorrelation (ICC/Intra-Class-Correlation) stellt ein Maß für die Ähnlichkeit zweier (oder mehr) Beurteiler hinsichtlich einer vorgenommenen Merkmalseinschätzungen dar. Sie eignet sich für intervallskalierte Ratings und kann Werte zwischen 0 und 1 annehmen, wobei ein Wert nahe 1 Ausdruck einer hohen Übereinstimmung ist (siehe auch ebd.: 157ff).

Astrid Jurecka, Jules Bündgens-Kosten & Daniela Elsner (2014), Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 78-99. Abrufbar unter http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Jurecka_et_al.pdf.