

## 8 Prä- und Posttest-Kongruenzen – Gemeinsamkeiten und Differenzen in der fallspezifischen Anforderungsstruktur

### *Una Dirks*

In diesem Kapitel werden zunächst die Unterschiede zwischen den fallspezifischen Anforderungsstrukturen näher herausgearbeitet, um die von den Probanden erreichten diagnostischen Kompetenzleistungen besser würdigen zu können (s. Kap. 8.1). Dabei nehme ich sowohl auf die Fallvignetten als auch auf die Item-Pools Bezug. Auf dem Hintergrund der fallspezifischen Differenzen in den Anforderungsstrukturen präsentiere ich dann einige Prinzipien, die sich für die Erstellung einer sozialtheoretisch fundierten Bewertungsgrundlage für unterschiedliche Schwierigkeitsgrade der Fallvignetten eignen (s. Kap. 8.2). Der letzte Abschnitt dieses Kapitels befasst sich mit der Einordnung der sozialtheoretisch fundierten Indikatoren für Schwierigkeiten fallbasierter Diagnosen in die derzeitige Diskussion und schließt mit der Beschreibung verschiedener Kompetenzniveaus (s. Kap. 8.3).

### 8.1 Ein Vergleich der Anforderungsstrukturen von Prä- und Posttest

Die Anforderungsstrukturen im Einstiegs- und Abschlusskompetenztest entsprechen in dreierlei Hinsicht dem klassischen Prä-/Posttest-Design (vgl. Leslie & Caldwell 2009: 415): Beide Fallvignetten weisen eine ähnliche Genrestruktur in Form eines Transskriptes auf, auch ist die Aufgabenstellung dieselbe, und die Vignetten fallen in dieselbe Domäne, die hier aus einer Unterrichtssituation besteht. Ein deutlicher Unterschied ergibt sich aus der Textlänge: So müssen für die Diagnose des ca. zwei Seiten langen Abschlusstestes vergleichsweise mehr Informationen verarbeitet werden als in dem kurzen Einstiegstest. Da den Probanden jedoch eine verdreifachte Bearbeitungszeit gewährt und die etwa verdreifachte Item-Anzahl durchweg anteilig bewertet wurde (vgl. Kap. II.7, Dirks et al. 2012), ließ sich dieser Unterschied weitgehend kompensieren. Hinzu kommt, dass beide Item-Pools – entsprechend des Kompetenzmodells – identisch aufgebaut sind und die Ergebnisse der beiden Kompetenztests systematisch miteinander in Beziehung gesetzt werden konnten.

#### 8.1.1 Der Prätest: Der ‚Handy‘-Fall

Im ‚Handy‘-Fall sind lediglich drei Akteure am Interaktionsgeschehen beteiligt. Die Unterrichtssequenz mit vier beschriebenen bzw. enacteden Praktiken ist – im Kontrast zum üblichen Unterrichtsgeschehen mit 20 oder mehr Schülern – vergleichsweise kurz und übersichtlich. Nicht zuletzt aufgrund des für Studierende im Allgemeinen bekannten Settings in einer Schulklasse und der für sie ebenfalls mutmaßlich vertrauten Auseinandersetzungen mit Mitschülern kann die fallbasierte Aufgabenstellung „Bitte analysieren Sie schriftlich in Einzelarbeit den folgenden

Fall“ für prinzipiell durchführbar und insbesondere zu Beginn eines Seminars für angemessen gehalten werden.

Dem algorithmischen Step V (vgl. verstehendes Erklären), der die Ergebnisse der Diagnose widerspiegelt, wurden acht Items zugeordnet. Verfolgen wir deren Aggregation entlang der deduktiven Clusterungs-Verläufe (vgl. Kap. II.7.2, Dirks et al. 2012), beziehen sich die Items im Wesentlichen auf die Reflexion eines etwaigen Fehlschlusses der Lehrerin, das Handy gehöre – wie von Schüler B behauptet – Schüler A, und auf ihr damit verbundenes Konfliktlösungshandeln (vgl. Kap. II.7.2, Dirks et al. 2012). Die Item-Cluster berücksichtigen somit ein für das Lehrerman-dat typisches Spannungsfeld, das zwischen der geplanten bzw. für sinnvoll erachteten Fortsetzung des Unterrichts und der – auch mittel- und langfristig effektiven – Beilegung eines Konfliktes angesiedelt ist. Von den Probanden wird hier die Re-konstruktion möglicher lehrerspezifischer Framings mithilfe verschiedener Brücken-hypothesen verlangt, die auf schulspezifischem Erfahrungswissen gründen und Inferenzen über schülerbezogene Verhaltenszuschreibungen durch die Lehrerin berücksichtigen (vgl. algorithmischer Step IVb). Darüber hinaus sind die Proban-den gefordert, das Framing der beiden Schüler A und B hinsichtlich ihrer Hand-lungsziele zu diagnostizieren, wozu Erfahrungswissen aus Alltag und Schule über typische Identitätsbehauptungskämpfe zwischen Schülern ebenso wie über die für sie wichtigen Prestigeobjekte wie ein Handy benötigt wird (vgl. algorithmischer Step IVd). Die im algorithmischen Step V enthaltenen acht Items fußen demnach auf 14 Items in Step IVd. Weitere für Step V relevante Items befinden sich in den situations-, handlungs- und wirkungsbezogenen Beschreibungen (vgl. algorithmi-sche Steps I bis III), die insgesamt 14 Items umfassen und sich durch Referenzen auf explizit lokalisierbare Informationen im ‚Handy‘-Fall auszeichnen. Dabei über-wiegen Angaben zur Sozialstruktur und zu den einzelnen Praktiken.

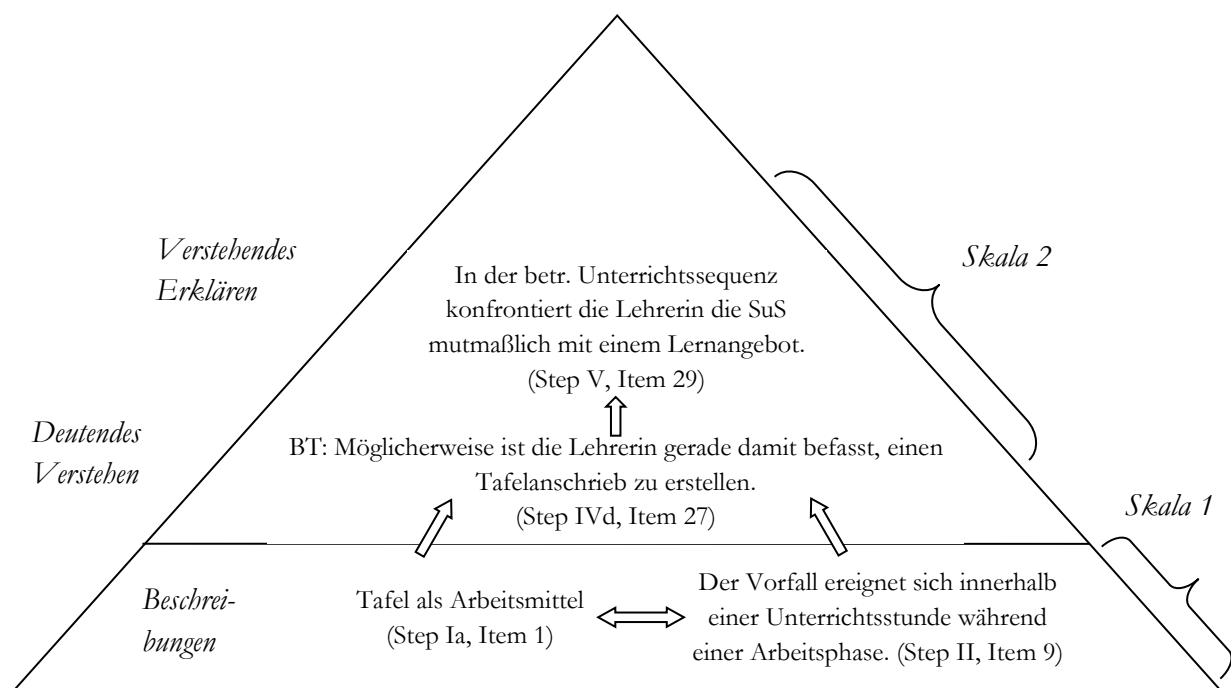


Abb. 1: Cluster A aus dem Item-Pool des ‚Handy‘-Falls im Rahmen des Kompetenzmodells

Inwiefern sich die einzelnen Cluster voneinander unterscheiden und damit auch potentiell unterschiedliche Schwierigkeitsanforderungen an die Probanden stellen, soll im Rahmen eines Cluster-Vergleichs verdeutlicht werden (s. Abb. 1 und 2). Die im Cluster A enthaltene Diagnose auf der Aggregatebene (vgl. algorithmischer Step V) ist sowohl durch ein Item auf der interpretativen Skala 2 (Step IVd) als auch durch zwei Items auf der deskriptiven Skala 1 (Steps Ia, II) abgesichert (vgl. Abb. 1). Die Diagnose stützt sich demnach auf rekonstruktiv erschließbare und auf explizit benannte Informationen. Letztere sind im Cluster I (s. Abb. 2) lediglich durch *ein* Item abgedeckt. Die Anzahl deskriptiver Items in Bezug auf interpretative Items kann bereits erste Hinweise auf die Schwierigkeit einer fallbasierten Diagnose liefern.

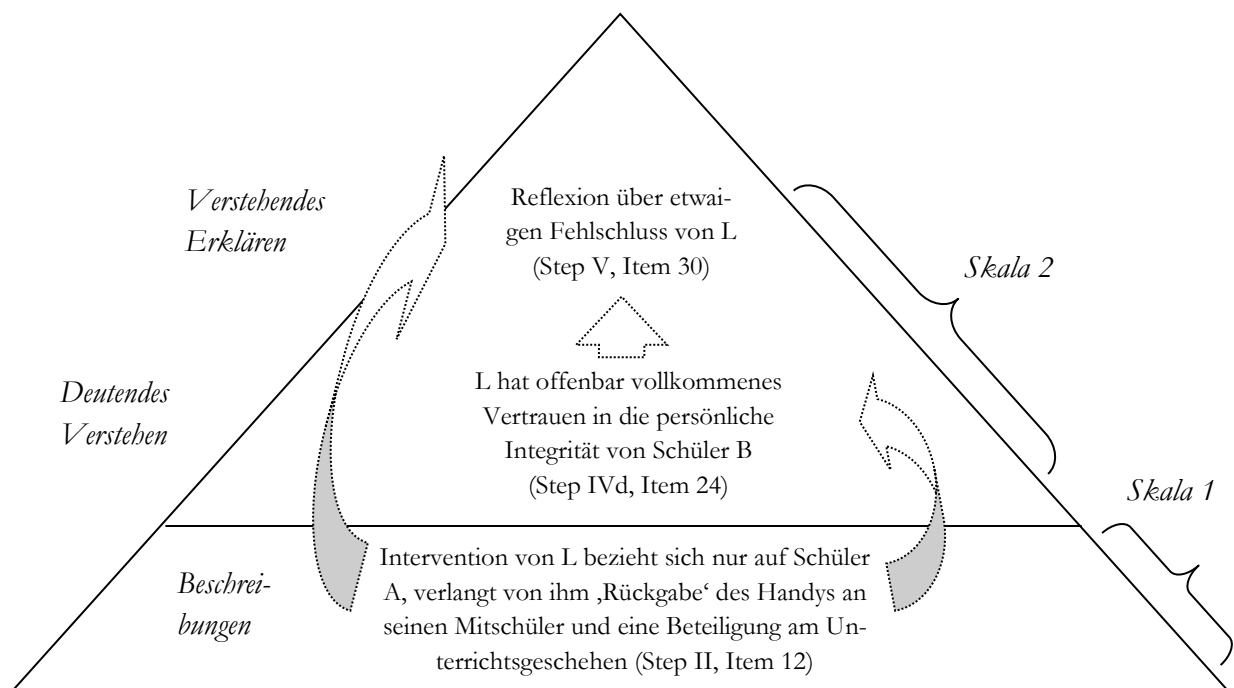


Abb. 2: Cluster I aus dem Item-Pool des ‚Handy‘-Falls im Rahmen des Kompetenzmodells

Darüber hinaus ist aber von entscheidender Bedeutung, ob ein deskriptives Item mit dem auf der Aggregatebene verorteten Erklärungsmuster *direkt* verknüpft werden und als *typisch* gelten kann oder zumindest *plausibel* erscheint. Vor diesem Hintergrund muss die im Cluster A enthaltene Informationsverknüpfung für vergleichsweise leicht und übersichtlich gehalten werden. Denn solange die Institution ‚Schule‘ aus Unterrichtsräumen mit Tafeln o.ä. und den am Unterricht unmittelbar beteiligten Lehrkräften sowie Schülerinnen und Schülern in Face-to-Face-Kommunikation besteht (also z.B. kein überwiegend computergestütztes Lernen vorherrscht), liegt es nahe, dass die Tafel als Medium für die Unterbreitung eines Lernangebotes genutzt wird. Die Plausibilität entsprechender Inferenzen wird anhand der durchgehend gestrichenen Pfeile verdeutlicht (vgl. Abb. 1).

Etwas anders verhält es sich mit nicht institutionalisierten, informellen und daher prinzipiell weniger transparenten Prozessstrukturdimensionen – wie z.B. in Bezug auf die in Interaktionsgeschichten eingebetteten Konflikte zwischen Schülern. Zwar können einzelne Aspekte der diagnostischen Anforderungsqualität jener des

Clusters A ähneln (vgl. Abb. 1): So interveniert die Lehrerin – trotz Unkenntnis der gesamten Vorgeschichte – und trägt offenbar entsprechend ihres hoheitsstaatlichen Mandates für eine ungestörte Fortsetzung des Unterrichts Sorge (vgl. Cluster B, Kap. II.7.2). Diesbezügliche Items beziehen sich auf typisches Lehrerhandeln und können direkt aufeinander bezogen werden. Dass das Handeln der Lehrerin jedoch mehrfache Kontingenzen vermissen lässt, weil es einfach hermeneutisch der Situationsdefinition des anklagenden Schülers folgt (vgl. Cluster I), verlangt von den Studierenden Erklärungen, die nur mithilfe komplexer Inferenzen formuliert werden können. So mag es für Rangeleien zwischen Schülern als typisch gelten, dass sie um den Besitz von Prestigeobjekten streiten (vgl. Cluster F, Kap. II.7.2). Weniger offenkundig erscheint jedoch die Möglichkeit einer gezielten Camouflage des anklagenden Schülers, der die Wegnahme ‚seines‘ Handys behauptet, wenngleich er es seinem Mitschüler zuerst entwendet haben könnte (vgl. Cluster H). Diese Art doppelt hermeneutischen Denkens lässt sich im Handeln der Lehrerin aber nicht erkennen, weshalb es als potentiell fehlschlüssig bewertet werden muss (vgl. Cluster I). Um die Möglichkeit eines Fehlschlusses im Lehrerhandeln zu inferieren und diesbezügliche Ursache-Wirkungsbeziehungen zu reflektieren, müssen die Probanden also einen Blick auf die Hinterbühne des im Fall appräsentierten overt Handlung werfen und eine ganze Reihe anderer Items aus weiteren inhaltlich nahen Clustern in ihre Überlegungen mit einfließen lassen. Mit anderen Worten: Die Probanden sind in besonderer Weise gefordert, Informationen aus verschiedenen Clustern derart miteinander zu verknüpfen und mithilfe kontextbezogener Inferenzen zu plausibilisieren, dass sie einen Fehlschluss der Lehrerin über den Handy-Besitz für möglich erachten. Um zu dieser Erkenntnis durchzudringen, ist davon auszugehen, dass sich die Probanden – sofern sie nicht schlicht auf ähnliche eigene Erfahrungen als Betroffene zurückgreifen – v.a. auf die im Cluster E enthaltenen Informationen hinsichtlich der unbekanntem Vorgeschichte zu der fallbasiert appräsentierten Schülerinteraktion stützen und/oder auf jene der Cluster H und F mit Bezug auf den ungeklärten Besitzstand des Handys und dessen Eigenschaft als ein Prestigeobjekt, dessen Aneignung erstrebenswert erscheint. Auch das gedankliche Durchspielen von Handlungsalternativen – wie im Cluster J abgebildet – könnte den Weg für eine differenziertere Diagnose hinsichtlich eines potentiell fehlschlüssigen Lehrerhandeln bereiten (zu den Clustern vgl. Kap. II.7.2, Dirks et al. 2012). Die wechselseitigen Beziehungen zwischen den einzelnen Clustern verdeutlichen systematische Zusammenhänge zwischen ihnen. Aus diesem Grunde können sie auch nicht per se als statistisch unabhängig angenommen werden.

### **8.1.2 Der Posttest: Beginn einer Erdkundestunde**

Die im Abschlusstest eingesetzte Fallvignette ist vergleichsweise umfangreicher und komplexer als die erste. Der größere Umfang wird bereits an der Anzahl der beteiligten Akteure (22 statt nur 3) und der einzelnen Sprechbeiträge bzw. „Turns“ (73 statt 4) deutlich, die insbesondere mit Bezug auf den Lehrer (35) oft auch mehrere Sprechakte bzw. Handlungen umfassen. Darüber hinaus handelt es sich bei ca. 40

Äußerungen um indirekte bzw. implizite Sprechakte<sup>1</sup>, welche die Probanden aufgrund ihrer Mehrdeutigkeiten vor eine große Herausforderung stellen. Insbesondere jene indirekten Sprechakte, die in den Kontext bestimmter Dimensionen eines Kernproblems fallen, bezeichne ich als diskurskritisch. Sie bedürfen i.d.R. eines höheren Analyseaufwands und können als wichtiger Indikator für die Schwierigkeit eines Falls bzw. Textes dienen.

Zur Zusammensetzung eines Lehrer-„Turns“ und zu den semantisch-pragmatischen Bedeutungsgehalten indirekter Sprechakte ein Beispiel: Wenn der Lehrer auf die Frage der Schülerin Sabine, ob sie die Stephanie hereinholen soll (Cluster F, Step II, Item 26), erwidert: „Ich glaub, die kann die Uhr genauso lesen wie wir alle (kurze Pause)“ (Cluster F, Step II, Item 27) und sich kurz darauf an die gesamte Klasse wendet, um diese mit den Worten zu begrüßen: „So, jetzt erst mal einen guten Morgen!“ (vgl. Kap. II.7.3, Dirks et al. 2012, Anhang, Z. 10-12), handelt es sich insgesamt um *einen* „Turn“ (Levinson 1983) des Lehrers, der aus zwei verschiedenen direktiven Sprechakten besteht. Der erste direktive Sprechakt ist zugleich ein indirekter Sprechakt, der die Schülerin lediglich implizit zum Verbleib im Klassenraum auffordert (vgl. „indirect question answer pair“ n. Asher & Lascarides 1998a); auf der Textoberfläche erscheint dieses indirekte Direktivum jedoch auch als ein direkter Sprechakt in Gestalt eines Assertivums (Searle 1969), indem der Lehrer einen Analogieschluss von der literalen Kompetenz des Uhrenlesens auf Seiten aller im Klassenraum Anwesenden, die pünktlich erschienen sind, auf die einzelne Schülerin außerhalb des Klassenraums herstellt. Da sich der Analogieschluss offenbar auf Jugendliche bezieht, die – anders als Kinder – in der Tat keine Probleme mit dem Uhrenlesen haben dürften, schwingt in seiner Aussage auch etwas Ironie mit. Indem der Lehrer auf eine für Jugendliche eigentlich banale Grundkompetenz verweist, versucht er wohl nicht nur, den Wahrheitsgehalt seiner Aussage zu plausibilisieren, sondern Sabine zugleich eine Begründung zu liefern, warum ihr Hilfsangebot entbehrlich ist. Möglich ist auch, dass er ihr unterstellt, nur nach einem Vorwand zu suchen, um selber dem Unterricht fernzubleiben (vgl. „Shirking“-Strategie, Cluster F, Step IVd, Item 79).

Diese Interaktionssequenz zwischen dem Lehrer und der Schülerin Sabine kann von den Probanden als ein erster Unterrichtsverhinderungsversuch bzw. als eine

<sup>1</sup> Die von John Searle (1975: 60f.) vorgeschlagene Definition für einen indirekten Sprechakt lautet: „In indirect speech acts the speaker communicates to the hearer more than he actually says by way of relying on their mutually shared background information, both linguistic and nonlinguistic, together with the general powers of rationality and inference on the part of the hearer“. Im Anschluss an Searle verstehe ich unter einem indirekten Sprechakt v.a. jene Äußerungen, die – neben einer sprechaktsspezifischen Repräsentation auf der Textoberfläche und ihrer literalen Bedeutung – zumindest einen weiteren Sprechakt i.S. einer weitgehend distinkten Sinneinheit implizieren. Wird die implizierte Bedeutung vernachlässigt, entsteht keine Kohärenz zwischen den verschiedenen „Turns“, weshalb „Reparaturen“ notwendig werden oder es auch zu Missverständnissen, wenn nicht zu Gesprächsabbrüchen kommt. Zur Diskussion weiterer Sprechakttyp-Kategorisierungen, die v.a. die Herstellung diskursiver Kohärenzen berücksichtigen, s. Asher & Lascarides (2006).

Unterrichtsstörung gewertet werden (Cluster J, Step IVd, Item 81), was sich auch an den temporalen Modalpartikeln „jetzt erst mal“ im zweiten Sprechakt ablesen lässt: Sie sollen den Schülerinnen und Schülern offenbar signalisieren, dass der Lehrer den Stundenbeginn mit dem Begrüßungsritual und nicht mit einer Ablenkung von diesem Skript einleiten möchte.<sup>2</sup> Im Gegensatz zum ersten Sprechakt erscheint der zweite Sprechakt – nicht zuletzt aufgrund seiner Direktheit – vergleichsweise leichter verständlich: Die Entrichtung des Lehrergrußes an die Klasse zielt skriptgemäß auf die reziproke Begrüßung durch die Schülerinnen und Schüler und kann von den Probanden als eine unterrichtstypische Praktik beschrieben werden. So ganz ‚glatt‘ stellt sich die zu erwartende Reziprozität (vgl. „adjacency pair“, Schegloff 2007) allerdings nicht her: Das Begrüßungsritual wird durch den Schüler Tim gestört, der den Lehrergruß unaufgefordert und mit betont lauter Stimme wiederholt (Cluster E, Step II, Item 29). Dieses auffällige Schülerverhalten ist Bestandteil des Clusters E, die darauf Bezug nehmende Reaktion des Lehrers fällt in das Cluster M (vgl. Kap. II.7.3, Dirks et al. 2012). Während die Interaktionssequenz zwischen der Schülerin Susanne und dem Lehrer dem Cluster F zugeordnet wurde, handelt es sich bei den anschließenden Handlungssträngen nunmehr um neue inhaltliche Facetten, die denn auch jeweils einem eigenen Cluster zugehören. Kleinste Diskurseinheiten appäsentieren demnach unterschiedliche Sinneinheiten, deren Relevanz von den Probanden für die zu erstellende Diagnose jeweils geprüft werden muss. Angesichts der Fülle miteinander verschränkter oder eigenständiger Handlungsstränge in dem Unterrichtstransskript stellt dieser Fall hohe Anforderungen an ihr Arbeitsgedächtnis. Wenn dann noch Unterbrechungen in einem Handlungsstrang entstehen, also die thematische Kohärenz im Handlungsablauf gestört wird, erschwert dies zwangsläufig die Identifikation und Zusammenführung der für die Diagnose benötigten Informationen. Diskohärenzen zeigen sich z.B., wenn Albert mehrfach beim Vorlesen durch andere Interaktionssequenzen unterbrochen wird oder der Lehrer erst nach einem neuerlichen Ereignis (Zuschlagen der Türe durch einen weiteren verspäteten Schüler) und seinem anschließenden Disziplinierungsaufforderung an die gesamte Klasse auf die verspätet eintreffenden Schülerinnen reagiert.

Indirekte Sprechakte können von den Probanden besondere Inferenzleistungen verlangen. Die o.e. Äußerung des Lehrers, in der er die Schülerin Sabine lediglich implizit zum Verbleib im Klassenzimmer auffordert, ist dafür nur ein Beispiel. Im Transskript finden sich diverse andere indirekte Sprechakte, deren Interpretation verschiedener Kontextinferenzen und damit eines größeren Analyseaufwands bedarf. Vergleichsweise einfach erscheinen dabei jene Sequenzen, in denen der zunächst nur implizierte Bedeutungsgehalt eines indirekten Sprechaktes kurz darauf expliziert wird, wie in dem folgenden Beispiel:

L: Mark, weißt du noch, wie ein Erdkundebuch aussieht? Hast du bestimmt schon mal irgendwo gesehen, oder ... (ca. 3 Sekunden Pause) So sieht ein Buch aus!

---

<sup>2</sup> Zum illokutionären Ausdrucksgehalt von Modalpartikeln s. Thurmair (1989: 2).

Mark: Ja.

L: Solltest auch mal eines dabei haben. (Z. 18-22, Anhang v. Kap. II.7)

Ähnlich wie in der vorausgegangenen Sequenz, in welcher der Lehrer ironisierend auf eine – eigentlich banale, da bei Jugendlichen mutmaßlich selbstverständlich verfügbare – Grundkompetenz abhebt, verweist er hier auf das ebenso für selbstverständlich zu erachtende Wissen eines ‚schulerefahrenen‘ Schülers, das sich auf die Identifikation eines Schulbuchs bezieht. Während zu diesem Zeitpunkt noch unklar ist, aufgrund welcher Situationswahrnehmung sich der Lehrer zu seiner Bemerkung veranlasst sieht (Mark könnte z.B. ein anderes Buch als das vom Lehrer erwähnte Erdkundebuch vor sich liegen haben), verdeutlicht sein anschließender Appell „Solltest auch mal eines dabei haben“, dass es ihm um die Beseitigung eines chronischen (vgl. der Modalpartikel „mal“) Missstands im Schülerverhalten geht (keine Bereithaltung des Fachbuchs). Der Modalpartikel im Lehrerappell legt die Inferenz bezüglich einer bereits seit längerem bestehenden Interaktionsgeschichte zwischen Lehrer und Schüler nahe (vgl. Cluster D, Step IVd, Item 82), die zudem durch den temporalen Modalpartikel „noch“ in der anschließenden Evaluation des Lehrers („Als Mark mir noch viel Freude bereitet hat“, Z. 22f.) gestützt wird. Nach den in der BRD geltenden Regelungen kann davon ausgegangen werden, dass der Schüler Mark durchaus im Besitz des Fachbuchs ist. Warum er es offensichtlich aber wiederholt nicht zum Unterricht mitbringt, bleibt unklar. Diesbezüglich konnten die Probanden lediglich Vermutungen anstellen, die – nicht zuletzt aufgrund ihres spekulativen Charakters – in unseren Daten jedoch kaum ausgeschöpft und auch nicht in ein zu erfüllendes Item zusammengefasst wurden.

Umso eindeutiger und leichter nachvollziehbar erscheinen hingegen jene Effekte, die aus fehlendem Arbeitsmaterial resultieren: So kann der Schüler Mark in Unterrichtspassagen, in denen das Erdkundebuch benötigt wird, nicht mitarbeiten und dadurch Störungen im Unterrichtsablauf verursachen (vgl. Cluster I, Step IVd, Item 69; Cluster J, Step IV d, Item 81) – es sei denn, er hat einen Banknachbarn, der ihn in sein Buch mit hineinschauen lässt.

Das ironische Lehrerhandeln lässt sich ebenfalls vergleichsweise leicht als eine bestimmte Praktik identifizieren, die auf dem Bloßstellen einzelner Schüler beruht: Sowohl der Verweis auf die Grundkompetenz des Uhrenlesens (Cluster F, Step II, Item 27) als auch jener auf das Wissen, wie ein Erdkundebuch aussieht, kongruieren mit einem Bloßstellen der betreffenden Schülerinnen und Schüler (Cluster M, Step II, Items 30f.; vgl. Cluster H, Step IVa, Item 61; Cluster C, Step IVd, Item 70). Das Ziel dieses Lehrerhandelns scheint allzu vertraut: Um sich des vielfach respektlosen Schülerverhaltens zu erwehren, greift der Lehrer zu verschiedenen Strategien der Identitätsbehauptung (Cluster M, Step IVd, Item 86; Cluster M, Step V, Item 92) – mit dem übergeordneten Ziel, seinem Mandat der Unterrichtserteilung nachkommen zu können (Cluster I, Step V, Item 100).

Exemplarisch sei auf einen weiteren indirekten Sprechakt verwiesen, der im Verlauf der aufeinanderfolgenden „Turns“ jedoch nicht ‚aufgelöst‘ wird und mutmaßlich Verständnisschwierigkeiten verursacht:

Ich war früher auch mal ein Mädchen. (Z. 39)

Diese Aussage wird von den Probanden z.T. ignoriert oder in den Kontext ironischer, generationsübergreifender Handlungsstrategien des Lehrers eingeordnet (Cluster M, Step IVd, Items 83, 86).<sup>3</sup>

Wie sich am Beispiel der vielen indirekten Sprechakte in diesem Transskript bereits andeutet, ist eine sorgsame Diagnose immer auch darum bemüht, jene *Informationslücken* zu benennen (s.u. „Anamnese“), die für diagnoserelevante Inferenzen benötigt werden. Bezogen sich diese im *Handy-Fall* v.a. auf die unmittelbare und weitere Interaktionsvorgeschichte zwischen den Schülern A und B sowie auf etwaige Vorerfahrungen der Lehrerin mit den Schülern und daraus resultierende Rollenzuschreibungen, vervielfacht sich im *Abschlussfall* die Anzahl fehlender Informationen in Relation zur Anzahl der an der Interaktion Beteiligten. Der Beginn der Erdkundestunde lässt u.a. Informationen zur Infrastruktur (Step Ia) hinsichtlich vorangegangener Unterrichtsinhalte und -stunden sowie der Schulform, Klassenstufe und Anzahl der Schülerinnen und Schüler vermissen. Darüber hinaus ist unbekannt, welchen Lerngewinn die Schülerinnen und Schüler aus dem betreffenden Unterrichtsausschnitt gezogen haben (vgl. Cluster L, Step V, Item 91). Angesichts des fehlenden Hintergrundwissens können die Probanden z.B. nur Vermutungen darüber anstellen, warum sich der Schüler Franz sehr selten am Unterricht beteiligt hat (Cluster D, Step IVd, Item 89). Die Hypothese, dass der Schüler Albert möglicherweise an einer Lese-Rechtschreib-Schwäche leidet, kann lediglich anhand seiner Schwierigkeiten beim Lesen eines einfachen Textes inferiert, aber nicht umfassend belegt werden (Cluster D, Step IVa, Item 63). Ob es sich um eine Hauptschulklasse handelt, erscheint mit Bezug auf diverse Indikatoren im Fall wahrscheinlich, ist aber ebenfalls nicht hinreichend belegt (vgl. Cluster E, Step IVd, Item 74). Selbiges gilt für das Alter der Schüler und Schülerinnen, das aufgrund ihres Verhaltens (z.B. Schminken der Mädchen, lautes Rufen der Jungen in die Klasse) in der Pubertät verortet werden kann (vgl. Item 76). Darüber hinaus fällt auf, dass der Lehrer auf eine Suche nach Gründen für das Fernbleiben der Schülerin Stephanie verzichtet (Step II, Item 28) und ihm zwei zu spät erschienene Schülerinnen keinen Grund für ihre Verspätung nennen und dieser auch nicht vom Lehrer erfragt wird (Item 35). Die Anamnese fehlender Informationen über Themen der vorausgegangenen Stunden (Step Ia, Item 1) wird zusätzlich durch die Beobachtung indirekt gestützt,

---

<sup>3</sup> Einige Probanden ließen sich durch diese Aussage auch dazu verleiten, die Lehrkraft als Frau zu identifizieren, obwohl der Lehrer zu Beginn der Fallvignette explizit als „Herr Mustermann“ benannt wird. Um einen Anstieg der ohnehin hohen Item-Anzahl zu vermeiden, wurden diese Ausnahmefälle zwar bei der Auswertung ignoriert. Sie schlugen aber dennoch ‚zu Buche‘, weil die geschlechtsbezogene Fehlannahme die Erfüllung bestimmter Items mit Bezug auf geschlechtsspezifische Merkmale und Interaktionsweisen teilweise verhindert hat (Step Ib, Item 23; Step II, Item 34; Step IVd, Item 84; Step V, Items 92, 95).



der Lehrer beginne den eigentlichen Unterricht ganz unvermittelt ohne Verweis auf ein zurückliegendes Thema, also ohne einen Unterrichtseinstieg (Item 39; alle Items in Kap. II.7.3, Dirks et al. 2012).

Solche Informationslücken dürfen nicht einfach ignoriert werden, da sie für die Re-/Konstruktion fallrelevanter Ursache-/Wirkungszusammenhänge von entscheidender Bedeutung sein können. Dabei geht es allerdings keineswegs um ein unbeschränktes Zusammenführen beliebig vieler, für möglich erachteter Dimensionen. Diese finden – wie sich auch in den Testantworten gezeigt hat – vielmehr im Hinblick auf ihre Relevanz für das Verstehen und Erklären des Falls Erwähnung und damit Eingang in diesbezügliche Items. Die Anamnese fehlender Informationen (vgl. Skala 1) scheint also von der Suche nach plausibel erscheinenden Inferenzen (Skala 2) geleitet zu sein, was sich auch in der deduktiven Item-Clusterung widerspiegelt (vgl. Kap. II.7, Dirks et al. 2012). Dass v.a. die Probanden der Versuchsgruppen bei der Bildung angemessener Inferenzen reüssieren, könnte sich darauf zurückführen lassen, dass sie in ein sorgfältiges ‚Durchspielen‘ der algorithmischen Steps eingeführt wurden.

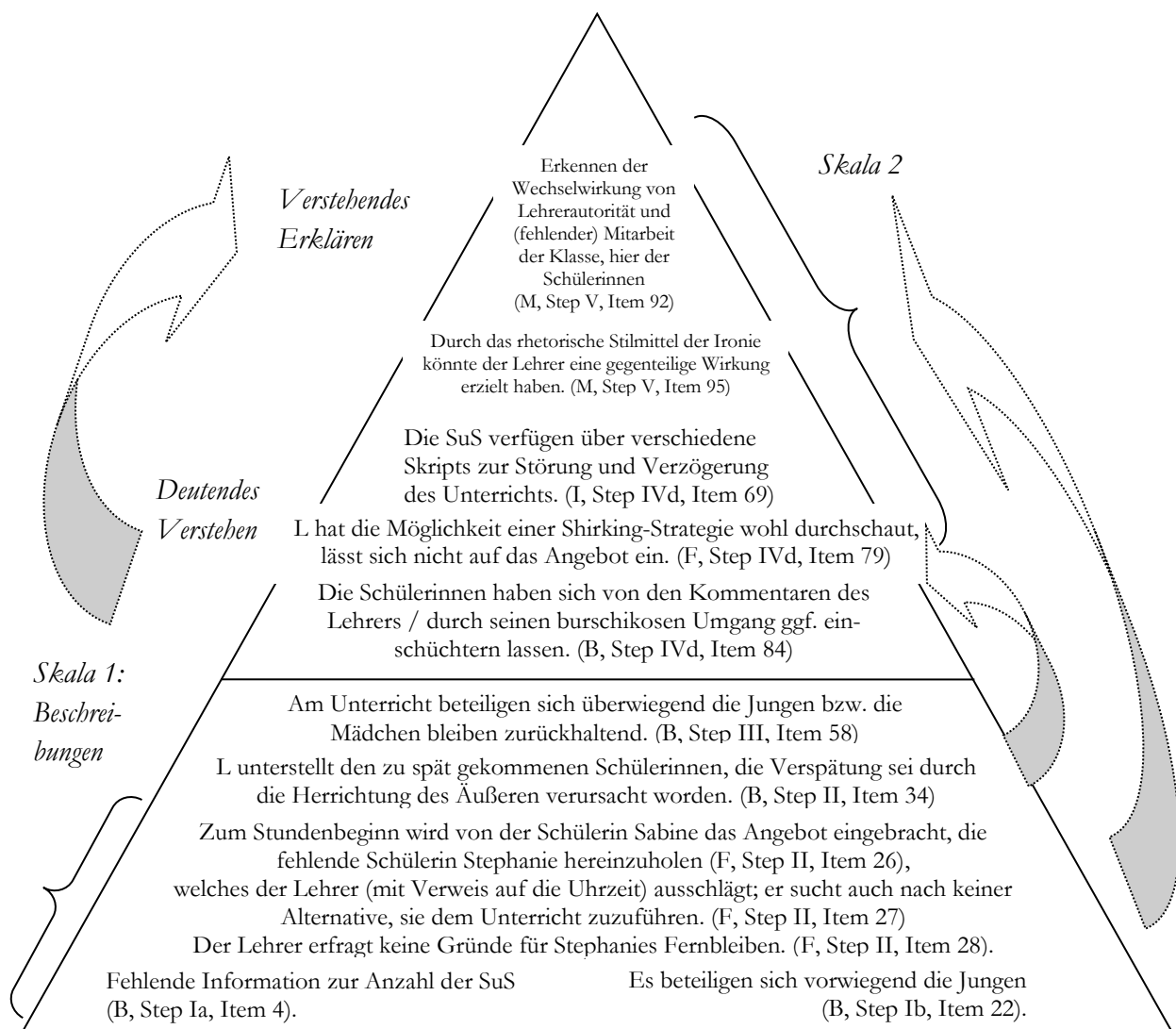


Abb. 3: Cluster-Verknüpfung bezüglich mangelnder Unterrichtsbeteiligung der Schülerin Sabine

In dem Maße, wie Informationen bezüglich der im Fall appräsenzierten Dimensionen des Kernproblems fehlen und inferiert werden müssen, erschwert dies meist auch die Diagnose. Dementsprechend können Notwendigkeiten der Informationsüberbrückung auch als Indikatoren für die Aufgabenschwierigkeit dienen.

Die gesteigerte Komplexität des Abschlussfalls wird sowohl im Hinblick auf die darin enthaltene *Informationsdichte*, aber auch hinsichtlich der *Informationslücken* mit Bezug auf die weitgehend fehlende Beteiligung der Mädchen am eigentlichen Unterrichtsgeschehen in der Abbildung 3 veranschaulicht. Anders als im Handy-Fall, dessen Anforderungsstruktur in stärker distinkte, wenngleich auch miteinander vernetzte Item-Cluster untergliedert wurde, liegt in Abbildung 3 eine komplexe Cluster-Verknüpfung (Cluster B, F, I, M) vor. Die mehr oder weniger typischen Dimensionen des Kernproblems „Herstellung eines Arbeitsbogens im Unterricht trotz Schulzwangs“ werden hier durch neue Problemaspekte von Schule und Unterricht erweitert, die mit dem Kernproblem eng verschränkt sind, aber ggf. als eigenes Problem definiert werden können. Hier geht es v.a. um die Ermöglichung bzw. Behinderung gleicher Bildungschancen für Jungen und Mädchen im koedukativen Schulsystem. Die Fallvignette bietet diesbezüglich diverse Anhaltspunkte, die v.a. auf geschlechterstereotyp divergierenden Verhaltensweisen (vor-/laute Störmanöver durch die Jungen, Hilfsangebot und Kooperation sowie Zuspätkommen durch die Mädchen), aber auch auf geschlechtsspezifischen Verhaltenszuschreibungen (sich schminken, frisieren) beruhen, die von den Mädchen möglicherweise als Diskriminierung v.a. gegenüber den Jungen in der Klasse und als Einschüchterung empfunden worden sind – mit entsprechenden Auswirkungen auf ihre Mitarbeit im Unterricht. Um solch eine Ursache-/Wirkungsverknüpfung abzusichern, müssten v.a. die folgenden Fragen geklärt und diesbezügliche Informationslücken geschlossen werden: Besteht eine geringe Unterrichtsbeteiligung der Mädchen während der *gesamten* Unterrichtsstunde? Lässt sich diese nur bei dem Erdkundelehrer, also auch in vorausgegangenen oder späteren Erdkundestunden, oder auch bei anderen Lehrkräften beobachten? Falls ja, ist deren Unterricht ebenfalls von geschlechterstereotypen, eher desavouierenden Zuschreibungen durch die jeweiligen Lehrkräfte geprägt?

## 8.2 Vergleich der fallspezifischen Anforderungsstrukturen

Eine Kontrastierung der beiden Fallvignetten bietet sich v.a. mit Bezug auf die Domänenspezifität sowie auf narratologische Genremerkmale und das Schließen von Informationslücken mithilfe von Hypothesen an. Die Ergebnisse der miteinander kontrastierten Anforderungsstrukturen des Einstiegs- und Abschlussfalls werden in einer Tabelle zusammengefasst (s.u.).

### *Fallvergleich hinsichtlich der Domänenspezifität und des Vorwissens der Probanden*

In beiden Fällen steht v.a. das pädagogische Handeln der Lehrkräfte und ihre Interaktion mit den Schülerinnen und Schülern im Vordergrund. Dabei zielt das Lehrerhandeln jeweils auf die Beseitigung von Unterrichtsstörungen und damit auf den

<i>Mögliche Indikatoren für unterschiedliche Schwierigkeitsgrade</i>	<b>Einstiegsfall (Prätest)</b>	<b>Abschlussfall (Posttest)</b>
Genre & Genremerkmale	Fallvignette: Transskript/Aktualtext; Narrative Genremerkmale: Mündliche Rede, antagonistische Akteurskonstellation in dramaturgisch sich zuspitzender Konfliktsituation unter Berücksichtigung von Raum und Zeit sowie non-verbale Praktiken; kurze Sätze, z.T. mit intonatorischen Hinweisen; aber: keine „turn“-übergreifenden Absatzmarkierungen	
Domäne & domänenspezifisches Sprachregister	Bezüge zu einem pädagogischen Kernproblem; Alltagssprache	mögliche Bezüge zu mehreren Kernproblemen; Alltagssprache, z.T. alltagsgebräuchliche Fachsprache
Vorwissen der Probanden	Gesetz zum Schulzwang, Schulordnung, unterrichtstypische Lehr-/Lernformen, schul- und jahrgangsspezifisches Schülerverhalten	
Textlänge	Kurze Sequenz (5 Zeilen)	ca. zwei Seiten
Anzahl der Akteure	3	22
Anzahl der „Turns“	4	73 (davon L: 35)
Anzahl wechselnder Gesprächsanlässe	1 (Wegnahme bzw. Rückgabe des Handy)	17 (v.a. Unterrichtsstörungen und Suche nach Fachbegriffen)
Dis-/Kohärenzen: Räumliche Verteilung aufeinander bezogener „Turns“ (vgl. „turn construction units“)	„Turns“ sehr dicht beieinander	Viele aneinander gefügte Dialoge zwischen L und SuS/ Klasse; zwei Unterbrechungen beim Vorlesen Alberts; Klassenbucheintrag verspäteter Schülerinnen erst nach Disziplinierung der gesamten Klasse; Unterrichtsgespräch durch Suche nach Schreibmitteln unterbrochen
Anzahl indirekter Sprechakte	0	ca. 40
Cluster/ Item-Anzahl	10 Cluster, 36 Items pro Cluster: 3 - 5 Items, Ø 3,6	13 Cluster, 100 Items pro Cluster: 3 - 10 Items, Ø 7,7
Algorithmischer Step: Item-Anzahl	Ia: 2, Ib: 6, II: 5, III: 1, IVd: 14, V: 8	Ia: 8, Ib: 15, II: 29, III: 7, Iva: 5, IVb: 3, IVd: 23, V: 10
Algorithmischer Step: Anzahl Hypothesen	IVd: 13, V: 7; Total: 20	IVa: 2, IVd: 14, V: 5 Total: 21
Absolute $K_{i(\max)}$ -Werte	Skala 1: 6,10; Skala 2: 7,66	Skala 1: 14,33; Skala 2: 12,15

Tab. 1: Mögliche Indikatoren für Schwierigkeiten im Prä- und Posttest

Vollzug regulären Unterrichts bzw. auf das Herstellen der für das Unterrichten erforderlichen sozialen Ordnung. Die in diesem Kontext erstellten Diagnosen lassen sich letztendlich demselben Kernproblem zuordnen, der „Herstellung eines Ar-

beitsbogens im Unterricht trotz Schulzwangs“ (Dirks 2000: 233ff.). Dabei eröffnet v.a. der Abschlussfall jedoch auch Optionen für Diagnosen mit Bezug auf andere, z.B. aus Koedukation und Geschlechterstereotypen resultierende Kernprobleme.

Unabhängig von dieser problembezogenen Mehrschichtigkeit ist festzuhalten, dass der Einstiegs- und Abschlussfall hinsichtlich ihrer Domänenspezifität nah beieinander liegen. Beide Fälle appäsentieren ein pädagogisches Problem, das für alle Fächer relevant sein kann. Die mit dem Kernproblem verknüpften Unterrichtsstörungen lassen sich insbesondere spieltheoretisch hinsichtlich des sog. Prinzipal-Agenten-Problems und der damit verbundenen Effekte in der Lehrer-Schüler-Interaktion herausarbeiten. Das heißt, die Studierenden sind gefordert, Deutungen und Erklärungen für das Handeln der jeweiligen Lehrkraft und der einzelnen Schülerinnen und Schüler sowie der daraus resultierenden Effekte zu erzeugen. Dabei können sie sich auf ihr Vor- bzw. Erfahrungswissen über den in Deutschland gesetzlich fixierten Schulzwang sowie über die in einer Schulordnung aufgeführten Regeln und unterrichtstypische Lehr-/Lernformen beziehen, ebenso wie auf das ihnen mutmaßlich vertraute schul- und jahrgangsspezifische Schülerverhalten. Für schulerprobte Probanden sollten fallbasierte Diagnosen, die im pädagogischen (also weniger fachsprachlich ausgerichteten) Kontext von Schule und Unterricht angesiedelt sind, daher keine allzu hohe Anforderung darstellen. Auch können sie jene diagnostischen Kompetenzen, die sie im Einstiegsfall eingesetzt hatten, nunmehr im Rahmen eines proximalen Transfers (vgl. Mähler & Stern 2006) auf den Abschlussfall anwenden.

Erleichtert wird die Ausschöpfung von Transfermöglichkeiten aber mutmaßlich nicht nur durch die Domänenähnlichkeit, sondern auch durch dasselbe Genreformat (Transskript). Dies mag v.a. für die Probanden der Versuchsgruppen gelten, die vielfältige Lernangebote zum systematischen Aufbau ihrer Diagnosekompetenz nutzen konnten, während jene der Kontrollgruppen zum Zeitpunkt des Posttests am Semesterende möglicherweise längst die im Prätest eingesetzten Diagnosekompetenzen bzw. die erstellten Diagnosen ‚vergessen‘ hatten.

Transferschwierigkeiten könnten dadurch entstehen, dass anstelle des im Einstiegsfall fachlich nicht näher spezifizierten Kontextes im *Abschlussfall* Fachbezüge (hier: Erdkunde) hergestellt und diese mit dem Kernproblem verknüpft werden müssen. Eine noch bedeutsamere Transferschwierigkeit resultiert aber vermutlich v.a. aus der höheren Anzahl von Akteuren und Handlungen, da die unterschiedlichen Interaktionen eine Vielzahl von Deutungs- und Erklärungsmöglichkeiten zulassen. Der Fallauszug kann somit unübersichtlicher erscheinen, was von den Probanden gesteigerte Arbeitsgedächtnisleistungen verlangt und eine angemessene Diagnose erschwert.

#### *Fallvergleich aus narratologischer Perspektive*

Das Genre des aktualtextuellen Transskriptes weist einige Besonderheiten auf, die sich v.a. mithilfe erzähltheoretisch fundierter Erkenntnisse weiter ausdifferenzieren lassen.

Zunächst ist festzuhalten, dass sich die Unterrichtstranskripte aus dem für Narrationen typischen Sprachmittel der mündlichen Rede zusammensetzen und der Reindehalt überwiegend von alltagssprachlichen Ausdrucksweisen geprägt ist. Mündliche Rede im Register der Alltagssprache ist mutmaßlich leichter anschlussfähig an die menschliche Informationsverarbeitung als ein per se schriftliches Genre, insbesondere, wenn es Domänenbezüge in einer Fachsprache aufweist. Gehen wir zurück in die menschheitsgeschichtliche Entwicklung von Kognition und Schriftzeugnissen (z.B. Höhlenmalereien, Sagen, Märchen), fällt auf, dass Strukturmerkmale des „Story Telling“ an erster Stelle menschlicher Kommunikationsformen stehen (vgl. Brooks 1992 [1984]: 3ff., 37ff.; Bruner 1986, Schlieben-Lange 1983, Virtanen 1992). Daher ist es auch nicht weiter verwunderlich, dass narrative Bausteine zentraler Bestandteil ganz unterschiedlicher Genres – also polygenerisch verwendbar – sind und die Erschließung der in narrative Strukturen eingekleideten Bedeutungsgehalte meist für leichter befunden wird als beispielsweise jene in einem deskriptiven oder argumentativen Genre (vgl. „expository texts“, Corkill et al. 1988, Coté et al. 1998, Graves & Graves 2003: 12). Die für diese Textformate erforderlichen Abstraktionsleistungen haben sich evolutionsgeschichtlich offensichtlich erst später herausgebildet. Vor diesem Hintergrund sollte es den Studierenden eigentlich leicht fallen, die Transskript-Inhalte zu verstehen und auf deren Basis eine Diagnose zu erstellen. Als vorteilhaft erweisen sich dabei mutmaßlich auch die insgesamt kurzen und kaum ineinander verschachtelten Sätze in beiden Fallvignetten. Während narrative, v.a. schriftliche Genres durchaus komplexe Satzstrukturen enthalten können, zeichnen sich mündliche Diskurse im Alltag und eben z.T. auch in der Schule durch eine vergleichsweise einfach strukturierte Syntax aus. Dies trifft auch für die vorliegenden Fallausschnitte zu und kann schwierigkeitsmindernd sein.

Allerdings sind Studierende i.d.R. nicht mit der monotonen Textstruktur von Transskripten vertraut und empfinden diese zunächst als etwas befremdlich. Eine wesentliche Abweichung von narrativen Genres besteht offensichtlich in der fehlenden Strukturierung durch inhaltlich begründete Absatzmarkierungen oder auch Kapitelüberschriften, die sich zumindest als Einleitung, Hauptteil und Schluss identifizieren lassen. Während Unterrichtsstunden durchaus analoge Strukturen aufweisen (z.B. Unterrichtseinstieg, Erarbeitungs- und Übungsphase, Ergebnissicherung), ermöglichen Fallvignetten, die ja stets nur einen kleinen Ausschnitt aus einer gesamten Unterrichtsstunde wiedergeben, i.d.R. nur Wahrscheinlichkeitsannahmen über deren zeitlich-räumliche Einordnung in den Stundenverlauf. Darüber hinaus erlauben Fallvignetten keine – für die Qualität von Unterricht wesentlichen – Aussagen zu der Frage, ob die betreffende Stunde – ähnlich einer Geschichte – ‚rund‘ ist und damit alle für den Unterrichtsgegenstand und für effektives Lernen angemessenen Schritte der Er- und Bearbeitung beinhaltet hat.

Das heißt, die narrationstheoretischen Prämissen bezüglich kohärenter Ereignisabfolgen wie ‚Orientierung‘, ‚Konflikt‘ und ‚Auflösung‘ (Labov & Waletzky 1967)<sup>4</sup>

<sup>4</sup> Die narrationstheoretischen Prämissen lassen sich weiter ausdifferenzieren und auch hinsichtlich ihrer Verbindlichkeit kritisch diskutieren (vgl. Dirks 2010: 396f.), worauf hier aber

werden stets nur teilweise in den Fallausschnitten manifest: Im ‚Handy‘-Fall ereignet sich die Störung offenbar erst im Unterrichtsverlauf; die Fallvignette bietet Einblick in eine Teilphase des Konfliktes und in die – von der Lehrerin offensichtlich intendierte – Beilegung des Konfliktes. Die Probanden können also nur Mutmaßungen darüber anstellen, in welcher Unterrichtsphase der Fall angesiedelt ist und welche Ereignisse dem angeblichen Handy-Diebstahl vorausgegangen sind. Der *Abschlussfall* setzt gleich zum Stundenbeginn ein. Entgegen didaktischer Grundprinzipien lässt dieser aber einen klassischen Unterrichtseinstieg mit Verweisen auf Thema und Arbeitsstand der vorausgegangenen Stunde einerseits und Ziel der aktuellen Stunde andererseits (vgl. ‚Orientierung‘) vermissen. Dabei werden gegenläufige Tendenzen zwischen dem offiziellen Stundenbeginn und dem eigentlichen Unterrichtsbeginn in multiplizierter Form deutlich. Die sich im Lehrer-Schüler-Verhältnis abzeichnende antagonistische Akteurskonstellation und die im Akteurshandeln auftretenden Interferenzen erhalten angesichts der sie rahmenden Zeit-/ Raumdimension besondere Brisanz: Je länger sich der eigentliche Unterrichtsbeginn aufgrund stets neuerlicher Unterrichtsstörungen verzögert, desto weniger Zeit verbleibt dem Lehrer, die Schülerinnen und Schüler entsprechend seines hoheitsstaatlichen Mandates zu unterrichten. Seine Versuche, einen fachlichen Handlungsstrang durch das Vorlesen des Schülers Albert zu initiieren, sind von Diskontinuitäten geprägt, die weniger mit klassischen Genremustern einer Erzählung übereinstimmen. Umso dominanter wird hier aber eine dramaturgische Zuspitzung der Spannungen im Lehrer-Schüler-Verhältnis, die sich schließlich in einen Wutausbruch des Lehrers entlädt. Das für narrative Kohärenzen typische Merkmal bipolarer Akteurskonflikte bildet hier quasi das ‚tool kit‘ für den ersten Teil der Fallvignette und vermag durchaus auch fehlende Kohärenzmerkmale zu kompensieren.<sup>5</sup>

Interessanterweise enthält der dramaturgische Höhepunkt auch einen ‚clue‘ zur Erschließung des Kernproblems aus Sicht des Lehrers (vgl. ‚recipient design‘, Schegloff & Sacks 1973): Er unterstellt den Schülerinnen und Schülern, sie wollten ihm die ‚Erdkunde[stunde] ... versauen‘. Das heißt, er bietet den ihm Anbefohlenen seine Deutung bzw. sein Framing ihres Handelns (Unterrichtsstörungen) an, welches auf die Verhinderung von Unterricht zielt. Das Katz-und-Maus-Spiel zwischen der Lehrkraft, die sich darum bemüht, Fassung zu bewahren und Störaktionen v.a. mit Ironie abzuwehren, und einigen Schülerinnen und Schülern, die dem äußeren Anschein nach kooperieren, dabei aber Regeln der Höflichkeit und Schulordnung missachten, hat nun ein Ende; der Konflikt ist vorerst ‚beigelegt‘. Nach der vom Lehrer enaktierten Decamouflage der gezielten Unterrichtsverhinderung ist offensichtlich ein Tabu gebrochen, ein ‚So-Tun-Als-Ob‘ nicht mehr möglich – ohne die vom Lehrer angedrohten Sanktionen zu riskieren. Es liegt in der Natur schülerspe-

---

verzichtet wird. Zu weiteren Eigenschaften typischer Erzählstrukturen s. auch Mandler & Johnson (1977).

<sup>5</sup> Ein ähnliches Phänomen konnte am Beispiel von Medientexten herausgearbeitet werden. Antagonistische Handlungen und Seins-Zuschreibungen firmieren hier als ‚tool kit‘ für anderweitig fehlende Kohärenzmerkmale (vgl. Dirks 2010, Kap. 11).

zifischer Kosten-Nutzen-Analysen, sich erst einmal nicht mehr ‚aus der Deckung zu wagen‘, bis die Drohgebärden des Lehrers ‚verpufft‘ sind und er sich wieder beruhigt hat. Nach einer derartigen Situationsbereinigung kann dann in der Tat der Unterricht und damit die zweite Fallsequenz beginnen.

Allein die Text- und Genrestruktur der Transskripte bietet keine Hinweise auf entsprechende Strukturierungen des Stundenverlaufs; sie enthalten keinerlei inhaltlich motivierte, textstrukturelle Markierungen. Diese werden lediglich aufgrund eines Sprecherwechsels vorgenommen. Die Transskripte zeichnen sich nämlich durch eine Kompartimentalisierung (Pieterse 1995, Baumann 1996: 173ff.) von ‚Turns‘ aus, d.h. durch das textstrukturell gleichförmig erscheinende Aneinanderfügen habitualisierter Genremusterroutinen. Um deren Bedeutung zu erschließen, sind die Probanden daher gezwungen, die einzelnen ‚Turns‘ per se und in ihrer Sequenzierung sowie im Gesprächskontext genau zu analysieren. Je mehr Sprecherwechsel und damit ggf. einhergehende Themenwechsel ein Fall appräsentiert (vgl. Gesprächsanlässe in Tab. 1), umso größere diagnostische Anstrengungen verlangt dies von ihnen. Fehlen dann auch noch die für eine ‚runde‘ Unterrichtsstunde typischen Strukturierungen – wie z.B. ein Unterrichtseinstieg im Erdkundefall – und nehmen Interaktionsformen – wie im Handy-Fall das potentiell fehlschlüssige Konfliktlösungshandeln der Lehrerin oder die wiederholte Verhinderung von Unterricht trotz Stundenbeginns im Erdkundefall – die Gestalt eines ‚diffusen Hindurchprozessierens‘ (Reckwitz 2003: 295) an, stellt dies umso höhere Anforderungen an die diagnostischen Informationsverarbeitungsprozesse der Probanden.

#### *Fallvergleich hinsichtlich Anforderungen an die Hypothesenbildung*

Trotz Unterschieden in der Item- und Cluster-Anzahl sowie in den absoluten Kompetenzwerten werden für die einsteigs- und abschlussfallbezogenen Diagnosen fast gleich viele Hypothesen benötigt. Die im Genre der Fallvignette appräsentierten Informationen werden dabei mit weiteren, für plausibel gehaltenen Dimensionen sozialer Realitäten verknüpft (vgl. Inferenzen) – und zwar offenbar entsprechend fast gleich vieler teildiagnostischer Performanzen. Dass diesbezüglich – zumindest rein quantitativ – kaum Differenzen zwischen den Fallvignetten bestehen, mag auch als weiterer Legitimationshinweis für den Prä- und Posttest-Vergleich dienen.<sup>6</sup>

Die sehr viel höhere Anzahl indirekter Sprechakte im Abschlussfall kann diese Ähnlichkeit kaum erschüttern, da sich lediglich einer dieser Sprechakte weder durch den Ko- noch den Kontext aufklären lässt. Es ist vielmehr davon auszugehen, dass Anforderungsstrukturen, deren Erfüllung – wie im Falle der indirekten Sprechakte – anhand leicht zugänglicher Kontextinferenzen weitgehend gewährt ist, eher in den Aufmerksamkeitsfokus der Probanden gelangen als ausschließlich explizite Informationen bzw. direkte Sprechakte (vgl. z.B. McNamara et al. 1996). Hinzu

<sup>6</sup> Auch variiert die Anzahl der Item-Cluster in den beiden zur Testung eingesetzten Fallvignetten nur unwesentlich, was ebenfalls als Hinweis auf die Angemessenheit der Fallauswahl für einen Prä- und Posttest dienen kann.

kommt, dass in den Fallexzerpten kaum fachsprachliche Begriffe verhandelt werden (s.o.) und die schulerfahrenen Probanden mit den Doppel- und Mehrdeutigkeiten von Sprache in Schule und Unterricht vertraut sein müssten.

*Fallvergleich mit Bezug auf die maximalen Kompetenzwerte<sup>7</sup>*

Die Anzahl der Items mit Bezug auf den Einstiegs- und Abschlussfall bietet lediglich einen ersten Anhaltspunkt für die Beurteilung der Schwierigkeit eines Falls, da der Anzahl eine Gleichgewichtung der einzelnen Item-Schwierigkeiten unterliegt. Das heißt, dass z.B. eine mehr oder weniger komplexe Inferenzbildung in unmittelbarem Bezug zu einer höheren oder niedrigeren, als fix betrachteten Item-Schwierigkeit gesetzt wird. Während sich solch ein Verfahren meist auf Expertenurteile stützt, orientieren wir uns – ähnlich wie die internationale PISA-Lesestudie, wenngleich wir nicht die Item-Response-Theorie verwenden – an den in den Testdaten erhobenen Bewältigungshäufigkeiten und leiten daraus eine Bewertung der Item-Schwierigkeit ab. Diese wird – getrennt nach den Subskalen – in einem Gesamtwert  $K_{\text{imax}}$  zusammengefasst (vgl. Baumbach 2012, Kap. II.6). Ausgehend von dem Idealfall, dass ein Proband alle Items auf einer Subskala erfüllt hat, entspricht der  $K_{\text{imax}}$ -Wert einem absoluten – also nicht anteiligen – Kompetenzwert. Er zeigt die absolut höchstmögliche Punktezahl auf einer Subskala in unserem Test an. Indem er die Antworthäufigkeiten für alle Items mit den Fähigkeiten der Probanden kombiniert, diese summarisch für jedes einzelne Cluster erfasst und die Fallvignette als Produkt aus den Dimensionen der im Fall appräsentierten sozialen Realitäten interpretiert, kann er als eine wichtige Maßeinheit für die Schwierigkeit des Tests bzw. der Fallvignette gelten. Vor diesem Hintergrund gestattet der  $K_{\text{imax}}$ -Wert ferner den Vergleich verschiedener Fallvignetten und gibt Auskunft über deren Proportionen. Um sicherzugehen, dass die Bewältigungshäufigkeiten mit unserer Modellierung pädagogisch soziologischer Diagnosekompetenzen korrelieren, wurden die Gesamtwerte zusätzlich hinsichtlich der kompetenzmodelltheoretisch fundierten Anforderungen an die Wiedergabe und Verknüpfung der Dimensionen sozialer Realitäten überprüft und nahezu ausnahmslos bestätigt.

Schauen wir uns nun die  $K_{\text{imax}}$ -Werte der beiden Fälle näher an (s.o. Tabelle 1), fällt auf, dass diese im Einstiegsfall auf beiden Skalen niedriger ausfallen als im Abschlussfall. Der deutliche Unterschied resultiert letztendlich aus der größeren Anzahl von Items und Item-Gruppen im Cluster innerhalb des Abschlussfalls. In beide Werte gehen die Länge des Falltransskripts sowie die Offenheit der Situation ein, welche sich im Spektrum möglicher Inferenzen widerspiegelt. Der fallspezifische Maximalwert steigt also in Relation zur Anzahl der Items und Item-Gruppen im Cluster. Schauen wir uns zudem die Proportionen der  $K_{\text{imax}}$ -Werte auf den beiden Subskalen für beide Fallvignetten etwas genauer an, finden wir auf der Skala 1 keine größeren Abweichungen von der Skala 2. Dennoch ist zu konstatieren, dass die Testdaten für den Einstiegsfall eine Dominanz der zweiten Subskala und jene

---

<sup>7</sup> Dieses Kapitel ist unter maßgeblicher Mitarbeit von Hendrik Baumbach entstanden. Für seine hilfreichen Hinweise sei ihm an dieser Stelle herzlich gedankt



für den Abschlussfall umgekehrt eine Dominanz der ersten Subskala verdeutlichen. Zweifelsohne erfordert eine vollständige diagnostische Erfassung des betreffenden Falls von den Probanden ein Zusammenwirken aller diagnostischen Teilkompetenzen. Die  $K_{\text{imax}}$ -Werte legen nahe, dass von ihnen im Einstiegsfall – gemäß der verwendeten Kompetenzdefinition – mit leicht höherer Gewichtung die Fähigkeiten des Deutenden Verstehens und Verstehenden Erklärens verlangt werden, während die zweite Testung das Ermitteln der Informationen (Steps I bis III) etwas mehr in den Vordergrund rückt. Denkbar ist, dass insbesondere das sehr kurze Transskript der ersten Erhebung den Probanden analytischen Freiraum für die Hypothesenbildung angeboten hat, während diese in der zweiten Fallvignette durch eine große Vielzahl von explizit entnehmbaren Informationen zur Situation und den handelnden Akteuren zunehmend kanalisiert wurde.

Aus den unterschiedlich proportionalen  $K_{\text{imax}}$ -Werten für beide Fälle lässt sich jedoch nicht automatisch die Schlussfolgerung ableiten, das erste Testinstrument eigne sich v.a. für einen Test der Teilkompetenzen auf der zweiten Subskala und die zweite Vignette umgekehrt für die erste Subskala. Vielmehr ist davon auszugehen, dass der Einstiegstest die für den Fall relevanten Dimensionen sozialer Realitäten unterkomplex appräsentiert hat und damit vergleichsweise weniger authentisch war als der Abschlussfall. Denn unter der Prämisse, dass jeder Fall in eine Interaktionsgeschichte und Aktuelsituation mit vielfältigen Bedingungsfaktoren eingebunden ist, besteht die besondere diagnostische Anforderung an die Probanden in einer realitätsangemessenen Verknüpfung ganz unterschiedlicher Informationen zum Fall. Im Gegensatz zum Einstiegstest ermöglicht der Abschlusstest daher eine stärker realitätsbezogene und evidenzbasierte Diagnose. Wie solch eine Diagnose zukünftig auch in höheren  $K_{\text{imax}}$ -Werten ihren Niederschlag finden könnte, bedarf noch der weiteren Klärung.

Da zur Messung pädagogisch soziologischer Diagnosekompetenz der theoriegeleiteten Formulierung gefolgt werden musste, wurden die Teilkompetenzen der Probanden nach Subskalen getrennt untersucht. Des Weiteren wurde die Bearbeitungszeit für beide Tests variiert, um die abweichende Item-Anzahl zu berücksichtigen (36 Items im Einstiegstest, 100 Items im Abschlusstest). Die Verdreifachung der Bearbeitungszeit für den Abschlussfall hat die Verwendung von anteiligen Kompetenzwerten für die Berechnung der Testergebnisse ebenso wie für die Ausweisung von Kompetenzniveaus zugelassen (s. Kap. II.8.3.5). Dazu sei erwähnt, dass die absoluten  $K_i$ -Werte der Probanden stärkere Zuwächse zwischen Einstiegs- und Abschlusstestung zeigen, die aber unter Berücksichtigung der Bearbeitungszeit in den anteiligen  $K_i$ -Werten reichlich vermindert worden sind. Dieser Minderungseffekt wäre bei einer Verdopplung der Bearbeitungszeit wohl geringer ausgefallen, was mit Blick auf das Verhältnis der  $K_{\text{imax}}$ -Werte im Einstiegs- und Abschlussfall nachträglich plausibler erscheint. Für den Vergleich mehrerer Kompetenztests sollte die Anpassung der Bearbeitungszeit daher stärker vom  $K_{\text{imax}}$ -Wert als von der Item-Anzahl geleitet sein. Durch die parallele Entwicklung von Testinstrumentari-

um und Datenerhebung war dies im Rahmen unseres Forschungsprojektes jedoch nicht möglich.

### **8.3 Indikatoren zur Beurteilung schwieriger Texte in der internationalen Forschung**

Nach unserer Kenntnis gibt es derzeit keine Vorbilder für das Erfassen diagnostischer Kompetenzniveaus in Relation zur Schwierigkeit fallbasierter Aktualtexte. Auch liegen insgesamt nur wenige Daten mit Bezug auf diagnostische Kompetenzen von Studierenden im Allgemeinen oder im Lehramt im Besonderen vor (vgl. Kap. I.3, Dirks & Hansmann 2012). Wollen wir dennoch Anschlüsse an die internationale Kompetenzforschung herstellen, finden sich diese am ehesten in Studien über das Text- und Leseverständnis sowie über Prinzipien menschlichen Lernens. Die bereits andernorts näher ausdifferenzierten Schnittmengen zwischen Lese- und Diagnosekompetenz (s. Kap. V.2.2, Dirks 2012b) werden nunmehr hinsichtlich möglicher Indikatoren für Textschwierigkeiten mit Bezug auf die Domänenspezifität, höher- (und nieder-) schweligen Informationsverarbeitungsprozesse und der dabei erforderlichen Herstellung von Kohärenzen und der Bildung von Inferenzen diskutiert. Besondere Aufmerksamkeit gilt hier den Erkenntnissen über die Graduierung von Kompetenzausprägungen in Relation zu unterschiedlichen Anforderungsstrukturen der jeweils administrierten Aufgaben bzw. Texte. Vor diesem Hintergrund schließt das Kapitel mit einer näheren Beschreibung der in unserem Forschungsprojekt zugrunde gelegten diagnostischen Kompetenzniveaus. Zunächst sollen jedoch einige Überlegungen zu der Frage angestellt werden, inwieweit kompetenzrelevante Indikatoren für die Schwierigkeit eines Textes bzw. einer Testaufgabe *a priori* definiert werden können.

#### **8.3.1 Indikatoren für schwierige Anforderungsstrukturen: *Ex ante* oder *post hoc*?**

Mit Bezug auf den Lesekompetenztest der OECD monieren Alderson & Huhta (2011: 40): „neither the framework nor the described scales are explicitly tied to features that predict the difficulty of individual reading items in PISA.“ Das Testdesign sieht nämlich erst im Anschluss an die Testung die Ausweisung von Schwierigkeitsgraden vor. Gemeinhin wird also – wie übrigens auch in unserem Forschungsprojekt – *post-hoc* von der nachgewiesenen Häufigkeit eines erfüllten Items auf dessen Schwierigkeit geschlossen:<sup>8</sup> Wird ein Item sehr oft erfüllt, wird es als leicht bewertet und einer entsprechenden Kompetenzstufe zugeordnet („post-hoc-proficiency-scaling“). Welche Konsequenzen solch ein Vorgehen haben kann, soll – mangels unzureichend gewährter Einblicke in die PISA-Daten – mit Bezug auf ein anderes internationales Bildungsforschungsprojekt (TEDS-LT) mit Lehramts-

---

<sup>8</sup> Nur am Rande sei erwähnt, dass sich ein Item in der PISA-Lesestudie i.d.R. auf jeweils eine (meist geschlossene) Aufgabe bezieht, während die offenen Aufgaben in unserem Forschungsprojekt jeweils eine Vielzahl von Items umfassen.

studierenden verdeutlicht werden: Die Testung enthält eine Vielzahl reiner Wissensfragen, die in den Bereich deklarativen Wissens fallen. Verfügten überdurchschnittlich viele Probandengruppen nicht über das für eine Testantwort benötigte Begriffswissen (z.B. „Präsupposition“), wurden die betreffenden Aufgaben als schwer eingeordnet und diesbezügliche Kenntnisse vergleichsweise hoch bewertet (z.B. bei Bremerich-Vos et al. 2011: 64, 66; vgl. Kap. V.2.3.2, Dirks 2012b). Aus bildungstheoretischer Perspektive verkennen derart konstruierte Kompetenzniveaus die Qualifikationen der Probanden (vgl. Dirks 2006, Rosebrock 2009). So scheint die Validität der Testergebnisse nicht hinreichend gesichert, wenn die Testaufgabe den Probanden keine Gelegenheit bietet, das hier interessierende linguistische Phänomen unabhängig von ihrer Kenntnis des jeweiligen Fachbegriffs zu identifizieren. Diese Kritik verweist v.a. auf Defizite in der Testnormierung, die in dem jeweils herangezogenen Kompetenzmodell begründet ist und in obigem Fall möglicherweise nicht hinreichend berücksichtigt wurde.

Vor diesem Hintergrund sind neue Forschungen, die sich um eine *a priori* Definition von Item-Schwierigkeiten bemühen, höchst willkommen. Mit Bezug auf die PISA-Lesestudie 2009 hat z.B. ein von Alderson geleitetes Forschungsteam eine Reihe möglicher Indikatoren für schwierige Aufgaben bzw. Texte ex ante definiert und mithilfe von Regressionsanalysen eine moderate Prädiktorqualität herausgefunden. Dabei handelt es sich um die folgenden Indikatoren: „Number of features and conditions“, „proximity of pieces of required information“, „competing information“, „structural prominence of target information“, „reinforcement and repetition“, „semantic match between task and target information“, „concreteness of information“, „familiarity of information needed to answer the question“, „register of the text“ und „the extent to which information from outside the text is required to answer the question“ (Alderson & Huhta 2011: 41). Ein anderes Projekt der Forschungsgruppe ist – wie auch unser Forschungsprojekt – im Hochschulkontext angesiedelt, allerdings unter einer gänzlich anderen Zielsetzung: Anhand des „Pearson Test of English Academic“ (PTEA) sollen – ähnlich wie im IELTS oder TOEFL – jene Studierenden herausgefiltert werden, die über hinreichend englische Sprachkenntnisse verfügen, um an einer englischsprachigen Universität erfolgreich studieren zu können (a.a.O.: 44).

An diesen Bemühungen zeigt sich bereits, dass die Schwierigkeit einer Aufgabe stets in Relation zu den jeweiligen Adressaten bzw. Probanden<sup>9</sup> und z.B. ihren Vorkenntnissen beurteilt werden muss (vgl. Graves & Graves 2003, Kap. 9). Da sich diesbezügliche Variablen i.d.R. aber nur schwer kontrollieren lassen, muss sicherlich jede *a priori* Definition von Indikatoren für schwierige Texte und diesbe-

<sup>9</sup> Von besonderem Interesse sind die Ziele, unter denen Probanden eine Aufgabe bearbeiten. Zum Beispiel bilden Probanden, die zur Textlektüre für einen Test angehalten wurden, mehr kohärenzbasierte Inferenzen als Probanden, die den Text nur zur Unterhaltung lesen sollten (van den Broek et al. 2001). Daraus ergab sich die Schlussfolgerung, dass – zumindest in dem betreffenden Forschungs-Setting – eine grundsätzliche Bereitschaft zu textbasierten Bedeutungskonstruktionen im Falle einer Testsituation bestand.

zügliche Items kontinuierlich empirisch überprüft werden. Dennoch erscheint das Ziel einer *a priori*-Indikatorisierung, die sich am jeweiligen *state of the art* orientiert, erstrebenswert: Könnte doch auf diese Weise die derzeit gängige, aus forschungsökonomischer Perspektive aber ineffektive Praxis revidiert oder zumindest eingeschränkt werden, dass erst *nach* erfolgter Datenerhebung ungeeignete Items aussortiert werden.

Nur am Rande sei erwähnt, dass in den USA zwischen 1923 und Ende der 1960er Jahre mehr als 80 v.a. entwicklungspsychologisch fundierte „readability formulas“ vorgelegt wurden, welche die Angemessenheit eines Textes hinsichtlich der Schülerinteressen und -kenntnisse *ex ante* bestimmen sollten (Pearson 2009: 7). Im Zuge der kognitiven Wende mussten derartige Listen aber umfassend revidiert werden, ohne dass diese jemals hinreichend theoretisch und empirisch abgesichert worden wären (a.a.O.: 15f., Kintsch & Kintsch 2005: 85).

### **8.3.2 Domänenspezifisches Vorwissen als Indikator für Textschwierigkeit**

Sowohl Forschungsergebnisse aus der pädagogischen Psychologie als auch aus der Psycholinguistik dokumentieren mit Bezug auf ganz unterschiedliche Adressatengruppen und Forschungsdesigns übereinstimmend die Bedeutung domänenspezifischer Kenntnisse für erfolgreiches Lernen (Neubauer & Stern 2007) ebenso wie für (das dafür auch benötigte) Text- und Leseverständnis: „Comprehension is easy when the domain knowledge is high“ (Kintsch & Kintsch 2005: 84, vgl. Fink 2006). Leserinnen und Lesern fällt es im Allgemeinen leichter, neue Informationen angemessen zu verarbeiten, wenn sie diese in bereits bestehende Wissensstrukturen bzw. erfahrungsbedingte Vorkenntnisse integrieren können. Sind vertraute Themen und Gegenstandsbereiche einer Domäne auch noch in ein dazu passendes, ebenso vertrautes Genre eingebettet, werden diese im Allgemeinen schneller und besser verarbeitet als in einer Verknüpfung mit einem weniger vertrauten Genre (vgl. O'Reilly & McNamara 2007, Shapiro 2004). Denn das Erinnern von Textinhalten wird maßgeblich vom Arbeitsgedächtnis ebenso wie vom Vor- bzw. Weltwissen in sprachlicher und sachlicher Hinsicht bestimmt (vgl. z.B. Braasch & Goldman 2010, van den Broek et al. 1995, Wolfe & Goldman 2005, Zwaan & Singer 2003). Während also die in Fallvignetten apprizierte Domäne ‚Schule‘ schulerfahrenen Probanden, die zudem als Lehramtsstudierende eingeschrieben sind, kaum Verständnisschwierigkeiten bereiten dürfte, gilt dies nicht unbedingt auch für das ihnen meist weniger vertraute Genreformat des Transskriptes. Inwieweit die Nähe eines Transskriptes zu einer authentischen Situation (hier: Unterrichtliche Realsituation) und die das Transskript dominierenden – für die menschliche Informationsverarbeitung eher vorteilhaften – narrativen Sprachmittel (s.o.) dennoch kompensatorisch wirken können, ist eine offene Forschungsfrage.

Weitere Hinweise für die Schwierigkeit eines Textes und diesbezügliche Kompetenzanforderungen ergeben sich bei der Bearbeitung von Aufgaben, die den Transfer domänenspezifischen Wissens in eine neue Domäne erfordern. Da aus der pä-

dagogisch psychologischen Wissenstransferforschung bekannt ist, dass Transfers einmal erworbenen Wissens und Könnens in eine neue Domäne – selbst bei vergleichbarer Anforderungsqualität – meist gezielter Hilfe bedürfen (vgl. Mähler & Stern 2006), sollte in Prä-/Posttest-Forschungsdesigns darauf geachtet werden, dass sich die Testaufgaben auf weitgehend ähnliche Domänen beziehen. Dementsprechend ließen sich in unserem Forschungsprojekt sowohl der Einstiegs- als auch der Abschlussfall demselben Kernproblem einer ähnlichen Domäne zuordnen.

Der folgende Regelmechanismus soll als vorläufige Leitlinie für die Bestimmung eines mehr oder weniger schwierigen Textes im Hinblick auf seine Domänenspezifität gelten:

- In dem Maße, wie Textadressaten mit einer Domäne und ggf. einem Genre vertraut sind, bereitet ihnen das Verständnis eines thematisch entsprechend ausgerichteten Textes weniger Probleme.

Vor diesem Hintergrund erscheinen jene Forschungsergebnisse nur zu plausibel, die anstelle semantischer Kenntnisse eines Begriffs v.a. die Erschließung seiner pragmatisch kontextualisierten Tiefenbedeutung als Garant für ein angemessenes Textverständnis betonen (Ouellet 2006; Tannenbaum et al. 2006). Ob die für die Aufgabenbearbeitung erforderlichen Informationen im Text auch jeweils gefunden werden, mag ebenfalls durch Domänenkenntnisse beeinflusst sein.

### 8.3.3 Indikatoren für höherschwellige Informationsverarbeitungsqualitäten

Unser Forschungsinteresse an einer kompetenzmodell- und aufgabenangemessenen sowie adressatengerechten Definition unterschiedlicher Niveaus für Beschreibungen, Deutungen und Erklärungen ähnelt einem Kernanliegen der Lesekompetenzforschung (Alderson & Huhta 2011, Beck et al. 1989, Chi et al. 1994, Singer & Remillard 2004, Wilson & Anderson 1986). Aus kognitionswissenschaftlicher Perspektive zählen Diagnosekompetenzen nämlich – wie auch das Text- bzw. Leseverständnis – zu den höherschwelligen Informationsverarbeitungsprozessen (vgl. „higher-level comprehension processes“), die allerdings stets niederschwellige Informationsverarbeitungsprozesse (vgl. „lower-level information processes“) mit einschließen (Landi 2010; vgl. Perfetti et al. 2005, 2008). Textverständnis im weitesten Sinne ist daher als ein „multi-level process“ zu betrachten (Koda 2005), in dem im Übrigen die jeweilige Verarbeitungsebene nicht automatisch für leichter oder schwieriger befunden werden kann (vgl. Grabe & Stoller 2011: 13). Da diese Mehrschichtigkeit unterschiedlicher Bedeutungsebenen i.d.R. im Verlauf alters- und erfahrungsbedingter Entwicklungsprozesse erworben wird, erscheint es nur allzu plausibel, dass die Verfügbarkeit höherschwelliger Fähigkeiten weitgehend davon abhängt, ob die Probanden mehr oder weniger geübte Leserinnen und Leser sind (McNamara & O’Reilly 2012).

Alleine die als niederschwellig definierte Fähigkeit, automatisch Begriffe und basale Satzstrukturen zu identifizieren, würde also i.d.R. kein hinreichendes Textverständ-

nis ermöglichen (vgl. Paris & Hamilton 2009: 33). Eine fallbasierte Anforderungsstruktur, die auf die Erfassung von Textverständnis oder diagnostischen Leistungen zielt, sollte also auch eine höherschwellige Informationsverarbeitung mit Bezug auf Text und Kontext unter Zuhilfenahme von Inferenzen ebenso wie metakognitiven (vgl. „comprehension monitoring“) und selbst-regulatorischen Strategien (vgl. Landi 2010) ermöglichen, ohne dass diese aufgrund niederschwelliger kognitiver Prozesse grundlegend behindert würde – etwa durch unbekannte fach- oder fremdsprachliche Begriffe (vgl. „language threshold hypothesis“, Grabe 2009: 146ff). Diese Voraussetzung scheinen insbesondere alltagssprachliche, zudem in der Muttersprache der Probanden generierte Texte – wie die vorliegenden Unterrichtstranskripte – zu erfüllen.<sup>10</sup>

Für die Auslotung individueller Kompetenzausprägungen in der höherschwelligen Informationsverarbeitung sind diesbezügliche Indikatoren bis dato nur unzureichend entwickelt. Dies mag in dem Umstand begründet sein, dass die Leseverständnisforschung vorrangig Kinder und deren *niederschwellige* Informationsverarbeitungsprozesse untersucht hat (vgl. Landi 2010, Paris & Hamilton 2009). Wurden *höberschwellige* Informationsverarbeitungsprozesse erforscht, bestanden die Probandengruppen meist aus Jugendlichen, aber wiederum kaum aus Erwachsenen (wie z.B. Studierenden). Die aus diesen Forschungen hervorgegangenen Erkenntnisse können also keine altersunabhängige Geltungsrelevanz beanspruchen und daher nicht ungeprüft für Textschwierigkeitsindikatoren für eine andere Altersgruppe genutzt werden. Hinzu kommt, dass in den Forschungs-Settings meist mit kürzeren Texten gearbeitet wurde, das Verständnis von Kurztexen aber – so die Kritik von Susan Goldman (2004) – kaum Ergebnisse über Kompetenzen im Bereich der Integration und Verknüpfung vielfältiger Informationen enthalten kann, wie sie für höherschwellige kognitive Prozesse typisch sind. Nicht zuletzt aufgrund fehlender Vorbilder hatten wir uns in unserem Forschungsdesign deshalb zunächst auf die Verarbeitungsprozesse des Deutens und Erklärens beschränkt, eine Erweiterung des Kompetenzmodells mit Bezug auf metakognitive Diagnosekompetenzen aber bereits angedacht (vgl. Kap. II.5.1, Dirks 2012a).

Unter o.g. Vorbehalt hinsichtlich einer altersunabhängigen Generalisierbarkeit seien im Folgenden einige Forschungsergebnisse präsentiert, die möglicherweise Hinweise für die Einordnung schwieriger Texte und diesbezügliche v.a. höherschwellige Kompetanzanforderungen bieten. Zur besseren Übersichtlichkeit wird hier die Bildung von Inferenzen und Kohärenzen separat behandelt, wenngleich sich diese auch wechselseitig beeinflussen.<sup>11</sup>

---

10 Vorsicht ist wohl v.a. bei der Administration von Fällen aus mathematisch-naturwissenschaftlichen Fächern geboten, wenn in den betreffenden Unterrichtsausschnitten viele – für das Textverständnis benötigte – Fachbegriffe verhandelt werden (Snow 2010).

11 Inwieweit die von Goldman (2004) formulierte Kritik an Forschungsergebnissen über höherschwelliges Verständnis anhand von Kurztexen auf die folgenden Studien zutrifft, konnte nicht im Einzelnen überprüft werden. Ich beziehe mich daher auf alle nachfolgend

*Bedingungsfaktoren für eine lokale und globale Kohärenzbildung*

Insgesamt scheint die Kohärenz eines Textes umso bedeutsamer zu sein, je geringer das Vorwissen der Probanden ausgeprägt ist (vgl. McNamara & Kintsch 1996, McNamara et al. 1996). Im Umkehrschluss ist davon auszugehen, dass sich Diskohärenzen in einem Text mithilfe domänenspezifischen Wissens kompensieren lassen, was noch einmal die Bedeutung von Indikatoren zum Vorwissen, aber auch die enge Verzahnung von nieder- und höherschwelligen Informationsprozessen unterstreichen würde (s.o.). Einer den Probanden vertrauten Themenwahl ist demnach umso mehr Gewicht beizumessen.

Welche linguistischen Phänomene werden gemeinhin mit einer *lokalen Kohärenzbildung* in Verbindung gebracht? Diese Frage soll hier als Erstes erörtert werden, da lokale Kohärenzbildungen eine zentrale Prämisse für globale Kohärenzbildungen darstellen.

Im Hinblick auf die – meist genrespezifisch geprägte – *Textstruktur* sei zunächst folgendes Forschungsergebnis erwähnt: Leserinnen und Leser, die gute Zusammenfassungen produzieren, orientieren sich an der Textstruktur (Meyer et al. 1980). Dieses Ergebnis erscheint auch vor dem Hintergrund plausibel, dass die korrekte Lokalisierung der jeweils benötigten Informationen durch ihre textstrukturelle Einbettung (Cataldo & Oakhill 2000, Lorch et al. 2001) und die Betitelung einzelner Abschnitte (Lorch et al. 2001) gestützt wird. Dass die Paragraphierung eines Textes verständnisförderliche Effekte erzeugt, wird auch in den Forschungen von Khalifa & Weir (2009: 52), Gernsbacher (1990), Meyer & Wijekumar (2007), Ozuru et al. (2008) sowie Williams (2007) bestätigt.

Nutzen wir die Befunde zu Auswirkungen der Textstruktur auf das Textverständnis für eine kritische Betrachtung der in unserem Forschungsprojekt eingesetzten Transskripte, muss wohl ihrer monotonen Struktur umso mehr Gewicht beigemessen werden: Fehlende Strukturierungshilfen könnten die Erfassung des Hauptgedankens (s.u. zur Herstellung globaler Kohärenz) und weitere Elaborierungen bezüglich fallimmanenter Informationen erschweren (vgl. auch Kap. II.8.2 zur narratologischen Begründung von Textschwierigkeiten).

Die Bedeutung lokaler Kohärenzen beim Lesen manifestiert sich zudem in der Fähigkeit, anaphorische Referenzen im Text herstellen zu können (Oakhill & Yuill 1986, Yuill & Oakhill 1988). Wenn Probanden in unserem Forschungsprojekt den im Transskript als „Herr Mustermann“ eingeführten Lehrer aufgrund seines – später im Text erfolgten – ironischen Kommentars als weiblich identifizieren, ist dies ein Indikator für eine unzureichend ausgebildete anaphorische Referenz.

Weitere signifikante Kohärenzmerkmale für das Lese- bzw. Textverständnis wurden u.a. mit Bezug auf bestimmte Konnektoren, syntaktische Konstruktionen, die Anordnung von Wörtern und Sätzen sowie Lücken im Text gefunden (vgl. Fletcher

---

aufgeführten Forschungsergebnisse unter dem Vorbehalt ihrer Geltungsrelevanz für ein höherschwelliges Textverständnis.

et al. 1995, Goldman & Rakestraw 2000, Graesser et al. 1997, 2004; Halliday & Hasan 1976, Harley 2008, Hobbs 1990, Kamalski et al. 2008, Kehler 2004, Rickheit & Habel 1995, Lorch & O'Brien 1995, Myers 1990). Zum Beispiel scheinen komplexe syntaktische Strukturen die Kohärenz eines Textes zu beeinträchtigen und sich negativ auf das Textverständnis auszuwirken (Rowe et al. 2006). Auch werden semantische Wechselbezüge offenbar v.a. von weniger erfahrenen Lesern für das Herstellen von Kohärenzen genutzt, während geübte Leser dazu eher kausale Bezugnahmen für wichtig erachten (Torado et al. 2010). Insbesondere in Fällen, in denen kausale Textbezüge verdeckt bleiben, werden kausale Konnektoren wie „weil“, „da“ für ein gelungenes Textverständnis benötigt (vgl. Graesser et al. 1997: 180).

Da die in unserem Forschungsprojekt eingesetzten Unterrichtstranskripte kaum explizite Kohärenzmerkmale enthalten, erfordern sie von den Probanden mutmaßlich ein umso genaueres und kognitiv aufwändigeres Erfassen der Bedeutungsinhalte in Raum und Zeit.

Die *globale Kohärenzbildung* betrifft etwa die Fähigkeit, Zusammenhänge zwischen sukzessiv aufeinanderfolgenden Themen herzustellen (Lorch et al. 1987) oder den Hauptgedanken aus der Informationsvielfalt eines Textes herauszuarbeiten (vgl. Palincsar & Brown 1984, Stevens 1988, Williams 1986). Wie bereits weiter oben erwähnt, sind beide Kompetenzen mit textstrukturellen Bedingungen eng verstrickt. Analogien ergeben sich mit Bezug auf jene Probanden in unserem Forschungsprojekt, die trotz vieler, z.T. unterbrochener Sprecherwechsel und wechselnder Gesprächsanlässe im Abschlussfall das dort virulente Thema einer Unterrichtsverzögerung im Spannungsfeld von Schulzwang und hoheitsstaatlichem Mandat des Lehrers diagnostizieren. Werden eigentlich zusammengehörende „Turn Construction Units“ (Sacks et al. 1974: 702, Ford et al. 1996) auseinandergerissen, wird der Text unübersichtlicher und erhöht sich seine Schwierigkeit. Prinzipiell ist davon auszugehen:

- Je enger die Sinneinheiten eines Textes und die entsprechenden Bedeutungsdimensionen miteinander verknüpft sind, je kohärenter ein Text also erscheint, desto erfolgreicher vollzieht sich vermutlich die jeweils erforderliche Informationsverarbeitung.

Nahezu alle der im Fallvergleich herausgearbeiteten potentiellen Indikatoren für die Textschwierigkeit könnten auch für eine textverständnisförderliche Kohärenzbildung relevant sein, insbesondere im Bereich „Dis-/Kohärenzen: Räumliche Verteilung aufeinander bezogener ‚Turns‘“ (s.o., Kap. II.8.2). Dementsprechend erscheint es unumstritten, dass Textkohäsion sowie textbezogene Bedeutungspotentiale und -konstruktionen für das Textverständnis von entscheidender Bedeutung sind (vgl. Gernsbacher 1997, McNamara et al. 1996).

#### *Anforderungen an die Bildung von Inferenzen: Informationslücken und -fehler*

Unter der Voraussetzung, dass die für das Textverständnis erforderlichen Domänenkenntnisse gesichert sind (s.o.), erscheint es nur allzu plausibel, dass explizit erwähnte Informationen leichter prozessiert werden als jene Informationen, die erst



mithilfe semantischer oder pragmatischer Implikationen (Bach 1994, Krifka 2009, Levinson 1983 u.a.) bzw. Präsuppositionen (vgl. Levinson 1983, Palmer 1985, Stalnaker 2002, Tiemann et al. 2011) inferiert werden müssen (vgl. Josephson & Josephson 1994). Mit anderen Worten: Geübte Interpreten ‚von Welt‘ sind in der Lage, nicht zuletzt aufgrund ihrer Domänenkenntnisse, fehlende Informationen zu inferieren oder ungenaue, wenn nicht falsche Sachverhaltsangaben zu korrigieren, während weniger geübte Leser bestehende Informationslücken gar nicht erkennen und sich leichter zu Fehlannahmen verleiten lassen (Garnham et al. 1982; Long et al. 1994, Magliano & Millis 2003; Magliano et al. 2002, Oakhill 1984, Oakhill & Yuill 1996; Oakhill et al. 1990; Yuill et al. 1989).

Welche Inferenzen jeweils für relevant befunden werden, wird meist mit Bezug auf verschiedene Aspekte des Text- und Diskursverständnisses bzw. aus diskurs- und kohärenztheoretischer Perspektive diskutiert (z.B. Chalhoub-Deville et al. 2006, Graesser & Bower 1990, Graesser et al. 1997, Irmer 2011, Rickheit & Strohner 1985, Roeper 2011, Schmalhofer & Perfetti 2007, Vellutino 2003; zu sprechaktbezogenen „Relevance Topics“ s. Repp 2011). Dabei wird – auch im Anschluss an das sog. Prinzip „Maximize Discourse Coherence“ (Asher & Lascarides 1998b: 107) – ein gelungenes Textverständnis eher auf kausale erklärungsbasierte Inferenzen als auf lediglich additiv aneinandergefügte oder vermutete (kausale) Konsequenzen und assoziierte elaborierte Inferenzen zurückgeführt (mit Bezug auf Kurz-/Geschichten s. Graesser et al. 1994, Trabasso & Magliano 1996, Wolfe et al. 2005, Zwaan & Brown 1996; mit Bezug auf technische Gebrauchstexte s. Chi et al. 1994 und auf Presstextberichte s. Sanders & Noordman 2000).

Dementsprechend liegt der Analyseschwerpunkt von Ursache-Wirkungs-Beziehungen u.a. auf Erklärungen, Erwartungsinterferenzen und „denials of preventer“, aber auch auf Ähnlichkeitsbeziehungen (Parallelen, Kontraste, Exemplifizierung, Generalisierung, Ausnahmen, Elaborierung) und Kontiguitätsbeziehungen in Zeit und Raum (z.B. Inferenzen aufgrund des Wandels bestimmter Phänomene hinsichtlich ihres Anfangs- oder Endstadiums; vgl. Kehler 2004: 246ff., zu weiteren Beziehungskonzepten s. Sanders & Spooren 2006). Psycholinguistische Lesetestarrangements, die in besonderer Weise die für das Textverständnis erforderlichen Inferenztypen auf der Diskursebene berücksichtigen, finden sich z.B. in den sog. MOCCA-Tests („Multiple-choice, Open-ended, Cloze Comprehension Assessment“, Carlson et al. 2011, vgl. auch van den Broek et al. 2001, Rapp et al. 2007).

Für die Prozessierung von Inferenzen sind – neben expliziten Informationen – eine Vielzahl linguistischer Phänomene von zentraler Bedeutung. Eine entsprechend enge Verknüpfung von nieder- und höherschweligen Informationsverarbeitungsprozessen finden wir v.a. im Rahmen der bereits oben angedeuteten epistemischen Modalitäten sowie der ko- und kontextspezifischen Anaphora (Irmer 2011, Kaiser 2011, Pu 2011; zum ‚Centering‘ im natürlichen Sprachgebrauch s. Walker et al. 1998). Ziehen wir informationstheoretische Arbeiten über die syntaktische Unterscheidung zwischen „givenness/newness“ bzw. „focus“ und „topic“ (s. den Über-

blick bei Hempelmann et al. 2005)<sup>12</sup> und die dabei auftretenden Konkurrenzen (vgl. Cook & Bildhauer 2011) zu Rate, finden wir hier weitere Hinweise auf die (domänenspezifische) höher- und niederschwellige Vernetztheit der für ein adäquates Textverständnis erforderlichen Inferenzen.

Abschließend sei auf die Forschung über „self-explanation effects“ und die dabei vollzogenen kausalen Inferenzen verwiesen (Chi 2000; Chi et al. 1989, 1994; McNamara 2004, 2006). In diesen Forschungs-Settings werden die Leserinnen und Leser eines Textes angehalten, die in dem Text enthaltenen Erklärungen eigenständig zu erarbeiten. Im Prinzip bietet der Ansatz umfassenden Einblick in viele der bereits skizzierten Phänomene, insbesondere in die Art und Weise, wie Inferenzen gebildet, verschiedene – textimmanente mit kontextspezifischen oder vertraute mit neuen – Informationen verknüpft oder auch falsche Angaben korrigiert werden (Roy & Chi 2005). Die Fähigkeit, die Bedeutung eines Textes quasi naturwüchsig erklären zu können, lässt offenbar auch Aussagen über eine höhere Problemlösekompetenz und ein insgesamt tieferes Textverständnis der betreffenden Probanden zu (Chi et al. 1989, 1994).

Kompetenzen des Erklärens, die ja auch für unseren Ansatz von zentraler Bedeutung sind, werden hier einer übergeordneten Verstehenskompetenz subsumiert (Chi 2000, McNamara et al. 2006). Deren Erfassung ist jedoch – ähnlich wie in unserem Forschungsdesign – mit einem immensen Zeit- und Analyseaufwand verbunden. So können die Testantworten sehr lang ausfallen und müssen von verschiedenen Codierern per Hand ausgewertet werden (Magliano et al. 2002; Magliano et al. 2005; McNamara et al. 2004, 2006; Millis et al. 2004). Bis dato sind aus diesen Forschungen keine Modelle für Graduierungen der jeweils erfassten Kompetenzausprägungen hervorgegangen.

Dass sich die für ein angemessenes Textverständnis benötigten Inferenzen auf verschiedene Ebenen herunterbrechen lassen (Cain et al. 2001), scheint unbestritten. Mit unserem Vorschlag, sich dabei an der Logik sozialer Realitäten auf der Basis sozialtheoretisch fundierter Expertise zu orientieren (vgl. Kap. II.5, Dirks 2012a), betreten wir aber offensichtlich Neuland.

### **8.3.4 Ein Graduierungsbeispiel aus der internationalen Lesekompetenzforschung (PISA)**

Welche Indikatoren für die Beurteilung der Schwierigkeit von Texten herangezogen werden können, soll am Beispiel des PISA-Lesekompetenztests (vgl. „reading literacy“) verdeutlicht werden. Aufgrund seiner weltweiten Rezeption kann er als Normativitätsfolie für den internationalen Forschungsstand gelten. Da die nationalen Expertengruppen z.T. variierende Lesekompetenzmodelle herangezogen haben,

---

<sup>12</sup> Um textbezogene Bedeutungsgehalte in Relation zur Informationsstruktur eines Textes zu erschließen und daraus mögliche Indikatoren für Item-Schwierigkeiten abzuleiten, könnte auch die von Asher & Lascarides (2003) entwickelte „Segmented Discourse Representation Theory“ (SDRT) weitere Hinweise liefern.

sei vorab erwähnt, dass ich mich auf das von der OECD (2009) definierte Test-Setting beziehe, wie es in den USA mit überwiegend induktiv generierten Konstruktvariablen aus einer primär pragmatischen Perspektive umgesetzt wurde (Kirsch & Mosenthal 1990; Kirsch et al. 2002: 16ff., 23ff.). Neben der Abstraktheit von Informationen (Mosenthal & Kirsch 1998) findet v.a. die textspezifische Informationsvielfalt, -komplexität und -in-/transparenz sowie das Vorwissen der Probanden Berücksichtigung:

- Je mehr Informationen ein Text enthält, je vielfältiger und ungeordneter die Beziehungen konkurrierender Informationen im Text zueinander stehen und je weniger vertraut die thematischen Kontexte erscheinen oder je mehr Vor- und Kontextwissen aktiviert werden muss, desto voraussetzungsvoller werden Reflexions- und Bewertungsaufgaben bspw. hinsichtlich des Bildens und Überprüfens von Hypothesen (vgl. Inferenzen) und desto schwieriger wird das Text- bzw. Leseverständnis (vgl. OECD 2009: 45, 75 sowie Crossley et al. 2008, Green et al. 2010).

Entsprechende Wechselbeziehungen wurden bereits in den vorangegangenen Abschnitten II.8.2 und II.8.3 diskutiert und scheinen – zumindest für 15-jährige Schülerinnen und Schüler – empirisch abgesichert zu sein. Mit Bezug auf „kontinuierliche“ Texte, zu denen Transskriptionen gezählt werden können, konkretisiert die PISA-Expertengruppe der USA die folgenden Einflussfaktoren für schwierige Texte:

... the length of the text; the explicitness and transparency of its structure; how clearly the parts are related to the general theme; and whether there are text features such as paragraphs or headings, and discourse markers such as sequencing words. (OECD 2009: 45)

Die Lesekompetenzanforderungen werden innerhalb der folgenden Skalen näher ausdifferenziert:

In *access and retrieve* tasks, difficulty is conditioned by the number of pieces of information that the reader needs to locate, by the amount of inferencing required, by the amount and prominence of competing information, and by the length and complexity of the text.

In *integrate and interpret* tasks, difficulty is affected by the type of interpretation required (for example, making a comparison is easier than finding a contrast); by the number of pieces of information to be considered; by the degree and prominence of competing information in the text; and by the nature of the text: the less familiar and the more abstract the content and the longer and more complex the text, the more difficult the task is likely to be.

In *reflect and evaluate* tasks, difficulty is affected by the type of reflection or evaluation required (from least to most difficult, the types of reflection are: connecting; explaining and comparing; hypothesising and evaluating); by the nature of the knowledge that the reader needs to bring to the text (a task is more difficult if the reader needs to draw on narrow, specialised

knowledge rather than broad and common knowledge); by the relative abstraction and length of the text; and by the depth of understanding of the text required to complete the task (OECD 2009: 45).

„Level	Score points on the PISA scale
5	More than 625
4	553 to 625
3	481 to 624
2	408 to 480
1	335 to 407
Below level 1	Less than 335“

Tab. 2: Schwierigkeitsstufen der Lesekompetenzskalen in PISA 2009 (OECD 2009: 76)

Um die Schwierigkeit einer Aufgabe besser erfassen zu können, wurde jede der drei Subskalen in fünf Stufen unterteilt (OECD 2009: 75f., vgl. Tab. 2). Je höhere Anforderungen ein Text und eine Leseaufgabe an die Probanden stellen, desto höher wurde die erbrachte Leistung bewertet (vgl. Tab. 2). Die Stufen beruhen demnach auf der Annahme einer systematischen Progression. Das heißt, es wird im Prinzip davon ausgegangen, dass jene Probanden, die bereits über eine höher bewertete Kompetenzausprägung verfügen, zugleich den darunter liegenden Kompetenzanforderungen genügen (OECD 2009: 76). Eine solche Regelmäßigkeit konnte in unserem Forschungsprojekt nur teilweise beobachtet werden. Dennoch halten wir eine Graduierung der von uns ausdifferenzierten zwei Kompetenzskalen für sinnvoll (s. Kap. II.8.3.5).

### 8.3.5 Die Graduierung diagnostischer Kompetenzen in unserem Forschungsprojekt<sup>13</sup>

Zur Beurteilung der Diagnosekompetenz von Lehramtsstudierenden haben wir im Rahmen unseres Forschungsprojektes Kompetenzniveaus ausdifferenziert, die sich im Wesentlichen am theoretischen Kompetenzmodell mit seinen zwei Skalen und den für die Skalen maßgeblichen Referenzbereichen orientieren (vgl. Kap. II.5, Dirks 2012a). Letztere werden als Handlungs-/Strukturdimensionen sozialer Realitäten definiert, die sich in eine Situations-, Selektions- und Aggregatlogik untergliedern lassen. Vor diesem Hintergrund weist die folgende Tabelle Deskriptoren für fünf Kompetenzniveaus auf, die sich – neben dem Kompetenzmodell – auch auf die in den vorangegangenen Abschnitten erörterten Erkenntnisse aus anderen Forschungszusammenhängen stützen. Die Niveaubeschreibungen enthalten demnach v.a. Deskriptoren bezüglich der Wiedergabe und des Inferierens von Informationen

<sup>13</sup> Dieses Kapitel ist unter Mitarbeit von Hendrik Baumbach entstanden. Für seine Hinweise sei ihm an dieser Stelle herzlich gedankt.

unter Berücksichtigung ihrer Komplexität und Kohärenz im Transskript. Auf nähere Angaben zum (domänenspezifischen) Vorwissen der Probanden wurde verzichtet, da dieses im Rahmen der Testarrangements vorausgesetzt werden konnte.

Dass die ausgewerteten Testdaten letztendlich nur den Kompetenzniveaus 1 bis 3 zugeordnet wurden, ist wohl zum Einen auf die literalen Grundfähigkeiten der Probanden zurückzuführen, weshalb KN 0 ausgeschlossen werden konnte, und zum Anderen auf die begrenzte Bearbeitungszeit, die Performanzen entsprechend des KN 3+ verunmöglicht hat.

Kompetenz-niveau	$K_i$	Skala 1: Beschreibungen	Skala 2: Verstehen & Erklären
<b>KN 0:</b> Bereich der marginal-zufälligen Item-Treffer	[0, 20)	Diese Personen zeigen kaum deskriptive Fähigkeiten. Sie haben der zu bearbeitenden Fallvignette nur eine geringe Anzahl von Informationen entnommen, ggf. nur mit Bezug auf einzelne situations- oder handlungs- oder effektspezifische Dimensionen. Die Informationswiedergabe lässt keine inhaltsangemessene (sozial- und handlungstheoretisch begründbare) Zuordnung zu bestimmten Dimensionen sozialer Realitäten bzw. des Falls erkennen.	Diese Personen zeigen kaum analytische Fähigkeiten. Ihre Diagnosen enthalten allenfalls fall-spezifisch unangemessene Inferenzen, d.h. die inferierten Kontextannahmen werden in keinen logisch nachvollziehbaren (sozial- und handlungstheoretisch begründbaren) Zusammenhang gebracht.
<b>KN 1:</b> Unteres Kompetenz-niveau	[20,40)	Diese Personen können nur wenige, im Fall appräsenierte (explizit erwähnte) Dimensionen sozialer Realitäten beschreiben, so dass diese weitgehend unvollständig bleiben. Die wiedergegebenen Informationen entstammen aber durchaus verschiedenen Handlungs-/Strukturdimensionen, die in einer – dem Fall angemessenen – Art und Weise erfasst werden.	Diese Personen können einzelne fehlende Informationen durch Inferenzen ergänzen. Die Inferenzen zeichnen sich jedoch überwiegend durch Unterkomplexität aus: Sie basieren auf einfach kontingenten Verknüpfungen oder lassen eine einseitige Fokussierung bestimmter Realitätsdimensionen erkennen (z.B. ausschließlich mit Bezug auf das Akteurs-Framing unter Vernachlässigung der Aggregatlogik).
<b>KN 2:</b>	[40,60)	Diese Personen können der Fallvignette zahlreiche, z.T. im Text verstreute Informationen	Diese Personen können naheliegende Inferenzen vornehmen und dabei mehrere Dimensionen

<b>Mittleres Kompetenz-niveau</b>		mit Bezug auf verschiedene Realitätsbereiche entnehmen. Diesbezügliche Handlungs-/Strukturdimensionen werden von ihnen in einen sinnvollen Zusammenhang gebracht, sowohl anhand einfacher als auch komplexer Verknüpfungen.	fallbezogener Realitäten analytisch einbinden. Die jeweiligen Dimensionen werden sowohl situations- als auch selektions- und aggregationslogisch in angemessener Weise miteinander verknüpft, bleiben jedoch in einigen zentralen Bereichen unvollständig.
<b>KN 3: Oberes Kompetenz-niveau</b>	[60,80)	Diese Personen können fast alle der im Fall appräsentierten Dimensionen sozialer Realitäten beschreiben und sinnvoll miteinander verknüpfen, auch wenn diese im Text weit zerstreut oder nicht durch bestimmte Sprachmittel (Kohäsionsmittel) verknüpft sind.	Diese Personen können fehlende Informationen inhaltsangemessen ergänzen und umfassend kontextualisierte Inferenzen unter Berücksichtigung der Situations-, Selektions- und Aggregatlogik bilden. Dabei werden einzelne kausal-funktionale Verknüpfungen detailliert analysiert.
<b>KN 3+: Bereich oberhalb der erwarteten Performanz</b>	[80,100)	Diese Personen können die im Fall appräsentierten Informationen praktisch lückenlos, ggf. anhand der algorithmischen Steps beschreiben und sinnvoll miteinander verknüpfen.	Diesen Personen gelingt es, eine vollständige und mit Bezug auf den Fall umfassend begründete, ggf. algorithmisch strukturierte Diagnose zu erstellen, die sich auf zahlreiche Detailanalysen stützt.

Tab. 3: Deskriptoren der Kompetenzniveaus mit Bezug auf Diagnosekompetenzen

Der maximale Kompetenz-Wert ( $K_{i,max}$ ) wurde in unserem Forschungsprojekt mit 100% angenommen und der von den einzelnen Probanden erreichte  $K_i$ -Wert dazu ins Verhältnis gesetzt. Damit sind beide Tests auf 100 normiert; auch werden die Spannweiten der Kompetenzniveaus durch die gleichen  $K_i$ -Werte in beiden Tests definiert (vgl. Kap. II.6, Baumbach 2012). Wir folgen damit der Annahme, dass die Kompetenzwerte auf einer Skala zwischen 0 und 100 gehalten werden können, wenn die Schwierigkeit des Tests und die Arbeitszeit in ein proportionales Verhältnis gesetzt werden. Weiterhin erhöht die Arbeitszeit die Textmenge der Analyse, was i.d.R. eine höhere (zufällige) Item-Treffer-Anzahl bedingt. Zugleich ermöglicht eine längere Fallvignette, zu der meist auch ein größerer Item-Pool generiert wird, mehr Item-Treffer. Aus diesem Grund wurde das untere Kompetenzniveau (KN 1) ab einem anteiligen  $K_i$ -Wert von 20 angenommen und nicht etwa bei 0 angesetzt. Dass diese Niveaueinteilung für anteilige und damit leichter vergleichbare Kompetenzwerte sinnvoll ist, zeigen u.a. die in der Kontrollgruppe 3 (KG3) erzielten Kompetenzergebnisse (s. Kap. III.1, Baumbach & Hansmann 2012).

Abschließend sei nochmals daran erinnert, dass das Kompetenzmodell in zukünftigen Forschungen auch metakognitive Dimensionen enthalten soll (vgl. Kap. II.5, Dirks 2012a). Die in obiger Tabelle 3 näher beschriebenen Kompetenzniveaus müssten dann entsprechend ausdifferenziert werden. Für die Auswertung der Testdaten in unserem Forschungsprojekt war jedoch noch keine Modellerweiterung notwendig, da die Testaufgaben von den Probanden keine diesbezüglichen Performanzen verlangten.

## Literatur

- Alderson, Charles J. & Ari Huhta (2011). Can research into the diagnostic testing of reading in a second or foreign language contribute to SLA research? *EUROSLA Yearbook 11*, 30-52.
- Asher, Nicholas & Alex Lascarides (1998a). Questions in Dialogue. *Linguistics and Philosophy*, Vol. 23, Issue 3, 237–309.
- Asher, Nickolaus & Alex Lascarides (1998b). Bridging. *Journal of Semantics*, Jg. 15, No. 1, 83-113.
- Asher, Nicholas & Alex Lascarides (2003). *Logics of Conversation*. Cambridge, MA: CUP.
- Asher, Nicholas & Arthur Lascarides (2006). Indirect Speech Acts. S. 1-37, [URL: <http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&ved=0CDgQFjAB&url=http%3A%2F%2Fciteseerx.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.106.2978%26rep%3Drep1%26type%3Dpdf&ei=kKVcUMuRNY3YyAHAIIgDw&usq=AFQjCNGxMcoMsziB1jDVHBien30bvo6t6g>, Zugriff 20.09.2012].
- Bach, Kent (1994). Conversational Implicature. *Mind and Language* 9, 124–162.
- Baumann, Gerd (1996). *Contesting Culture: Discourses of identity in multi-ethnic London*. Cambridge: CUP.
- Baumbach, Hendrik (2012). Das Messmodell. Messverfahren zur Analyse fallbasierter Diagnosekompetenz [21 Seiten]. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), *Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium*. (Kap. II.6, Online-Publikation der Philipps-Universität Marburg). [URL: <http://archiv.ub.uni-marburg.de/es/2012/0022/view.html>].
- Baumbach, Hendrik & Wilfried Hansmann (2012). Entwicklung pädagogisch soziologischer Diagnosekompetenz von Lehramtsstudierenden. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), *Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium*. (Kap. III.1, Online-Publikation der Philipps-Universität Marburg). [[http://archiv.ub.uni-marburg.de/opus/schriftenreihen\\_ebene2.php?sr\\_id=30&la=de](http://archiv.ub.uni-marburg.de/opus/schriftenreihen_ebene2.php?sr_id=30&la=de)].

- Beck, Isabel; Margaret McKeown & Erika Gromoll (1989). Learning from social studies texts. *Cognition and Instruction*, 6, 99–158.
- Braasch, Jason L.G. & Susan Goldman (2010). The role of prior knowledge in learning from analogies in science texts. *Discourse Processes*, 47, 447–479.
- Bremerich-Vos, Albert; Jutta Dämmer, Heiner Willenberg & Knut Schwippert (2011). Professionelles Wissen von Studierenden des Lehramts Deutsch. In: Sigrid Blömeke; Albert Bremerich-Vos, Helga Haudeck, Gabriele Kaiser, Günter Nold, Knut Schwippert & Heiner Willenberg (Hrsg.), *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen – Erste Ergebnisse aus TEDS-LT* (S. 47-76). Münster: Waxmann.
- Brooks, Peter (1992 [1984]). *Reading for the Plot. Design and Intention in Narrative*. Cambridge, Mass: Harvard UP.
- Bruner, Jerome (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.
- Cain, Kate; Jane V. Oakhill, Marcia A. Barnes & Peter E. Bryant (2001). Comprehension skill, inference-making ability, and the relation to knowledge. *Memory & Cognition*, 29, 850–859.
- Carlson, Sarah; Ben Seipel & Kristen McMaster (2011). Using a New Reading Comprehension Assessment to Measure Discourse Representations and Identify Types of Comprehenders. Society for Research on Educational Effectiveness (SREE) [URL: [http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?\\_nfpb=true&&ERICExtSearch\\_SearchValue\\_0=ED528950&ERICExtSearch\\_Search\\_Type\\_0=no&accno=ED528950](http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&&ERICExtSearch_SearchValue_0=ED528950&ERICExtSearch_Search_Type_0=no&accno=ED528950), Zugriff 30.10.2012].
- Cataldo, Maria G. & Jane V. Oakhill (2000). The effect of text organization (original vs. scrambled) on readers' ability to search for information. *Journal of Educational Psychology*, 92, 791-799.
- Chalhoub-Deville, Micheline; Carol A. Chapelle & Patricia A. Duff (2006). *Inference and Generalizability in Applied Linguistics: Multiple Perspectives*. Amsterdam: Benjamins.
- Chi, Michelene T. H. (2000). Self-explaining expository texts: The dual process of generating inferences and repairing mental models. In: Robert Glaser (Hrsg.), *Advances in instructional psychology* (S. 161–238). Mahwah, NJ: Erlbaum.
- Chi, Michelene T. H.; Miriam Bassok, Matthew W. Lewis, Peter Reimann & Robert Glaser (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145–182
- Chi, Michelene T. H.; Nicholas De Leeuw, Mei-Hung Chiu & Christian LaVancher (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439–477.



- Cook, Philippa & Felix Bildhauer (2011). Annotating information structure: The case of topic. In: Stefanie Dipper & Heike Zinsmeister (Hrsg.), *Beyond Semantics: Corpus-based Investigations of Pragmatic and Discourse Phenomena* (S. 45-56). Bd. 3 der *BLA (Bochumer Linguistische Arbeiten)*.
- Corkill, Alice J.; John A. Glover & Roger H. Bruning (1988). Advance organizers: Concrete versus abstract. *Journal of Educational Research*, 82(2), 76-81.
- Coté, Nathalie; Susan R. Goldman & Elizabeth U. Saul (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, 25, 1–53.
- Crossley, Scott A.; Jerry Greenfield & Danielle S. McNamara (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42(3), 475-493.
- 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 (2000). *Wie werden EnglischlehrerInnen professionell? Eine berufsbio-graphische Untersuchung in den neuen Bundesländern*. Münster: Waxmann.
- Dirks, Una (2010). *Der Irak-Konflikt in den Medien. Eine sprach-, politik- und kommunikationswissenschaftliche Analyse*. Konstanz: uvk.
- Dirks, Una (2012a). Pädagogisch soziologische Diagnosekompetenz im Modell [22 Seiten]. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), *Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium*. (Kap. II.5, Online-Publikation der Philipps-Universität Marburg). [URL: <http://archiv.ub.uni-marburg.de/es/2012/0018/view.html>].
- Dirks, Una (2012b). 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), [URL: <http://archiv.ub.uni-marburg.de/es/2012/0004/view.html>].
- Dirks, Una & Wilfried Hansmann (2012). Forschungsstand zur Analyse von Diagnosekompetenz in der Lehrerbildung [21 Seiten]. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), *Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium*. (Kap. I.3, Online-Schriftenreihe der Philipps-Universität Marburg). [URL: <http://archiv.ub.uni-marburg.de/es/2012/0014/view.html>].

- Dirks, Una; Wilfried Hansmann & Hendrik Baumbach (2012). Item Pools der administrierten Fälle [28 Seiten]. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), *Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium*. (Kap. II.7, Online-Schriftenreihe der Philipps-Universität Marburg: Professionalisierung und Diagnosekompetenz.), [URL: <http://archiv.ub.uni-marburg.de/es/2012/0021/view.html>].
- Fink, Rosalie (2006). *Why Jane and Johnny couldn't read – and how they learned*. Newark, DE: International Reading Association.
- Fletcher, Charles R.; Susan T. Chrysler, Paul van den Broek, Jennifer A. Deaton & Charles P. Bloom (1995). The role of co-occurrence, coreference, and causality in the coherence of conjoined sentences. In: Robert F. Lorch & Edward J. O'Brien (Hrsg.). *Sources of coherence in reading* (S. 203-218). Hillsdale, NJ: Erlbaum.
- Ford, Cecilia E.; Barbara A. Fox & Sandra A. Thompson (1996). Practices in the construction of turns. The "TCU" revisited. *Pragmatics* 6 (3), 427-454.
- Garnham, Alan; Jane V. Oakhill & Philip N. Johnson-Laird (1982). Referential continuity and the coherence of discourse. *Cognition*, 11, 29–46.
- Gernsbacher, Morton A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates
- Gernsbacher, Morton A. (1997). Two decades of structure building. *Discourse Processes*, 23, 265-304.
- Goldman, Susan R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In: N. Stuart-Ferris & D. M. Bloome (Hrsg.), *Uses of intertextuality in classroom and educational research* (S. 317–351). Greenwich, CT: Information Age Publishing.
- Goldman, Susan R. & John A. Rakestraw (2000). Structural aspects of constructing meaning from text. In: Michael L. Kamil, Peter B. Rosenthal, P. David Pearson & Rebecca Barr (Hrsg.), *Handbook of reading research* (S. 311-335). Mahwah, NJ: Erlbaum.
- Grabe, William (2009). *Reading in a Second Language. Moving from Theory to Practice*. Cambridge: CUP.
- Grabe, William & Fredricka L. Stoller (2011). *Teaching and Research Reading. Second Edition*. Harlow, U.K.: Pearson Education Limited.
- Graesser, Arthur C. & Gordon H. Bower (1990). *Inferences and Text Comprehension*. San Diego, CA: Academic Press.
- Graesser, Arthur C.; Danielle S. McNamara, Max M. Louwerse & Zhiqiang Cai (2004). Coh-Metrix: Analysis of text cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36, 193-202.

- Graesser, Arthur C.; Keith. K. Millis & Rolf A. Zwaan (1997). Discourse comprehension. *Annual Review of Psychology*, Vol. 48, 163-189.
- Graesser, Arthur C.; M. Singer & T. Trabasso (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371-395.
- Graves, Michael F. & Bonnie B. Graves (2003)<sup>2</sup>. Assessing text difficulty and accessibility. In: Dies., Scaffolding reading experiences: Designs for student success (Ch. 9). Norwood, MA: Christopher-Gordon.
- Green, Anthony, Aylin Ünalı & Cyril Weir (2010). Empiricism versus connoisseurship: Establishing the appropriacy of texts in tests of academic reading. *Language Testing* 27(2), 191-211.
- Halliday, Michael A. K. & Ruqayia Hasan (1976). Cohesion in English. London: Longman.
- Harley, Trevor A. (2008).<sup>3</sup> The Psychology of Language: From Data to Theory. Hove: Psychology Press.
- Hempelmann, Christian F.; David Dufty, Philip M. McCarthy, Arthur C. Graesser, Zhiqiang Cai & Danielle S. McNamara (2005). Using LSA to Automatically Identify Givenness and Newness of Noun Phrases in Written Discourse. Institute for Intelligent Systems/FedEx Institute of Technology, University of Memphis, TN 38152.
- Hobbs, Jerry R. (1990). Literature and Cognition. CSLI Lecture Notes 21.
- Irmer, Matthias (2011). Bridging Inferences. Constraining and Resolving Under-specification in Discourse Interpretation. Berlin: De Gruyter Mouton.
- Josephson, John R. & Susan G. Josephson (1994). Abductive Inference, Computation, Philosophy, Technology. New York: CUP.
- Kaiser, Elsi (2011). On the relation between coherence relations and anaphoric demonstratives in German. In: Ingo Reich, Eval Horch & Dennis Pauly (Hrsg.), Proceedings of Sinn & Bedeutung 15 (S. 337–351). Saarbrücken: Universaar – Saarland University Press.
- Kamalski, Judith; Ted Sanders & Leo Lentz (2008). Coherence Marking, Prior Knowledge, and Comprehension of Informative and Persuasive Texts: Sorting Things Out. *Discourse Processes*, 45, 323-345.
- Kehler, Andrew (2004). Discourse Coherence. In: Laurence R. Horn & Gregory Ward (Hrsg.), The Handbook of Pragmatics (S. 241-265). Malden, MA: Blackwell.
- Khalifa, Hanan & Cyril J. Weir (2009). Examining Reading. Research and practice in assessing second language reading. Cambridge, UK: CUP.
- 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.

- Kintsch, Walter & Eileen Kintsch (2005). Comprehension. In: Scott G. Paris & Steven A. Stahl (Hrsg.), *Children's reading comprehension and assessment* (S. 71-92). Mahwah, NJ: Erlbaum.
- Kirsch, Iwar (2001). *The International Adult Literacy Survey: Understanding What Was Measured*. Educational Testing Service, Princeton, NJ.
- 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].
- Kirsch, Iwar & Peter B. Mosenthal (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly*, 25(1), 5-30.
- Koda, Keiko (2005). *Insights into second language reading: A cross-linguistic approach*. New York: Cambridge University Press.
- Krifka, Manfred (2009). Implicatures. In: Uli Sauerland & Kazuko Yatsushiro (Hrsg.), *Semantics and pragmatics: From experiment to theory* (S. 3-15). New York: Palgrave Macmillan.
- Landi, Nicole (2010). An examination of the relationship between reading comprehension, higher-level and lower-level reading sub-skills in adults. *Reading & Writing* 23, 701-717.
- Leslie, Lauren & JoAnne Cladwell (2009). Formal and Informal Measures of Reading Comprehension. In: Susan E. Israel & Gerald G. Duffy (Hrsg.), *Handbook of Research on Reading Comprehension* (403-427). New York: Routledge.
- Levinson, Stephen C. (1983). *Pragmatics*. Cambridge, UK: CUP.
- Long, Debra L.; Brain J. Oppy & Mark R. Seely (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1456-1470.
- Lorch, Robert F., Jr. & Edward J. O'Brien (Hrsg.) (1995). *Sources of coherence in reading*. Hillsdale, NJ: Erlbaum.
- Lorch, Robert F., Jr.; Elizabeth P. Pugzles-Lorch & Ann M. Morgan (1987). Task effects and individual differences in on-line processing of the topic structure of a text. *Discourse Processes*, 10, 63-80.
- Lorch, Robert F.; Elizabeth Pugzles-Lorch, Kristin Ritchey, Lisa McGovern & Deana Coleman (2001). Effects of headings on text summarization. *Contemporary-Educational-Psychology*, 26, 171-191.
- Mähler, Claudia & Elsbeth Stern (2006). Transfer. In: Detlef H. Rost (Hrsg.). *Handwörterbuch Pädagogische Psychologie*, (3., überarb. u. erw. Auflage; 782-793). Berlin: Beltz.

- Magliano, Joseph P. & Keith K. Millis (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction*, 21, 251–283.
- Magliano, Joseph P.; Stacey Todaro, Keith K. Millis, Katja Wiemer-Hastings, H. Joyce Kim & Danielle S. McNamara (2005). Changes in reading strategies as a function of reading training: A comparison of live and computerized training. *Journal of Educational Computing Research*, 32, 185-208.
- Magliano, Joseph P.; Katja Wiemer-Hastings, Keith K. Millis, Brenton D. Muñoz & Danielle S. McNamara (2002). Using latent semantic analysis to assess reader strategies. *Behavior Research Methods, Instruments & Computers*, 34, 181–188.
- Mandler, Jean M. & Nancy S. Johnson (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9, 11–151.
- Mani, Kannan & Philip N. Johnson-Laird (1982). The mental representation of spatial descriptions. *Memory and Cognition*, 10, 181-187.
- McNamara, Danielle S. (2004). SERT: Self explanation reading training. *Discourse Processes*, 38, 1–30.
- McNamara, Danielle S.; Eileen Kintsch, Nancy B. Songer & Walter Kintsch (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14, 1-43.
- McNamara, Danielle S.; Irwin B. Levinstein & Chutima Boonthum (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, & Computers*, 36, 222–233.
- McNamara, Danielle S.; Tenaha O'Reilly, Rachel Best & Yasuhiro Ozuru (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, 34, 147–171.
- McNamara, Danielle S. & Tenaha O'Reilly (2012). Theories of comprehension skill: Knowledge and strategies versus capacity and suppression. In: F. Columbus (Hrsg.), *Progress in experimental psychology research*. Hauppauge, NY: Nova Science Publishers, Inc. [URL: <ftp://129.219.222.66/pdf/NOVA.pdf>, Zugriff 30.10.2012] (forthcoming).
- Meyer, Bonnie J. F.; David M. Brandt & George J. Bluth (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly*, 16, 72–103.
- Meyer, Bonnie J. F. & Kausalai K. Wijekumar (2007). A Web based tutoring system for the structure strategy: theoretical background, design, and findings. In: Danielle S. McNamara (Hrsg.), *Reading comprehension strategies: Theory, interventions, and technologies* (S. 347–374). Mahwah, NJ: Erlbaum.
- Millis, Keith; H. Joyce Kim, Stacey Todaro, Joseph P. Magliano, Katja Wiemer-Hastings & Danielle S. McNamara (2004). Identifying reading strategies us-

- ing latent semantic analysis: Comparing semantic benchmarks. *Behavior Research Methods, Instruments, & Computers*, 36, 213-221.
- Mosenthal, Peter & Irwin S. Kirsch (1998). A new measure of assessing document complexity: The PMOSE/IKIRSCH document readability formula. *Journal of Adolescent & Adult Literacy*, 41(8), 638-657.
- Myers, Jerome L. (1990). Causal relatedness and text comprehension. In: David A. Balota, G.B. Flores d'Arcais & Keith Rayner (Hrsg.), *Comprehension processes in reading* (S. 361-375). Hillsdale, NJ: Erlbaum.
- Neubauer, Aljoscha & Elsbeth Stern (2007). *Lernen macht intelligent. Warum Begabung gefördert werden muss*. München: Deutsche Verlags-Anstalt.
- Oakhill, Jane V. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, 54, 31-39.
- Oakhill, Jane V. & Nicola Yuill (1986). Pronoun resolution in skilled and less-skilled comprehenders: Effects of memory load and inferential complexity. *Language and Speech*, 29, 25-37.
- Oakhill, Jane V. & Nicola Yuill (1996). Higher order factors in comprehension disability: Processes and remediation. In: Cesare Cornoldi & Jane V. Oakhill (Hrsg.), *Reading comprehension difficulties: Processes and intervention* (S. 69-92). Mahwah, NJ: Erlbaum.
- Oakhill, Jane V.; Nicola Yuill & Morag L. Donaldson (1990). Understanding of causal expressions in skilled and less skilled text comprehenders. *British Journal of Developmental Psychology*, 8, 401-410.
- OECD (2009). *PISA 2009. Assessment Framework. Key competencies in reading, mathematics and science*. Paris: OECD.
- O'Reilly, Tenaha & Danielle S. McNamara (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, 43(2), 121-152.
- Ouellet, Gene P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554-566.
- Ozuru, Yasuhiro; Michael Rowe, Tenaha O'Reilly & Danielle S. McNamara (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods*, 40, 1001-1015.
- Palincsar, Annemarie S. & Ann L. Brown (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117-175.
- Palmer, Humphrey (1985). *Presupposition and Transcendental Inference*. Sydney, Australia: Palmer.

- Paris, Scott G. & Ellen E. Hamilton (2009). The development of children's reading comprehension. In: Susan E. Israel & Gerald G. Duffy (Hrsg.), *Handbook of Research on Reading Comprehension* (S. 32-53). New York: Routledge
- Pearson, P. David (2009). The Roots of Reading Comprehension Instruction. In: Susan E. Israel & Gerald G. Duffy (Hrsg.). *Handbook of Research on Reading Comprehension* (S. 3-31). New York: Routledge.
- Perfetti, Charles; Nicole Landi & Jane Oakhill (2005). The acquisition of reading comprehension skill. In: Margaret J. Snowling & Charles Hulme (Hrsg.), *The science of reading: A handbook* (S. 227-247). Oxford: Blackwell.
- Perfetti, Charles; Chin-Lung Yang & Franz Schmalhofer (2008). Comprehension skill and word-to-text integration processes. *Applied Cognitive Psychology* 22 (3), 303-318.
- Pieterse, Jan N. (1995). Globalization as hybridisation. In: Mike Featherstone, Scott Lash & Roland Robertson (Hrsg.), *Global Modernities* (S. 45-68). London: Sage.
- Pu, Ming-Ming (2011). *Discourse Anaphora: A Cognitive-Functional Approach*. (LINCOM Studies in Theoretical Linguistics 47). München: LINCOM.
- Rapp, David M.; Paul van den Broek, Kristen L. McMaster, Panayiota Kendeou & Christine A. Espin (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading* 11(4), 289-312.
- Reckwitz, Andreas (2003). Grundelemente einer Theorie sozialer Praktiken. Eine sozialtheoretische Perspektive. *Zeitschrift für Soziologie* 32 (4), 282-301.
- Repp, Sophie (2011). Relevance Topics. In: Ingo Reich, Eval Horch & Dennis Pauly (Hrsg.), *Proceedings of Sinn & Bedeutung* 15 (S. 483-498). Universaar – Saarland University Press: Saarbrücken.
- Rickheit, Gert & Hans Strohner (Hrsg.) (1985). *Inferences in Text Processing*. Amsterdam: North Holland.
- Rickheit, Gert & Christopher Habel (Hrsg.) (1995). *Focus and coherence*. Berlin: de Gruyter.
- Roeper, Tom (2011). How the Emergence of Propositions Separates Strict Interfaces from General Inference. In: Ingo Reich, Eval Horch & Dennis Pauly (Hrsg.), *Proceedings of Sinn & Bedeutung* 15 (S. 55-77). Universaar – Saarland University Press: Saarbrücken.
- Rosebrock, Cornelia (2009). Lesekompetenz als Mehrebenenkonstrukt. In: Andrea Bertschi-Kaumann & Cornelia Rosebrock (Hrsg.). *Literalität. Bildungsaufgabe und Forschungsfeld* (S. 59-72). Weinheim: Juventa.
- Rowe, Michael; Yasuhiro Ozuru & Danielle S. McNamara (2006). An analysis of a standardized reading ability test: What do questions actually measure? In:

- Sasha A. Barab, Daniel T. Hickey, Kenneth E. Hay & Nancy Butler Songer (Hrsg.), *Making a difference, Volume I and II: The Proceedings of the Seventh International Conference of the Learning Sciences* (S. 627-633). Mahwah, NJ: Erlbaum.
- Roy, Marguerite & Michelene T. H Chi (2005). The self-explanation principle in multi-media learning. In: Richard E. Mayer (Hrsg.), *Cambridge handbook of multimedia learning* (S. 271-286). Cambridge: Cambridge University Press.
- Sacks, Harvey; Emanuel Schegloff & Gail Jefferson (1974). A simplest systematic for the organization of turn-taking for conversation. *Language* 50 (4), 696-735.
- Sanders, Ted & Leo G.M. Noordman (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, 29, 37-60.
- Sanders, Ted & Wilbert Spooren (2006). Discourse and Text Structure. In: D. Geeraerts & H. Cuykens (Hrsg.). *Handbook of Cognitive Linguistics* (S. 916-943). Oxford: OUP.
- Schegloff, Emanuel A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis I*. Cambridge: CUP.
- Schegloff, Emanuel A. & Harvey Sacks (1973). Opening up closings. *Semiotica*, Jg. 8, 289-327.
- Schlieben-Lange, Brigitte (1983). *Traditionen des Sprechens. Elemente einer pragmatischen Sprachgeschichtsschreibung*. Stuttgart: Kohlhammer.
- Schmalhofer, Franz & Charles A. Perfetti (2007). *Higher Level Language Processes in the Brain. Inference and Comprehension Processes*. Mahwah, NJ: Erlbaum.
- Searle, John (1969). *Speech Acts*. Cambridge: CUP.
- Searle, John (1975). Indirect speech acts. In: Peter Cole & Jerry L. Morgan (Hrsg.), *Syntax and Semantics, 3: Speech Acts* (59-82). New York: Academic Press.
- Shapiro, Amy M. (2004). How including prior knowledge as a subject variable can change outcomes of learning research. *American Educational Research Journal*, 41, 159-189.
- Singer, Murray & Gilbert Remillard (2004). Retrieving text inferences: Controlled and automatic influences. *Memory & Cognition*, 32, 1223-1237.
- Snow, Catherine E. (2010). Academic Language and the Challenge of Reading for Learning About Science. *Science*, Jg. 328, H. 5977, 450-452.
- Stalnaker, Robert (2002). Common Ground. *Linguistics and Philosophy* 25(5-6), 701-721.
- Stevens, Robert J. (1988). Effects of strategy training on the identification of the main idea of expository passages. *Journal of Educational Psychology*, 80, 21-26.



- Tannenbaum, Kendra R.; Joseph K. Torgesen & Richard Wagner (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, 10(4), 381–398.
- Thurmair, Maria (1989). Modalpartikeln und ihre Kombinationen. Tübingen: Niemeyer.
- Tiemann, Sonja; Mareike Schmid, Nadine Bade, Bettina Rolke, Ingo Hertrich, Hermann Ackermann, Julia Knapp & Sigrid Beck (2011). Psycholinguistic Evidence For Presuppositions: On-line and Off-line Data. In: Ingo Reich, Eval Horch & Dennis Pauly (Hrsg.), *Proceedings of Sinn & Bedeutung 15* (S. 581–595). Universaar – Saarland University Press: Saarbrücken.
- Torodo, Stacey; Keith Millis & Srikanth Dandotkar (2010). The Impact of Semantic and Causal Relatedness and Reading Skill on Standards of Coherence. *Discourse Processes*, 47, 421-446.
- Trabasso, Tom & Joseph P. Magliano (1996). Conscious understanding during comprehension. *Discourse Processes* 23, No. 1, 255-287.
- van den Broek, Paul; Kirsten Risdén & Elizabeth Husebye-Hartmann (1995). The role of readers' standards for coherence in the generation of inferences during reading. In: Robert F. Lorch, Jr. & Edward J. O'Brien (Hrsg.), *Sources of coherence in reading* (S. 353-373). Mahwah, NJ: Erlbaum.
- van den Broek, Paul; Robert F. Lorch, Tracy Jr. Linderholm & Mary Gustafson (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, 29, 1081–1087.
- Vellutino, Frank R. (2003). Individual differences as sources of variability in reading comprehension in elementary school children. In: A. P. Sweet & C. E. Snow (Hrsg.), *Rethinking reading comprehension*. New York: Guilford Press, 51-81.
- Virtanen, Tuija (1992). Issues of text typology: Narrative - a 'basic' type of text? *Text* 12, 293-310.
- Walker, Marilyn A.; Aravind K. Joshi & Ellen F. Prince (1998). Centering in naturally occurring discourse: An overview. In: Marilyn A. Walker, Aravind K. Joshi & Ellen F. Prince (Hrsg.), *Centering theory in discourse* (S. 1-28). Oxford: Clarendon Press.
- Williams, Joanna P. (1986). Teaching children to identify the main idea of expository texts. *Exceptional Children*, 53, 163–168.
- Williams, Joanna P. (2007). Literacy in the curriculum: Integrating text structure and content area instruction. In: Danielle S. McNamara (Hrsg.), *Reading comprehension strategies: Theory, interventions, and technologies* (S. 199–219). Mahwah, NJ: Erlbaum.
- Wilson, Paul T. & Richard C. Anderson (1986). What they don't know will hurt them: The role of prior knowledge in comprehension. In: Judith Orasanu

- (Hrsg.), Reading comprehension: From research to practice (S. 31–48). Hillsdale, NJ: Erlbaum.
- Wolfe, Michael B. & Susan R. Goldman (2005). Relations between adolescents' text processing and reasoning. *Cognition & Instruction*, 23, 467-502.
- Wolfe, Michael B.; Joseph P. Magliano & Benjamin Larsen (2005). Causal and semantic relatedness in discourse understanding and representation. *Discourse Processes*, 39, 165-187.
- Yuill, Nicola & Jane V. Oakhill (1988). Understanding of anaphoric relations in skilled and less skilled comprehenders. *British Journal of Psychology*, 79, 173–186.
- Yuill, Nicola; Jane V. Oakhill & Alan Parkin (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology*, 80, 351–361.
- Zwaan, Rolf A. & Carol M. Brown (1996). The influence of language proficiency and comprehension skill on situation model construction. *Discourse Processes*, 21, 289-327.
- Zwann, Rolf A. & Murray Singer (2003). Text comprehension. In: Arthur C. Graesser, Morton A. Gernsbacher & Susan R. Goldman (Hrsg.), Handbook of discourse processes (S. 83-121). Mahwah, NJ: Erlbaum.

## Zitation

Dirks, Una (2012). Prä- und Posttest-Kongruenzen. Gemeinsamkeiten und Differenzen in der Anforderungsstruktur der Fallvignetten. [43 Seiten]. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium. (Online-Schriftenreihe der Philipps-Universität Marburg). [URL: [http://archiv.ub.uni-marburg.de/opus/schriftenreihen\\_ebene2.php?sr\\_id=30&la=de](http://archiv.ub.uni-marburg.de/opus/schriftenreihen_ebene2.php?sr_id=30&la=de)].

**Anhang: Diskurskritische indirekte Sprechakte im Abschlussfall** (vgl. Anm. 1)

- L: „Ich glaub, die kann die Uhr genauso lesen wie wir alle.“ (Z. 11)
- Tim: „Guten Morgen, Herr Mustermann.“ (*ruft nochmal extra laut*; Z. 14)
- L: „Mark, weißt du noch, wie ein Erdkundebuch aussieht? Hast du bestimmt schon mal irgendwo gesehen, oder (...) So sieht ein Buch aus!“ (Z. 18f.)
- L: „Als Mark mir noch viel Freude bereitet hat (...) Mark, so langsam muss wieder was kommen. Merkst du das?“ (Z. 22ff.)
- L: „Wir fangen an zu lesen, damit (...) der Dennis sich auch mal bisschen die Wörter in den Ohren zergehen lassen kann.“ (Z. 25f.)
- Klaus: „Ganz laut lesen!“ (Z. 28)
- L: „Dennis weiß anscheinend nicht, wie man ein Buch aufschlägt und weiß nicht, wie die Zahl 14 aussieht.“ (Z. 30f.)
- L: „Schön, dass ihr auch da seid! Sitzt die Frisur? Schminke klar?“ (Z. 35)
- L: „Hat heute Morgen ein bisschen länger gedauert?“ (Z. 37)
- L: „Ich war früher auch mal ein Mädchen.“ (Z. 39)
- L: „Stephan, ich möchte dich bitten, wenn du reinkommst, den Türgriff zu benutzen. Du bist gerne willkommen bei mir in der Klasse. (2 Sek. Pause) Die Tür hat einen Griff. (2 Sek. Pause) Stephan, ist klar, oder? Die Tür hat einen Griff.“ (Z. 41ff.)
- Stephan: „Bei mir nicht.“ (Z. 44)
- L: „Anscheinend nicht, deswegen mache ich dich auch darauf aufmerksam! (4 Sek. Pause) Und Stephan, jetzt aber wirklich. (...) Jetzt haben wir unseren Kredit erschöpft. (3 Sek. Pause) Ey Leute, wir sollten langsam mal anfangen, sonst holen wir es am Mittag nach! Ich hab eh schon eine Klassenkonferenz wegen euch, und dann kann ich euch auch nebenbei hier ... Bin heut Mittag eh da wegen euch, Klassenkonferenz, ich weiß nicht, wie lange das dauert. Ihr werdet sehen: ... Überlegt es euch gut, ob ihr heut‘ morgen Erdkunde machen wollt oder heut Mittag. So! (...) Es ist nur für euren Klassenlehrer, dass er weiß, dass, ähm es so schön ist, wie es sein sollte. Ok?“ (Z. 47ff.)
- L: „So. ... Ihr habt nichts verpasst. (...) Alles wunderbar.“ (Z. 60)
- L: „Warte mal, nicht alle sind soweit.“ (Z. 66)