

Zeitschrift für Interkulturellen Fremdsprachenunterricht

Didaktik und Methodik im Bereich Deutsch als Fremdsprache

ISSN 1205-6545 Jahrgang 19, Nummer 2 (Oktober 2014)

Aktuelle Entwicklungen in der Messung von Sprachkompetenzen: Einführung

Gabriele Kecker

TestDaF-Institut
Universitätsstr. 134
44799 Bochum
Tel +49 (0) 234-3229720
E-Mail: Gabriele.Kecker@testdaf.de

Der thematische Schwerpunkt dieses Heftes umfasst Testverfahren zur Ermittlung von Sprachkompetenzen in der Erst-, Zweit- und Fremdsprache im schulischen und außerschulischen Kontext mit unterschiedlicher Zielsetzung: Sprachdiagnostik im Elementarbereich, schulische und außerschulische Kompetenzermittlung und Ermittlung des Lernfortschritts sowie länderübergreifende Lernstandserhebungen. Fragen der Selbst- oder Peer-Evaluation im Unterricht werden lediglich am Rande erörtert.

Die ausgewählten Beiträge spiegeln wesentliche Tendenzen in der Leistungsmessung seit Beginn des 21. Jahrhunderts wider, die durch mehrere Faktoren geprägt sind: die Einführung von kompetenzorientiertem Testen, die Einführung von Standards bei der Kompetenzermittlung (Mindeststandards oder Regelstandards mit mittlerem Anforderungsniveau) und die damit einhergehende Output-Orientierung im Unterricht. Internationale Lernstandserhebungen der OECD wie beispielsweise PISA 2000 (*Programme for International Student Assessment*) oder die Internationale Grundschul-Lese-Untersuchung IGLU, aber auch die Einführung des Gemeinsamen europäischen Referenzrahmens für Sprachen (GER) (Europarat 2001) haben wesentlich zu dieser Neuorientierung beigetragen. Obwohl die Beschreibungen fremdsprachlicher Kompetenzen des GER auf sechs verschiedenen Niveaustufen ursprünglich von den Autoren eher als beispielhafte Illustration gedacht waren und erwachsene Lernende als Zielgruppe berücksichtigen, dienen die GER-Stufen als Ausgangspunkt für die Entwicklung von Bildungsstandards, Curricula und Sprachprüfungen auch jüngerer Lernender im Primar- und Sekundarschulbereich und werden darüber hinaus von vielen als Norm für sprachliche Kompetenzen angesehen.

Die Ergebnisse der oben erwähnten internationalen Lernstandserhebungen der OECD haben in Deutschland die Bedeutung der Unterrichtssprache Deutsch für den Lernerfolg und die Teilhabe am Bildungssystem von Lernenden mit nicht-deutscher Herkunftssprache in den Fokus der Aufmerksamkeit gerückt. Mängel in der Zweitsprache Deutsch (DaZ) wurden nach PISA 2000 für den im Vergleich mit Lernenden deutscher Herkunftssprache geringeren Lernerfolg verantwortlich gemacht. Die Erkenntnis, dass eine möglichst frühe Sprachstandsdiagnose und frühzeitige Fördermaßnahmen auch bei Kindern mit deutscher Herkunftssprache, die Defizite in der Sprachentwicklung aufweisen, langfristig bessere Ergebnisse ermöglichen, hat in den folgenden Jahren zu der Entwicklung einer Reihe von diagnostischen Verfahren für den Elementarbereich geführt. Neugebauer & Becker-Mrotzek (2013; siehe auch die Rezension von Hoffmann & Böhme in diesem Heft) haben solche Verfahren unlängst im Hinblick auf die Einhaltung von Qualitätsstandards (vgl. Mercator-Institut 2013) analysiert und kommentiert.

Die Qualitätsstandards des Mercator-Instituts (2013) für Testverfahren im Elementarbereich gehen u.a. auf Testgütekriterien zurück, die auf internationaler Ebene die Grundlage im Bereich psychologischen und pädagogischen

Testens bilden (vgl. die sog. *Joint Standards* der AERA, APA & NCME 1999 oder Internationale Richtlinien für die Testanwendung, ITC 2000). Die klassischen Gütekriterien wie Validität, Reliabilität und Objektivität werden seit Beginn des 21. Jahrhunderts zunehmend auch von Sprachtestanbietern in Europa bei der Testentwicklung für unterschiedliche Zielgruppen (z.B. SchülerInnen, StudienbewerberInnen, Zuwanderer und Zuwanderinnen, Berufstätige) und Zielsetzungen (z.B. Sprachstandsdiagnose, Zulassung zum Hochschulstudium, Zuerkennung der Staatsbürgerschaft, Zugang zu medizinischen Berufen) zugrunde gelegt. Als Anzeichen für diese Tendenz können die Bestrebungen europäischer Sprachtestverbände wie der *Association of Language Testers in Europe* (ALTE) gelten, die Einhaltung von Qualitätsstandards durch ihre Mitglieder mithilfe von Audits zu überprüfen. Bestrebungen des Europarats, Qualitätsstandards in der Sprachtestentwicklung in einem stärkeren Maße zu implementieren, sind aus Projekten ersichtlich, die PraktikerInnen in der Testentwicklung wissenschaftliche Methoden näherbringen sollen. Als Beispiele seien hier die Entwicklung eines methodischen Ansatzes zur Zuordnung von Sprachprüfungen zum GER genannt, das *Manual for Relating Language Examinations to the Common European Framework of Reference* (Council of Europe 2009) oder die Veröffentlichung des *Manual for Language Test Development and Examining* (ALTE 2011). Beide Veröffentlichungen betonen die Notwendigkeit, statistische Methoden und empirische Forschung zur Qualitätssicherung von Testverfahren einzusetzen.

Den Auftakt des thematischen Schwerpunktes in diesem Heft bilden zwei Beiträge, die sich mit der Sprachdiagnostik für Kinder im Elementar- und Primarbereich befassen. Im Mittelpunkt steht die Diagnostik des Sprachstands in DaZ mit dem Ziel, Kindern durch geeignete Fördermaßnahmen den Schuleinstieg zu erleichtern und ihre Kompetenzentwicklung in der Unterrichtssprache Deutsch und in den verschiedenen Schulfächern zu fördern.

Kapica, Klages & Pagonis stellen in ihrem Beitrag „Narrative Kompetenz: Anforderungen an ein Verfahren der Sprachstandsfeststellung im Elementarbereich“ Überlegungen zur Entwicklung eines sprachdiagnostischen Instruments an, das die kindliche Erzählkompetenz im Vorschulbereich als zentralen Indikator für frühkindliche Sprachentwicklung verwendet. Die Autoren untersuchen, welche Merkmale für die narrative Kompetenz charakteristisch sind und welche Bedeutung sie für den kindlichen Spracherwerbsprozess und für die Teilhabe am Bildungsprozess haben. Sie berufen sich dabei auf Untersuchungen der Spracherwerbsforschung zur Entwicklung der narrativen Kompetenz bei Kindern unterschiedlicher Altersstufen in der Erstsprache. Ziel ist es, ein förderdiagnostisches Verfahren zu entwickeln, das einerseits zu einem möglichst frühen Zeitpunkt präzise Daten zur Ermittlung des Sprachstands in DaZ liefert und andererseits eine zuverlässige Grundlage für daran anschließende passgenaue Fördermaßnahmen bietet.

Die Untersuchung von **Böhme & Hoffmann** „Sprachstandsdiagnostik bei mehrsprachigen Grundschulkindern – Empirische Befunde zum Einsatz diagnostischer Verfahren in Deutschland“ befasst sich gleichfalls mit der Sprachkompetenz von Kindern nicht-deutscher Herkunftssprache, allerdings im Grundschulbereich. Vor dem Hintergrund verschiedener Studien zur Mehrsprachigkeit heben Böhme & Hoffmann die Bedeutung der Erstsprache für den Erwerb der Zweitsprache und geeignete Förderinstrumente hervor. Die Autoren verwenden Daten einer bundesweiten Erhebung des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) an Grundschulen aus dem Jahr 2011, um zu ermitteln, in welchem Umfang diagnostische Verfahren eingesetzt werden, die auch den Sprachstand der Kinder in der Erstsprache berücksichtigen. Sie werten dazu die Angaben von 1227 SchulleiterInnen an Grundschulen aus, die mithilfe eines Fragebogens erhoben wurden.

DaZ in einem mehrsprachigen Kontext ist Gegenstand eines Forschungsprojekts in der Schweiz, das von **Peyer & Studer** in ihrem Beitrag „Schreibkompetenz in der Zielsprache Deutsch in einem mehrsprachigen Schweizer Kontext: Zur Entwicklung von Basisstandards am Beispiel eines bilingualen Schulmodells (Romanisch-Deutsch)“ vorgestellt wird. Am Beispiel DaZ und anhand der Teilkompetenz Schreiben wird der Einsatz qualitativer und quantitativer Forschungsmethoden zur Entwicklung von Basisstandards, Testaufgaben, analytischen Beurteilungsskalen und Bestehensgrenzen demonstriert. Mithilfe der Multifacetten-Rasch-Analyse entwickeln Peyer & Studer ein 4-dimensionales Modell der Beurteilungskriterien. Die Bestehensgrenzen der Basisstandards wird von ihnen in einem Standard-Setting-Verfahren ermittelt, das sich methodisch am Manual des Europarats (Council of Europe 2009) *Relating Language Examinations to the Common European Framework of Reference* orientiert.

Lepage, Marty, Perlmann-Balme & Stelter beschreiben in ihrem Projektbericht „Évaluation de la production écrite dans le cadre de l’Enquête européenne sur les compétences linguistiques (ESLC)“ den methodischen Ansatz für die Bewertung schriftlicher Leistungen in der europaweiten Lernstandserhebung *European Survey on Language Competences* (ESLC). Das Bewertungsverfahren ist darauf ausgerichtet, möglichst standardisierte und vergleichbare Beurteilungen zu erzielen und kann auch in anderen Kontexten angewendet werden. Die ESLC wurde 2008 vom Europarat mit dem Ziel ins Leben gerufen, anhand einer repräsentativen Stichprobe (53 000 SchülerInnen aus 14 Ländern) die Sprachkompetenz junger Lernender in der ersten und zweiten Fremdsprache an Sekundarschulen zu ermitteln und den Bildungsbehörden der teilnehmenden Länder Daten für ein Bildungsmonitoring in den Fremdsprachen zur Verfügung zu stellen. Ausgehend vom handlungsorientierten Ansatz kommunikativer Kompetenz des GER wurden in den Sprachen Deutsch, Englisch, Französisch, Italienisch und Spanisch äquivalente Testaufgaben für die Niveaustufen A1-B2 entwickelt und in der Erhebung 2011 eingesetzt.

Jurecka, Bündgens-Kosten & Elsner untersuchen in ihrem Beitrag „Die Eignung von Interview- und Peer-to-Peer Test-Settings zur Erfassung fremdsprachlicher Interaktion bei Grundschulkindern“ die Konstruktvalidität von Test-Settings in mündlichen Prüfungen im Englischunterricht der Primarstufe. Die Autorinnen geben einen umfassenden Überblick über bisherige Studien zur Erfassung rezeptiver und produktiver Teilkompetenzen im Primarschulunterricht, in denen zumeist Teilaspekte der vorliegenden Untersuchung berücksichtigt worden sind. Ziel ihrer empirischen Studie ist es herauszufinden, inwieweit sich asymmetrische (Lehrkraft – SchülerIn) und symmetrische (SchülerIn - SchülerIn) Rollenverteilungen in der Kommunikation mündlicher Prüfungen qualitativ und quantitativ auf die Verwendung von interaktionalen Sprachhandlungen in der Fremdsprache auswirken. Die Untersuchung fand an drei Grundschulen in Hessen mit insgesamt 38 SchülerInnen der 4. Klasse nach einem Jahr Sprachunterricht statt. Die Autorinnen analysieren die von den SchülerInnen verwendeten interaktionalen Sprachhandlungen anhand eines Codierschemas, das sich an den Skalen des GER für mündliche Interaktion orientiert.

Die Untersuchung von **Hahn, Koch & Fromme** „Der Online-Test BULATS Deutsch – ein geeignetes Instrument zur Messung von Lernfortschritten? Erste Ergebnisse im Rahmen des Projekts *Tablet-Klasse*“ bezieht sich im Gegensatz zu den Zielgruppen der zuvor dargestellten Beiträge auf Deutschlernende im außerschulischen Kontext, d.h. in Goethe-Instituten. Die AutorInnen beschäftigen sich mit der Frage, ob durch den gezielten Einsatz von Tablets (iPads) im Unterricht tatsächlich ein messbarer Lernfortschritt im Hörverstehen erzielt werden kann. Der adaptive Test BULATS (*Business Language Testing Service*) deckt alle sechs GER-Stufen ab und wurde als Vor- und Nachtest eines Sprachkurses eingesetzt, um den Einfluss der im Kurs genutzten Tablets auf den Lernfortschritt der Kurs teilnehmenden zu erfassen. Der Lernzuwachs wurde anhand der Punktwerte im BULATS-Prüfungsteil Hörverstehen ermittelt und hinsichtlich seiner statistischen Signifikanz analysiert.

Schließlich widmen sich **Dobrić & Sigott** in ihrem Beitrag „Towards an Error Taxonomy for Student Writing“ einer Forschungsfrage, die im Zusammenhang mit Beurteilungen von schriftlichen Leistungen Lernender entstanden ist: der Beschaffenheit einer Fehlertaxonomie und Fehlerannotierung. Ziel ist es, Beurteilungsinstrumente zu entwickeln, die von Beurteilenden zuverlässig und konsistent gehandhabt werden und gleichzeitig eine verständliche und hilfreiche Anleitung zur Fehlerbehebung durch Lernende beinhalten. Die Autoren geben einen Überblick über bisherige Klassifizierungen von Fehlern und setzen sich kritisch mit den meist indirekten Fehler- oder Defizitbeschreibungen in den Niveaustufen-Beschreibungen des GER auseinander. Die Autoren entwerfen ihr Modell zur Fehlertaxonomie und -annotierung ausgehend von den Kategorien *scope* und *substance* auf Wort-, Satz- und Textebene und folgen dabei dem Grammatik-Modell von Quirk, Greenbaum, Leech & Svartvik (1985; siehe Dobrić & Sigott in diesem Heft). Das Modell wird von den Autoren zukünftig anhand von Leistungsbeispielen aus dem nationalen Standardtest in Englisch, der in Österreich durchgeführt wird, erprobt.

Literatur

ALTE (2011), *Manual for Language Test Development and Examining*. Strasbourg: Council of Europe, Language Policy Division.

Council of Europe (Hrsg.) (2009), *Relating Language Examinations to the Common European Framework of Reference. A Manual*. Strasbourg.

Gabriele Kecker (2014), Aktuelle Entwicklungen in der Messung von Sprachkompetenzen: Einführung. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19: 2, 1-4. Abrufbar unter <http://zif.spz.tu-darmstadt.de/jg-19-2/beitrag/Kecker.pdf>.

- [AERA, APA & NCME] American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999), *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Europarat (2001), *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Straßburg.
- [ITC] International Test Commission (2000), *Internationale Richtlinien für die Testanwendung* [Online unter <http://www.intestcom.org/Guidelines/%5BTranslations%5D.php>. 31.08.2014].
- Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache (Hrsg.) (2013), *Qualitätsmerkmale für Sprachstandsverfahren im Elementarbereich. Ein Bewertungsrahmen für fundierte Sprachdiagnostik in der Kita*. Köln.
- 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.