



Leseprobe aus Schwabe, McElvany, Bos und Holtappels, Jahrbuch der Schulentwicklung.
Band 20, ISBN 978-3-7799-3930-6

© 2018 Beltz Juventa in der Verlagsgruppe Beltz, Weinheim Basel

<http://www.beltz.de/de/nc/verlagsgruppe-beltz/gesamtprogramm.html?isbn=978-3-7799-3930-6>

Bildungssprachliche Anforderungen, Lesbarkeit und Textverständlichkeit von Sachtexten für die Grundschule

Ruven Stahns, Anke Hußmann, Daniel Kasper

1 Einleitung

Ergebnisse von Schulleistungsstudien im Grundschulbereich haben wiederholt gezeigt, dass Schülerinnen und Schüler mit Migrationshintergrund über geringere Lesekompetenzen verfügen als Schülerinnen und Schüler ohne Migrationshintergrund (vgl. Schwippert, Hornberg, Freiberg & Stubbe, 2007; Schwippert, Wendt & Tarelli, 2012; Wendt & Schwippert, 2017). Zudem zeigen sich leistungsbezogene Disparitäten im Lesen zwischen Schülerinnen und Schülern aus soziokulturell sowie sozioökonomisch schwachen und starken Elternhäusern (vgl. Bos, Schwippert & Stubbe, 2007; Wendt, Stubbe & Schwippert, 2012; Hußmann, Stubbe & Kasper, 2017). Aus dem Versuch, diese Befunde zu erklären, resultierte die Hypothese, dass Schülerinnen und Schüler, die in Kompetenztestungen im sprachlichen Bereich (aber auch in Mathematik und anderen Fächern) mit geringem Erfolg abschneiden, nicht über ausreichende Kompetenzen im Register Bildungssprache verfügen, um den sprachlichen Anforderungen, mit denen sie in der Testsituation konfrontiert werden, begegnen zu können (vgl. Berendes, Dragon, Weinert, Heppt & Stanat, 2013; Gogolin & Duarte, 2016; Gogolin & Lange, 2011; Gogolin & Schwarz, 2004; Heppt, Haag, Böhme & Stanat, 2014a; Wiater, 2015). Aus diesem Grund sind die bildungssprachlichen Anforderungen des Testmaterials von Kompetenztestungen in den Forschungsfokus gerückt (vgl. Haag, Heppt, Stanat, Kuhl & Pant, 2013; Haag, Heppt, Roppelt & Stanat, 2015; Heppt et al., 2014a; Heppt, Henschel & Haag, 2016; Heppt, Stanat, Dragon, Berendes & Weinert, 2014b; Stahns, Walzebug & Kasper, 2016). Mittlerweile legen die Ergebnisse einiger Studien nahe, dass bildungssprachliche Anforderungen in Leistungstests Schülerinnen und Schüler unabhängig von deren Zugehörigkeit zu einer bestimmten Gruppe bei der Bearbeitung von Testaufgaben herausfordern können (vgl. Abschnitt 3).

Ein Kriterienkatalog zur Einschätzung bildungssprachlicher Anforderungen von Erzähl- und Sachtexten sowie dazugehöriger Testaufgaben, die bei der Evaluation der Bildungsstandards für die Grundschule eingesetzt wurden, stammt von Heppt et al. (2014a). Hier werden Merkmale auf der lexikalischen Ebene (z. B. der Einsatz bildungssprachlichen Wortschatzes) und der grammatischen Ebene (z. B. die syntaktische Komplexität von Sätzen) berücksichtigt. Stahns (2016) kommt zu der Einschätzung, dass durch ein Rating auf dieser Grundlage nicht ausschließlich Anforderungen beurteilt werden, die mit dem Konstrukt Bildungssprache in Verbindung stehen. Stattdessen werde zum Teil die Lesbarkeit der Texte bestimmt. Dies wäre hinsichtlich der Schärfung des Konstrukts Bildungssprache nicht wünschenswert. Zudem müssten die Einschränkungen hinsichtlich der Aussagekraft, von denen für Lesbarkeitsbewertungen auszugehen ist, auch für Urteile auf der Grundlage dieser Rating-Kategorien gelten (vgl. die Abschnitte 4 und 5). Die Operationalisierung von Heppt et al. (2014a) wird von Stahns et al. (2016) in einer Untersuchung aufgegriffen, in der sprachliche Anforderungen von Items aus der *Internationalen Grundschul-Lese-Untersuchung* (IGLU) 2011 eingeschätzt werden. Dort kann kein Zusammenhang zwischen den sprachlichen Anforderungen und der Aufgabenschwierigkeit¹ der Items nachgewiesen werden. In der Diskussion der Ergebnisse greifen Stahns et al. (2016) Ergebnisse der Verständlichkeitsforschung auf (vgl. Groeben, 1982; Langer, Schulz von Thun & Tausch, 2011). Sie vermuten, dass die Berücksichtigung weiterer Dimensionen, die sich auf die Textverständlichkeit auswirken, notwendig sei, um den Beitrag sprachlicher Anforderungen zur Textverständlichkeit angemessen abzubilden. Das sei eine Voraussetzung dafür, die empirisch ermittelte Schwierigkeit von Aufgaben aus Large-Scale-Assessments zu erklären. Eine Überprüfung dieser Annahmen steht noch aus.

Im vorliegenden Beitrag stehen bildungssprachliche Anforderungen der Sachtexte (N = 8) und der dazugehörigen Leseaufgaben (N = 99) im Fokus, die im Rahmen der Studie IGLU zwischen 2001 und 2011 zur Testung von Lesekompetenzen eingesetzt wurden. In der Auseinandersetzung mit dem Konstrukt Bildungssprache (vgl. Abschnitt 2) und seiner Operationalisierung (vgl. Abschnitt 3) werden die Lesbarkeit (vgl. Abschnitt 4) und die

1 Bei der berichteten Aufgabenschwierigkeit handelt es sich um die international berichtete Itemschwierigkeit, wie sie für jedes in IGLU eingesetzte Item von der internationalen Studienleitung zur Verfügung gestellt wird. Details zur Skalierung (bspw. hinsichtlich der verwendeten Stichprobe oder des Linkings) finden sich etwa in Mullis, Martin, Foy und Drucker (2012) sowie Martin und Mullis (2011).

Textverständlichkeit (vgl. Abschnitt 5) als weitere potenziell schwierigkeitsbestimmende Merkmale von Texten neben bildungssprachlichen Anforderungen eingeführt und diskutiert. Nach einer Beschreibung der Forschungsfragen (vgl. Abschnitt 6) und des methodischen Zugangs (vgl. Abschnitt 7) werden die Ergebnisse einer Kodierung beziehungsweise eines Ratings bildungssprachlicher Merkmale, der Lesbarkeit und der Verständlichkeit der Sachtexte beziehungsweise Items vorgestellt (vgl. Abschnitt 8) und diskutiert (vgl. Abschnitt 9). Der Beitrag schließt mit einer Zusammenfassung der wesentlichen Erkenntnisse und einem Ausblick ab (vgl. Abschnitt 10).

Ein Ziel des vorliegenden Beitrags ist es, das Verhältnis der bildungssprachlichen Anforderungen und der Lesbarkeit zueinander und deren Verhältnis zu Einschätzungen der Textverständlichkeit zu untersuchen. Die Textverständlichkeit wird dazu auf Grundlage von vier Dimensionen nach Langer et al. (2011) eingeschätzt. Überprüft wird, ob die Ergebnisse der Kodierung bildungssprachlicher Merkmale sowie die Lesbarkeitswerte hinsichtlich der Textschwierigkeit in dieselbe Richtung weisen wie das Rating der Textverständlichkeit.

Anschließend wird untersucht, ob mit dem Lesbarkeitsindex *Lix* ein kompensatorisches Messinstrument für die Einschätzung der bildungssprachlichen Anforderungen der Texte existiert (vgl. Bamberger & Vanecek, 1984, S. 62 ff.; Rosebrock, Nix, Rieckmann & Gold, 2013, S. 71 ff.). Weisen die Einschätzungen bildungssprachlicher Merkmale und die Lesbarkeitswerte auf der Grundlage der *Lix*-Formel in dieselbe Richtung, könnte diese Lesbarkeitsformel verwendet werden, um eine objektive Messung der sprachlichen Anforderungen von Texten vorzunehmen. Fraglich wäre allerdings, ob in diesem Fall von bildungssprachlichen Merkmalen gesprochen werden sollte. Zudem wird im vorliegenden Beitrag überprüft, ob Zusammenhänge zwischen der Aufgabenschwierigkeit der in IGLU eingesetzten Items und den Einschätzungen bildungssprachlicher Anforderungen, der Lesbarkeit und der Verständlichkeit der in IGLU eingesetzten Texte und Aufgaben nachzuweisen sind.

2 Das Konstrukt Bildungssprache

In der Diskussion um Bildungssprache wird im deutschsprachigen Raum mit unterschiedlichen Schwerpunktsetzungen Bezug auf Forschungsergebnisse aus der Soziologie, der (Sozio-)Linguistik und der Zweitspracherwerbsforschung genommen. Unter anderem dienen Bernsteins (1977) Unterscheidung eines restringierten und eines elaborierten Sprachcodes und

Arbeiten Bourdieus (1990) zur Macht der Sprache als Instrument sozialer Distinktion als Ausgangspunkt für Überlegungen zur Bildungssprache. Zudem wird häufig Bezug auf die Arbeiten Hallidays (1994) genommen, die es erlauben, Bildungssprache als ein funktionales Sprachregister zu beschreiben. Des Weiteren wird Kochs und Oesterreichers (1985) Unterscheidung zwischen konzeptioneller und medialer Mündlichkeit und Schriftlichkeit sowie kommunikativer Nähe und Distanz in der Auseinandersetzung mit Bildungssprache aufgegriffen. Auf die Arbeiten von Cummins (2008), dessen Unterscheidung von *Basic Interpersonal Communication Skills* (BICS) und *Cognitive Academic Language Proficiency* (CALP) verschiedene Grade des Erlernens einer (Fremd-)Sprache aufzeigt, wird häufig zur Abgrenzung von Bildungs- und Alltagssprache Bezug genommen. Zudem werden im deutschsprachigen Raum mitunter Ausführungen zur *language of schooling* von Schleppegrell (2004) berücksichtigt, die am Beispiel des Englischen beschreibt, welche Sprache im Kontext von Schule Anerkennung findet und welche nicht.

Die Entwicklung und Verwendung des Begriffs Bildungssprache ist andernorts aufgearbeitet (vgl. Berendes et al., 2013, S. 18 ff.; Gogolin & Duarte, 2016, S. 480 ff.). Im vorliegenden Beitrag wird auf Ausführungen zur Begriffsgenese verzichtet und der Bestimmung von Bildungssprache in der Arbeit von Stahns et al. (2016, S. 59) gefolgt. Die Autoren verstehen Bildungssprache als ein in der Schule beziehungsweise in Bildungskontexten verwendetes, konzeptionell schriftliches, aber medial mündliches oder schriftliches, domänenspezifisches und im Verlaufe der Schulstufen zunehmend ausdifferenziertes sprachliches Register. Der Einsatz von Bildungssprache in der Schule erfolgt funktional. Bildungssprachliche Kompetenzen stellen eine Voraussetzung für den Wissenserwerb dar, Bildungssprache dient aber auch der Präsentation von Wissen in Bildungsinstitutionen. Bildungssprache findet zudem außerhalb der Schule Verwendung und ein Beherrschen der Mittel des Registers wird in zahlreichen Kontexten vorausgesetzt. Dem vorliegenden Beitrag wird damit eine Fassung von Bildungssprache zugrunde gelegt, die in ähnlicher Form in einer Vielzahl von Forschungsarbeiten und Überblicksartikeln nachzuweisen ist, die in den letzten Jahren entstanden sind, ohne dass behauptet werden könnte, der Begriff sei eindeutig definiert (vgl. Berendes et al., 2013; Gogolin, 2014, S. 409 ff.; Gogolin & Duarte, 2016; Gogolin & Lange, 2011, S. 107 ff.; Gogolin & Schwarz, 2004, S. 836 f.; Heppt et al., 2014a, S. 62; Heppt et al., 2014b, S. 140; Heppt et al., 2016, S. 1 f.; Lange, 2012, S. 123 ff.).

Um Bildungssprache von Alltags- und Fachsprache abzugrenzen, werden auf unterschiedlichen sprachlichen Ebenen Merkmale bestimmt, die

Bildungssprache in höherem Maße als andere Register kennzeichnen. In den bislang vorliegenden empirisch ausgerichteten Arbeiten zur Bildungssprache werden vor allem lexikalische und grammatische Merkmale auf den Ebenen Wort und (zusammengesetzter) Satz aufgeführt. Dabei wird darauf hingewiesen, dass eine klare Grenze zwischen Bildungs-, Alltags- und Fachsprache nicht gezogen werden kann (vgl. Berendes et al., 2013, S. 24). Um Bildungssprache von anderen Registern abzugrenzen, geht es also um ein Mehr oder Weniger an bestimmten Merkmalen. In den einschlägigen Aufstellungen werden der häufige Gebrauch von Fremdwörtern, Fachvokabular und fachübergreifend eingesetzter bildungssprachlicher Wortschatz, eine fachspezifische Verwendung alltagssprachlichen Wortschatzes, die Verwendung von Nominalisierungen und längeren Attributen, die die Komplexität von Nominal- und Präpositionalphrasen erhöhen, Passivkonstruktionen und Konjunktivformen, Ableitungen und Komposita, Funktionsverbgefügen und syntaktisch komplexen Satzkonstruktionen (u. a. Partizipialkonstruktion, Passivkonstruktion, Satzgefüge, Schachtelung von Nebensätzen) als bildungssprachliche Merkmale genannt (vgl. Berendes et al., 2013, S. 26; Gogolin & Duarte, 2016, S. 489 f.; Gogolin & Lange, 2011, S. 113 f.; Morek & Heller, 2012, S. 73). Listen bildungssprachlicher Merkmale, die zum Teil an Umfang deutlich über den dieser Aufstellung hinausgehen (vgl. z. B. Gogolin & Duarte, 2016), werden zum Teil unter Einbezug internationaler Forschungsergebnisse gewonnen (vor allem solchen aus dem US-amerikanischen Raum), wobei selbstverständlich Anpassungen vorgenommen werden müssen (vgl. Heppt et al., 2014a). Auffällig in den Aufstellungen ist der Fokus auf „Oberflächenmerkmale“ (Morek & Heller, 2011, S. 73) auf der Wort- und Satzebene und häufig das weitgehende Ausblenden von Merkmalen der Textebene (ebd.).

Es ist auch möglich, sich dem Konstrukt Bildungssprache über die Bestimmung von Merkmalen anzunähern, die der Leichten Sprache zuzuordnen sind – einem Konstrukt, das der Bildungssprache als „Schwerer Sprache“ (Oomen-Welke, 2015, S. 24) gegenübergestellt wird. Auch in der Diskussion um Leichte Sprache stehen lexikalische und grammatische Merkmale im Fokus: Nach Oomen-Welke (ebd., S. 25) ist Leichte Sprache gekennzeichnet „durch einfache Wörter, möglichst Konkreta“ (ebd.), die Verwendung kurzer Hauptsätze – „möglichst nicht über acht Wörter, Stellung nach dem SVO(Adv)-Prinzip“ (ebd.) – „durch Verzicht auf Zusammensetzungen, Ableitungen, Fremdwörter und Abkürzungen“ (ebd.) sowie „durch Vermeiden des Genitivs, des Konjunktivs, des Passivs, impersonaler Subjekte“ (ebd.) und vergleichbare Merkmale.

Auch wenn ein Überschneidungsbereich dessen erkennbar ist, was in den bislang vorliegenden Studien unter Bildungssprache verstanden wird, kann nicht von einem deutlich abgrenzbaren Konstrukt ausgegangen werden. Konkretisiert werden die Vorstellungen, die hinter dem Begriff Bildungssprache stehen, in studienspezifischen Operationalisierungen. Im Fall der vorliegenden Untersuchung sind Arbeiten von Interesse, in denen bildungssprachliche Anforderungen in medial schriftlichen Texten bestimmt wurden.

3 Bildungssprachliche Anforderungen von Texten

Ein Rating-Instrument zur Einschätzung bildungssprachlicher Anforderungen von Texten haben Heppt et al. (2014a) beschrieben und zur Analyse von literarischen Texten und Sachtexten sowie Leseaufgaben eingesetzt. Berücksichtigt werden die Wort- und Satzanzahl, die mittlere Satzlänge, die Anzahl an drei- oder mehrsilbigen Wörtern, der Einsatz fachspezifischen und fachunspezifischen bildungssprachlichen Wortschatzes, die Anzahl an Nominal- und Präpositionalphrasen und Komposita sowie die Anzahl einfacher und zusammengesetzter Sätze (ebd., S. 80 f.). Der Fokus der Arbeit von Heppt et al. liegt auf dem Zusammenhang von bildungssprachlichen Anforderungen und *Differential Item Functioning* von Items, die bei der Evaluation der Bildungsstandards im Kompetenzbereich „Lesen“ für die Grundschule eingesetzt wurden. Überprüft wurde, ob sich die Items für bestimmte Schülergruppen (Muttersprachlerinnen und Muttersprachler sowie Nicht-Muttersprachlerinnen und Nicht-Muttersprachler mit hohem und niedrigem sozioökonomischen Status) als unterschiedlich schwierig erwiesen haben – auch unter Berücksichtigung von bildungssprachlichen Merkmalen der Items. Für die Items, für die sich hinsichtlich der Schülerinnen und Schüler, die Deutsch nicht als Muttersprache sprechen, *Differential Item Functioning* gezeigt hat, ist vor allem ein Zusammenhang mit der Länge und Komplexität auf der Wortebene und der mittleren Satzlänge nachzuweisen. Insbesondere diese Merkmale scheinen sich für diese Schülerinnen und Schüler verständniserschwerend auszuwirken. Allerdings belegen die Ergebnisse nicht, dass sich bildungssprachliche Anforderungen nur auf die Leseleistungen der Schülerinnen und Schüler auswirken, die Deutsch nicht als Muttersprache sprechen.

Ausgehend von Forschungsergebnissen, die belegen, dass Schülerinnen und Schüler mit Migrationshintergrund aus unterschiedlichen Herkunftsländern bei Kompetenztestungen in unterschiedlichem Ausmaß schlechter

abschneiden als Schülerinnen und Schüler ohne Migrationshintergrund, haben Heppt et al. (2016) überprüft, inwiefern Schülerinnen und Schüler türkischer Herkunft bei der Bewältigung bildungssprachlicher Anforderungen in einer Kompetenztestung schlechter abschneiden als Schülerinnen und Schüler mit einem anderen Sprachhintergrund (Schülerinnen und Schüler mit einer anderen Muttersprache als Türkisch oder Deutsch und deutschsprachige Schülerinnen und Schüler). Überprüft wird, ob das Verständnis bildungssprachlicher Texte in höherem Maße mit der Leistung von Grundschülerinnen und Grundschulern in einem Mathematiktest einhergeht als das Verständnis alltagssprachlicher Texte. Des Weiteren wird überprüft, ob Schülerinnen und Schüler türkischer Herkunft bei einem Verständnistest bildungssprachlich geprägter Texte größere Schwierigkeiten haben als andere Sprachgruppen. Die Lesetexte (Erzähl- und Sachtexte), die bei der Evaluation der Bildungsstandards für die Grundschule 2011 verwendet worden sind, werden hinsichtlich ihrer bildungssprachlichen Anforderungen auf Grundlage der Rating-Kategorien beurteilt, die Heppt et al. (2014a) vorschlagen. Auf der Grundlage der Ergebnisse werden die Texte als eher bildungs- beziehungsweise eher alltagssprachlich eingeteilt. Als Maß für die Mathematikleistung werden die Ergebnisse der Mathematiktestung aus der Evaluation der Bildungsstandards für die Grundschule verwendet. Zum einen zeigt sich, dass das Verständnis bildungssprachlicher Texte in einem stärkeren Zusammenhang mit der Mathematikleistung steht als das Verständnis der alltagssprachlichen Texte. Zum anderen wird nachgewiesen, dass alle Schülerinnen und Schüler bei der Bearbeitung eher bildungssprachlich geprägter Texte größere Schwierigkeiten haben als bei der Bearbeitung eher alltagssprachlich geprägter Texte. Zudem zeigt sich, dass die Leistungen türkischstämmiger Schülerinnen und Schüler sowohl beim Verständnis alltags- als auch bildungssprachlich geprägter Texte hinter den Leistungen der anderen Gruppen zurückbleiben. Dieser Effekt ist aber bei einer Kontrolle von zentralen Hintergrundvariablen nicht mehr nachzuweisen.

Die Annahme, bildungssprachliche Anforderungen würden vor allem das Abschneiden von Schülerinnen und Schülern mit Migrationshintergrund oder nicht monolingual deutschsprachigen Schülerinnen und Schülern in Kompetenztestungen beeinflussen, ist unzureichend empirisch belegt (vgl. auch Heppt et al., 2014b). Das könnte mit dem Konstrukt oder der Operationalisierung von Bildungssprache zusammenhängen. Auf die fehlende Eindeutigkeit des Konstrukts Bildungssprache wurde bereits hingewiesen. Das kann dazu führen, dass Unterschiedliches erhoben wird, wenn in empirischen Arbeiten von Bildungssprache gesprochen wird (vgl.

Abschnitt 2). Insofern besteht weiterer Klärungsbedarf hinsichtlich der Operationalisierung von Bildungssprache in medial schriftlichen Texten. Einen Beitrag dazu stellt die Arbeit von Stahns et al. (2016) dar. Die dort unter Rückgriff auf die Operationalisierung von Bildungssprache nach Heppt et al. (2014a) präsentierten Analysen sprachlicher Anforderungen von Leseaufgaben, die im Rahmen von IGLU 2011 eingesetzt wurden, zeigen, dass sich für keines der Merkmale bei Berücksichtigung der in IGLU angesetzten Leseverstehensprozesse ein signifikanter statistischer Zusammenhang mit der Aufgabenschwierigkeit nachweisen lässt. Das dürfte unter anderem damit zusammenhängen, dass die Text- und die Aufgabenschwierigkeit sich nicht alleine durch eine Analyse der sprachlichen Schwierigkeit einschätzen lassen. Es ist zu vermuten, dass Einschätzungen bildungssprachlicher Merkmale durch Ratings ergänzt werden müssen, die es ermöglichen, die Text- und Aufgabenschwierigkeit umfassender einzuschätzen. Dazu bedürfe es unter anderem der Einschätzung der Texte auf Grundlage weiterer Kriterien neben der sprachlichen Schwierigkeit, die die Verständlichkeitsforschung ermittelt hat, so Stahns et al. (2016). Zudem weisen Stahns (2016) sowie Stahns et al. (2016) hinsichtlich der von Heppt et al. (2014a) aufgeführten Kategorien *Wortanzahl*, *Satzanzahl*, *Mittlere Satzlänge* und *Anzahl drei- und mehrsilbiger Wörter* auf die Überschneidung mit Lesbarkeitsberechnungen von Texten hin: Auszählungen der Wort- und Satzanzahl beziehungsweise die Berechnung der Wort- und Satzlänge sind die Grundlage für die Berechnung einiger Lesbarkeitsformeln (vgl. Abschnitt 4). Aus diesem Grund ergänzen Stahns et al. (2016) das Rating der bildungssprachlichen Anforderungen durch die Berechnung des Lesbarkeitsindizes *Lix*. Das Ziel war zu untersuchen, ob das Rating bildungssprachlicher Anforderungen und die *Lix*-Werte in dieselbe Richtung weisen, um zu überprüfen, ob ein Mehrwert in der Durchführung des Ratings bildungssprachlicher Anforderungen nach Heppt et al. (2014a) im Vergleich zu einer Berechnung der Lesbarkeit zu erkennen ist. Auf Grundlage der Berechnung bivariater Korrelationen auf Aufgabenebene ließen sich nicht viele Zusammenhänge zwischen den Merkmalen, die zur Einschätzung bildungssprachlicher Anforderungen eingesetzt wurden, und den Lesbarkeitswerten der Items nachweisen. Trotz der inhaltlichen Überschneidung bei der Kodierung bildungssprachlicher Anforderungen und der Berechnung der Lesbarkeit stützen die Ergebnisse nicht die Annahme, die Berechnung des *Lix*-Werts könnte das Rating der bildungssprachlichen Anforderungen ersetzen. Allerdings weisen die Autoren darauf hin, dass das Textmaterial der Aufgaben zu kurz ist, um hinsichtlich des *Lix*-Werts durchgängig zu plausiblen Ergebnissen zu kommen. Bei dieser Befundlage

ist zu konstatieren, dass es weiterhin aussichtsreich ist, sich im Zusammenhang mit bildungssprachlichen Anforderungen in Kompetenztestungen im Bereich „Lesen“ mit Ergebnissen der Lesbarkeits- und Textverständlichkeitsforschung auseinanderzusetzen.

4 Lesbarkeit

Ein Operationalisierungsvorschlag für bildungssprachliche Anforderungen wie der von Heppt et al. (2014a) kann die Grundlage für eine Einschätzung sprachlicher Anforderungen von Texten darstellen. Sprachliche Anforderungen von Texten können aber auch durch eine Berechnung der Lesbarkeit der Texte ermittelt werden (vgl. zusammenfassend Klare, 1963). Der Begriff Lesbarkeit kann inhaltlich unterschiedlich gefasst werden (vgl. Bamberger & Vanecek, 1984, S. 15 f.). Im Folgenden wird er im Sinne von sprachlicher Einfachheit beziehungsweise Schwierigkeit verwendet.

Steht die Lesbarkeit von Texten in diesem Sinne im Fokus, werden Merkmale der sprachlichen Oberfläche der Texte im Hinblick darauf beurteilt, inwiefern sie den Prozess der erfolgreichen Verarbeitung des Textes beeinflussen könnten. Von besonderem Interesse für die Lesbarkeitsforschung sind Arbeiten, in denen Befunde zu sprachlichen Merkmalen von Texten (Prädiktorvariablen) mit Ergebnissen von Verständnistests oder Lesegeschwindigkeitstests (Zielkriterien) verbunden worden sind (vgl. Groeben, 1982, S. 173 ff.; Groeben & Christmann, 1989, S. 166 f.; Klare, 1963, S. 33 ff.). Resultat dieser Arbeiten sind Lesbarkeitsformeln wie Fleschs „reading ease“ (Flesch, 1948, S. 225; vgl. auch Flesch, 1951) oder der *Lix* (vgl. Bamberger & Vanecek, 1984, S. 62 ff.; vgl. auch die umfassende Darstellung und Bewertung von Lesbarkeitsformeln bei Klare, 1963, S. 37 ff.). Deren Berechnung liegen Faktoren auf der Wort- und Satzebene zugrunde. Im Fall der Formel von Flesch handelt es sich um die mittlere Wortlänge, die über die Silbenzahl pro 100 Wörter berechnet wird (vgl. Flesch, 1948, S. 223), und die mittlere Satzlänge, die sich aus der Anzahl der Wörter und Sätze ergibt (vgl. ebd., S. 229; vgl. auch Flesch, 1951). Im Fall der *Lix*-Formel werden die mittlere Satzlänge und der prozentuale Anteil an Wörtern mit mehr als sechs Buchstaben berücksichtigt (vgl. Bamberger & Vanecek, 1984, S. 62). Die Berechnung der Lesbarkeit von Texten auf der Grundlage dieser Formeln erlaubt eine objektive Messung einiger Anforderungen der Textoberfläche. Es wurde argumentiert, dass mit den Formeln trotz der Beschränkung auf einen Wort- und einen Satzfaktor weiterreichende Schlüsse über die sprachliche Schwierigkeit auf der Wort- und Satzebene möglich

seien, weil Merkmale auf diesen Ebenen hoch miteinander korrelierten – zum Beispiel die Wortlänge und die Auftretenshäufigkeit von Wörtern oder die Satzlänge und die Satzkomplexität (vgl. Amstad, 1978, S. 49; Bamberger & Vanecek, 1984, S. 37 ff.; vgl. auch Abschnitt 9). Die prädiktive Validität von (einigen) Lesbarkeitsformeln ist für die Lesegeschwindigkeit und seltener für das Leseverständnis nachgewiesen (vgl. Groeben, 1982, S. 178 f.; Klare, 1963, S. 121 ff.).

Einige der Merkmale, die in der Lesbarkeitsforschung betrachtet wurden und die in Lesbarkeitsformeln wie den oben genannten berücksichtigt werden, finden sich in der Diskussion um Bildungssprache wieder. Stahns (2016) und Stahns et al. (2016) weisen hinsichtlich der Merkmale *Wortanzahl*, *Satzanzahl*, *Mittlere Satzlänge* und *Anzahl drei- und mehrsilbiger Wörter* aus dem Operationalisierungsvorschlag für bildungssprachliche Anforderungen von Heppt et al. (2014a) darauf hin, dass sich die Zuordnung dieser Merkmale zum Konstrukt Bildungssprache nicht unmittelbar ergibt und von Lesbarkeit gesprochen werden sollte. Der Überschneidungsbereich zwischen den Merkmalen, die in Aufstellungen bildungssprachlicher Merkmale berücksichtigt werden, und Merkmalen von Texten, die mit der Lesbarkeit in Verbindung stehen, ist allerdings größer. In einer Aufstellung bildungssprachlicher Merkmale führen Gogolin und Duarte (2016) auf lexikalischer Ebene unter anderem „Fachvokabular“, „nichtfachliche Fremdwörter“, „differenzierende und abstrahierende Ausdrücke“, „Nominalisierungen“, „nominale Zusammensetzungen“, „selten verwendete Wörter“, „Abkürzungen und Akronyme“, „viele und seltener verwendete Strukturwörter“ (ebd., S. 489) und auf grammatischer Ebene unter anderem „Satzgefüge“, „Passiv“ und „Passiversatzformen“ sowie „Junktionen“ (ebd., S. 490) als bildungssprachliche Merkmale auf. Hier zeigen sich Überschneidungen mit Merkmalen, die in Studien zur Lesbarkeit von Texten berücksichtigt wurden. Neben der Wort- und Satzlänge handelt es sich um das Auftreten hoch- beziehungsweise wenig frequenter Wörter, den Gebrauch von Fremd- und Fachwörtern, die Abstraktheit beziehungsweise Konkretheit von Wörtern, die syntaktische Komplexität von Sätzen (v. a. das Auftreten von Satzreihen und Satzgefügen, die Zahl der Nebensätze, das Verhältnis von Haupt- und Nebensätzen, die Art der Nebensätze, die Verwendung von Passivformen) und die Art der Satzverbindung (durch Konjunktionen etc.) (vgl. Amstad, 1978, S. 33 ff.; Bamberger & Vanecek, 1984, S. 37 ff.; Groeben, 1982, S. 185; Klare, 1963, S. 164 ff.).

Diese Überschneidungen werfen die Frage auf, ob von bildungssprachlichen Merkmalen gesprochen werden sollte, wenn die Frage nach dem Beitrag dieser Merkmale für die Textschwierigkeit im Fokus steht. Unabhängig

davon, ob in Bezug auf diese Merkmale von Aspekten der Lesbarkeit oder von bildungssprachlichen Anforderungen gesprochen wird, ist darauf hinzuweisen, dass die Berücksichtigung von sprachlichen Oberflächenmerkmalen für Einschätzungen der Textschwierigkeit keine ausreichende Grundlage bieten dürfte. In der Aufarbeitung der Ergebnisse der Lesbarkeitsforschung im Zusammenhang mit der Diskussion um Textverständlichkeit wird dies deutlich.

5 Textverständlichkeit

Von Vertretern der Verständlichkeitsforschung wurde der Ansatz der Lesbarkeitsforschung kritisiert: Unter anderem wurde angemerkt, dass in der Lesbarkeitsforschung individuelle Rezeptionsprozesse und die Leser-Text-Interaktion ausgeblendet werden und ausschließlich die sprachlich-stilistische Einfachheit von Texten fokussiert wird (vgl. Groeben, 1978, 1982; Groeben & Christmann, 1989). Groeben und Christmann (1989) bezeichnen die Lesbarkeitsforschung als „abgeschlossenen geltenden Ansatz“ (ebd., S. 167) und Groeben (1982) bezeichnet sie als „Vorstufe zur eigentlichen Verständlichkeitsforschung“ (ebd., S. 173). Von dieser „eigentlichen“ Verständlichkeitsforschung wurden Modelle entwickelt, in denen unter anderem Textmerkmale jenseits sprachlicher Anforderungen bei der Einschätzung der Textverständlichkeit Berücksichtigung finden (vgl. Groeben, 1978, 1982; Groeben & Christmann, 1989; Langer et al., 2011).

Ein Modell zur Einschätzung der Textverständlichkeit stammt von Langer et al. (2011). Die Autoren schlagen ein Rating auf Grundlage der Dimensionen „Gliederung/Ordnung“ (ebd., S. 32), „Kürze/Prägnanz“ (ebd.), „Anregende Zusätze“ (ebd.) und „Einfachheit“ (ebd.) vor. Die Dimensionen wurden durch ein induktives Vorgehen gewonnen: Auf Grundlage von Experteneinschätzungen wurden unterschiedliche Texte hinsichtlich aus der Forschungsliteratur gewonnener Merkmale eingeschätzt, die Korrelationen der Merkmale untereinander wurden berechnet und durch Faktorenanalysen wurden die vier Dimensionen ermittelt. Abgesichert wurde die Relevanz der vier Dimensionen durch Verständnis- und Behaltentests (vgl. die zusammenfassende Darstellung des Verfahrens in Langer & Schulz von Thun, 2007, S. 39 ff.; Langer et al., 2011, S. 189 ff.). Langer und Schulz von Thun (2007, S. 39 f.) plädieren für ein Rating als Verfahren zur Messung der Textverständlichkeit wegen der begrenzten Aussagekraft der auszählungsbasierten Verfahren, die die Grundlage für die Arbeit mit Lesbarkeitsformeln darstellen. Texte seien dann sprachlich einfach, wenn

darin „geläufige, anschauliche Wörter [...] zu kurzen, einfachen Sätzen zusammengefügt“ (Langer et al., 2011, S. 22) werden. Schwierig zu verstehen seien syntaktisch unnötig komplexe und lange Sätze mit ungeläufigen, nicht anschaulichen Wörtern (vgl. ebd.). Hinsichtlich eines Urteils über die *Gliederung/Ordnung* wird zwischen den Aspekten „Innere Ordnung“ (ebd., S. 24) und „Äußere Gliederung“ (ebd.) unterschieden. Texte könnten als geordnet angesehen werden, wenn die Beziehungen zwischen den Sätzen erkennbar seien – die Sätze seien dann „folgerichtig aufeinander bezogen.“ (Ebd.) Zudem sei zu überprüfen, ob der Textinhalt in einer sinnvollen Reihenfolge präsentiert werde (vgl. ebd.). Gegliedert werde ein leicht verständlicher Text durch die Gruppierung zusammengehöriger Teile und die dazugehörigen Überschriften sowie „Vor- und Zwischenbemerkungen“ (ebd.). Ebenfalls von Bedeutung seien Zusammenfassungen und Hervorhebungen von Wichtigem. Für die Beurteilung des Aspekts *Kürze/Prägnanz* sei die Passung von Textlänge und der zu vermittelnden Inhalte entscheidend (vgl. ebd., S. 26). Texte seien leichter verständlich, beschränkten sich die Autoren auf Wesentliches und vermittelten es so umfangreich wie nötig. Anregende Zusätze seien Mittel, mit denen eine Autorin beziehungsweise ein Autor „Interesse, Anteilnahme, Lust am Lesen und Zuhören hervorrufen will“ (ebd., S. 27) – zum Beispiel der Einsatz wörtlicher Rede oder direkte Leseransprachen. *Einfachheit* betrachten Langer et al. (2011) als die wichtigste Dimension für die Textverständlichkeit – sie plädieren für maximale Einfachheit und setzen hinsichtlich der sprachlichen Anforderungen keine „Banalitätsschwelle“ (Bremerich-Vos, 1991, S. 153) an. Ob Merkmale der Dimension *Anregende Zusätze* die Verständlichkeit erhöhten, hänge von Merkmalen der Dimension *Gliederung/Ordnung* ab – unübersichtliche Texte erschwerten Zusätze eher (vgl. Langer et al., 2011, S. 32). Zudem könnten *Kürze/Prägnanz* und *Anregende Zusätze* in Konkurrenz zueinander treten, so die Autoren (vgl. ebd.).

Neben diesem Modell ist im deutschsprachigen Raum ein Verständlichkeitsmodell von Groeben (1978, 1982) populär. In dem Modell werden nicht nur textimmanente Merkmale berücksichtigt, sondern dem Anspruch nach auch die Interaktion zwischen der Leserin beziehungsweise dem Leser und dem Text. Groeben (1978) setzt die Dimensionen „Stilistik“, „Kognitive Strukturierung“ und „Konzeptueller Konflikt“ beziehungsweise in einer neueren Version des Modells (vgl. Groeben, 1982, S. 223 ff.) die Dimensionen „Sprachliche Einfachheit“, „semantische Kürze/Redundanz“, „Kognitive Gliederung/Ordnung“ und „Stimulierender kognitiver Konflikt“ (ebd.) an. Einen Vergleich und eine Kritik der Modelle von Groeben und Langer et al. hat unter anderem Tergan (1981) vorgenommen (vgl. auch

Bremerich-Vos, 1991, S. 150 ff.; Groeben, 1982, S. 188 ff.; Groeben & Christmann, 1989, S. 173 f.). Unterschiede zwischen den Modellen zeigen sich hinsichtlich der Einschätzung der Relevanz der Dimensionen. Während Langer et al. (2011) der Dimension *Einfachheit* die größte Bedeutung für die Textverständlichkeit zusprechen, weist Groeben (1982) nach, dass die kognitive Strukturierung das bedeutendste Merkmal ist, während sich die sprachliche Einfachheit für die Textverständlichkeit als weniger bedeutsam erweist (vgl. auch Groeben & Christmann, 1989, S. 169 ff.; Tergan, 1981, S. 343).²

Eine erste Annäherung an die Textschwierigkeit der in IGLU eingesetzten Lesetexte ist unter Rückgriff auf die Dimensionen von Langer et al. (2011) erfolgt (vgl. Bremerich-Vos, Tarelli & Valtin, 2012, S. 72 f.). Im vorliegenden Beitrag wird der Ansatz von Langer et al. (2011) verfolgt, weil das Rating, das dafür vorgesehen ist, dem auszählungsbasierten Ansatz der Lesbarkeitsforschung und auch der Bestimmung bildungssprachlicher Anforderungen in neueren Arbeiten mit dem Fokus auf textimmanente Merkmale ähnelt.

Zwischen der Einfachheit im Sinne des Modells von Langer et al. (2011) und den Merkmalen, die den einschlägigen Aufstellungen zufolge als kennzeichnend für Bildungssprache angesehen werden können (vgl. die Abschnitte 2 und 3), sowie den von der Lesbarkeitsforschung ermittelten schwierigkeitsgenerierenden Merkmalen (vgl. Abschnitt 4) sind Überschneidungen festzustellen. Bei der Beschreibung der Dimension *Einfachheit* stehen „Wortwahl und Satzbau“ (Langer et al., 2011, S. 22) im Fokus: Zur Einschätzung der sprachlichen Einfachheit sind auf der Wortebene Abstraktheit beziehungsweise Konkretheit sowie Anschaulichkeit und Geläufigkeit der Wörter sowie der Einsatz von Fachwörtern, die nicht erläutert werden, und auf der Satzebene die syntaktische Komplexität zu berücksichtigen (vgl. ebd.). Insofern wird ein Rating im engeren Sinne – keine Auszählung – von Merkmalen vorgenommen, die aus der Tradition der Lesbarkeitsforschung stammen und die in der Diskussion um Bildungssprache

2 Des Weiteren können Textverarbeitungsmodelle aus der kognitionspsychologischen Forschung verwendet werden, um Anforderungen von Texten für Leserinnen und Leser zu bestimmen. Dem Modell der Textverarbeitung von van Dijk und Kintsch (1983) zufolge sind drei Ebenen zu unterscheiden: Die Ebene der oberflächlichen Textrepräsentation, die Ebene einer semantischen Textbasis und diejenige eines Situationsmodells, in das Vor- und Weltwissen der Leserin/des Lesers einfließen. Um Aussagen über die Verständlichkeit eines Textes zu treffen, kann dessen propositionale Struktur aufgeschlüsselt werden. Das dürfte bei komplexen Texten sehr aufwändig sein (vgl. für ein Beispiel Bremerich-Vos, 1991, S. 157 ff; vgl. auch Groeben, 1982, S. 48 ff.).

zum Teil wieder aufgegriffen wurden. Zudem werden in dem Modell von Langer et al. drei weitere Dimensionen berücksichtigt, die für die Textverständlichkeit von Bedeutung sind beziehungsweise sein könnten.

6 Forschungsfragen

Mit der Bestimmung bildungssprachlicher Anforderungen auf Grundlage von Katalogen wie denen von Heppt et al. (2014a) oder Stahns et al. (2016) und der Berechnung der Lesbarkeit existieren zwei auszählungsbasierte Verfahren, um Textschwierigkeit einzuschätzen. Bei den Merkmalen von Bildungssprache sowie von Lesbarkeit, die in der einschlägigen Literatur diskutiert werden, gibt es einen breiten Überschneidungsbereich. Ein Ergebnis der Lesbarkeitsforschung ist die Entwicklung von Lesbarkeitsformeln, die die Berechnung der Lesbarkeit erheblich vereinfachen, weil nicht mehr alle Merkmale, die einen Einfluss auf die sprachliche Schwierigkeit eines Textes haben könnten, berücksichtigt werden müssen. Ergebnisse der Lesbarkeitsforschung belegen vor allem die Auswirkungen von weniger einfach lesbaren Texten auf die Lesegeschwindigkeit, während in der jüngeren Forschung, in der bildungssprachliche Anforderungen medial schriftlicher Texte fokussiert werden, das Textverständnis beziehungsweise die Aufgabenschwierigkeit im Fokus stehen (vgl. Heppt et al., 2014a; Stahns et al., 2016). Ebenfalls mit der Textschwierigkeit beschäftigt sich die Textverständlichkeitsforschung. Im Falle des Verständlichkeitsmodells von Langer et al. (2011) wird zur Einschätzung der Verständlichkeit ein Rating durchgeführt. In diesem Rating werden bei der Einschätzung der Dimension *Einfachheit* Merkmale berücksichtigt, die auch der Ermittlung der bildungssprachlichen Anforderungen und der Berechnung der Lesbarkeit zugrunde liegen können, allerdings findet keine Auszählung der einzelnen Merkmale statt. Des Weiteren werden in dem Modell von Langer et al. (ebd.) Dimensionen von Verständlichkeit berücksichtigt, die in der Lesbarkeitsforschung und bei der Bestimmung von bildungssprachlichen Anforderungen unberücksichtigt bleiben. Das Modell ist umfassender, wenn es darum geht, Merkmale zu bestimmen, die potenziell das Verständnis und die Schwierigkeit von Lesetexten beeinflussen.

Die vorliegende Studie stellt einen Beitrag zur Klärung der Frage dar, in welchem Verhältnis bildungssprachliche Anforderungen von Sachtexten zu weiteren potenziell schwierigkeitsbestimmenden Merkmalen von Texten stehen. Analysiert wird, ob bildungssprachliche Anforderungen oder andere sprachliche beziehungsweise sonstige Textmerkmale mit der Aufgabenschwierigkeit

schwierigkeit von Aufgaben aus der Studie IGLU zusammenhängen. Die Analysen der Textverständlichkeit beziehen sich nur auf die Sachtexte, nicht auf die Aufgaben zu den Sachtexten, da die Einschätzung der in dem Modell von Langer et al. (2011) angesetzten Dimensionen wegen der Kürze der Aufgaben nicht möglich ist. Im Einzelnen stehen folgende Forschungsfragen im Fokus des Beitrags:

- 1.) In welchem Ausmaß werden Leserinnen und Leser mit sprachlichen Anforderungen – gemessen an bildungssprachlichen Merkmalen und der Lesbarkeit – in den Sachtexten aus IGLU und den dazugehörigen Leseaufgaben konfrontiert?
- 2.) Wie sind die Sachtexte aus IGLU hinsichtlich der vier Dimensionen der Textverständlichkeit nach Langer et al. (2011) zu beurteilen?

Im Anschluss daran wird überprüft, wie sich die bildungssprachlichen Merkmale, die Lesbarkeit und die Einschätzung der Textverständlichkeit zueinander verhalten:

- 3.) Lassen sich auf Grundlage der Ergebnisse des Ratings der Textverständlichkeit Zusammenhänge zwischen der Verständlichkeit sowie der Lesbarkeit und den bildungssprachlichen Merkmalen nachweisen?
- 4.) Existiert mit dem *Lix* ein kompensatorisches Messinstrument, das die Kodierungen der sprachlichen Anforderungen, insbesondere die, die häufig als bildungssprachlich bezeichnet werden, ersetzen kann?
- 5.) Wie verhalten sich sprachliche Anforderungen auf Text- und Aufgabenebene zueinander?
- 6.) Zeigen sich Zusammenhänge zwischen den bildungssprachlichen Anforderungen, der Lesbarkeit, der Textverständlichkeit und der Aufgabenschwierigkeit?

7 Methoden und Daten

IGLU ist eine international vergleichende Schulleistungsstudie, in der alle fünf Jahre das Leseverständnis von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe im internationalen Vergleich untersucht wird (vgl. zuletzt Hußmann et al., 2017). Deutschland beteiligt sich seit 2001 an IGLU.