

2 Eine Kritik der Lese-/Testkonstrukte in PISA, DESI und TEDS-LT aus sozialtheoretischer Perspektive

Una Dirks

Unter der Prämisse, dass fallbasierte Diagnosen auf der Grundlage schriftlicher und/oder visueller Texte erfolgen, sind diagnostische Kompetenzen mit jenen des Text- bzw. Leseverständnisses¹ eng verschränkt. Aus diesem Grunde widmet sich der erste Abschnitt dieses Kapitels der – für diagnostische und lesespezifische Kompetenzen gleichermaßen relevanten – Genregebundenheit der jeweils zu bearbeitenden Texte (s. Kap. 2.1). Die kompetenzspezifischen Referenzbereiche weisen aber auch Unterschiede auf, die für den Erhalt beider Kompetenzen als eigenem Forschungsgegenstand sprechen (s. Kap. 2.2). Eine Diskussion der im Rahmen von PISA („Program of International Student Assessment“), DESI (Deutsch Englisch Schülerleistungen International) und TEDS-LT („Teacher Education and Development Study: Learning to Teach“) herangezogenen Kompetenzmodelle (s. Kap. 2.3) soll v.a. verdeutlichen, inwiefern diese von sozialtheoretischer Expertise profitieren und ein entsprechend dimensioniertes Kompetenzmodell – wie es im Forschungsprojekt „Entwicklung der Diagnosekompetenz in der ersten Phase der hessischen Lehrerbildung“ angewendet wurde – zur Bearbeitung der vorgetragenen Kritikpunkte beitragen könnte (vgl. Kap. II.5, Dirks/Hansmann 2012). Dieser Beitrag versteht sich daher zugleich als eine Antwort auf die Frage, warum wir zur Erfassung von Diagnosekompetenz nicht die bekannten Lese- und Wissenskompetenzmodelle herangezogen haben.

Auf weitere Modellanalysen – z.B. im Zusammenhang mit Problemlösekompetenzen in PISA 2003 (vgl. Klieme et al. 2005, Leutner et al. 2004, OECD 2004) – wird hier trotz einzelner Kompetenzüberschneidungen verzichtet. Einige der unten aufgeführten Kritikpunkte hätten sich lediglich wiederholt: Dies betrifft u.a. den nicht näher erläuterten Verstehens-Begriff bei Missachtung nomologischer Zusammenhänge des Erklärens (z.B. OECD 2004: 28ff.) ebenso wie die vernachlässigte Definition der Beziehungen zwischen sozialen Phänomenen, wenngleich diese für die Ausweisung von Schwierigkeitsgraden benötigt wird.

2.1 Die Genregebundenheit des Leseverständnisses

Texte – seien es schriftliche, mündliche oder visuelle Texte – sind Konstrukte sozialer Realitäten, die zugleich einem Genre oder mehreren Genres zugeordnet werden können. Die Genrezugehörigkeit ist insofern von zentraler Bedeutung, als ein Genre eine spezifische Anforderungsstruktur an die Leserinnen und Leser stellt und deren Informationsverarbeitung von den jeweiligen textbasierten Genremerkmalen maßgeblich beeinflusst wird (vgl. Graesser et al. 1997). Die mit den Gen-

¹ Die Begriffe „Text-“ und „Leseverständnis“ verwende ich synonym.

rekenntnissen verbundenen Erfahrungen und Erwartungen sind somit für das Ergebnis der Textverarbeitung von maßgeblicher Bedeutung. Erhalten Studierende eine Aufgabe, die sich auf ein ihnen vertrautes Genre wie ‚Märchen‘, ‚Bericht‘, ‚Witze‘ oder ‚weblog‘ bezieht, werden sie diese Aufgabe in aller Regel schneller und besser bearbeiten, als wenn ihnen das Genre gänzlich unbekannt ist oder ihnen gar ein ‚falsches‘ Genre genannt wird (vgl. Zwaan 1994) und sich keine ‚gute‘ Passung zwischen den erwarteten formalen und den – meist eng damit verknüpften – inhaltlichen Textdimensionen herstellen lässt (zur Passung s. Rickheit/Strohner 1999: 275f.). Wie auch immer das beim Lesen evozierte Genre- bzw. Gattungswissen (n. Thomas Luckmann) prozessiert wird: In letzter Konsequenz kann es Lesern umfassenden Einblick in die kommunikativen Möglichkeiten und Grenzen „des Aufbaus, der Aufrechterhaltung und des Wandels gesellschaftlicher Ordnungen“ bieten (Luckmann 1986: 205) und damit einen bedeutenden Beitrag zur Orientierung über textbezogene Sinngehalte leisten. Deshalb sind auch die fallbasierten Genremerkmale, mit denen die Studierenden im Rahmen des Forschungsprojektes konfrontiert wurden, im Abschnitt II.4 näher erläutert worden (Dirks 2012).

2.2 Die wechselseitige Verwiesenheit fallbasierter Lese- und Diagnosekompetenz

Lesekompetenzen wurden im Rahmen der PISA- und DESI-Studien als vergleichsweise basale Kompetenzen definiert – und zwar als Fähigkeit, „am gesellschaftlichen Leben und seinen genrespezifischen Ausdrucksweisen“ teilzunehmen, Wissensstrukturen aufzubauen und zu erweitern sowie Perspektivenwechsel zu erproben (Artelt et al. 2001: 69). Im Allgemeinen wird davon ausgegangen, dass die jeweils individuellen Sinnkonstruktionen aus einer aktiven Verbindung der Aussagen im Text mit dem Vor-, Welt- und Sprachwissen der Leser/-innen entstanden sind (a.a.O.: 70). Ähnliches gilt beim Erstellen von Diagnosen. Denn sowohl Probanden von Lesekompetenztests als auch jene von Diagnosekompetenztests vollziehen Konstruktionen, die Graesser et al. (1994) mit Bezug auf das Textverstehen als andauernde „search after meaning“ (371) beschreiben. Trotz dieser Gemeinsamkeit bestehen auch deutliche Unterschiede zwischen den Kompetenzen: So setzen fallbasierte Diagnosen die Verfügbarkeit von Lesekompetenzen voraus, da ein angemessenes Leseverständnis als Grundprämisse für die jeweilige Diagnose fungiert. Das jeweilige Leseziel wird allerdings nicht zwangsläufig in den Dienst bestimmter Diagnosen gestellt. Die Referenzbereiche des Lesens bzw. von Lese(test-)aufgaben können sowohl fiktionale als auch nicht-fiktionale Genres umfassen, sind also vergleichsweise breit gefächert; die Anwendung diagnostischer Kompetenzen zielt dagegen auf die Analyse bestimmter sozialer Phänomene, die vorrangig anhand nicht-fiktionaler Genres diagnostiziert werden und – das ist wohl der wichtigste Unterschied – in sehr viel stärkerem Maße einer evidenzbasierten Fundierung verpflichtet sind. Denn eine Diagnose ist in der Regel an ein individuell und/oder gesellschaftlich relevantes Ziel gebunden: Wenn die Studierenden z.B. diagnostizieren sollen, wie sich eine soziale Ordnung im Klassenraum herstellt, die das Unterrichten überhaupt erst möglich macht, oder über welche Vorkenntnisse

die Schüler/-innen mutmaßlich verfügen, eignen sie sich zugleich professionstheoretisch relevante Kenntnisse an, die für ihr späteres Lehrerhandeln und für die Qualität der von ihnen angeleiteten Lehr-/ Lernprozesse von zentraler Bedeutung sein können. Ähnlich wie in der Medizin stehen Diagnosen nämlich am Beginn einer Kette von Folgehandlungen, die im schulischen Bereich aus pädagogischen oder fachspezifischen Interventionen bestehen, die wiederum Gegenstand von Evaluationen sind, in denen überprüft wird, inwieweit die Interventionen zu besseren Outcomes geführt haben.

Unabhängig von den Unterschieden erscheint eine Verknüpfung von Lese- und Diagnosekompetenzen insofern legitim und mit dem Lehrerberuf gut vereinbar, als es sich in pädagogischen Berufen um eine übliche Praktik handelt, Akten und andere Dokumente als erste Informationsquellen *vor* einer Intervention im Feld zu konsultieren, also eine evidenzbasierte Anamnese und Diagnose zu erstellen. Das dazu erforderliche Leseverständnis wird im Forschungsprojekt durch die Konfrontation mit verschriftlicht vorliegenden Fällen, also durch eine zugleich diagnose- und lesekompetenzspezifische Anforderungsstruktur erhoben.

Angesichts der skizzierten Kompetenzüberschneidungen ist es nicht weiter verwunderlich, dass der Diagnosebegriff in der Pädagogischen Psychologie gleichermaßen im Zusammenhang mit der Identifikation und Einordnung von Informationen während der Textlektüre benutzt wird: Demnach schließt die Aktivierung bzw. Genese eines Lese-Situationsmodells – sozialtheoretisch gewendet geht es um das beim Lesen vollzogene *Framing* von Text und Kontext in Verbindung mit den individuellen Wissensressourcen – stets auch diagnostische Textinformationen mit ein (vgl. Dutke et al. 2011). Textinformationen, die nicht nur einem Situationsmodell entsprechen, sondern sich allen möglichen Situationsmodellen zuordnen lassen, werden als nicht-diagnostisch begriffen (Hecker/Dutke 2004); Letzteren fehlt i.d.R. die für das spezifische Textverständnis erforderliche Relevanz. Interessant ist die diagnostische Qualität von Informationen sowohl für die Psychologie als auch für die Kognitionsoziologie bzw. für soziokognitive Konzeptualisierungen: So können die beim Lesen diagnostizierten neuen Informationen in das Situationsmodell bzw. den Frame integriert werden, wodurch das textbezogene kognitive Modell einem „Update“ (Blanc et al. 2008, Graesser et al. 1994, Kintsch 1998, van Oostendorp et al. 2002, Zwaan/Radvansky 1998) bzw. „Re-Framing“ (Esser 2001: 274ff., 2002) unterzogen wird. Kommt es im Rahmen der Informationsdiagnose z.B. nicht nur zur Konkretisierung einer Situation,² sondern gar zur Identifikation von Widersprüchen zwischen einem bereits erzeugten Situationsmodell/Framing und den textimmanenten Informationen über eine Situation, bestimmte Praktiken und/oder

² Solche textbasierten Veränderungen, die eine Engführung oder Fokussierung auf ein bestimmtes Situationsmodell erfordern, werden in psycholinguistischen Studien im englischsprachigen Raum meist als „constraints“ diskutiert. Im Vergleich zu Situationsmodellanalysen, die sich mit dem Einwirkungspotential von Ursache-Wirkungs-Mechanismen und Widersprüchen im Textmodell befassen, sind diese „constraints“ bisher eher selten ein Forschungsgegenstand (vgl. auch Dutke et al. 2011).

Effekte, und erweisen sich diesbezügliche Informationen als hinreichend abgesichert und relevant für das Textverständnis, handelt es sich per definitionem um diagnostische Informationen (Hecker/Dutke 2004), die im Rahmen des „Update“ bzw. „Re-Framing“ auch neue Ursache-Wirkungsverknüpfungen nach sich ziehen können. Exemplarisch sei auf ein Hauptargument für den Beginn des Irak-Krieges im Jahre 2003 verwiesen: Hieß es zunächst in den offiziellen Verlautbarungen der anglo-amerikanischen Regierungen, der Irak verfüge über Atomwaffen, deren Einsatz es beizeiten abzuwenden gelte, musste das kriegslegitimierende Bedrohungs-Framing infolge der späteren Unauffindbarkeit solcher Waffen revidiert werden. Solch eines Re-Framings hätte es möglicherweise aber gar nicht bedurft, wären die bereits damals vorliegenden Expertisen (Dirks 2010: 65ff.) nur mit der hinreichenden Sorgfalt und Unvoreingenommenheit gelesen und in der Öffentlichkeit verbreitet worden. Die Qualität von Lesekompetenzen kann also – neben jener der Hörkompetenz – von ganz zentraler Bedeutung für eine realitätsangemessene Situationsbewertung sein, die zudem die eigenen Logiken verschiedener Handlungsfelder (z.B. der politisch-militärischen und massenmedialen Informationsaufbereitung) einkalkulieren muss.

2.3 Lesekompetenzmodelle auf dem Prüfstand

Welche Lesekompetenzen werden für ein angemessenes, hier: diagnostisches Textverständnis benötigt? Welches Testkonstrukt ist für die systematische Erfassung der Kompetenzen geeignet? In diesem Abschnitt werden einige theoretische und methodologische Aspekte der bekannten Lesekompetenzmodelle aus den PISA- und DESI-Studien näher betrachtet, die sich aus sozialtheoretischer Perspektive als problematisch erwiesen haben. Dabei orientiere ich mich an der folgenden Definition einer erfolgreichen Informationsverarbeitung beim Lesen, wie sie zugleich für das Diagnostizieren benötigt wird:

- ‚Gute‘ Leserinnen versuchen, die Deutungen und Praktiken von Textautoren und den im Text erwähnten Akteuren sowie die aus dem jeweiligen Akteurshandeln hervorgehenden Effekte soweit wie möglich unter systematischer Berücksichtigung genre-, situations- und domänenspezifischer Kenntnisse zu verstehen und zu erklären.

Demnach ist die jeweils generierte Prozessierung von Wissensressourcen – wie das Erkennen und Verknüpfen von Informationen mithilfe bestimmter Inferenzen – immanenter Bestandteil des Text- bzw. Leseverständnisses. Dass die dabei generierten bzw. aktivierten Bedeutungskonstruktionen auf unterschiedlichen, meist miteinander verknüpften Ebenen verortet werden können, erscheint unstrittig. Die im Zuge der kognitiven Wende der 1970er-Jahre entstandenen und seitdem weiter entwickelten Konstruktionsmodelle, Schematheorien und mentalen Modelle gehen denn auch von der Grundannahme aus, dass die im Leseprozess generierte Interaktion zwischen Text und Lesern als Synthese von ‚bottom-up‘- und ‚top-down‘-Prozessen betrachtet (vgl. Rumelhart 1977, Grabe 1991) und den metaphorischen Modellierungen zugerechnet werden kann (Grabe/Stoller 2011: 25ff.). Bevor weite-

re Detaillierungen erfolgen, seien zunächst die in den PISA- und DESI-Studien herangezogenen, vorrangig kognitionspsychologisch fundierten Lesetestkonstrukte kurz benannt:

- Das Internationale PISA-Konsortium der OECD orientiert sich v.a. an den Vorarbeiten von Kirsch und Mosenthal (vgl. OECD 1999). Obwohl diese mit überwiegend induktiv generierten Konstruktvariablen operieren, erfolgt die Begründung der Lesekompetenzen aus einer primär pragmatischen Perspektive: Die Lesestrategien und kognitiven Prozesse werden mit den jeweiligen Anwendungssituationen korreliert. Besondere Bedeutung wird daher Sachtexten und einer an Alltagsanforderungen ausgerichteten Aufgabenformulierung beigemessen (Kirsch et al. 2002: 16ff., 23ff.).
- Das Deutsche PISA-Konsortium stützt sich auf das kognitionspsychologisch fundierte „construction-integration“-Modell von Kintsch (1988, 1998) und van Dijk & Kintsch (1983), das zwischen Sinneinheiten auf der Textoberfläche („surface representation“), den propositionalen Inhalten im Text („text representation“) und einem Situationsmodell („situation modell“) unterscheidet (vgl. Schaffner et al. 2004b: 199f.). In Ergänzung zur internationalen Studie erhebt das in Deutschland verwendete Testkonstrukt auch die Behaltensleistungen der Schüler.³ Im DESI-Projekt wird das Modell mithilfe zusätzlicher Wechselbeziehungen zwischen den verschiedenen Repräsentationsebenen weiter ausdifferenziert (Nold 2003: 375, Nold/Rossa 2007, Nold/Willenberg 2007).

PISA und DESI erheben die Kompetenzen von 15-jährigen Schülerinnen und Schülern. TEDS-LT stellt einen der – in Deutschland, aber auch international – bis dato wenigen Versuche dar, die Kompetenzen von Studierenden zu erfassen (Blömeke et al. 2011). Diese studentische Adressatengruppe stammt – wie in unserem Forschungsprojekt – aus dem Lehramtsstudium. Zudem überschneiden sich die kognitiven Anforderungen fallbasierter Diagnosen mit den in TEDS erfassten Kompetenzdimensionen überschneiden, deren Performanz wiederum Lesekompetenzen voraussetzt. Vor diesem Hintergrund erschien es sinnvoll, das Testkonstrukt exemplarisch im Rahmen der Studienfächer Germanistik und Anglistik zu diskutieren.

- Als Kompetenzmodell der Wahl für eine Vielfalt unterschiedlicher Kompetenzen („one for all“) nutzen die Deutsch- (Bremerich-Vos et al. 2011) und Englisch-Experten (Roters et al. 2011) von TEDS-LT das zweidimensionale Modell von Anderson und Krathwohl (2001). Es unterscheidet zwischen dekla-

³ Für PISA 2000 s. Artelt et al. (2001), für die Modellierung nicht-kontinuierlicher Texte in PISA 2003 s. Schaffner et al. (2004a), zur unveränderten Kompetenzmodellierung für PISA (2006) s. Drechsel & Artelt (2007). Wohl nicht zuletzt, um Vergleiche zu den vorangegangenen Testdurchläufen herstellen zu können, wurde die Lesekompetenzmodellierung im Rahmen von PISA 2009 nur geringfügig verändert. So sind v.a. die hierarchieniedrigeren und -höheren Kompetenzniveaus stärker ausdifferenziert worden (vgl. Naumann et al. 2010: 26ff.).

rativem Fakten- und Begriffswissen sowie prozeduralem bzw. metakognitivem Wissen einerseits und der Anwendung des prozeduralen Wissens anhand verschiedener kognitiver Prozesse des Erinnerns, Verstehens, Anwendens, Analysierens, Bewertens und Kreierens andererseits. Die zu bearbeitenden Aufgaben sind in der Linguistik, Literaturwissenschaft und Fachdidaktik angesiedelt.

Keinem der genannten Testkonstrukte gelingt es, textbezogene pragmatische, hermeneutisch und nomologisch inferierte ‚Tiefen‘-Kenntnisse entsprechend des sozialtheoretischen „state-of-the-art“ zu erfassen. Meist überwiegt das Testen deklarativer Wissensbestände im Rahmen textbezogener Sinngehalte zulasten prozeduralisierten Könnens, dessen Erfassung vergleichsweise aufwändigerer Auswertungsverfahren bedarf.⁴ Auch die empirisch fundierte Erkenntnis, das Leseverständnis müsse als übergreifendes Sprachverstehen begriffen werden (z.B. Catts & Kamhi 2005), hat bisher nicht zur Entwicklung entsprechend elaborierter Testverfahren geführt, die den durch Sprache konstruierten Bedeutungshorizonten und sozialen Realitäten gerecht würden.

Welche Probleme die skizzierten Testkonstrukte und deren Anwendung aufwerfen, sei am Beispiel einiger zentraler Kritikpunkte kurz zusammengefasst. Ich beziehe mich hierbei u.a. auf meine Sekundäranalysen der öffentlich zugänglichen PISA-, DESI- und TEDS-LT-Aufgaben.⁵ Zunächst werde ich die Ausdifferenzierung der

⁴ In seinen Sekundäranalysen der Aufgaben aus der *internationalen* PISA 2000-Studie gelangt Grzesik (2003: 149) zu ähnlichen Befunden: Die erfassten Lesekompetenzen reduzierten sich überwiegend auf die „Lokalisation von punktuell explizitem Textsinn, d.h. der Suche von Stellen in der sogenannten ‚Textoberfläche‘“, obwohl das zugrunde gelegte Testkonstrukt auch tiefergehende Verständnisebenen berücksichtigt. Im Kontrast dazu bleibt das in der *deutschen* PISA-Ergänzungsstudie herangezogene Testkonstrukt nach Kintsch und van Dijk *von vornherein* zu stark auf der Textoberfläche verhaftet (zur Kritik vgl. Bremerich-Vos/Wieler 2003: 16ff., Grzesik 2003: 157) und ist auch in seiner Weiterentwicklung zum episodischen Textmodell (vgl. Kintsch 1998) hinsichtlich sprachlich symbolisierter Bedeutungen und etwaiger Mehrdeutigkeiten ebenso wie mit Bezug auf verschiedene Interaktionsformen zwischen textspezifischen Repräsentationsformaten und dem Vor-/Wissen der Leser nicht hinreichend geklärt. Während die internationale Studie also ganz offensichtlich versucht, die durch das Testkonstrukt ermöglichte Komplexität in der Aufgabenstellung und Kompetenzmessung soweit wie möglich zu reduzieren, weist jenes der deutschen Ergänzungsstudie von vornherein stark vereinfachende Merkmale auf. Trotz der unterschiedlichen Zugangsweisen zeichnet sich die Konstruktion der bislang öffentlich einschubaren Testaufgaben-Items durch große Ähnlichkeiten aus.

⁵ Es handelt sich um öffentlich zugängliche Beispielaufgaben, die für die PISA-Studie 2000 über das Leibniz-Institut für die Pädagogik der Naturwissenschaften, Universität Kiel, eingesehen werden können: <http://pisa.ipn.uni-kiel.de>. Anlässlich der PISA-Studien 2003 und 2006 wurden keine weiteren Aufgaben zur gezielten Erfassung des Leseverstehens publiziert. Im Rahmen der internationalen PISA-Studie 2009 sind jedoch eine Reihe neuer Leseaufgaben online zugänglich gemacht worden (vgl. OECD 2009: 48ff.). Für das DESI-Projekt wurden die Publikationen von Beck & Klieme (2007), Nold & Willenberg (2007) und Willenberg (2006, 2007a-c) herangezogen. Einige der in TEDS-LT eingesetzten Aufgaben konnten in einer Publikation erster Forschungsergebnisse eingesehen werden (Bremerich-Vos et al. 2011, Roters et al. 2011).

Anforderungsstrukturen diskutieren und danach offene Fragen zur Testfairness und Normativität der Testkonstrukte formulieren.

2.3.1 Kritik an der Untergliederung der Anforderungsstrukturen

- *Keine Unterscheidung zwischen Ausbildungs- und Bildungskompetenzen:* In den PISA- und DESI-Studien werden Lesekompetenzen mehr oder weniger willkürlich sowohl den Ausbildungs- als auch Bildungskompetenzen zugerechnet. Notwendig ist jedoch eine systematische Unterscheidung beider Kompetenzprofile, um die Qualität der jeweils getesteten Fähigkeiten näher bestimmen und individuell anschlussfähige verwendungstheoretische Konsequenzen aus den Leseergebnissen ziehen zu können. In TEDS-LT bleibt eine bildungstheoretische Rahmung weitgehend ausgespart, wenngleich das zugrunde gelegte Modell von Anderson & Krathwohl (2001: 12ff.) sehr wohl an ein von Ralph Tyler (1949) ausgearbeitetes Bildungstheorem angelehnt ist, das gerade im Zusammenhang mit prozeduralem Wissen und dessen Anwendung ein Bewusstsein für Bildungspotentiale erkennen lässt. Diese bedürfen allerdings einer weiteren Ausdifferenzierung (s.u.): Denn Aufgaben, die zum Nachweis von „educational objectives“ wie „The ability to interpret various types of social data“ dienen (Anderson/Krathwohl 2001: 16), verlangen den Probanden möglicherweise hohe Transferleistungen zwischen verschiedenen Wissensdomänen und Handlungsfeldern ab. Ohne eine sowohl ausbildungs- als auch bildungstheoretische Fundierung des herangezogenen Kompetenzmodells würden sich die Antworten der Probanden weder angemessen würdigen noch hinreichend fair bewerten lassen.
- *Mangelnde analytische Trennung der Subskalen:* Das für die PISA 2000-Studie herangezogene und 10 Jahre später nur leicht veränderte Lesetestkonstrukt basiert auf einer unzureichenden Ausdifferenzierung zwischen den – zudem theoretisch nicht fundierten – Subskalen „Informationen ermitteln“, „textbezogenes Interpretieren“, „Reflektieren und Bewerten“, sowie zwischen den jeweils zugeordneten Dimensionen für die unterschiedlichen Kompetenzstufen (Artelt et al. 2001: 86ff.). Warum nicht auch für das Ermitteln von Informationen und für das textbezogene Interpretieren Reflexionen erforderlich sein sollen, bleibt unklar (zur Kritik s. auch Rost & Schilling 2006: 455f.). Obwohl das Lesetestkonstrukt der PISA-Studie 2003 diese Skalierung aufhebt, gänzlich auf den Reflexionsbegriff verzichtet und stattdessen nur noch eine textsortenspezifische Unterscheidung in kontinuierliche (z.B. Erzählungen, Berichte, Kommentare) und nicht-kontinuierliche Texte vorsieht (z.B. Graphiken, Tabellen, Bilder), die mit jeweils fünf Kompetenzstufen verknüpft werden (Schaffner et al. 2004a: 96), ist im Hinblick auf Grenzziehungen zwischen den einzelnen Stufen nicht viel gewonnen: Sind doch viele der gewählten Begriffe unpräzise und die angeblichen Unterschiede zwischen den Stufen ‚Zwei‘ und ‚Fünf‘ kaum erkennbar (vgl. zur Kritik auch Bremerich-Vos/Grotjahn 2007: 158f.). Anderson und Krathwohl (2001: 89) gehen davon aus, dass – je nach Aufgaben-

stellung – die einzeln beschriebenen kognitiven Prozesse eng miteinander verknüpft sind. Dabei kommt es – wie mit Bezug auf das Verstehen und Erklären deutlich wird (s.u.) – auch zu Überschneidungen, die den Nachweis der jeweils vollzogenen Kompetenzen deutlich erschweren und ein gravierendes Defizit des Kompetenzmodells darstellen. Weitere kompetenzverzerrende Überschneidungen ergeben sich v.a. zwischen dem Verstehen, Analysieren, Bewerten und Kreieren. Beim Kreieren (a.a.O.: 84f.) handelt es sich – im Anschluss an den sozialtheoretischen ‚State-of-the-Art‘ – um nichts Anderes als um ein Verstehendes Erklären – mit dem Unterschied, dass ein Verstehendes Erklären auch ohne eine Klärung des Normativitätsproblems („educators must define what is original or unique.“ a.a.O.: 85) auskommt. Die in einer früheren Fassung der Taxonomie benutzte Basiskategorie des Synthetisierens ist gleichermaßen von solch einem normativen Ballast befreit und erscheint daher für die Abbildung von Kompetenzen besser geeignet als „Kreieren“ (a.a.O.: 308). Geht es doch bei dem für ein Verstehendes Erklären erforderlichen Synthetisieren darum, soweit wie möglich evidenzbezogene, plausible Informationsverknüpfungen zu vollziehen, die umso schwerer erscheinen, je mehr Vorkenntnisse aktiviert und vom Ausgangstext weiter entfernte, distale Inferenzen generiert werden müssen. Gänzlich unhaltbar, da vielfach verquickt mit anderen kognitiven Basisprozessen, erscheint schließlich das dem Analysieren untergeordnete „attributing“; es soll sich auf die Re-/ Konstruktion akteursbezogener Situationsdefinitionen beziehen: „Attributing occurs when a student is able to ascertain the point of view biases, values, or intention underlying communications“ (a.a.O.: 82). Sozialtheoretisch gewendet, sind solche Framings immanenter Bestandteil des Verstehens bzw. Interpretierens, weshalb für diese keine eigene Kategorie benötigt wird. Schauen wir uns die Nutzung der Kategorie des Analysierens in TEDS-LT an, irritiert zudem, dass sie offenbar mit Aufgaben in Verbindung gebracht wird, die kaum dem eigentlichen Anwendungszweck (vgl. Anderson/Krathwohl 2001: 79ff.) dienen – etwa, wenn es lediglich darum geht, die betonten Silben einzelner Begriffe zu kennzeichnen (Bremerich-Vos et al. 2011: 65).

- *Fehlende Unterscheidung zwischen (Lese-)Verstehen und Erklären:* Die Testkonstrukte und Aufgaben-Items unterscheiden nicht zwischen (Lese-) Kompetenzen des Verstehens und Erklärens (als eigenständiger oder übergeordneter Basiskategorie), obwohl diese im Falle akteursbezogener Handlungen und Wirkungen stets miteinander verknüpft sind und im Kontext nomologischer, z.B. physikalischer Zusammenhänge – wie bei der Berechnung der Erdumdrehung um die Sonne – einer prinzipiell anderen Wissens- und Kompetenzlogik folgen als das akteurs- und handlungsbezogene Verstehen. Aus diesem Grunde lassen die Testergebnisse auch keine fundierten Aussagen über die jeweils ermittelten Kompetenzstände im Verstehen *und* Erklären zu.

Das Kompetenzmodell von Anderson und Krathwohl bemüht sich dagegen um eine größere Ausdifferenzierung des Verstehensbegriffs. Diesem werden die kognitiven Prozesse „Interpretieren“, „Exemplifizieren“, „Klassifizieren“, „Zu-

sammenfassen“, „Inferieren“, „Vergleichen“ und „Erklären“ subsumiert (Anderson/Krathwohl 2001: 67, 70ff.). Das Erklären wird dem Verstehen untergeordnet, obwohl aus sozialtheoretischer Sicht ein Differenzverhältnis, wenn nicht gar eine umgedrehte Hierarchisierung angenommen werden müsste.

- *Vermischung textspezifischer Sinnebenen:* Die mangelnde Trennschärfe von Subskalen und Stufen findet ihre Fortsetzung in einer Vermischung textimmanenter, ‚deutend verstehender‘ und z.T. ‚verstehend erklärender‘ Sinnebenen. Einige der von Anderson und Krathwohl (2001) angeführten Beispiele zeigen, dass es ihnen um das verstehende Erklären von Handlungs-/Struktureffekten geht, die par excellence mithilfe sozialtheoretischen Know-Hows erläutert werden könnten. Anstelle des Erklärens als eigenständiger oder gleichwertig vernetzter Basis-kategorie subsumieren die Autoren ihre Aufgabenbeispiele jedoch dem „Verstehen“ (z.B. „Explain the causes of important 18th-century events in France“, a.a.O.: 67; „Why does air enter a bicycle tire pump when you pull up on the handle?“, a.a.O.: 76; „explain the consequences of the Parliamentary Acts for different colonial groups“, a.a.O.: 172). Eine vergleichsweise größere Komplexität des Erklärens zeigt sich auch in dem Umstand, dass die dem Verstehen zugeordneten kognitiven Prozesse des Interpretierens, Exemplifizierens (z.B. „locate an inorganic compound on a field trip and tell *why* it is inorganic (i.e., specify the defining features)“, a.a.O.: 72, Herv. v. U.D.), Inferierens, Zusammenfassens und Vergleichens ebenfalls Erklärungen enthalten können (vgl. a.a.O.: 70ff.).

Kritikwürdig erscheinen auch Zuordnungen von Testfragen und Antworten, wie sie mithilfe des DESI-Testkonstrukts erfolgt sind: Auf die im Textbeispiel „Quastenflosser“ gestellte Testfrage: „Wovon handelt der Text?“ lautet die Antwort: „Nachweis einer zoologischen Sensation“. Sie wird dem Niveau A „Identifikation sinntragender Wörter ...“ subsumiert (Willenberg 2007a: 112f.). Während die „zoologische Sensation“ durchaus als lokales Textthema firmieren kann, entspricht der „Nachweis ...“ jedoch dem globalen Textthema, weshalb die Antwort eher in den Bereich des Kompetenzniveaus C fällt: „Verknüpfende Lektüre, d.h. die Verbindung auseinander liegender Textstellen und die Öffnung von allgemeinem Wissen bzw. von Textwissen“ (a.a.O.: 110).

- *Ungeklärte Reichweiten verschiedener Inferenzarten:* Die unzureichende Ausdifferenzierung hermeneutischer und nomologischer Verknüpfungsformen sowie die fließenden Übergänge zwischen textimmanent gestützten, lokalen Inferenzen (vgl. Halldorson & Singer 2002) und einer darüber hinausgehenden globalen Wissensaktivierung führen zu Ambivalenzen in den Aufgaben-Konstruktionen. So lassen sich die von den DESI-Autoren mit Bezug auf einzelne Aufgaben ausgewiesenen Inferenzen teilweise nur aufgrund weiterreichender Verknüpfungen i.S. eines ‚deutenden Verstehens‘ oder ‚verstehenden Erklärens‘ herstellen. Diesbezügliche Ungenauigkeiten zeigen sich bspw. an dem von Willenberg (2006: 294) erläuterten Aufgaben-Item „Der Taugenichts“ von Joseph von Eichendorff (1822/23). Zum besseren Verständnis folgt ein Gesprächsauszug

zwischen dem „Taugenichts“ und der vermeintlichen Gräfin, die er heiraten möchte:

„Aber was nennst du mich denn Gräfin?“ Ich sah sie groß an. – „Ich bin ja gar keine Gräfin“, fuhr sie fort, „unsere gnädige Gräfin hat mich nur zu sich aufs Schloß genommen, da mich mein Onkel, der Portier, als kleines Kind und arme Waise mit hierherbrachte.“ Nun war’s mir doch nicht anders, als wenn mir ein Stein vom Herzen fiel.

Die kognitive Leseanforderung konzentriert sich hier v.a. auf den letzten Satz, wonach der individuelle Effekt emotionaler Erleichterung auf seine Ursachen zurückgeführt werden muss. Die dazu erforderlichen Lesekompetenzen bestehen aus einer Aktivierung oder ggf. aus dem Inferieren kontextspezifischen historischen Hintergrundwissens und aus einer Verknüpfung dieses Kontextwissens mit der Situation im Text. Der von den DESI-Autoren angebotene Vorschlag lautet: „Eine Ständegesellschaft würde eine Beziehung zwischen einer Gräfin und dem Taugenichts (Müllerssohn) sehr erschweren“ (Willenberg 2006: 294). Dabei handelt es sich um nichts anderes als um den Versuch, verstehend erklärende Inferenzen zu tätigen, die von den DESI-Autoren aber lediglich unter „Allgemeinwissen“ verbucht werden (ebd.).

In dem Maße wie „Informationen“ (vgl. OECD 2009: 34ff.), Wissen und kognitive Basisprozesse in Kompetenzmodellierungen hinsichtlich ihrer situations-, handlungs- oder wirkungsbezogenen Referenzen ungeklärt bleiben, lassen sich die jeweils performierten Kompetenzen nicht hinreichend voneinander unterscheiden, bleiben Teilkompetenzen verdeckt und werden die von den Probanden erbrachten Leistungen weniger fair beurteilt.

2.3.2 Offene Fragen zur Testfairness und Normativität

- *Unvereinbarkeiten zwischen input-basiertem Testkonstrukt und output-basierter Kompetenzmessung:* Es besteht ein Widerspruch zwischen der von Schülerinnen und Schülern bei der Aufgabenbearbeitung erwünschten Konstruktion *eigener* Konzepte und deren Nutzung für Kompetenztests (vgl. auch die Kritik von Karg 2005). Die Erfassung leserspezifischer Sinndeutungen unter input- und through-put-basierten Prämissen, die u.a. mit Lernprinzipien der konstruktivistischen Spracherwerbsforschung korrelieren (z.B. Swain 2001: 281), kann nicht zugleich als output-basiertes Testkonstrukt fungieren, das die Mechanismen und Bestandteile der Sinndeutungen für die Diagnose vermeintlich objektiv messbarer Lesekompetenzen nutzt. Solange dieser Widerspruch nicht aufgelöst ist, wird das für pädagogisch psychologische Forschung maßgebliche Gütekriterium der Fairness gegenüber Testpersonen verletzt. Aus erziehungs- und sozialwissenschaftlicher Perspektive liegt hier das Problem einer nicht hinreichend geklärten Normativitätsfolie und ihrer Kriterien vor. Es steht in engem Zusammenhang mit den bereits vor einem Vierteljahrhundert diskutierten Differenzen zwischen psycholinguistischen und logischen Inferenzen (z.B. Rickheit et al. 1985: 5ff.).

- *Begriffliche bzw. kategoriale Unschärfen in den Aufgaben-Items:* Die öffentlich zugänglichen Aufgabenformate enthalten eine Reihe semantisch-pragmatischer Ambiguitäten zwischen den in der Aufgabenstellung gewählten Begriffen und den eigentlichen Textinhalten sowie Zeichnungen (vgl. auch Karg 2005). So hat es z.T. den Anschein, dass nicht die inhalts- und formbezogenen Lesekompetenzen der Schüler, sondern ihre *Test*-Kompetenzen getestet werden: Sind die Textinterpreten in der Lage zu ergründen, worauf die Testautoren hinauswollen (vgl. „testing to the test“)?

In TEDS-LT (Fach Deutsch) fällt zudem auf, dass auch andere Antworten als die jeweils anerkannten für richtig hätten befunden werden können: Ein Textausschnitt über das Phänomen, dass „ein so großer Teil der Menschen ... zeit lebens unmündig bleiben“ [sic], soll einer Epoche zugeordnet werden (Bremrich-Vos et al. 2011: 63). Neben der als richtig anerkannten „Aufklärung“ (ebd.) hätte hier auch die „Moderne“ akzeptiert werden sollen, steht doch die Aufklärung am Beginn der Moderne. In einem anderen Item erscheint die als richtig anerkannte Feststellung: „Eine Lesekompetenz muss in Stufen aufgebaut sein, bei der die höhere die niedrigere umfasst“ problematisch. Nur am Rande sei erwähnt, dass die mit dem Item intendierte „Erfassung literaturdidaktischen Wissens“ durchaus auch in den Bereich der (Angewandten) Linguistik fällt. Für weitaus bedeutsamer halte ich aber den Einwand, dass die – u.a. im Rahmen der PISA-Lesestudien vertretene – Auffassung, dass hierarchiehöhere Kompetenzstufen quasi automatisch hierarchieniedrige Stufen umfassen (vgl. Naumann et al. 2010: 26), nicht unbedingt richtig sein muss. Entscheidend ist, auf welches Kompetenzmodell Bezug genommen wird und was mit „höher“ und „niedriger“ gemeint ist. So hat sich z.B. im Marburger Forschungsprojekt gezeigt, dass Probanden dazu in der Lage sind, auf eine ‚höhere‘ Stufe i.S. einer komplexeren Informationsverarbeitung durchzusteigen, aber nicht quasi automatisch auch die auf einer ‚niedrigeren‘ Stufe verorteten kognitiven Prozesse der Informationssuche und -entnahme zu beherrschen scheinen (vgl. Kap. III.2, Hansmann/Baumbach 2012). Ähnliche Vorbehalte ergeben sich aus der Lesekompetenzforschung: Das gemeinhin als höhere Informationsverarbeitung gewertete Inferieren und Transferieren von Vorwissen bei der Herausbildung eines Text- oder Situationsmodells schließt nicht automatisch alle ‚hierarchieniedrigen‘ Prozesse der „more automated linguistic processes“ mit ein, wie „lexical access, syntactic parsing, semantic proposition formation“ (Grabe/Stoller 2011: 15, 32). Auch erscheinen Annahmen der Art, dass eine als höher kategorisierte Verarbeitungsform zwangsläufig schwieriger sei als eine wenig hoch angesiedelte, in keiner Weise gerechtfertigt (a.a.O.: 15). Es ist wohl vielmehr von individuell ganz unterschiedlichen „bottom-up“- und „top-down“-Informationsverarbeitungsprozessen auszugehen (a.a.O.: 26f., 32).

Unklarheiten in den Items sind mit dem Testgütekriterium einer fairen Aufgabenstellung nicht vereinbar.

- *Zweifel an einer methodisch kontrollierten Kodierung offener Aufgaben-Items:* In Bezug auf die Subskala „Reflektieren und Bewerten“ kann insbesondere im Falle offener

Fragen bezweifelt werden, ob sich diesbezügliche Antworten überhaupt methodisch kontrolliert kodieren lassen; berücksichtigen doch die für eine Kodierung definierten Kriterien niemals den gesamten Spielraum potenziell relevanter Reflexionen oder Bewertungen. Angesichts der von Schülern verlangten eigenständigen Denkleistungen, die auf unmittelbar textbezogenen (Präsuppositionen, Implikationen) und textübergreifenden Inferenzen (Brückenhypothesen des ‚deutenden Verstehens‘, Transformationsregeln des ‚verstehenden Erklärens‘) beruhen und mithilfe ihres Vorwissens oder aufgrund der bereitgestellten Informationen erbracht werden sollen, erscheint es unangemessen, die Antworten zu den Gruppen „vollständige Lösung“, „akzeptierte Teillösung“ und „nicht gelöste Aufgabe“ zusammen zu fassen: Impliziert der Begriff „Lösung“ doch eine mehr oder weniger lineare Beantwortungsmöglichkeit mit ontologischen, konkret definierbaren Elementen. Und selbst im Falle einer anderen Begriffswahl – im Englischen werden z.B. die Begriffe „full credit“, „partial credit“ und „no credit“ benutzt –, ändert dies nichts am Problem, dass eine angemessene Kodierung offener Aufgaben-Items ungeklärt ist. Damit auch multikontingente Denkopoperationen hinreichend berücksichtigt werden, müssten v.a. die Anforderungsstrukturen zur Qualität und Komplexität unterschiedlicher Reflexions- und Bewertungsniveaus stärker abstrahiert werden. Diese Kritik bezieht sich nicht nur auf die Bearbeitung literarischer Texte, sondern auch auf die Bearbeitung nicht-fiktionaler Texte (s. z.B. die PISA 2000-Aufgabe zur Grippeimpfung⁶ oder die PISA 2009-Aufgabe „Student Opinions“, OECD 2009: 54ff.). Durch eine mangelnde theoretische Fundierung der normativen Fixierungen entsteht unter Umständen ein verzerrtes Bild der Schülerkompetenzen. Dass sich in den „Large-Scale Assessment“-Tests nur sehr wenige offene Aufgaben-Items befinden, mag mit den ungeklärten Normativitätsfragen und dem Arbeitsaufwand für eine angemessene Auswertung zusammenhängen.⁷ Allerdings steht die Ausklammerung offener Aufgaben-Items – meist zugunsten konkreter Referenzfragen und Multiple-Choice-Tests (zu TEDS-LT s. Bremerich-Vos et al. 2011: 55ff., Roters et al. 2011: 85) – im Gegensatz zu Genreanforderungen in Alltag, Schule, Studium und Beruf: In Zeiten der fortgeschrit-

⁶ Das Item bezieht sich auf das folgende Zitat aus einem Informationsblatt zur Grippe-schutzimpfung: „Wer sollte sich impfen lassen? Jeder, der sich vor dem Virus schützen möchte.“ Im Anschluss wird auf die Kritik von einem Betriebsmitarbeiter verwiesen, der meinte, man hätte „den Satz ‚Jeder, der sich vor dem Virus schützen möchte‘ weglassen sollen, weil er irreführend sei.“ Die zu bearbeitende offene Frage lautet: „Bist du auch der Meinung, dass dieser Satz irreführend ist und hätte weggelassen werden sollen? Begründe deine Antwort.“ (OECD 2000: 8)

⁷ Ein weiteres Beispiel für die Vermeidung offener Fragen findet sich in den weltweit eingesetzten IELTS-Tests (International English Language Testing System). So berichten Khalifa & Weir (2009: 55) von der gängigen Praxis, dass „item writers are advised to focus Short Answer Questions on factual information as otherwise the items end up too open to be workable. This also applies to other constructed response IELTS reading tasks such as completing gaps in notes, tables, flow-charts, sentences, or summaries using words from the reading text.“

tenen Moderne wird von den Adressaten zunehmend eigenständiges Formulieren und Kombinieren erwartet.

- *Verfehlung bildungsspezifischer Bedeutungsdimensionen/Sexismen*: Die Aufgaben-Items zielen allzu oft auf die Erschließung textimmanenter – semantisch-lexikalischer und grammatikalisch-formaler – sowie expressiver Sinndimensionen, vernachlässigen jedoch die ‚verstehend erklärenden‘ Sinnhorizonte (vgl. Kap. III.3). Eine zentrale Kategorie des Bildungswissens bleibt somit ausgeklammert. Andernfalls hätten die Testautoren möglicherweise die in ihren Aufgaben-Items „Amanda und die Herzogin“ (PISA 2000) und „Großmutter“ (DESI, zu den Testfragen s. Willenberg 2007a: 114ff.) enthaltenen Sexismen nicht billigend in Kauf genommen und in DESI gar als „lustige Stellen“ für eine Aufgabe beschrieben (a.a.O.: 114). Vielmehr hätte man erwarten können, dass solche Propositionsgehalte zum Anlass für emanzipatorische Überlegungen bspw. bezüglich der Gleichstellung der Geschlechter früher und heute genommen werden. Ein Exzerpt aus dem Basistext „Amanda und die Herzogin“ und eine der daran anknüpfenden Aufgaben mögen die Kritik verdeutlichen:

Zusammenfassung: ... Die Herzogin will, dass Amanda [Verkäuferin, U.D.] ihr hilft, den Prinzen [Sohn der Herzogin, U.D.] von seinen quälenden Erinnerungen [Tod seiner Frau, die Amanda ähnlich sieht, U.D.] zu befreien. (OECD 2000: 37)

Zu diesem Zweck arrangiert die Herzogin eine Begegnung zwischen Amanda und dem Prinzen, die in der folgenden Multiple Choice-Aufgabe als „Trick“ bezeichnet wird:

Wovon handelt dieser Auszug?

Die Herzogin überlegt sich einen Trick,

A: um den Prinzen dazu zu bringen, sie öfter zu besuchen.

B: um den Prinzen dazu zu bringen, sich endlich zum Heiraten zu entschließen.

C: um Amanda dazu zu bringen, den Prinzen seinen Schmerz vergessen zu lassen.

D: um Amanda dazu zu bringen, mit ihr zusammen im Schloss zu wohnen. (OECD 2000: 39)

Die Antwort C wurde als richtig gewertet. Bei dem von den Schülern und Schülerinnen zu leistenden textbezogenen Interpretieren (vgl. PISA-Lesetestkonstrukt) handelt es sich um ein Deutendes Verstehen, d.h. um die Erfassung einer akteursspezifischen Deutung (Framing der Herzogin), die den Handlungen der Herzogin zugrunde liegt. Die Aufgabe soll wohl zum Einen an Alltagswissen über die Fürsorge von Müttern für das Wohlergehen ihrer Söhne anschließen, wengleich es sich hier doch ‚nur‘ um die Tante handelt (vgl. Antworten B, C). Zum Anderen setzen die Antworten A und D Alltagswissen über die Situation von Müttern voraus, deren Kinder das Haus verlassen haben und die alles dafür tun würden, um ihre Kinder wieder an sich zu binden und in ihrer Nähe zu haben. Diese Präsupposition beruht ebenfalls auf einer Erweiterung

stereotyper Familienrollen auf eine Verwandte. Stellt sich die Frage: Inwieweit gehört solch ein Wissen zu den Wissensressourcen von 15-jährigen? Und selbst für den Fall, dass sie über entsprechendes Erfahrungswissen verfügen: Kann davon ausgegangen werden, dass die Probanden überhaupt dazu in der Lage wären, ein ggf. aus Familienbeziehungen vertrautes Wissen nicht nur auf weiter entfernte Verwandte, sondern auch auf einen anderen Kontext – wie jenem im Textbeispiel bzw. in der Aufgabe – zu transferieren? Denn Akteure wie eine Herzogin und ein Prinz oder das erwähnte Schloss verweisen auf einen nicht-alltäglichen Kontext und damit auf eine neue Wissensdomäne. Und dass sich ein Wissenstransfer zwischen verschiedenen Domänen quasi automatisch vollzieht, wäre eher die Ausnahme als der Regelfall (vgl. Mähler/Stern 2006). Nun könnte man argumentieren, dass die Kontextferne des Textbeispiels durch den Einbau vertrauter Handlungsorte und Akteure – wie ein Geschäft mit einer Verkäuferin – oder vertrauter Fortbewegungsformen – wie das Fahrradfahren des Prinzen – kompensiert wird. Dem ist entgegenzuhalten, dass frame- und script-atypische Elemente einer Wissensdomäne das Leseverständnis auch irritieren, wenn nicht erschweren können (vgl. Interferenzen): Warum erscheint der Prinz z.B. nicht hoch zu Roß? Warum basiert keine der Multiple-Choice-Optionen auf dem für monarchistisch strukturierte Gesellschaften typischen Handlungsmuster „... damit der Prinz wieder heiratet und einen Thronfolger zeugt“?

Zur Vermeidung der hier nur angedeuteten möglichen Verständnisschwierigkeiten wäre es wohl gegenüber den Probanden fairer und für die Erfassung eines bildungstheoretisch fundierten Leseverständnisses aufschlussreicher gewesen, danach zu fragen, auf welche Annahmen sich Idee und Handlungsplan der Herzogin stützen und wie sich die Annahmen begründen lassen. Denkbar wäre zudem eine anschließende Kontrastierung der – wie auch immer substantiierten – Situationsdefinition der Herzogin mit eigenem Erfahrungs- und Weltwissen. In diesem Zusammenhang hätten die Probanden dann auch Gelegenheit erhalten können, auf die Möglichkeit gegenteiliger oder anderer transintentionaler Effekte zu verweisen. Auch hätte es aus bildungstheoretischer Perspektive nahe gelegen, gerade die frame- und script-spezifischen Interferenzen zwischen Habitusmerkmalen einer feudalen Ständegesellschaft und modernen demokratischen Gesellschaftsformen zu thematisieren. Beide Veränderungsvorschläge hätten jedoch die prinzipiell geschlossene Multiple-Choice-Aufgabe in eine offene Aufgabe verwandelt, die sich vergleichsweise schwieriger auswerten lässt. Schließlich richtet sich meine Hauptkritik an der Testaufgabe auf die sexistische Rollenverteilung: Die Verkäuferin Amanda wird hier zur Simulation einer angeblich authentischen Lebensfigur i.S. eines psychischen Hilfsangebotes für den Prinzen instrumentalisiert. Das „hidden curriculum“ der Aufgabe kann zur folgenden symbolischen Wissensordnung verdichtet werden: FRAUEN WERDEN FREMDEN MÄNNERN ALS TROSTPFLASTER ANGEDIENT oder ggf. FRAUEN WERDEN ALS TROSTPFLASTER AN FREMDE MÄNNER VERKUPPELT. Dass derart stereotype Rollenklischees in Testaufgaben selbst

in der fortgeschrittenen Moderne unhinterfragt ‚um die Welt gehen‘, erscheint in hohem Maße anstößig.

Ein anderes potentiell sexistisches Beispiel findet sich in TEDS-LT. Um linguistisches Wissen mit Bezug auf das Erkennen einer Präsupposition im Lehramtsfach Deutsch zu testen, wurde die folgende Aufgabe konstruiert:

Staatsanwalt zum Angeklagten: „Haben Sie aufgehört, Ihre Frau zu schlagen? Antworten Sie nur mit ‚Ja‘ oder ‚Nein‘.“ – Verteidiger: „Das ist unfair.“

Worauf bezieht sich der Verteidiger mit seinem Kommentar, linguistisch betrachtet?

- A. auf eine Entscheidungsfrage
- B. auf das Zeigen auf eine Person (Personendeixis)
- C. auf eine Präsupposition
- D. auf eine indirekte Aufforderung. (Bremerich-Vos et al. 2011: 66)

Das Item knüpft offensichtlich an bekannte Gerichtssituationen an: Frauen suchen Schutz vor Gewalt in der Ehe, verfügen aber nicht über hinreichend Beweismittel, um die Täter ihrer gerechten Strafe zu überführen, so dass diese straffrei ausgehen. Wenngleich es auch Einzelfälle der Vortäuschung von Gewalt gibt, fehlen in dem Item entsprechende Hinweise. Der Standard-Fall ist für die betroffenen Frauen meist sehr belastend und ggf. doppelt demütigend. Warum haben die Testautoren solch eine Aufgabe konstruiert? Um zu testen, ob Studierende des Lehramts für das Fach Deutsch in der Lage sind, das linguistische Phänomen einer Präsupposition zu identifizieren, steht ihnen ein Universum unterschiedlicher Diskurssphären zur Verfügung, aus denen sie problemlos nicht-sexistische Items hätten generieren können.

Interessant ist das Item aber auch aus einem anderen Grund: Die als richtig gewertete Lösung „Präsupposition“ wurde nur von 65% der Studierenden angekreuzt und deshalb für eher schwierig befunden (ebd.). Das Item soll pragmatisches Wissen mithilfe von „Verstehen/ Anwenden“ testen, was auch in Verbindung mit der konstatierten Schwierigkeit auf das Erfassen von Bildungswissen verweist. Um dieser kognitiven Anforderung gerecht zu werden, müsste allerdings vorab sichergestellt sein, dass den Studierenden bekannt ist, welches linguistische Phänomen überhaupt als Präsupposition identifiziert werden kann. Eine Definition hätte also gleich mitgeliefert werden müssen. Da dies nicht geschehen ist, lässt das schlechte Resultat lediglich Aussagen über fehlendes linguistisches Fakten- bzw. Ausbildungswissen zu. Nur am Rande sei darauf verwiesen, dass auch die Antwort A (am besten in Verbindung mit C) als richtig hätte bewertet werden können – und zwar unter der Prämisse, dass sich die in „Das“ enthaltene Anapher auf das Stellen einer – von Amts wegen autorisierten, aber aus Sicht des Verteidigers nicht gerechtfertigten, da unbegründeten (vgl. Präsupposition zum Schlagen der Frau) – Entscheidungsfrage bezieht.

Unter der Prämisse, dass Items nicht ein „hidden curriculum“ undemokratischer Realitätskonstruktionen und -erfahrungen von Geschlechterungleichheit perpetuieren sollen, erstaunt nicht nur die mangelnde Sensibilität der – meist männli-

chen – Testautoren für die Konstruktion gender-neutraler oder zumindest nicht-diskriminierender Items. Vielmehr erstaunt, dass es offenbar keine Kontrollmechanismen bei der Konstruktion und Auswahl von Test-Items gibt, die auf einer umfassenden Evaluation etwaiger ‚Fehl-‘/Bildungspotentiale einzelner Items beruhen.

- *Ungeklärter Bildungsauftrag des Fachs Englisch:* Wenn die DESI-Autoren von vornherein weniger komplexe Aufgaben für das Leseverständnis in englischer Sprache vorsehen als in deutscher Sprache (Nold & Willenberg 2007: 39f.), stellt sich die Frage, welche Berechtigung das Fach Englisch im Bildungskanon hat. Nach fast fünf Jahren Englischunterricht sollten die in der Studie befragten 15-jährigen Schülerinnen und Schüler in der Lage sein, auch den Nachweis tiefergehender Verständnisdimensionen zu erbringen.

2.3.3 Beispiele für einen sozialtheoretisch fundierten Modellierungsbedarf

Testkonstrukte wie jene der DESI- und PISA-Studien⁸ bilden die theoretische Ausgangsgrundlage für die Erfassung von Lesekompetenzen mithilfe von Aufgaben-Items. Aus Sicht der Testpersonen bzw. Schüler repräsentieren die jeweiligen Aufgaben dann spezifische Anforderungsstrukturen, die sich stets auf soziale Realitätsausschnitte beziehen. Daraus folgt, dass eine erfolgreiche Aufgabenbearbeitung umfassende Kenntnisse über die – für soziale Realitäten typischen – Eigenlogiken, ihren Strukturprinzipien und Wechselwirkungsmechanismen einschließt:

- Sei es, dass die an der DESI-Studie beteiligten 15-jährigen Schülerinnen und Schüler am Beispiel eines Textauszugs aus dem „Taugenichts“ von Joseph von Eichendorff mit der Frage konfrontiert werden, warum dem Taugenichts nach Auflösung des Rätsels über die Herkunft der Gräfin „ein Stein vom Herzen“ fällt (Willenberg 2006: 294). Der individuelle Effekt emotionaler Erleichterung soll also von den Testpersonen auf seine Ursachen zurückgeführt werden. Dazu bedarf es der Aktivierung kontextspezifischen historischen Hintergrundwissens und einer Verknüpfung dieses Wissens mit der Situation im Text (vgl. Kap. 2.3.1). Dementsprechend lautet die ‚verstehend erklärende‘ Lösung der Testautoren: „Eine Ständegesellschaft würde eine Beziehung zwischen einer Gräfin und dem Taugenichts (Müllerssohn) sehr erschweren“ (ebd.).
- Sei es, dass die PISA-Aufgaben-Items „Erwerbstätige Bevölkerung“, „Grippe“ oder „Polizei“ umfassende Situationsbeschreibungen in Gestalt von Sachtexten beinhalten, die an Alltagssituationen anknüpfen. Die einzelnen Fragen beziehen sich überwiegend auf das in den Texten enthaltene fachliche Spezialwissen und

⁸ TEDS-LT bleibt hier ausgeklammert, da die Anzahl der bisher veröffentlichten Aufgaben zu gering ist (vgl. Blömeke et al. 2011). Besonders aufschlussreich in sozialtheoretischer Hinsicht könnten jedoch die von Roters et al. (2011: 93f.) lediglich erwähnten „Test-Aufgaben zur Syntax“ sein, zu deren Bearbeitung „auch semantisches und pragmatisches Wissen“ benötigt wird.

zielen auf Antworten im Makro-Genreformat von Beschreibungen und Erklärungen.

Um die Antworten der Beteiligten im Hinblick auf die jeweilige Situation, die Situationsdeutungen, Prozessverläufe und Wirkungen zu verobjektivieren, wird ein Testkonstrukt benötigt, das die mikro- und makrotheoretischen Zugänge zu textvermittelten sozialen Realitäten systematisch miteinander verbindet und die sozialwissenschaftlich begründeten, methodologisch und methodisch fundierten Normen transparent macht. Diesem Anspruch werden Testkonstrukte gerecht, die dem von Max Weber (1988/1922) und Alfred Schütz (1971, 1972) geprägten und von Hartmut Esser (1999-2001) weiter entwickelten ‚verstehenden Erklären‘ verpflichtet sind (vgl. Dirks 2006). Sie ermöglichen nicht nur Antworten auf die Frage, inwieweit Leserinnen und Leser zu inhaltlich-deskriptiven Aussagen gelangen, sondern auch dazu, ob sie in der Lage sind, tiefergehende Erkenntnisse aus Texten zu gewinnen. Dementsprechend interessiert bei der Diagnose des Leseverständnisses, inwieweit die jeweils Beteiligten herausfinden,

- über welche situationsspezifischen Bedingungen ein Text Auskunft gibt,
- welche ‚deutend verstehenden‘ Ausdruckspotentiale in den sprachlich-visuellen Gebilden enthalten und
- welche ‚verstehend erklärenden‘ Verknüpfungen angemessen sind sowie
- mithilfe welcher sprachlichen und bildlichen Ressourcen diesbezügliche Informationen strukturiert werden.

2.4 Zusammenfassende Kritik

Insgesamt zeichnet sich der Entwicklungsstand der PISA-, DESI- und TEDS-LT-Testkonstrukte und Aufgabenformate für die Diagnose von Lese-/Kompetenzen durch eine erkenntnishinderliche Abgrenzung disziplinspezifischer Diskurse aus: Favorisiert werden kognitionspsychologische Konzepte, während bildungs- und sozialwissenschaftliche Zugänge weitgehend ausgespart bleiben oder – wie im Falle des funktionalen Bildungsbegriffs des PISA-Konsortiums – nicht hinreichend mit dem Testkonstrukt und der Kompetenzmodellierung verbunden werden. Die Vernachlässigung text- und pragmalinguistischer sowie sozialwissenschaftlicher Expertise hat offenbar eine theoretisch wenig fundierte Ausdifferenzierung des Textbegriffs zur Folge, obwohl diese beim Leseverständnis eine zentrale Rolle spielen. ‚Gemessen‘ an den mit Verstehen und Erklären verbundenen situations-, handlungs- und aggregationslogischen Implikationen erweist sich die für Testzwecke konzipierte ‚theoretische‘ Struktur vom Textverständnis denn auch als wissenschaftstheoretisch, methodologisch und methodisch inkonsistent: Zum Einen fehlt eine systematische Trennung zwischen den Logiken des Verstehens und Erklärens, zum Anderen eine analytische Trennung von Ausgangs- und Endsituation und den dazwischen liegenden Praktiken bzw. Entwicklungsprozessen. Infolgedessen geraten die für Realitätsverzerrungen typischen, in öffentlichen Diskursen weit verbreiteten Verwechslungen von Situationsdeutungen der Akteure und dem eigentlichen Handlungsprozess sowie zwischen Ursache und Wirkung kaum in den Blick. Das

Erkennen, Vermeiden oder Bearbeiten diesbezüglicher Fehlschlüsse könnte jedoch als eine zentrale Lesekompetenz identifiziert werden, die – ganz i.S. der Kantianischen Aufklärungspostulate – Aussagen zur Verfügbarkeit vernunftbasierter Rationalitätsprinzipien und damit zur Bildungsqualität der ermittelten Testergebnisse ermöglicht.

Last but not least, sei auf das Problem mangelnder Vergleichbarkeit der Testergebnisse aus den bis dato vorliegenden Lesestudien verwiesen; beruhen diese doch auf teilweise stark differierenden Testkonstrukten (vgl. auch die Kritik bei Grzesik 2003: 154-159). Eine Hauptschwierigkeit für die Bereitstellung eines gemeinsamen Testkonstrukts liegt m.E. in der unzureichenden Klärung, welche Art von Informationen und Verknüpfungen bzw. Transfers von den Beteiligten geleistet werden sollen und welche Teilkompetenzen beim Lesen folglich benötigt werden. Entwicklungsbedarf besteht somit hinsichtlich der Bereitstellung einer zunächst theoretisch fundierten und sukzessive empirisch substantiierten Ausdifferenzierung von „levels of performance sufficiently generalized to have validity across assessments and groups“ (Kirsch 2001: 45).

Weitere Hindernisse für eine angemessene Würdigung von Lesekompetenzen offenbaren sich in fehlenden Antworten auf die folgenden Fragen: Wie erfolgt die Prozessierung von Textinhalten? Welche Art von Information muss von den Probanden identifiziert werden? Welche Informationsverknüpfungen müssen sie vollziehen? Inwieweit müssen sie Informationsverzerrungen korrigieren und Fehlschlüsse vermeiden? Inwiefern kann ihr Vorwissen dabei mehr oder weniger hilfreich sein? Um welche Art von Vorwissen handelt es sich jeweils? Sozialtheoretisches Know-How, das zur Klärung des Stellenwertes von Wissen und Informationen ebenso wie zu deren Identifikation und Verknüpfung einen entscheidenden Beitrag leisten kann, findet in den bekannten Lesekompetenzmodellen bisher keine Berücksichtigung.

Auch die neueren „multiplen Textverständnismodelle“ – wie das „Mental Representation Model of Multiple Texts Comprehension“ bzw. „Document Model“ (Perfetti et al. 1999), das „Model of Skill and Strategy Use“ (Pressley & Afflerbach 1995) bzw. „Constructively Responsive Reading Model“ (Afflerbach & Cho 2009) oder ein „Information Processing Model“ wie das „Multiple Document-TRACE-Model“ (Rouet & Britt 2011) – helfen bei der Suche nach den tieferen Bedeutungsverknüpfungen kaum weiter (vgl. Kim 2011). Zudem wurden bis heute keine bedeutungsorientierten Teilkompetenzen des Lesens theoretisch und/oder empirisch fundiert (vgl. summarisch Nold/Willenberg 2007: 25f.). Der „componential skills“-Ansatz trägt ebenfalls nicht zu einer bedeutungsorientierten Modellprofilierung bei, da hier ganz unterschiedliche, mehr oder weniger leserelevante Fähigkeiten erfasst werden (vgl. Grabe/Stoller 2011, Huey, 1908/1968: 6, Levy/Carr 1990), um z.B. fehlende Voraussetzungen für das Lesen an sich zu diagnostizieren (vgl. Strucker et al. 2007).

Vor diesem Hintergrund kann ein ‚verstehend erklärendes‘ Testkonstrukt – wie jenes pädagogisch soziologischer Diagnosekompetenz (vgl. Kap. II.5, Dirks/

Hansmann 2012) – als ein Vorschlag zur Bearbeitung der oben skizzierten Kritikpunkte betrachtet werden.

Literatur

- Afflerbach, Peter & Byeong-Young Cho (2009). Identifying and describing constructively responsive comprehension strategies in new and traditional forms of reading. In: Susan E. Israel & Gerald G. Duffy (Hrsg.), *Handbook of Research on Reading Comprehension* (S. 69-90). New York: Routledge.
- Anderson, Lorin W. & David R. Krathwohl (2001). *A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives. Complete Edition*. New York: Longman.
- Artelt, Cordula; Petra Stanat, Wolfgang Schneider & Ulrich Schiefele (2001). Lesekompetenz. Testkonzeption und Ergebnisse. In: Jürgen Baumert, Eckhart Klieme, Michael Neubrand, Manfred Prenzel, Ulrich Schiefele, Wolfgang Schneider, Petra Stanat, Jürgen Tillmann & Manfred Weiß (Hrsg.). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69-137). Opladen: Leske + Budrich.
- Beck, Bärbel & Eckhard Klieme (Hrsg.) (2007). *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)*. Weinheim: Beltz.
- Blanc, Nathalie; Panayiota Kendeou, Paul van den Broek & Denis Brouillet (2008). Updating situation models during reading of news reports: Evidence from empirical data and simulations. *Discourse Processes*, 45, 103-121.
- Blömeke, Sigrid; Albert Bremerich-Vos, Helga Haudeck, Gabriele Kaiser, Günter Nold, Knut Schwippert & Heiner Willenberg (Hrsg.) (2011), *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen – Erste Ergebnisse aus TEDS-LT*. Münster: Waxmann.
- Bremerich-Vos, Albert; Jutta Dämmer, Heiner Willenberg & Knut Schwippert (2011). Professionelles Wissen von Studierenden des Lehramts Deutsch. In: Blömeke et al. (Hrsg.), 47-76.
- Bremerich-Vos, Albert & Rüdiger Grotjahn (2007). Lesekompetenz und Sprachbewusstheit: Anmerkungen zu zwei aktuellen Debatten. In Beck & Klieme (Hrsg.), 158-177.
- Bremerich-Vos, Albert & Petra Wieler (2003). Zur Einführung. In: Ulf Abraham, Albert Bremerich-Vos, Volker Frederking & Petra Wieler (Hrsg.), *Deutschdidaktik und Deutschunterricht nach PISA* (S. 13-25). Fillibach: Freiburg i. B.
- Catts, Hugh W. & Alan G. Kamhi (Hrsg.) (2005). *The Connections Between Language and Reading Disabilities*. Mahwah, NJ: Erlbaum.

- Dijk, Teun A. van & Walter Kintsch (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.
- Dirks, Una (2006). Texte ‘verstehend erklären’: Ein bildungstheoretischer Beitrag zu einer sozialwissenschaftlich fundierten Textlinguistik. In: Carmen Spiegel & Rüdiger Vogt (Hrsg.), *Vom Nutzen der Textlinguistik für die Schule* (S. 33-49). Baltmannsweiler: Schneider-Hohengehren.
- Dirks, Una (2010). *Der Irak-Konflikt in den Medien. Eine sprach-, politik- und kommunikationswissenschaftliche Analyse*. Konstanz: uvk.
- Dirks, Una (2012). Aufgabenformate – Das Genre ‚Fallvignette‘. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), *Professionalisierung und Diagnosekompetenz. Kompetenzentwicklung und -förderung im Lehramtsstudium*. (Online-Publikation der Philipps-Universität Marburg, Kap. II.4).
- Dirks, Una & Wilfried Hansmann (2012). Pädagogisch soziologische Diagnosekompetenz im Modell. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), *Professionalisierung und Diagnosekompetenz. Kompetenzentwicklung und -förderung im Lehramtsstudium*. (Online-Publikation der Philipps-Universität Marburg, Kap. II.5).
- Drechsel, Barbara & Cordula Artelt (2007). Lesekompetenz. In: Manfred Prenzel, Cordula Artelt, Jürgen Baumert, Werner Blum, Marcus Hammann, Eckhard Klieme & Reinhard Pekrun (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 225-247). Münster: Waxmann.
- Dutke, Stephan; Christiane Baadte, Andrea Hähnel, Ulrich von Hecker & Mike Rinck (2010). Using Diagnostic Text Information to Constrain Situation Models. *Discourse Processes*, 47, 510-544.
- Esser, Hartmut (1999). *Soziologie. Spezielle Grundlagen Bd.1. Situationslogik und Handeln*. Frankfurt/ M.: Campus.
- Esser, Hartmut (2000). *Soziologie. Spezielle Grundlagen. Bd. 2: Die Konstruktion der Gesellschaft*. Frankfurt/M.: Campus.
- Esser, Hartmut (2001). *Soziologie. Spezielle Grundlagen. Bd. 6: Sinn und Kultur*. Frankfurt/Main: Campus.
- Esser, Hartmut (2002). Ehekrisen: Das (Re-)Framing der Ehe und der Anstieg der Scheidungsraten. *Zeitschrift für Soziologie, Vol. 31, H. 6*, 472-496.
- Grabe, William (1991). Current developments in second language reading research. *TESOL Quarterly* 25(3), 375-406.
- Grabe, William & Fredricka L. Stoller (2011). *Teaching and Research Reading. Second Edition*. Harlow, U.K.: Pearson Education Limited.
- Graesser, Arthur C.; Keith K. Millis & Rolf A. Zwaan (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163-189.

- Graesser, Arthur C.; Murray Singer & Tom Trabasso (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371-395.
- Grzesik, Jürgen (2003). Was testet der PISA-Test des Lesens? In: Ulf Abraham, Albert Bremerich-Vos, Volker Frederking & Petra Wieler (Hrsg.), *Deutschdidaktik und Deutschunterricht nach PISA* (S. 135-164). Fillibach: Freiburg i. B.
- Halldorson, Michael & Murray Singer (2002). Inference processes: Integrating relevant knowledge and text information. *Discourse Processes*, 34, 145-161.
- Hansmann, Wilfried & Hendrik Baumbach (2012). Entwicklung pädagogisch soziologischer Diagnosekompetenz von Lehramtsstudierenden. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), *Professionalisierung und Diagnosekompetenz. Kompetenzentwicklung und -förderung im Lehramtsstudium*. (Online-Schriftenreihe der Philipps-Universität Marburg, Kap. III.2).
- Hecker, Ulrich von & Stefan Dutke (2004). Integrative social perception: Individuals low in working memory benefit more from external representations. *Social Cognition*, 22, 336-365.
- Huey, Edmund B. (1908/1968). *The psychology and pedagogy of reading*. New York: Macmillan. (Neuaufgabe, Cambridge, MA: MIT Press, 1968).
- Karg, Ina (2005). *Mythos PISA. Vermeintliche Vergleichbarkeit und die Wirklichkeit eines Vergleichs*. Göttingen: V&R.
- Khalifa, Hanan & Cyril J. Weir (2009). *Examining Reading: Research and Practice in Assessing Second Language Reading*. New York: CUP.
- Kim, Jong-Yun (2011). A Comparison of Three Models of Multiple Text Comprehension. Poster presented at the 21st Annual Meeting of the Society of Text & Discourse, Université de Poitiers, July 11-13, 2011.
- Kintsch, Walter (1988). The role of knowledge in discourse comprehension. A construction-integration model. *Psychological Review*, 95, 163-182.
- Kintsch, Walter (1998). *Comprehension*. New York: CUP.
- Kirsch, Irwin (2001). *The International Adult Literacy Survey (IALS): Understanding what was measured*. Princeton, N.J.: Research Publications Office. [URL: <http://www.ets.org/Media/Research/pdf/RR-01-25-Kirsch.pdf>, Zugriff 19.04.2007]
- Kirsch, Irwin; John de Jong, Dominique Lafontaine, Joy McQueen, Juliette Mendelovits & Christian Monseur (2002). *Reading for Change: Performance and Engagement across Countries*. OECD. [URL: <http://www.oecd.org/dataoecd/43/54/33690904.pdf>, Zugriff 14.05.2012].
- Klieme, Eckhard; Johannes Hartig, Joachim Wirth (2005). Analytisches Problemlösen: Messansatz und Befunde zu Planungs- und Entscheidungsaufgaben. In:

- Eckhard Klieme, Detlev Leutner, Joachim Wirth (Hrsg.), Problemlösekompetenz von Schülerinnen und Schülern (S. 37-54). Wiesbaden: VS.
- Leutner, Detlev; Eckhard Klieme, Katja Meyer & Joachim Wirth (2004). Problemlösen. In PISA-Konsortium Deutschland (Hrsg.), PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs (S. 147-175). Münster: Waxmann.
- Levy, Betty Ann & Thomas H. Carr (1990). Component Process Analyses: Conclusions and Challenges. In: Thomas H. Carr & Betty Ann Levy (Hrsg.), Reading and Its Development. Component Skills Approaches (S. 423-438). San Diego, Ca.: Academic Press.
- Luckmann, Thomas (1986). Grundformen der gesellschaftlichen Vermittlung des Wissens: Kommunikative Gattungen. In: Friedhelm Neidhart, M. Rainer Lepsius & Johannes Weiß (Hrsg.). Kultur und Gesellschaft (S. 191-211, Kölner Zeitschrift für Soziologie und Sozialpsychologie: Sonderbd. 27). Opladen: Leske + Budrich.
- Mähler, Claudia & Elsbeth Stern (2006). Transfer. In: Detlef H. Rost (Hrsg.), Handwörterbuch Pädagogische Psychologie. (3., überarb. u. erw. Auflage, S. 782-793). Berlin: Beltz.
- Mani, Kannan & Philip N. Johnson-Laird (1982). The mental representation of spatial descriptions. *Memory and Cognition*, 10, 181-187.
- Naumann, Johannes; Cordula Artelt, Wolfgang Schneider & Petra Stanat (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In: Eckhard Klieme, Cordula Artelt, Johannes Hartig, Nina Jude, Olaf Köller, Manfred Prenzel, Wolfgang Schneider & Petra Stanat (Hrsg.). PISA 2009. Bilanz nach einem Jahrzehnt (S. 23-71). Münster: Waxmann.
- Nold, Günter M. (2003). DESI – a language assessment project in Germany and the pros and cons of large-scale testing. *Empirische Pädagogik*, 17(3), 368-379.
- Nold, Günter M. & Henning Rossa (2007). Leseverstehen. In Beck & Klieme (Hrsg.), 197-211.
- Nold, Günter M. & Heiner Willenberg (2007). Lesefähigkeit. In Beck & Klieme (Hrsg.), 23-41.
- OECD (1999). Measuring Student Knowledge and Skills: A New Framework for Assessment. Paris.
- OECD (2000). PISA 2000. Beispielaufgaben aus dem Lesekompetenztest. Ohne Ortsangabe.
- OECD (2004). Problem Solving for Tomorrow's World. First Measures of Cross-Curricular Competencies from PISA 2003. Paris: OECD.
- OECD (2009). PISA 2009. Assessment Framework. Key competencies in reading, mathematics and science. Paris: OECD.

- Oostendorp, Herre van; José Otero & Juan M. Campanario (2002). Conditions of updating during reading. In: Max Louwerse & Willie van Peer (Hrsg.). *Thematics: Interdisciplinary Studies*. Vol. 3 (S. 55-76). Amsterdam: Benjamins.
- Perfetti, Charles A.; Jean-Francois Rouett & M. Anne Britt (1999). Toward a theory of document representation. In: Herre van Oostendorp & Susan R. Goldman (Hrsg.). *The construction of mental representation during reading* (S. 99-122). Mahwah, NJ: Erlbaum.
- Pressley, Michael & Peter Afflerbach (1995). *Verbal reports of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum Associates.
- Rickheit, Gert & Hans Strohner (1999). Textverarbeitung: Von der Proposition zur Situation. In: Angela D. Friederici (Hrsg.). *Sprachrezeption* (S. 271-306; = *Enzyklopädie der Psychologie, Abteilung Sprache, Bd. 2*). Göttingen: Hogrefe.
- Roters, Bianca; Günter Nold, Helga Haudeck, Jörg-U. Keßler & Agnes Stancel-Piatak (2011). Professionelles Wissen von Studierenden des Lehramts Englisch. In: Blömeke (Hrsg.), 77-99.
- Rouett, Jean-Francois & M. Anne Britt (2011). Relevance Processes in Multiple Document Comprehension. In: Matthew T. McCrudden; Joseph P. Magliano & Gregory Schraw (Hrsg.), *Text Relevance and Learning from Texts* (S. 19-47). Greenwich, CT: Information Age Publishing.
- Rumelhart, David E. (1977): Toward an interactive model of reading. In Dornic, Stanislav (Hrsg.), *Attention and performance VI* (S. 573-603). Hillsdale, NJ: Erlbaum.
- Schaffner, Ellen; Ulrich Schiefele, Barbara Drechsel & Cordula Artelt (2004a). Lesekompetenz. In Manfred Prenzel, Jürgen Baumert, Werner Blum, Rainer Lehmann, Detlev Leutner, Michael Neubrand, Rainer Pekrun, Hans-Günter Rolff, Jürgen Rost & Ulrich Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 93-110). Münster: Waxmann.
- Schaffner, Ellen; Ulrich Schiefele & Wolfgang Schneider (2004b). Ein erweitertes Verständnis der Lesekompetenz: Die Ergebnisse des nationalen Ergänzungstests. In: Ulrich Schiefele, Cordula Artelt, Wolfgang Schneider & Petra Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (S. 197-241). Wiesbaden: VS.
- Schütz, Alfred (1971). *Gesammelte Aufsätze, Bd. 1: Das Problem der sozialen Wirklichkeit*. (Aus dem Amerikanischen übersetzt und mit einem Nachwort zur Übersetzung von Benita Luckmann und Richard Grathoff). Den Haag: Martinus Nijhoff.

- Schütz, Alfred (1972). Die soziale Welt und die Theorie der sozialen Handlung. In: Ders., *Gesammelte Aufsätze*, Bd. 2: Studien zur soziologischen Theorie (S. 3-21). Den Haag: Martinus Nijhoff.
- Strucker, John; Kentaro Yamamoto & Irwin Kirsch (2007). The Relationship of the Component Skills of Reading to IALS Performance: Tipping Points and Five Classes of Adult Literacy Learners. NCSALL Reports #29, National Center for the Study of Adult Learning and Literacy, Harvard Graduate School of Education, Cambridge, MA.
[URL: <http://www.eric.ed.gov/PDFS/ED495943.pdf>; Zugriff 14.05.2012].
- Swain, Merrill (2001). Examining dialogue: another approach to content specification and to validating inferences drawn from test scores. *Language Testing*, 18(3), 275-302.
- Tyler, Ralph W. (1949). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.
- Weber, Max (1988 [1922]). Über einige Kategorien der verstehenden Soziologie. In Ders., *Gesammelte Aufsätze zur Wissenschaftslehre* (7. Auflage, S. 427-474). Tübingen: UTB Fink.
- Willenberg, Heiner (2006). Vom Wort zum mentalen Modell. *PädForum*, 5, 293-296.
- Willenberg, Heiner (2007a). Lesen. In Beck & Klieme (Hrsg.), 107-117.
- Willenberg, Heiner (2007b). Argumentation. In: Beck & Klieme (Hrsg.), 118-129.
- Willenberg, Heiner (Hrsg.) (2007c). *Kompetenzhandbuch für den Deutschunterricht. Auf der empirischen Basis des DESI-Projekts*. Hohengehren: Schneider.
- Zwaan, Rolf A. (1994). Effect of genre expectation on text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 920-933.
- Zwaan, Rolf A. & Gabriel A. Radvansky (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185.

Zitation

Dirks, Una (2012). Eine Kritik der Lese-/Testkonstrukte in PISA, DESI und TEDS-LT aus sozialtheoretischer Perspektive [24 Seiten]. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), *Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium*. (Kap. V.2, Online-Publikation der Philipps-Universität Marburg).