

5 PIRLS 2016 und BIST D4 2015: Vergleich der Konstrukte und narratologische Analyse der literarischen Stimuli

Marcel Illetschko & Katrin Widauer

Der vorliegende Beitrag erörtert, inwieweit sich die beiden in Österreich durchgeführten Schülerleistungsstudien PIRLS 2016 und die nationale Standardüberprüfung Deutsch auf der 4. Schulstufe (BIST D4, 2015) aus inhaltlicher Perspektive vergleichen lassen. Dazu werden die Operationalisierungen der latenten Konstrukte analysiert, auf denen die Tests aufbauen: Um etwaige Gemeinsamkeiten und Unterschiede in Itempool und Testdesign darzustellen, werden alle Items der BIST D4 (117 Items) und von PIRLS (175 Items) in Bezug auf die grundlegenden Kategorien der jeweils anderen Studie klassifiziert. Zurückgegriffen wird dazu einerseits auf die vorliegenden Informationen zu den Items, andererseits auf eine Neukodierung der Stimuli und Items. Darüber hinaus werden alle literarischen bzw. narrativen Stimuli nach narratologischen Gesichtspunkten analysiert. Die Ergebnisse dieses Vergleichs zeigen deutliche Unterschiede in der Konzeption und Breite des Konstrukts: Die Textauswahl der BIST D4 ist vielfältiger, die testinterne Gewichtung der Kategorien aber nur bedingt an die internationale Leseforschung anzubinden. Nach narratologischen Gesichtspunkten ist das Konstrukt der BIST D4 generell facettenreicher, hinsichtlich der Gestaltung der Figuren ist allerdings PIRLS vielfältiger.

Einleitung

Der Lesefähigkeit kommt als elementarer Kulturtechnik individuell und gesellschaftlich eine zentrale Bedeutung zu. Zum Bildungsmonitoring existieren auf der 4. Schulstufe in Österreich zwei große Schülerleistungsstudien, welche die Lesekompetenz erfassen – allerdings mit unterschiedlichen Zielsetzungen bzw. Schwerpunkten:

Einerseits testet die internationale Studie Progress in International Reading Literacy Study (PIRLS) eine Stichprobe von Schülerinnen und Schülern in ca. 50 Teilnehmerländern (Österreich: N = 4360) mit standardisierten Testverfahren, die Testzeit beträgt 80 Minuten (Mullis & Martin, 2015; Widauer & Wallner-Paschon, 2017). Aus den aggregierten Ergebnissen von PIRLS lässt sich die Positionierung Österreichs im internationalen Vergleich ablesen.¹ Die teilnehmenden Länder verwenden „PIRLS data for system-level monitoring of education achievement in a global context. They compare their reading achievement levels and contexts for learning with those of other countries, and monitor progress in reading achievement over time“ (Mullis & Martin, 2015).

Auf der anderen Seite stellt die Überprüfung der österreichischen Bildungsstandards im Leseverständnis auf der 4. Schulstufe (BIST D4) eine nationale Vollerhebung dar (N = 75.297; Breit, Bruneforth & Schreiner, 2016). Die Bildungsstandardüberprüfung besteht im Gegensatz zu PIRLS aus mehreren Testteilen (Textproduktion, Prozessorientierte

Teilkompetenzen, Sprachbetrachtung, Lesefertigkeit auf Wortebene, Leseverständnis, Rechtschreiben, Hören und – nur in einer Stichprobe [N = 2820 Schüler/innen] – Sprechen), die im Laufe von zwei Vormittagen zu absolvieren sind. Die Testzeit für das Leseverständnis beträgt 30 Minuten (ebd.). Die gesamte Testzeit der Schüler/innen für alle Testteile ist höher als bei PIRLS.

Die Bildungsstandards sollen laut Verordnung (BGBl. II Nr. 1/2009) allerdings nicht vorrangig einer Überprüfung dienen, sondern „1. eine nachhaltige Ergebnisorientierung in der Planung und Durchführung von Unterricht bewirken, 2. durch konkrete Vergleichsmaßstäbe die bestmögliche Diagnostik als Grundlage für individuelle Förderung sicher stellen [sic] und 3. wesentlich zur Qualitätsentwicklung in der Schule beitragen“. In den Standardüberprüfungen „sind die von den Schülerinnen und Schülern bis zur 4. bzw. zur 8. Schulstufe erworbenen Kompetenzen objektiv festzustellen und mit den angestrebten Lernergebnissen zu vergleichen“ (ebd.). Die Überprüfungen sind „**zentral vorgegebene Leistungsmessungen** mit dem Ziel, die **Ergebnisse für Schul- und Unterrichtsentwicklung nutzbar zu machen** [und dienen] als Impuls für Qualitätsentwicklungsprozesse am jeweiligen Schulstandort sowie landes- und bundesweit“ (Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens [BIFIE], 2012, S. 5, Hervorhebungen im Original). Besonders die vielseitige Rückmeldung auf den Ebenen Bund, Bundesland (Schulbehörden), Schule, Klasse bzw. Gruppe und Schüler/in ist im deutschsprachigen Raum einzigartig.

1 Detaillierte Infos zur Konzeption finden sich im Technischen Bericht zu PIRLS 2016 (Wallner-Paschon & Salchegger, 2017) sowie in der Einführung im Erstbericht zu PIRLS 2016 (Wallner-Paschon & Widauer, 2017).

Doch inwiefern sind die Ergebnisse dieser Studien (bzw. von Large-Scale Assessments generell) mit ihren unterschiedlichen Zielsetzungen vergleichbar? Bereits seit Langem wird dieser Frage unter unterschiedlichen Gesichtspunkten nachgegangen (siehe etwa Voss et al., 2005; Bos et al., 2007; Artelt, 2009; Venn-Brinkmann, 2015). In vorliegender Untersuchung soll aus der Perspektive fachdidaktischer und textwissenschaftlicher Überlegungen sowie deskriptiver Statistiken untersucht werden, inwiefern die in PIRLS und BIST D4 eingesetzten Items und Textkorpora vergleichbar sind. Ausgangspunkt dazu ist die Analyse der Items und Stimulustexte vor dem Hintergrund der theoretischen Konzeptionierung von PIRLS und BIST D4. Zudem werden die Eigenschaften der literarischen bzw. narrativen Texte mittels außerhalb der publizierten Testkonstrukte liegenden narratologischen Analysekriterien dargestellt (etwa bezüglich der Erzählinstanz, der Komplexität der Handlung etc.). Verglichen wird also die konkrete Operationalisierung der Konstrukte.

Theoretischer Hintergrund

Zum Begriff „Konstrukt“

Im Vorfeld der Testkonstruktion sind eine Reihe verschiedener theoretischer Vorüberlegungen anzustellen, welche beispielsweise den Testzweck und -inhalt, die Zielpopulation oder Beschränkungen bei der Testdurchführung betreffen (Schmeiser & Welch, 2006; Itzlinger-Bruneforth, Kuhn & Kiefer, 2016). In der jeglichen Tests zugrundeliegenden Testtheorie werden Konstrukte als Variablen für Modelle genutzt, mithilfe derer die Testkonstruktion erfolgen kann. Unter einem Konstrukt versteht man ein „latentes Merkmal [...], [dessen] Ausprägungen anhand beobachtbarer Variablen geschätzt werden“ (Eid & Schmidt, 2014, S. 416). Latent ist ein solches Konstrukt, wenn es – im Gegensatz zu manifesten Konstrukten – nicht direkt beobachtet werden kann, sondern über bestimmte Merkmale (Indikatoren) operationalisiert werden muss, sodass der interessierende Sachverhalt auch tatsächlich messbar wird. Unter Validität versteht Messick (1989, S. 13) in diesem Zusammenhang Aussagen über das Ausmaß, in dem empirische Nachweise und theoretische Begründungen die Angemessenheit und Eignung von Schlussfolgerungen und Handlungen stützen, die aus Testleistungen und anderen Beurteilungsmethoden abgeleitet werden.

Mit anderen Worten geht man also beim Testen latenter Konstrukte davon aus, dass man einen bestimmten Ausschnitt der Wirklichkeit, im Fall von PIRLS bzw. im Testteil zum Leseverstehen der BIST D4 eine Fähigkeit, genauer: das Leseverstehen, mittels einzelner Items abbilden kann. Man nimmt an, das Abzubildende habe gewisse wichtige Eigenschaften, die mithilfe von Items bzw. einer bestimmten Zusammenstellung von Items (genauer: der Reaktion der zu Testenden auf diese Items) dargestellt werden können: Ei-

genschaft A sei zu a % in dem betreffenden Ausschnitt der Wirklichkeit vertreten, Eigenschaft B zu b %, Eigenschaft C zu c % usw. Repräsentierte man nun diese Eigenschaften in denselben Relationen in einer bestimmten Menge an Items, dann sei das in Testform gegossene Abbild des Wirklichkeitsausschnitts eine repräsentative Wiedergabe der Wirklichkeit und der Test somit konstruktvalide.

Nun hat die Wirklichkeit die Eigenschaft, viele Eigenschaften zu haben (zu dieser keineswegs trivialen Aussage siehe etwa Gabriel, 2013; Stekeler-Weithofer, 2014). Die Zeit und die Mittel von Testerstellern und Testerstellerinnen sowie der Kenntnisstand über den interessierenden Ausschnitt der Wirklichkeit etc. hingegen sind beschränkt, genauso wie die zur Verfügung stehende Testzeit und die Mittel für eine Testdurchführung bzw. Datenaufbereitung. Deshalb akzeptiert man gemeinhin, nicht alle wichtigen Eigenschaften des Wirklichkeitsausschnitts in einem Test abbilden zu können, sondern nur die (im Sinne des jeweils definierten Konstrukts) wichtigsten und die am besten (unter welchen Kriterien auch immer bewertet) messbaren. In unterschiedlichen Tests derselben Fähigkeit kann sich die theoretische Einschätzung dieser Wichtigkeiten unterscheiden, weshalb zwei von verschiedenen Personen(-gruppen) bzw. Institutionen entwickelte Tests zu ähnlichen Wirklichkeitsausschnitten (also etwa zwei Tests zum Leseverstehen von österreichischen Kindern auf der 4. Schulstufe) im Grunde kaum zu deckungsgleichen Ergebnissen führen werden.

Komplex wird der Vergleich von Testergebnissen zu ähnlichen Wirklichkeitsausschnitten vor allem dann, wenn die Zahl der als wichtig erachteten Eigenschaften sehr groß ist bzw. wenn umfangreiche Forschungen oder gar unterschiedliche Forschungsrichtungen zu einem Wirklichkeitsausschnitt existieren und den Wirklichkeitsausschnitt maßgeblich konstituierende Eigenschaften in den unterschiedlichen Forschungsrichtungen unterschiedlich bewertet werden. Denn der Vielzahl der theoretisch maßgeblichen Eigenschaften auf der einen Seite steht auf der anderen die sehr begrenzte Zahl an administrierbaren Testitems und inhaltlichen Eigenschaften der Testitems gegenüber.

Kurz gesagt: Bei Tests gibt es ein Konstrukt, von dem man weiß. Und sehr viele, von denen man nicht weiß. Anders, und hinsichtlich der Operationalisierung formuliert: Es gibt ein Konstrukt, das man plant. Und es gibt viele Konstrukte, die gleichzeitig passieren.

Was wird in PIRLS und BIST D4 geplant und was passiert unkontrolliert? Mögliche Antworten auf diese Frage sollen im Folgenden gegeben werden.

Zentrale Begriffe der Konstrukte von PIRLS und BIST D4

PIRLS kennt als Dimensionen der Lesekompetenz einerseits vier sogenannte „Verstehensprozesse“, die der Kategorisierung

Verstehensprozesse	Leseabsicht		PIRLS gesamt (%)
	literarische Erfahrung (%)	Informationsgewinnung (%)	
Erkennen und Wiedergeben expliziter Informationen	10 %	10 %	20 %
Einfaches Schlussfolgern	15 %	15 %	30 %
Interpretieren und Verknüpfen	15 %	15 %	30 %
Bewerten und Kritisieren	10 %	10 %	20 %
Gesamt	50 %	50 %	100 %

Tabelle 1: (Theoretische) Item- bzw. Punkteverteilung nach Leseabsicht und Verstehensprozessen in Prozent (PIRLS 2016)

der kognitiven Bedeutungskonstruktion beim Lesen dienen (Mullis & Martin, 2015, Kapitel 13 & 15; Wallner-Paschon & Widauer, 2017). Diese Verstehensprozesse ähneln aus theoretischer Perspektive den sogenannten „Teilkompetenzen“ bzw. „Kompetenzfeldern“ der BIST D4, die ihrerseits in weiter unten spezifizierte Can-Do-Statements bzw. Deskriptoren unterteilt sind (BIFIE, 2016 bzw. Wiesner, Pacher, George, Breit & Schreiner, 2017).²

Als zweite wesentliche Kategorisierung postuliert PIRLS die Bedeutung sogenannter „Leseabsichten“, nämlich das „Lesen, um literarische Erfahrungen zu machen“ und das „Lesen, um Informationen zu gewinnen“. In den bisher veröffentlichten Dokumenten über die theoretische Fundierung der BIST D4 ist keine vergleichbare Unterscheidung getroffen, wenngleich in einzelnen Deskriptoren Gattungs- bzw. Textsortenspezifika angesprochen sind (BIFIE, 2016; siehe unten).

Fragestellung und methodischer Zugang

Im Folgenden sollen Unterschiede in der inhaltlichen Konzeption zwischen den beiden Tests sichtbar gemacht werden, indem alle Items in Bezug auf die grundlegenden Kategorien der jeweils anderen Studie klassifiziert werden. Zurückgegriffen wird dazu auf die (für die BIST D4 unveröffentlichten) Iteminformationen aller Items von BIST D4 (117 Items) und PIRLS (175 Items). Alle PIRLS-Items werden den einzelnen Deskriptoren und Kompetenzfeldern der BIST D4 zugewiesen und vice versa alle BIST D4-Items den Leseabsichten und Verstehensprozessen von PIRLS, um zu vergleichen, wie die Konstrukte in den beiden Tests operationalisiert werden. Darüber hinaus werden narrative Stimuli auch nach narratologischen Gesichtspunkten analysiert. Die entsprechenden Kategorien finden sich in den Abschnitten

Studieninterne Mengengerüste von PIRLS und BIST D4 sowie PIRLS- und BIST-D4-Stimuli aus narratologischer Perspektive.

Alle Zuordnungen erfolgten durch einen einzelnen Rater. Dies ist methodisch unvoreilhaft, da keine Reliabilitätswerte für diese Zuordnung geboten werden können. Für etwaige Folgestudien wäre daher eine Kodierung durch mehrere Rater wünschenswert, zeigt sich doch, dass auch Experteneinschätzungen bezüglich Itemschwierigkeit und auch kategorialer Zuordnung keinesfalls homogen sind (Sigott & Cesnik, 2018).

Die Ergebnisse werden deskriptiv beschrieben und inhaltlich interpretiert. Eine genauere empirische Zusammenschau der ersten Befunde mit Leistungs- und Kontextdaten, wie Geschlecht, Sozialstatus etc., wäre der nächste Schritt, kann hier aber nicht geleistet werden.

Ergebnisse

Studieninterne Mengengerüste von PIRLS und BIST D

Die Verteilung der PIRLS-Items – und der zu erreichenden Punkte, denn vereinzelt gibt es offene Items, bei denen mehrere Punkte zu erreichen sind (Martin, Mullis & Foy, 2015, S. 63 f.) – auf Verstehensprozesse und Leseabsichten ist in Tabelle 1 dargestellt (Mullis, Martin & Sainsbury, 2015, S. 14).

Die Leseabsichten sind laut dem Framework gleichverteilt, die Verstehensprozesse ungleich, aber in einem klaren Verhältnis, das seit PIRLS 2001 (an dem Österreich nicht teilgenommen hat) konstant gehalten wird (Campbell, Kelly, Mullis, Martin & Sainsbury, 2001; Wallner-Paschon, 2007; Wallner-Paschon, Grafendorfer, Widauer & Suchań, 2012). Die tatsächlichen Zahlen in den technischen Berichten und

2 Die „Teilkompetenzen“ sind Bündel mehrerer Kompetenzen und werden deshalb z. B. in der Informellen Kompetenzmessung (IKM) auch „Kompetenzfelder“ genannt (Wiesner et al., 2017). Auf der anderen Seite werden sie als Dimensionen eines Konstrukts für die Lesekompetenz insgesamt angesehen, weshalb die Bezeichnung „Teilkompetenzen“ gewählt wird. Im Folgenden wird meist die Bezeichnung „Kompetenzfeld“ verwendet.

Verstehensprozesse	2001		2006		2011		2016	
	N	%	N	%	N	%	N	%
Erkennen und Wiedergeben expliziter Informationen	25 (29)	26 (22)	31 (36)	25 (22)	33 (38)	24 (22)	50 (57)	29 (26)
Einfaches Schlussfolgern	27 (31)	28 (23)	43 (47)	34 (28)	46 (49)	34 (28)	53 (59)	30 (26)
Interpretieren und Verknüpfen	31 (54)	32 (41)	34 (61)	27 (37)	38 (66)	28 (38)	47 (79)	27 (35)
Bewerten und Kritisieren	15 (19)	15 (14)	18 (23)	14 (14)	18 (22)	13 (12)	25 (28)	14 (13)
Gesamt	98 (133)	100	126 (167)	100	135 (175)	100	175 (223)	100

Tabelle 2: Absolute und prozentuale Verteilung der Items und Score-Punkte (in Klammern) nach Verstehensprozess (PIRLS 2001, 2006, 2011, 2016)

Kompetenzfeld, Deskriptor	Itemanzahl	Prozent
K1: Explizite Informationen ermitteln	35	30 %
2.2 Leseverständnis auf der Wort- und Satzebene	15	13 %
3.2 Informationen entnehmen	20	17 %
K2: Allgemeines Textverständnis	60	51 %
3.4 Informationen aus Texten miteinander vergleichen	4	3 %
3.5 Inhalte/Informationen aus Texten ordnen	11	9 %
3.6 Den Verlauf einer Handlung erschließen	6	5 %
3.7 Das Wesentliche eines Textes erfassen	9	8 %
6.1 Gestaltung und Aufbau von Texten erkennen	12	10 %
6.2 Textsorten nach wesentlichen Merkmalen unterscheiden	18	15 %
K3: Textbezogenes Interpretieren und Reflektieren	22	19 %
4.2 Nicht ausdrücklich genannte Sachverhalte verstehen	16	14 %
4.3 Zu Text Stellung nehmen und Meinung begründen	6	5 %

Tabelle 3: Itemverteilung nach Kompetenzfeld und Deskriptor in absoluten Zahlen und Prozent (BIST D4 2015)

weiteren Publikationen des Boston College zu PIRLS zeigen kleine Abweichungen von der theoretischen Verteilung, da diese lediglich einen ungefähren Rahmen darstellt (Mullis, Martin & Gonzales, 2004; Kelly, 2003; Kennedy & Sainsbury, 2007; Mullis, Martin, Kennedy & Foy, 2007; Martin & Mullis, 2012; Martin, Mullis & Hooper, 2017).³ Beispielhaft sei dazu in Tabelle 2 eine Gegenüberstellung angeführt, in der für PIRLS die Itemverteilung bezüglich der Verstehensprozesse von 2001 bis 2016 verzeichnet ist (Mullis, Martin, Gonzales & Kennedy, 2003; Kennedy & Sainsbury, 2007; Mullis, Martin, Kennedy & Foy, 2007; Martin & Mullis, 2012; Mullis & Prendergast, 2017; siehe auch Bremerich-Vos, Wendt & Bos, 2017, S. 87). Die Verteilung der Items wird durch die jeweils zu erreichenden Score-Punkte und deren prozentualen Anteil ergänzt.

Bezüglich der Verteilung der Items auf Verstehensprozesse sind für PIRLS somit kaum Schwankungen im Zeitverlauf zu beobachten, ebenso wenig auf Ebene der Subskalen (PIRLS fasst in den Rückmeldungen die ersten beiden Verstehensprozesse zur Subskala „Wiedergeben und Schlussfolgern“, die letzten beiden zu „Interpretieren und Bewerten“ zusammen, siehe Wallner-Paschon & Widauer, 2017). Erkennbar ist allerdings die sukzessive Vergrößerung der eingesetzten Itemmengen im Zeitverlauf.

Für die BIST D4 liegt kein technischer Bericht vor. Aus den (unveröffentlichten) Itemlisten bzw. dem Testdesign lässt sich aber ableiten, dass sowohl die Kompetenzfelder als auch die einzelnen Deskriptoren relativ ungleich verteilt sind (siehe Tabelle 3).

Das Kompetenzfeld 1 („Explizite Informationen ermitteln“, K1), das durch 35 Items repräsentiert ist, beschreibt gemäß

3 Bei einem Vergleich zu berücksichtigen sind die unterschiedlichen Zählweisen bzw. Größen: In internationalen PIRLS-Berichten findet man bei vergleichbaren Auflistungen etwa die Überschriften „percentages of the assessment devoted to each [purpose and process]“ (ohne dass die zugrunde liegenden Kenngrößen bezeichnet werden, also etwa theoretisch angenommene Testzeit, Punkte, Items ...) (PIRLS 2001; siehe Kelly, 2003, S. 5) oder „percentage of total score points“ (PIRLS 2006; siehe Kennedy & Sainsbury, 2007, S. 21). In einigen Berichten werden diese Mengenverteilungen als Verteilungen von Items aufgefasst (siehe Tabelle 1 bzw. Bremerich-Vos, Wendt & Bos, 2017, S. 87), obwohl das Wording in Bezug auf Mengenverteilungen im internationalen „Framework“ lautet: „the percentages of the test devoted to each for PIRLS, PIRLS Literacy, and ePIRLS“ (Mullis & Martin, 2015, S. 14), also nicht direkt auf eine Verteilung von Items verweist.

der „Konstrukt- und Kompetenzstufenbeschreibung“ zur BIST D4 (siehe BIFIE, 2016) „auf der Wort- und Satzebene die Fähigkeit, Informationen aus literarischen Texten sowie aus Sach- und Gebrauchstexten zu entnehmen“. Diese Definition lässt sich direkt auf Formulierungen der Can-Do-Statements aus der Bildungsstandardverordnung zurückführen, die im Folgenden nach ihrem Auftreten in der Verordnung nummeriert werden (BGBl. II Nr. 1/2009, BGBl. II Nr. 282/2011, BGBl. II Nr. 185/2012).

Eine Zwischenbemerkung: Aus der gesetzlichen Verordnung und den Veröffentlichungen des BIFIE lassen sich für die einzelnen Can-Do-Statements die Bezeichnungen „Kompetenzen“ bzw. „Deskriptoren“ herauslesen. Die Can-Do-Statements sind unter Zwischenüberschriften gebündelt, die nicht gleich den schon angesprochenen Kompetenzfeldern sind und für die den Dokumenten keine gesonderte Begrifflichkeit zu entnehmen ist. Sie werden im Folgenden als „den Can-Do-Statements übergeordnete Kategorie“ bezeichnet. Zudem ist zu erwähnen, dass in der BIST D4 nicht alle Deskriptoren der Verordnung getestet werden: In der Verordnung finden sich 21 Deskriptoren zu sieben übergeordneten Kategorien, getestet werden jedoch nur 10 Deskriptoren aus vier übergeordneten Kategorien.⁴ Eine detaillierte Begründung dieser Auswahl liegt nicht vor. Viele der nicht getesteten Deskriptoren sind im Rahmen von Large-Scale Assessments kaum bzw. nicht administrierbar (etwa jene aus der Kategorie „7. Literarische Angebote und Medien aktiv nutzen“, siehe Fußnote 4).

Die Definition von Kompetenzfeld 1 ist also zurückzuführen auf die Deskriptoren „Die Schülerinnen und Schüler verfügen über sicheres Leseverständnis auf der Wort- und Satzebene“ (2.2) aus der übergeordneten Kategorie „Über eine altersadäquate Lesefertigkeit und ein entsprechendes Leseverständnis verfügen“ sowie aus dem Can-Do-Statement „Schülerinnen und Schüler können Informationen aus literarischen Texten sowie aus Sach- und Gebrauchstexten entnehmen“ (3.2) aus der übergeordneten Kategorie „Den Inhalt von Texten mithilfe von Arbeitstechniken und Lesestrategien erschließen“.

Auffällig ist die relativ starke Repräsentation von Kompetenzfeld 2 („Allgemeines Textverständnis“, K2). Die Definition von Kompetenzfeld 2 lautet gemäß der „Konstrukt- und Kompetenzstufenbeschreibung“ zur BIST D4: „Die Teilkom-

petenz [bzw. das Kompetenzfeld] *Allgemeines Textverständnis* beschreibt die Fähigkeiten, Textinformationen bzw. Inhalte zu erkennen, wiederzugeben, miteinander zu vergleichen, zu ordnen und/oder zu verknüpfen. Des Weiteren kann das Wesentliche eines Textes erfasst sowie der Verlauf einer Handlung erschlossen werden. Die Unterscheidung von Textsorten nach wesentlichen Merkmalen und auch das Erkennen der sprachlichen und formalen Gestaltung bzw. des Aufbaus eines Textes werden dieser Teilkompetenz [bzw. diesem Kompetenzfeld] zugeordnet“ (BIFIE, 2016, S. 8). Auch diese Definition lässt sich direkt auf Formulierungen der Can-Do-Statements aus der Bildungsstandardverordnung zurückführen, und zwar auf die Statements „Die Schülerinnen und Schüler können Informationen aus Texten miteinander vergleichen“ (3.4), „Inhalte/Informationen aus Texten ordnen“ (3.5), „den Verlauf einer Handlung erschließen“ (3.6), „das Wesentliche eines Textes erfassen“ (3.7) aus der übergeordneten Kategorie „Den Inhalt von Texten mithilfe von Arbeitstechniken und Lesestrategien erschließen“ (3) sowie „Die Schülerinnen und Schüler können einfache sprachliche und formale Gestaltung sowie den Aufbau von Texten erkennen“ (6.1) und „Textsorten nach wesentlichen Merkmalen unterscheiden“ (6.2) aus der übergeordneten Kategorie „Formale und sprachliche Gegebenheiten in Texten erkennen“ (6).

Kompetenzfeld 3 („Textbezogenes Interpretieren und Reflektieren“, K3) beschreibt „die Fähigkeiten, den Sinn von Texten zu klären, nicht ausdrücklich genannte Sachverhalte zu verstehen bzw. zu einem Text Stellung nehmen zu können“. Diese Formulierung ist zurückzuführen auf die Can-Do-Statements „Die Schülerinnen und Schüler können ihr Textverständnis artikulieren und kommunizieren“ (4.1) und „den Sinn von Texten klären und auch nicht ausdrücklich genannte Sachverhalte verstehen“ (4.2) aus der übergeordneten Kategorie „Das Textverständnis klären und über den Sinn von Texten sprechen“ (4).

In Deskriptor 4.2 klingt neben der Idee einer allgemeinen Interpretationsleistung das in der Leseforschung und vielen Large-Scale Assessments etablierte Konzept der Inferenzbildung an (siehe etwa MacKoon & Ratcliff, 1992; West & Stanovich, 2000; Zwaan & Singer, 2003; Khalifa & Weir, 2009), das in den Bildungsstandards keine prominente Rolle spielt und nicht näher definiert wird. In PIRLS wird diese Fähigkeit als eigener

4 Nicht getestet werden die Deskriptoren „Die Schülerinnen und Schüler können beim Vorlesen interessiert und bewusst zuhören“ sowie „... Bücher und Texte nach eigenem Interesse in verschiedenen Medien selbst auswählen“ aus der übergeordneten Kategorie „1. Die Lesemotivation bzw. das Leseinteresse festigen und vertiefen“; „... können ihre Lesefertigkeit an einfachen Texten zeigen“ sowie „... können ihre Verlesungen korrigieren“ aus der übergeordneten Kategorie „2. Über eine altersadäquate Lesefertigkeit und ein entsprechendes Leseverständnis verfügen“; „... Arbeitstechniken und Lesestrategien zur Texterschließung anwenden“ sowie „... zur Klärung fehlender bzw. unzureichender Informationen zusätzliche Quellen, einschließlich elektronischer Medien, nutzen“ aus der übergeordneten Kategorie „3. Den Inhalt von Texten mithilfe von Arbeitstechniken und Lesestrategien erschließen“; „... ihr Textverständnis artikulieren und kommunizieren“ aus der übergeordneten Kategorie „4. Das Textverständnis klären und über den Sinn von Texten sprechen“; „... einen Text sinngestaltend vortragen bzw. ihn umgestalten“ sowie „... Sach- und Gebrauchstexte für die Ausführung bestimmter Tätigkeiten verstehen und nutzen“ aus der übergeordneten Kategorie „5. Verschiedene Texte gestaltend oder handelnd umsetzen“; „... Bücher und Medien zur Gewinnung von Information und zur Erweiterung ihres Wissens nutzen“ sowie „... literarische Angebote zur Erweiterung ihres Selbst- und Weltverständnisses sowie zur Unterhaltung nutzen“ aus der übergeordneten Kategorie „7. Literarische Angebote und Medien aktiv nutzen“.

Verstehensprozess geführt („make straightforward inferences“; Mullis & Martin, 2015), in der BIST D4 dem (in PIRLS als vom einfachen Schlussfolgern als verschieden zu denkenden) Interpretieren (Kompetenzfeld 3, siehe oben) zugeschlagen.

Die Definitionen der Kompetenzfelder werden also in der „Konstrukt- und Kompetenzstufenbeschreibung“ der BIST D4 nicht aus der Forschungstradition abgeleitet, sondern auf die Bildungsstandardverordnung zurückgeführt, deren theoretische Fundierung nicht offengelegt ist.

Abbildung 1 zeigt als Zusammenschau der bisherigen Ausführungen eine mögliche Anbindung der PIRLS-Verstehensprozesse an die Kompetenzfelder der BIST D4. Pfeile stellen hier exemplarische Zuordnungsmöglichkeiten dar (durchgehende für hauptsächliche Parallelen und gestrichelte für zusätzliche Anbindungen), welche sich aus den Beschreibungen der für die einzelnen Bereiche geforderten Kompetenzen der Schüler/innen ergeben.

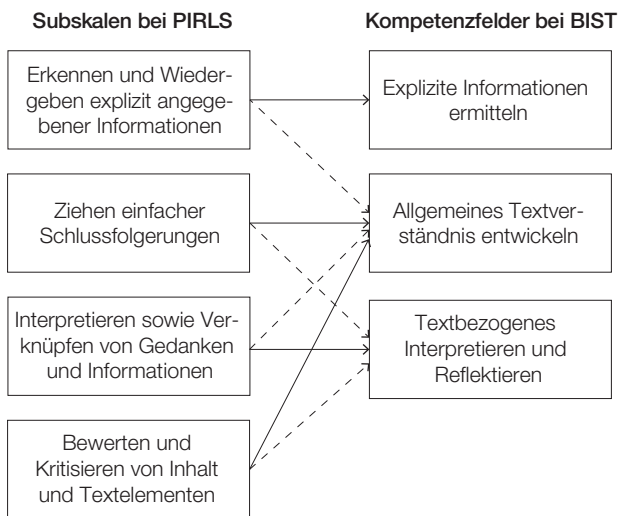


Abbildung 1: Anbindung der PIRLS-Verstehensprozesse an die BIST-D4-Kompetenzfelder

Unterschiedlich verstandene Begrifflichkeiten in BIST D4 und PIRLS (also etwa die schwache Trennung von expliziter Informationsentnahme und Inferenzbildung in der BIST D4, siehe oben) führen nun zu Unterschieden in den operationalisierten Konstrukten, die im Folgenden genauer dargestellt werden sollen.

PIRLS-Items aus Perspektive der BIST D4

Folgende Prozentverteilung (siehe Tabelle 4) ergibt sich bei einer Analyse der PIRLS-Items nach den eben vorgestellten BIST-D4-Kompetenzfeldern und Deskriptoren:

Auf **Ebene der BIST-D4-Kompetenzfelder** ist nach BIST-D4-Kriterien der Großteil aller PIRLS-Items den Kompetenzfeldern K1 („Explizite Informationen ermitteln“) und K3 („Textbezogenes Interpretieren und Reflektieren“) zuzuordnen. Auf K1 entfallen 46 % aller Items, auf K3 49 % aller Items.

Damit geht einher, dass besonders das Kompetenzfeld 2 der BIST D4, „Allgemeines Textverständnis“, – also jenes Kompetenzfeld, das in der BIST D4 mit 51 % aller Items relativ stark repräsentiert ist (siehe Tabelle 3) – von den PIRLS-Items kaum abdeckt wird. Nur 5 % aller PIRLS-Items sind dieser Kategorie zugeordnet (siehe Tabelle 4). Das liegt zum einen daran, dass die Deskriptoren, die in der BIST D4 dem allgemeinen Textverständnis zugeordnet werden (ähnlich der globalen Kohärenzbildung nach Rosebrock & Nix, 2017, bzw. mentalen Modellen etwa bei O’Brien & Albrecht, 1992), die Fähigkeit sehr differenziert aufteilen, sodass die zentrale Kompetenz dieses Kompetenzfelds „das Wesentliche eines Textes erfassen“ (3.7) ein (zudem mit 9 Items relativ schwach vertretener) Deskriptor neben anderen (3.4, 3.5, 3.6) ist.

Zum anderen werden nach dem theoretischen Konstrukt der BIST D4 auch die Fähigkeiten „Die Gestaltung und den Aufbau von Texten erkennen“ (6.1) und „Textsorten nach wesentlichen Merkmalen unterscheiden“ (6.2) diesem Kompetenzfeld zugeordnet und mit gesamt 30 Items intensiv getestet – Fähigkeiten also, die etwa im Lesekompetenzmodell von Rosebrock & Nix nicht der globalen Kohärenzbildung, sondern dem Erkennen von Superstrukturen und der Identifikation von Darstellungsstrategien zugeordnet werden (Rosebrock & Nix, 2017). Solche Items gibt es in PIRLS kaum.

Fragen nach Textsorten (wie Deskriptor 6.2) kennt PIRLS gar nicht. Items, in denen die Gestaltung und der Aufbau von Texten eine Rolle spielt, haben bei PIRLS immer weitere bzw. andere Komponenten (etwa Reflexionen zur Wahl eines bestimmten Satzes in einer Geschichte), sodass sie nach BIST-D4-Systematik meist dem „Textbezogenen Interpretieren und Reflektieren“ zugeordnet werden müssen. Der Schwerpunkt in PIRLS liegt vor dem Hintergrund der BIST-D4-Kriterien somit auf den Kompetenzfeldern „Explizite Informationsentnahme“ (46 % aller Items) und „Textbezogenes Interpretieren und Reflektieren“ (49 % aller Items).

Die Unterschiede der von den einzelnen Items abgedeckten Kompetenzen treten noch deutlicher zutage, wenn man die **Ebene der einzelnen BIST-D4-Deskriptoren** betrachtet: Es finden sich in PIRLS keine Items, die den BIST-Deskriptoren „Die Schülerinnen und Schüler verfügen über sicheres Leseverständnis auf der Wort- und Satzebene“ (2.2), „Die Schülerinnen und Schüler können Informationen aus Texten miteinander vergleichen“⁵ (3.4), „Die Schülerinnen und

5 Zu Eigenheiten der Itemkonstruktion für Deskriptor 2.2. siehe Abschnitt *BIST-D4-Items aus Perspektive von PIRLS*.

		ges.	K1	2.2	3.2	K2	3.4	3.5	3.6	3.7	6.1	6.2	K3	4.2	4.3
PIRLS gesamt		100	46	0	46	5	0	0	1	3	1	0	49	35	14
Absicht	literarisch	52	19	0	19	1	0	0	0	0	1	0	32	22	10
	Information	48	27	0	27	4	0	0	1	3	0	0	17	13	4
Verstehensprozess	Wiedergeben	28	28	0	28	0	0	0	0	0	0	0	0	0	0
	Schlussfolgern	31	12	0	12	1	0	0	0	1	0	0	18	17	1
	Wiedergeben & Schlussfolgern	59	40	0	40	1	0	0	0	1	0	0	18	17	1
	Interpretieren	27	6	0	6	2	0	0	1	1	1	0	19	12	7
	Bewerten	14	0	0	0	1	0	0	0	1	0	0	18	6	7
	Interpretieren & Bewerten	41	6	0	6	3	0	0	1	2	1	0	38	18	14

Tabelle 4: Verteilung der PIRLS-Items nach Leseabsicht, Verstehensprozess sowie BIST-D4-Kompetenzfeld und BIST-D4-Deskriptor in Prozent (PIRLS 2016)

Schüler können Inhalte/Informationen aus Texten ordnen“ (3.5) und „Die Schülerinnen und Schüler können Textsorten nach wesentlichen Merkmalen unterscheiden“ (6.2) zugeordnet werden können. Genauer: der Art und Weise, wie diese in den entsprechenden BIST-Items konkret operationalisiert werden, etwa in Form von Sequencing-Formaten für Deskriptor 3.5 oder im Erfragen von Textsortenbezeichnungen für Deskriptor 6.2.⁶

Der Deskriptor „Schülerinnen und Schüler können Informationen aus literarischen Texten sowie aus Sach- und Gebrauchstexten entnehmen“ (3.2) erscheint deshalb als stark ausgeprägt, wenn man die PIRLS-Items nach BIST-D4-Kriterien klassifiziert, da in der BIST D4 auch kleinräumige Inferenzen (etwa indem im Fragenstamm einer Aufgabe mittels eines Synonyms nach einer bestimmten Information im Stimulus gefragt wird) der „expliziten Informationsentnahme“ zugeordnet werden. PIRLS definiert die Grenze zwischen den Verstehensprozessen „Erkennen und Wiedergeben explizit angegebener Informationen“ und dem „Ziehen einfacher Schlussfolgerungen“, wie bereits erwähnt, strenger.

Ein ähnliches Bild ergibt sich, wenn man ausschließlich PIRLS-Items aus der **PIRLS-Leseabsicht „Literarische Erfahrungen“** betrachtet (siehe der hell eingefärbte obere Abschnitt von Tabelle 4). Wiederum ist eine schwache Repräsentation des BIST-D4-Kompetenzfelds K2, „Allgemeines Textverständnis“, zu beobachten (1 % aller Items), außerdem ist die relativ starke Abbildung des Kompetenzfelds „Textbezogenes Interpretieren und Reflektieren“ auffällig (32 % aller Items).

In Deskriptoren bildet sich dieser Trend folgendermaßen ab: Von den zehn Deskriptoren, die in der BIST D4 getestet werden, werden in den PIRLS-Items zur Leseabsicht „Literarische Erfahrungen“ nur vier abgebildet.

Betrachtet man nun die PIRLS-Items zur **PIRLS-Leseabsicht „Informationsgewinnung“**, zeigen sich andere Tendenzen: Es besteht nun ein Überhang aufseiten des BIST-D4-Kompetenzfelds K1, „Explizite Informationen ermitteln“ (27 % aller Items).

Klassifiziert nach BIST-D4-Deskriptoren zeigt sich, dass die Kompetenzen „Schülerinnen und Schüler können Informationen aus literarischen sowie aus Sach- und Gebrauchstexten entnehmen“ (3.2) und „Schülerinnen und Schüler können den Sinn von Texten klären und auch nicht ausdrücklich genannte Sachverhalte verstehen“ (4.2) dominieren.

Ähnliche Analysen können nun anstatt auf Basis der Leseabsichten auf Basis der PIRLS-Verstehensprozesse vorgenommen werden (siehe der dunkel eingefärbte untere Abschnitt von Tabelle 4). Nicht nur Items, die laut PIRLS-Systematik – wie sie im „Framework“ (Mullis & Martin, 2015) und vor allem in den „Item Writing Guidelines“ (Mullis & Martin, 2013) beschrieben wird⁷ – dem Verstehensprozess „Erkennen und Wiedergeben explizit angegebener Informationen“ zugeordnet sind, würden in der BIST D4 im Kompetenzfeld K1, „Explizite Informationen ermitteln“, administriert. Auch PIRLS-Items der Verstehensprozesse „Schlussfolgern“ und „Interpretieren“ wären nach

6 Die konkreten Operationalisierungen einzelner Deskriptoren sind aus der Gesamtschau der entsprechenden Items ersichtlich und können hier nicht im Detail dargestellt werden.

7 Eine Beschreibung des Prozesses der Itemerstellung der BIST D4 findet sich unter: https://www.bifie.at/wp-content/uploads/2017/05/TD_Itemerstellung_BiSt-UE.pdf (S. 6 ff.).

BIST-D4-Kriterien im Kompetenzfeld K1, „Explizite Informationen ermitteln“, verortet.

BIST-D4-Items aus Perspektive von PIRLS

Es soll nun der Fokus umgekehrt werden: Wie verteilen sich die Items der BIST D4, wenn sie nach den Einteilungskriterien von PIRLS klassifiziert werden?

	gesamt	Leseabsicht			Verstehensprozess					
		literarisch/nar.	Information	Satz	Wiedergeben	Schlussfolgern	Wiedergeben & Schlussfolgern	Interpretieren	Bewerten	Interpretieren & Bewerten
ges.	100	47	43	10	10	27	38	44	19	62
K1	30	12	9	10	8	17	25	5	0	5
2.2	13	3	0	10	0	11	11	2	0	2
3.2	17	9	9	0	8	6	14	3	0	3
K2	51	23	27	0	3	9	12	21	18	39
3.4	3	1	2	0	3	1	3	0	0	0
3.5	9	3	7	0	0	2	2	7	1	8
3.6	5	3	2	0	0	3	3	2	0	2
3.7	8	3	5	0	0	2	2	6	0	6
6.1	10	7	3	0	0	2	2	7	2	9
6.2	15	7	8	0	0	0	0	0	15	15
K3	19	12	7	0	0	1	1	17	1	18
4.2	14	9	5	0	0	1	1	12	1	13
4.3	5	3	2	0	0	0	0	5	0	5

Tabelle 5: Verteilung der BIST-D4-Items nach Kompetenzfeld, Deskriptor sowie PIRLS-Leseabsicht und PIRLS-Verstehensprozess in Prozent (BIST D4 2015)

Wie Tabelle 5 zeigt, bilden die Items der BIST D4 die **PIRLS-Leseabsichten** relativ ausgewogen ab: 47 % aller Items beziehen sich auf die Leseabsicht „Literarische Erfahrung“, 43 bzw. 53 % aller Items auf die Leseabsicht „Informationsentnahme“.

Der Deskriptor 2.2, „Die Schülerinnen und Schüler verfügen über sicheres Leseverständnis auf der Wort- und Satzebene“ wird in der BIST D4 hinsichtlich der *Satzebene* anhand von Items getestet, deren Stimuli nur ein bis zwei Sätze umfassen. Einige dieser Items, bei denen etwa Konnektoren (Konjunktionen, Pronomen etc.) in Satzlücken eingesetzt werden

müssen, könnten wohl statt der Lesekompetenz eher dem BIST-D4-Kompetenzbereich „Sprachbewusstsein“ zugeordnet werden. Sie sind in Tabelle 5 mit der Bezeichnung „Satz“ (10 % aller Items) verzeichnet, werden aber im Folgenden der Leseabsicht „Information“ zugerechnet (was durchaus diskussionswürdig ist). Die Wortebene wird mit einem gesonderten Speedtest abgedeckt (siehe BIFIE, 2016). In diesem Test gilt es, innerhalb von einer Minute möglichst viele Items zu lösen, bei denen einem Bild jeweils vier Wörter beige stellt sind, von denen eines inhaltlich dem Bild entspricht (Dekodierfähigkeit). Die Items dieses Speedtests sind nicht Gegenstand der vorliegenden Untersuchung.

Richtet man den Fokus auf die **PIRLS-Verstehensprozesse**, zeigt sich in der BIST D4 ein ungleiches Verhältnis der PIRLS-Subskalen: BIST-D4-Items testen in sehr hohem Ausmaß Kompetenzen ab, die im Sinne von PIRLS dem Interpretieren und Bewerten zuzuordnen sind (62 %). Das Wiedergeben und einfache Schlussfolgern im Sinne von PIRLS ist in BIST-D4-Items deutlich unterrepräsentiert (38 %).

So kommt es etwa, dass sich der Deskriptor 3.2 der BIST D4 auf mehrere PIRLS-Verstehensprozesse verteilt. Sogar Items, die nach den PIRLS-Kriterien dem Verstehensprozess „Interpretieren“ zugeordnet werden würden, sind in der BIST D4 im Kompetenzfeld „Explizite Informationen ermitteln“ verzeichnet.

Die BIST-D4-Items des in PIRLS nur schwach repräsentierten Kompetenzfelds K2, „Allgemeines Textverständnis“, sind nach PIRLS-Kriterien den Verstehensprozessen „Schlussfolgern“ bis „Evaluieren“ zuzuordnen, ebenso verhält es sich mit Items aus dem BIST-D4-Kompetenzfeld K3, „Textbezogenes Interpretieren und Reflektieren“.

Diese Unterschiede im Verständnis wesentlicher Kategorien können auch zu unterschiedlichen Diagnosen über die getesteten Schüler/innen führen, wie weiter unten diskutiert werden wird.

Zusammenschau: BIST-D4- und PIRLS-Items aus der Perspektive beider Tests

Die wichtigsten Ergebnisse obiger Ausführungen sind in Abbildung 2 noch einmal zusammengefasst. Zwecks eindrücklicherer Visualisierung der Gemeinsamkeiten und Unterschiede zwischen den Studien wird hier und im Folgenden die Darstellungsform des Profils verwendet, ohne dass dadurch ein Verlauf zwischen den einzelnen Merkmalen angedeutet werden soll. Die Ähnlichkeiten und Unterschiede hinsichtlich der Kategorien der jeweils anderen Studie treten deutlich zutage:

- Die Verteilungen der Items hinsichtlich der PIRLS-Leseabsichten sind vergleichbar (allerdings gibt es bei der BIST D4 Items mit einem einzelnen Satz als Stimulus).

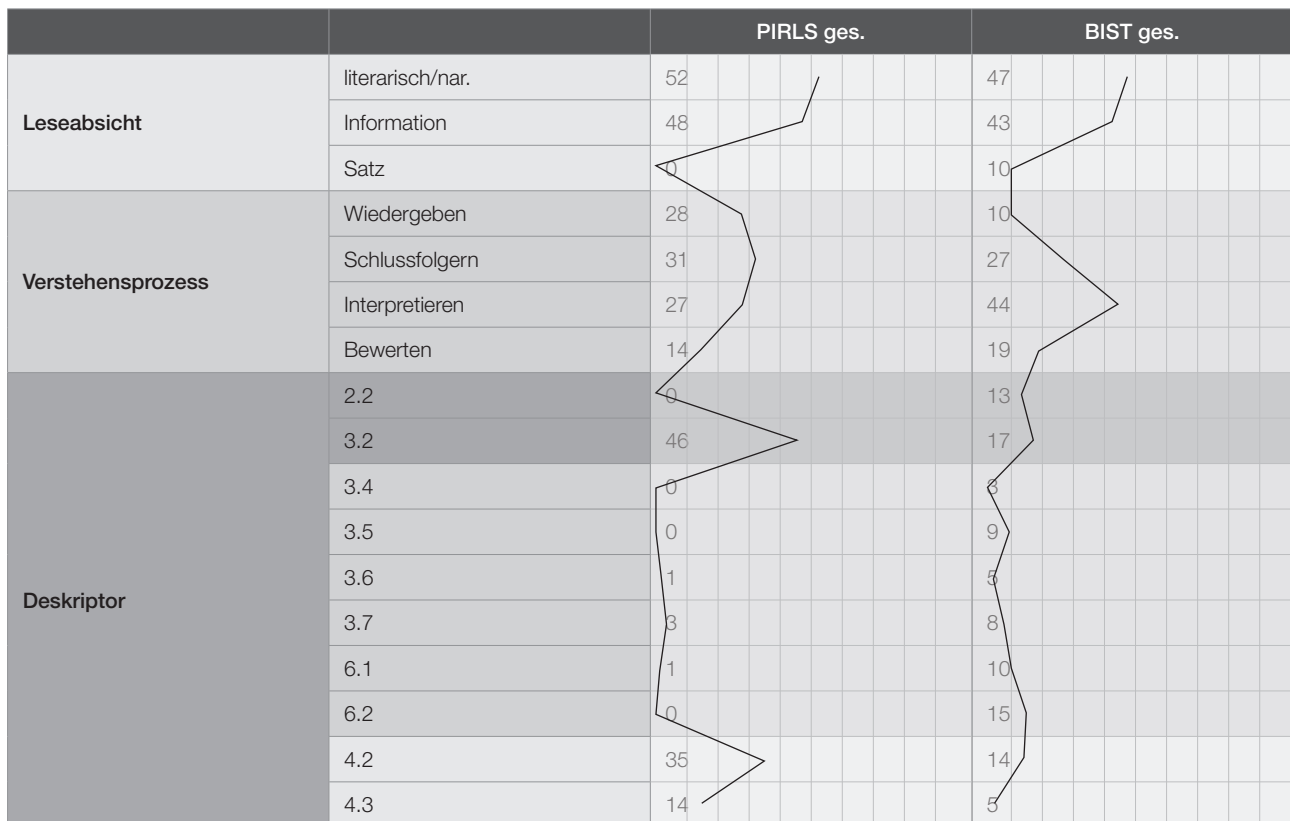


Abbildung 2: „Studienprofile“: Verteilung der BIST-D4- und PIRLS-Items nach Leseabsicht, Verstehensprozess, Deskriptoren (und Kompetenzfeldern) in Prozent (BIST D4 2015, PIRLS 2016)

- Die dominierenden PIRLS-Verstehensprozesse gemessen an der Anzahl zugeordneter Items sind in PIRLS vor allem das Wiedergeben und Schlussfolgern, in der BIST D4 das Interpretieren und Bewerten. Etwa zwei Drittel aller BIST-D4-Items gehören zu diesen beiden PIRLS-Kategorien.
- Von den zehn BIST-Deskriptoren werden in PIRLS sechs nicht bzw. kaum getestet. Dominierend sind „Informationen aus literarischen Texten sowie aus Sach- und Gebrauchstexten“ (3.2) sowie „den Sinn von Texten klären und auch nicht ausdrücklich genannte Sachverhalte verstehen“ (4.2). Etwa 80 % aller PIRLS-Items sind diesen beiden BIST-Kategorien zuzuordnen.
- In der BIST D4 sind vor allem die eben genannten Deskriptoren sowie das „sichere Leseverständnis auf der Wort- und Satzebene“ (2.2, Anmerkungen siehe oben) und „Textsorten nach wesentlichen Merkmalen unterscheiden“ (6.2.) relativ stark in den Items abgebildet.

PIRLS- und BIST-D4-Items nach Stimulus- und Itemzahlen, Leseload und Lesbarkeitsindices

Im Folgenden sollen die operationalisierten Konstrukte von PIRLS und BIST D4 nach weiteren Kriterien analysiert werden.

Dazu werden bei PIRLS jeweils Daten zum gesamten Bestand an Stimuli bzw. Items sowie Daten zu den Leseabsichten „Informationsgewinnung“ und „Literarische Erfahrungen“ dargestellt.

Bei der BIST D4 ist die Sachlage komplexer, da ein Testheft immer aus einem langen Text mit vier bis fünf Items (BIST lang) sowie 14 weiteren Items zu jeweils einem eigenen kurzen Stimulus (BIST kurz) besteht.

Dargestellt werden deshalb immer Daten zum Gesamtbestand an Stimuli bzw. Items der BIST D4 sowie zum Bestand an Stimuli bzw. Items, die bei PIRLS den Leseabsichten „Informationsgewinnung“ und „Literarische Erfahrungen“ zuzuordnen wären.

	PIRLS gesamt	PIRLS literarisch	PIRLS Information	BIST gesamt	BIST gesamt Info	BIST gesamt nar./lit.	BIST lang gesamt	BIST lang Info	BIST lang nar./lit.	BIST kurz gesamt	BIST kurz Info	BIST kurz nar./lit.	BIST kurz nar.
Anzahl Stimuli	12	6	6	102	59	43	4	1	3	98	58	40	33
Anzahl Items	175	93	88	117	62	55	19	4	15	98	58	40	33
Wörter Stim. (MW)	785	800	769	101	93	135	616	757	569	83	65	101	113
Wörter Item (MW)	26	25	26	38	45	33	36	33	36	39	45	39	33
Wörter Stim. Testh.	1569	1569	1569	1778	1919	1731	x	x	x	x	x	x	x
Wörter Items Testh.	741	741	742	701	678	726	x	x	x	x	x	x	x
Wörter ges. Testh.	2310	2310	2310	2479	2597	2457	x	x	x	x	x	x	x
Testzeit ges. (min.)	80	40	40	30	30	30	x	x	x	x	x	x	x
Wörter/min. Stim.	20	20	19	59	64	58	x	x	x	x	x	x	x
Wörter/min. Items	9	19	19	23	23	24	x	x	x	x	x	x	x
Wörter/min. ges.	29	29	29	83	87	82	x	x	x	x	x	x	x
Flesch Stimulus	66	73	58	66	62	74	72	69	72	68	62	74	70
Flesch Items	65	72	57	65	63	67	66	73	65	64	62	68	67
LIX Stimulus	33	29	37	36	40	31	32	34	32	35	40	30	32
LIX Items	31	26	36	36	37	35	34	32	35	36	37	35	34

Legende:

PIRLS gesamt: alle Stimuli aus PIRLS (mehrere Items pro Text)
PIRLS literarisch: die literarischen Stimuli aus PIRLS
PIRLS Information: die Informationsstimuli aus PIRLS

BIST lang gesamt: alle langen Stimuli der BIST D4 (mehrere Items pro Text)
BIST lang Info: die langen Informationsstimuli der BIST D4
BIST lang nar./lit.: die langen literarischen bzw. narrativen Stimuli der BIST D4
BIST kurz gesamt: alle kurzen Stimuli der BIST D4 (ein Item pro Text)
BIST kurz Info: die kurzen Informationsstimuli der BIST D4
BIST kurz nar./lit.: die kurzen literarischen bzw. narrativen Stimuli der BIST D4
BIST kurz nar.: die kurzen narrativen Stimuli der BIST D4

Tabelle 6: Ausgewählte Kennwerte der Konstrukte in absoluten Zahlen (PIRLS 2016 und BIST D4 2015)

Letztere umfasst in der BIST D4 literarische (und teils nicht narrative, z. B. Gedichte) und narrative (und oft nicht literarische, z. B. Erlebniserzählungen) Texte und wird deshalb mit „BIST gesamt nar./lit.“ abgekürzt. In vorliegender Untersuchung wurde die Einteilung in „literarisch“ bzw. „narrativ“ pragmatisch getroffen, da in den meisten Fällen keine authentischen, sondern speziell für den Test angefertigte Stimuli verwendet wurden und die Klassifikation „literarisch“ v. a. als Gegensatz zum „Informationstext“ aufzufassen ist (zur Diskussion der Begriffe Informationstext, Sachtext, expositorischer Text, nicht fiktionaler Text etc. siehe etwa Krieg-Holz, 2016; Fisher, Frey & Lapp, 2016).

Ebenso wird diese Aufteilung dargestellt für die speziellen Gruppen von Langtexten sowie Kurztexten, also der Kombinationen von einem Stimulus und einem einzelnen Item. Zusätzlich wird bei den Kurztexten noch die Subgruppe der

„literarischen“ Texte mit rein narrativem Charakter ausgewiesen („BIST kurz nar.“).

Vergleicht man in Tabelle 6 die absoluten Zahlen an eingesetzten Stimuli und Items, ergibt sich folgendes Bild: Die Menge an im Test eingesetzten Stimuli bzw. Items ist zwischen PIRLS und BIST D4 sehr unterschiedlich. In PIRLS kommen deutlich weniger Stimulustexte zum Einsatz als in der BIST D4 (12 im Vergleich zu 102), allerdings ist in PIRLS die Menge an Items höher (175 zu 117). In PIRLS sind jedem Stimulus im Durchschnitt 14,6 Items zugeordnet, bei der BIST D4 sind es 1,1.

In PIRLS befinden sich in jedem Testheft zwei Stimuli mit durchschnittlich 29 Items, welche in 80 Minuten bearbeitet werden. Bei der BIST D4 sind es im Mittel 15 Stimuli und 19 Items pro Testheft, für deren Bearbeitung die Schüler/innen

30 Minuten Zeit haben. Zudem fällt auf, dass die Konstrukt- abdeckung in Bezug auf lange Texte in der BIST D4 deutlich kleiner und ungleich gewichtet ist: Bei PIRLS lässt sich, wie bereits erwähnt, eine Verteilung der Leseabsichten von 50:50 erkennen, jeweils repräsentiert durch sechs Texte. Bei der BIST D4 kommen insgesamt vier lange Texte zum Einsatz, drei davon sind narrativ, einer eine Art Hybrid: sachtextähnliche Inhalte werden in Form eines Gesprächs zwischen mehreren Figuren verhandelt. Trotz dieser Uneindeutigkeit wurde der Text in vorliegender Studie aus pragmatischen Gründen (etwa dem Zählen und Vergleichen von Itemmengen) der Leseabsicht „Informationsgewinnung“ zugeordnet.

Vergleicht man die durchschnittliche Wortanzahl in Stimuli und Items zwischen PIRLS und BIST D4, so fällt auf, dass die Unterschiede zwischen „Informationsgewinnung“ und „Literarische Erfahrungen“ in PIRLS klein sind. Bei der BIST D4 ist der Informationstext deutlich länger (etwa 200 Wörter) als der durchschnittliche literarische/narrative lange Text. Bei den BIST-D4-Kurztexten ist das Bild umgekehrt: Informationstexte sind im Mittel deutlich kürzer als literarische/narrative insgesamt und viel kürzer als die Gruppe der narrativen Texte im Speziellen. Ebenso fällt auf, dass in der BIST D4 die Wortanzahl in den Items – also jenen Teilen der Aufgabe, die Frage, Antwortoptionen und sonstige Anweisungen beinhalten (kurz: alles außer dem Stimulus) – durchwegs höher ist als in PIRLS.

Richtet man nun den Fokus nicht auf die Menge an Stimuli und Items für beide Tests im Gesamten, sondern auf einzelne Testhefte, wie sie den Schülerinnen und Schülern vorgelegt werden, treten weitere Unterschiede zutage. Verglichen werden dazu fiktive prototypische Testhefte. Bei PIRLS besteht ein Testheft aus einem durchschnittlich langen Text zur Leseabsicht „Informationsgewinnung“ sowie einem durchschnittlich langen Text zur Leseabsicht „Literarische Erfahrungen“ sowie durchschnittlich 29,2 dazugehörigen Items. Bei der BIST D4 werden ein fiktives prototypisches Testheft mit einem Informationstext und 14 durchschnittlich langen Kurztexten (gemischt informativ und literarisch/narrativ) („BIST gesamt Info“) sowie ein fiktives prototypisches Testheft mit einem durchschnittlich langen literarischen/narrativen Text und 14 durchschnittlich langen Kurztexten (ebenfalls gemischt informativ und literarisch/narrativ) („BIST gesamt nar./lit.“) sowie der Mittelwert aus diesen beiden fiktiven Testheften verglichen („BIST gesamt“). Es zeigt sich, dass die durchschnittliche Wortanzahl in der BIST D4 etwas länger ist als in prototypischen PIRLS-Testheften und dass dieser Unterschied v. a. durch eine höhere Wortanzahl in den Stimuli – also den Ausgangstexten einer Aufgabe ohne Frage, Antwortoptionen und sonstige Anweisungen – entsteht.

Zieht man nun noch die unterschiedlichen Bearbeitungszeiten für den Test – 80 min. für PIRLS, 30 min. für den Lese- testteil der BIST D4 (Wallner-Paschon & Salchegger, 2017; Breit, Bruneforth & Schreiner, 2016) – ins Kalkül, ergibt sich folgender Befund: In der BIST D4 muss mit durchschnittlich 83 Wörtern in der Minute beinahe dreimal so schnell gelesen werden wie bei PIRLS (29 Wörter pro Minute). In prototypischen BIST-D4-Testheften mit einem Langtext zur Leseabsicht „Informationsgewinnung“ liegt das Verhältnis im Vergleich zu PIRLS tatsächlich bei drei zu eins (87 Wörter pro Minute). Dieser Unterschied ist zumindest auffällig, hängt allerdings wohl auch damit zusammen, dass PIRLS an internationalen Stichproben normiert wurde, die BIST D4 nur im über dem PIRLS-Durchschnitt liegenden Österreich.

Hinsichtlich bekannter Lesbarkeitsindizes wie Flesch- und LIX-Index, also einfacher Zahlen zur Schätzung der prinzipiellen Verständlichkeit von Texten auf Basis von Wort- und Satzlängen, Silbenanzahl etc. (DuBay, 2004; Zamanian & Heydari, 2012), sind die Texte in der BIST D4 nicht einfacher als in PIRLS (siehe Tabelle 6).

Es kann also festgehalten werden, dass zwischen BIST D4 und PIRLS sowohl hinsichtlich der studieninternen Kategorien Leseabsichten, Verstehensprozesse, Kompetenzfelder und Deskriptoren als auch hinsichtlich allgemeinerer Indikatoren wie Stimulus- und Itemzahlen, Wortanzahl sowie Anforderungen an die Lesegeschwindigkeit Unterschiede festzustellen sind. Einzig bezüglich etablierter Lesbarkeitsindizes ist das Bild sehr homogen.

PIRLS- und BIST-D4-Stimuli aus narratologischer Perspektive

Ebenso groß wie die eben beschriebenen Unterschiede sind die Gegensätze hinsichtlich der inhaltlichen Beschaffenheit der in den beiden Tests eingesetzten Stimulustexte (der Einfachheit halber und entsprechend der Tradition in der empirischen Leseforschung im Folgenden „Genres“ genannt; zur Diskussion des Begriffs siehe etwa Scheinpflug, 2014).

Während bei PIRLS, wie schon gesagt, jeweils sechs „klassische“ expository Texte und sechs narrative Texte zum Einsatz kommen, ist das Bild in der BIST D4 ein anderes. Dargestellt werden in Tabelle 7 die Unterschiede in vorliegender Studie anhand der narrativen Stimuli.⁸

Jene 50 % aller Stimuli, die bei PIRLS der Leseabsicht „Literarische Erfahrungen“ zugeordnet werden, sind klassische (und oft authentische) Kurzgeschichten. Vier davon erzählen realistisch, zwei „phantastisch“ (zur Diskussion dieser Kategorien siehe etwa Durst, 2010). Betrachtet man die drei narrativen Texte

8 Eine differenzierte Auseinandersetzung mit Texten zur PIRLS-Leseabsicht „Informationsgewinnung“ und vergleichbaren Kategorien in BIST D4 muss aus Zeit- und Platzgründen unterbleiben.

	PIRLS gesamt	PIRLS lit.	PIRLS lit. %	BIST gesamt	BIST ges. % Stimuli	BIST nar./lit. % Stimuli	BIST nar. lang	BIST nar. lang % Stimuli	BIST nar. lang Items	BIST nar. kurz	BIST nar. kurz % Stimuli	BIST nar./lit. Items gesamt	BIST nar./lit. Items gesamt %
Auszählvers	0	0	0	2	2	5	0	0	0	0	0	2	4
Gedicht	0	0	0	5	5	12	0	0	0	0	0	5	9
lit., nicht nar. Texte	0	0	0	7	7	17	0	0	0	0	0	7	13
Erlebniserz./-ber.	0	0	0	21	21	49	1	33	5	20	61	26	47
Kürzestgeschichte	0	0	0	1	1	2	0	0	0	1	3	1	2
Roman(-ausschnitt)	0	0	0	2	2	5	0	0	0	2	6	2	4
Kurzgesch. real.	4	4	67	3	3	7	1	33	5	2	6	7	13
Kurzgesch. phant.	2	2	33	6	6	14	1	33	5	5	15	10	18
Märchen	0	0	0	2	1	2	0	0	0	2	6	1	2
Dialog	0	0	0	1	1	2	0	0	0	1	3	1	2
narrative Texte	6	6	100	36	35	81	3	100	15	33	100	48	88
nar./lit. Texte	6	6	100	43	42	100	3	100	15	33	0	55	100

Legende:

PIRLS gesamt: alle Stimuli aus PIRLS (mehrere Items pro Text)
PIRLS lit: Anzahl der literarischen Stimuli in PIRLS
PIRLS lit. %: Anteil der literarischen Stimuli in PIRLS in Prozent

BIST gesamt: Anzahl aller narrativen bzw. literarischen Stimuli in der BIST D4
BIST ges. % Stimuli: Anteil an allen Stimuli der BIST D4

BIST nar./lit. % Stimuli: Anteil an allen nar. bzw. lit. Stimuli der BIST D4
BIST nar. lang: Anzahl aller narrativen langen Stimuli in der BIST D4
BIST nar. lang % Stimuli: Anteil an allen nar. langen Stimuli der BIST D4
BIST nar. lang Items: Anzahl aller Items zu nar. langen Stimuli in der BIST D4
BIST nar. kurz: Anzahl aller narrativen kurzen Stimuli in der BIST D4
BIST nar. kurz % Stimuli: Anteil an allen nar. kurzen Stimuli der BIST D4
BIST nar./lit. Items gesamt: Anzahl aller Items zu nar. kurzen Stimuli BIST D4
BIST nar./lit. Items gesamt %: Anteil an allen Items zu nar. bzw. lit. Stim. B D4

Tabelle 7: Verteilung der Genres in literarische bzw. narrative Stimuli in absoluten Zahlen (Anzahl der Stimuli, Anzahl der Items) bzw. in Prozent (Anteil an allen lit. bzw. nar. Stimuli, Anteil an Items zu lit. bzw. nar. Stimuli) (PIRLS 2016 und BIST D4 2015)

in der BIST D4, so findet sich eine der realistischen Erzählweise verpflichtete Kurzgeschichte, eine „phantastische“ und eine Art Erlebniserzählung mit stark berichtendem Charakter.

Noch unterschiedlicher ist das Bild bei den narrativen Kurztexten: 47 % aller kurzen Stimuli der BIST D4 sind den Kategorien Erlebniserzählung bzw. Erlebnisbericht zuzuordnen, 18 % „phantastischen“ Kurzgeschichten(-auszügen), 16 % realistischen Kurzgeschichten(-auszügen). Der Rest verteilt sich auf Auszählverse, Kürzestgeschichten, (nicht authentische, aber textstrukturell wohl als solche zu bezeichnende) Romane/Romanauszüge, Märchen und (dramatisiert anmutende) Dialoge (mit narrativem Charakter).

Schließlich soll nun noch ein genauerer Blick auf die Beschaffenheit der in PIRLS und BIST D4 eingesetzten narrativen Texte erfolgen. Die Anzahl der analysierten Stimuli ist Tabelle 6 (grob: PIRLS: N = 12, BIST: N = BIST lang nar./

lit. + BIST kurz nar. = 36), die analysierten Kategorien den Tabellen 8 bis 10 zu entnehmen. Diese Kategorien folgen einer verbreiteten Einführung in die Erzählanalyse (Lahn & Meister, 2016). Die reichhaltige und Jahrzehnte andauernde Debatte narratologischer Begriffe kann hier nicht einmal in Ansätzen wiedergegeben werden, ebenso wenig werden die verwendeten Termini vertiefend diskutiert, sondern als dem interessierten Leser bekannt vorausgesetzt. Alternative Vorgehensweisen sind denkbar und allenfalls künftig wünschenswert.

Was intendiert wurde, ist eine grobe Gegenüberstellung von narratologischen Eigenheiten der Texte in PIRLS und BIST D4. Um die Vergleichbarkeit zu gewährleisten, wurden in einem ersten Schritt alle Stimuli aus der Leseabsicht „Literarische Erfahrungen“ aus PIRLS sowie alle narrativen Lang- und Kurztexte aus der BIST D4 nach gewissen Merkmalen analysiert und mittels einer meist dichotomen Systematik

	PIRLS lit. %	BIST nar. lang % Stimuli	BIST nar. kurz % Stimuli	BIST nar. ges. % Items
Darstellung des Erzählers				
implizit	66,7 %	66,7 %	75,8 %	72,9 %
explizit	33,3 %	33,3 %	24,2 %	27,1 %
offen	50,0 %	33,3 %	39,4 %	37,5 %
verborgen	50,0 %	66,7 %	60,6 %	62,5 %
Angaben zu Erzähler vorhanden	33,3 %	33,3 %	30,3 %	31,3 %
Ontologische Bestimmung des Erzählers in Relation zur erzählten Welt				
homodiegetisch (autodiegetisch)	33,3 % (33,3 %)	66,7 % (33,3 %)	48,5 % (30,3 %)	54,2 % (31,3 %)
heterodiegetisch	66,7 %	33,3 %	51,5 %	45,8 %
Wechsel der Erzählerposition	0 %	0 %	0 %	0 %
Ich-Erzähler (nach Stanzel)	33,3 %	33,3 %	30,3 %	31,3 %
auktoriale Erzählsituation (nach Stanzel)	33,3 %	66,7 %	45,5 %	52,1 %
personale Erzählsituation (nach Stanzel)	33,3 %	0 %	24,2 %	16,7 %
Zeitlogische Bestimmungen				
Erzählen im Präteritum	100 %	66,7 %	57,6 %	60,4 %
Erzählen im Präsens	0 %	33,3 %	42,4 %	39,6 %
prospektives Erzählen	0 %	0 %	0 %	0 %
Darstellung des Adressaten				
implizit	66,7 %	66,7 %	97,0 %	87,5 %
explizit	33,3 %	33,3 %	3,0 %	12,5 %

Tabelle 8: Verteilung der Parameter der Erzählinstanz in narrativen Stimuli (PIRLS 2016 und BIST D4 2015)

verzeichnet (Text X hat Eigenschaft A: ja/nein). Nicht dichotom zu analysierende Eigenschaften (z. B. die Verteilung von diegetischer versus mimetischer Erzählerrede in Texten, siehe unten) wurden nach ihrer prozentuellen Verteilung in den einzelnen Texten verzeichnet.

In einem nächsten Schritt wurden alle absoluten und prozentuellen Werte in ihrer Gesamtheit, also für die Menge

aller Stimuli bzw. Stimulusgruppen in PIRLS und BIST D4, betrachtet. Um auch zu berücksichtigen, dass die Eigenschaften der langen Stimuli stärker wiegen, da an ihnen im Unterschied zu Kurztexten mehrere Items angeschlossen sind, wurden die Verhältnisse auch auf Ebene der einzelnen Items berechnet (à la: In der Gesamtheit aller PIRLS/BIST-D4-Items findet sich Eigenschaft A zu a Prozent). Um die Ergebnisse besser zu veranschaulichen, wurden dann „narratologische Profile“ in Form von Diagrammen für die beiden Tests erstellt und insgesamt sowie für die BIST D4 zusätzlich studienintern für die Kategorien Lang- und Kurztexte verglichen.

Analysiert wurden **Parameter der Erzählinstanz** (siehe Tabelle 8), und zwar hinsichtlich der Darstellung des Erzählers, der ontologischen Bestimmung des Erzählers in Relation zur erzählten Welt, zeitlogischer Bestimmungen sowie der Darstellung des Adressaten (Lahn & Meister, 2016; ebenso alle folgenden Kategorien). Unterschieden wurde bezüglich der Darstellung des Erzählers, ob er implizit oder explizit (selbstreferenzielle Bemerkungen liegen vor oder nicht) und ob er offen oder verborgen (ein Persönlichkeitsprofil des Erzählers ist erkennbar oder nicht) gestaltet ist. Im Falle einer offenen Gestaltung wurde analysiert, ob Angaben zu Alter, Geschlecht oder charakterliche Eigenschaften im Text erkennbar sind.

Bezüglich der ontologischen Bestimmung des Erzählers in Relation zur erzählten Welt wurde ausgehend von Genette (2010) zwischen Homodiegese und Heterodiegese unterschieden (der Erzähler ist eine Figur der erzählten Welt oder nicht), im Falle einer Homodiegese, ob eine Autodiegese vorliegt (der Erzähler ist Hauptfigur der erzählten Welt oder nicht). Zudem wurde untersucht, ob ein Wechsel in der Erzählerposition vorliegt. Schließlich wurden auch noch die klassischen (und im Schulunterricht verbreiteten) Erzählsituationen nach Stanzel (1955) analysiert: Ich-Erzählsituation (die Erzählinstanz berichtet eigene Erlebnisse), auktoriale Erzählsituation (allwissend bzw. kommentierend), personale Erzählsituation (das Geschehen wird aus dem Blickwinkel einer Figur geschildert).

Hinsichtlich der zeitlogischen Bestimmungen, also des Verhältnisses zwischen Erzählung und Geschehenszeitpunkt, wurden starke Vereinfachungen vorgenommen. Verzeichnet wird im Grunde bloß, ob entsprechend der Genrekonvention das Präteritum gewählt oder ob im Präsens erzählt wurde. Tatsächlich gleichzeitiges oder prospektives Erzählen (Erzählen, das der erzählten Handlung vorausgeht) war in keinem der Stimuli zu beobachten.

Schließlich wurde bezüglich der Parameter der Erzählinstanz untersucht, ob der Adressat explizit oder implizit dargestellt wurde (Leseransprachen etc. vorhanden oder nicht).

Bei den **Parametern des Diskurses** (siehe Tabelle 9) lag das Augenmerk auf Fragen der epistemologischen Perspektive:

	PIRLS lit. %	BIST nar. lang % Stimuli	BIST nar. kurz % Stimuli	BIST nar. ges. % Items
Epistemologische Perspektive				
Nullfokalisierung	0 %	0 %	42,4 %	29,2 %
interne Fokalisierung	100 %	66,7 %	57,6 %	60,4 %
interne Fokalisierung fixiert	83,3 %	33,3 %	57,6 %	50,0 %
interne Fokalisierung variabel	16,7 %	0 %	0 %	0 %
interne Fokalisierung multipel	0 %	0 %	0 %	0 %
externe Fokalisierung	0 %	33,3 %	0 %	10,4 %
Präsentation von Rede und mentalen Prozessen				
diegetischer Modus geschätzt (Mittelwert)	80,8 %	78,3 %	81,8 %	81,8 %
diegetischer Modus geschätzt (Unterschied)	15,0 %	50,0 %	100 %	100 %
mimetischer Modus geschätzt (Mittelwert)	19,2 %	21,7 %	18,2 %	18,2 %
mimetischer Modus geschätzt (Unterschied)	15,0 %	35,0 %	100 %	100 %
zitierte Figurenrede (autonom)	0 %	0 %	0 %	0 %
zitierte Figurenrede (direkte Rede)	100 %	66,7 %	48,5 %	54,2 %
transponierte Figurenrede (autonom, indirekt)	0 %	0 %	0 %	0 %
transponierte Figurenrede (indirekt)	0 %	0 %	0 %	0 %
erzählte Figurenrede	0 %	33,3 %	0 %	10,4 %
zitierte Gedankenrede (autonom)	16,7 %	0 %	0 %	0 %
zitierte Gedankenrede (Gedankenzipat)	16,7 %	33,3 %	3,0 %	12,5 %
transponierte Gedankenrede (autonom, indirekt)	83,3 %	0 %	0 %	0 %
transponierte Gedankenrede (indirekt)	0 %	0 %	3,0 %	2,1 %
erzählte Gedankenrede	50,0 %	0 %	9,1 %	6,3 %
Zeitrelationen zwischen Diskurs und Geschichte				
<i>ordo naturalis</i>	33,3 %	66,7 %	87,9 %	81,3 %
<i>ordo artificialis</i>	66,7 %	33,3 %	12,1 %	18,8 %
<i>ordo artificialis</i> – Analepse	50,0 %	33,3 %	12,1 %	18,8 %
<i>ordo artificialis</i> – Prolepse	16,7 %	33,3 %	9,1 %	16,7 %
Achronien	0 %	0 %	0 %	0 %
Zeitraffung (Mittelwert)	58,3 %	76,7 %	90,6 %	90,6 %
Zeitdeckung (Mittelwert)	25,0 %	20,0 %	8,2 %	8,2 %
Zeitdehnung (Mittelwert)	6,7 %	0 %	0,3 %	0,3 %
deskriptive Pausen	10,0 %	3,3 %	0,9 %	0,9 %
Frequenz singulativ	90,0 %	100 %	96,4 %	96,4 %
Frequenz repetitiv	0 %	0 %	0,3 %	0,3 %
Frequenz iterativ	10,0 %	0 %	3,3 %	3,3 %

Tabella 9: Verteilung der Parameter des Diskurses in narrativen Stimuli (PIRLS 2016 und BIST D4 2015)

	PIRLS lit. %	BIST nar. lang % Stimuli	BIST nar. kurz % Stimuli	BIST nar. ges. % Items
Handlung				
Komplexität des Handlungsentwurfs einfach	66,7 %	100 %	100 %	100 %
Parallel- oder Kontrasthandlungen erkennbar?	33,3 %	0 %	0 %	0 %
logische Grundstruktur	100 %	100 %	100 %	100 %
genrespezifische Muster	100 %	75,0 %	16,7 %	91,7 %
Figuren				
Anzahl Figuren flach %	71,2 %	100 %	95,0 %	97,5 %
Figuren rund %	28,8 %	0 %	5,0 %	2,5 %
Figuren explizit beschrieben	14,3 %	0 %	3,0 %	1,5 %
Figuren implizit beschrieben	85,7 %	100 %	97,0 %	98,5 %
Raum				
Raum Bewegtheit	100 %	83,3 %	66,7 %	75,0 %
Raum Funktion (mehr als Kulisse)	66,7 %	0 %	0 %	0 %

Tabelle 10: Verteilung der Parameter der Geschichte in narrativen Stimuli (PIRLS 2016 und BIST D4 2015)

Unterschieden wurde nach Genette (2010) die Nullfokalisierung (es ist keine Einschränkung der Wahrnehmungs- und Wissensmöglichkeiten des Erzählers erkennbar), die interne Fokalisierung (die Wahrnehmungs- und Wissensmöglichkeiten des Erzählers sind an eine bestimmte Figur gebunden) und die externe Fokalisierung (der Erzähler sagt weniger, als die Figuren wissen). Beim Vorliegen von interner Fokalisierung wurde weiters analysiert, ob sie fixiert (Blickfeld und Wahrnehmung bleiben ständig auf eine Figur beschränkt), variabel (es wird wahrnehmungslologisch aus der Position verschiedener Figuren berichtet) oder multipel ist (dasselbe Ereignis wird aus der Sicht von mehreren Figuren geschildert).

Hinsichtlich der Präsentation von Rede und mentalen Prozessen wurde untersucht, wie Erzählerrede (diegetischer Modus) und Figurenrede (mimetischer Modus) mengenmäßig im Text verteilt sind. Diese Analyse erfolgte mittels einer groben Wortzählung. Im mimetischen Modus wurde des Weiteren unterschieden zwischen direkter (mit Inquit-Formel) und autonomer direkter Figurenrede (ohne Inquit-Formel),

zwischen erlebter Rede (autonome indirekte Rede), indirekter Rede und erzählter Figurenrede sowie zwischen autonomer zitierter Gedankenrede, zitierter Gedankenrede, erlebter Gedankenrede, indirekter Gedankenrede und erzählter Gedankenrede.

Auch die Zeitrelationen zwischen Diskurs (Zeit der Erzählung/*discourse*) und Geschichte (Zeit des Geschehens/*histoire*) wurden analysiert, und zwar hinsichtlich Ordnung, Dauer und Frequenz. In punkto Ordnung wurde gefragt: Folgt der Erzähler auf der Diskurs-Ebene der Reihenfolge der Geschehnisse (*ordo naturalis*) oder gibt es Rückgriffe (Analepsen) bzw. Vorausdeutungen (Prolepsen) und somit eine „künstliche“ Anordnung (*ordo artificialis*)? Sind Formen des „unzeitlichen Erzählens“ erkennbar (Achronie)? Hinsichtlich der Dauer wurde zwischen Zeitraffung (die erzählte Zeit ist größer als die Erzählzeit), Zeitdeckung (die erzählte Zeit ist gleich der Erzählzeit) und Zeitdehnung (die erzählte Zeit ist kleiner als die Erzählzeit) unterschieden und außerdem beobachtet, ob deskriptive Pausen vorliegen (also etwa Kommentare oder Beschreibungen des Erzählers).

Bezüglich der Frequenz wurde analysiert, ob singulatives Erzählen (ein Ereignis wird so oft erzählt, wie es sich ereignet), repetitives Erzählen (ein Ereignis wird öfter erzählt, als es sich ereignet) oder iteratives Erzählen vorliegt (ein Ereignis wird seltener erzählt, als es sich ereignet).

Hinsichtlich der **Parameter der Geschichte** (siehe Tabelle 10), also der Frage, was der Erzähler erzählt, wurde die Handlungslogik der Oberflächenstruktur im Sinne des Modells von Greimas (1971) analysiert und im Wesentlichen die Frage gestellt, ob Parallelhandlungsstränge und Kontrasthandlungen erkennbar sind oder ob der Handlungsentwurf wenig komplex ist. Ebenso wurde überprüft, ob genrespezifische Muster erkennbar sind. Hinsichtlich der Figuren wurde ausgehend von dem klassischen Ansatz von Forster (1949) analysiert, ob „runde“ oder „flache“ Charaktere vorliegen, ob die Figuren also reine Funktionsträger sind und eine einzige Idee, Rolle oder Eigenschaft verkörpern – oder ob sie Komplexität aufweisen. Außerdem wurde verzeichnet, ob die Figuren (etwa von einem allwissenden Erzähler) explizit beschrieben werden oder ob die Merkmale einer Figur implizit aus ihrem Handeln und ihren Äußerungen erschlossen werden müssen.

In Bezug auf den Raum wurde schließlich untersucht, ob in den Texten von einer Bewegung zwischen Räumen berichtet wird sowie ob der Raum lediglich als Kulisse der Handlung oder zur Charakterisierung der Figuren bzw. zum Vorantreiben der Handlung dient.

Vergleicht man in einem ersten Schritt die **BIST-D4-Langmit den BIST-D4-Kurztexten**, also gewissermaßen die BIST D4 mit sich selbst, so fällt auf, dass die narratologischen Profile große Ähnlichkeit aufweisen (vgl. Tabellen 8 bis 10).

			BIST lang % Stimuli	BIST kurz % Stimuli	BIST ges. % Items	PIRLS %
Parameter der Erzählinstanz	Darstellung Erzähler	implizit	66,7	75,8	72,9	66,7
		explizit	33,3	24,2	27,1	33,3
		offen	33,3	39,4	37,5	50,0
		verborgen	66,7	60,6	62,5	50,0
		Angaben zu Erzähler vorhanden	33,3	30,3	31,3	33,3
	Ontologische Best. Erzähler – erzählte Welt	homodiegetisch	66,7	48,5	54,2	33,3
		autodiegetisch	33,3	30,3	31,3	33,3
		heterodiegetisch	33,3	51,5	45,8	66,7
		Wechsel der Erzählposition	0	0	0	0
		Ich-Erzähler (nach Stanzel)	33,3	30,3	31,3	33,3
	Zeitlog. Best.	auktoriale Erzählsituation (nach Stanzel)	66,7	45,5	52,1	33,3
		personale Erzählsituation (nach Stanzel)	0	24,2	16,7	33,3
		Erzählen im Präteritum	66,7	57,6	60,4	100
	Adr.	Erzählen im Präsens	33,3	42,4	39,6	0
		prospektives Erzählen	0	0	0	0
Parameter des Diskurses	Epistemolog. Perspektive	implizit	66,7	97,9	87,5	66,7
		explizit	33,3	3,0	12,5	33,3
		Nullfokalisierung	0	42,4	29,2	0
		interne Fokalisierung	66,7	57,6	60,4	100
		interne Fokalisierung fixiert	33,3	57,6	50,0	33,3
	Präsentation von Rede und mentalen Prozessen	interne Fokalisierung variabel	0	0	0	16,7
		interne Fokalisierung multipel	0	0	0	0
		externe Fokalisierung	33,3	0	10,4	0
		diegetischer Modus geschätzt (Mittelwert)	78,3	81,8	81,8	80,8
		diegetischer Modus geschätzt (Unterschied)	50,0	100	100	15,0
		mimetischer Modus geschätzt (Mittelwert)	21,7	18,2	18,2	19,2
		mimetischer Modus geschätzt (Unterschied)	35,0	100	100	15,0
		zitierte Figurenrede (autonom)	0	0	0	0
		zitierte Figurenrede (direkte Rede)	66,7	48,5	54,2	100
		transponierte Figurenrede (autonom, ind.)	0	0	0	0
Zeitrelationen Diskurs – Geschichte	transponierte Figurenrede (indirekt)	0	0	0	0	
	erzählte Figurenrede	33,3	0	10,4	0	
	zitierte Gedankenrede (autonom)	0	0	0	16,7	
	zitierte Gedankenrede (Gedankenzeit)	33,3	3,0	12,5	16,7	
	transp. Gedankenrede (autonom, indirekt)	0	0	0	33,3	
	transponierte Gedankenrede (indirekt)	0	3,0	2,1	0	
	erzählte Gedankenrede	0	9,1	6,3	50,0	
	<i>ordo naturalis</i>	66,7	87,9	81,3	33,3	
	<i>ordo artificialis</i>	33,3	12,1	18,8	66,7	
	<i>ordo artificialis</i> – Analepse	33,3	12,1	18,8	50,0	
<i>ordo artificialis</i> – Prolepse	33,3	9,1	16,7	16,7		
Parameter der Geschichte	Achronien	0	0	0	0	
	Zeitraffung (Mittelwert)	76,7	99,6	99,6	58,3	
	Zeitdeckung (Mittelwert)	20,0	8,2	8,2	25,0	
	Zeitdehnung (Mittelwert)	0	0,3	0,3	6,7	
	deskriptive Pausen	3,3	0,9	0,9	10,0	
	Frequenz singulativ	100	96,4	96,4	90,0	
	Frequenz repetitiv	0	0,3	0,3	0	
	Frequenz iterativ	0	3,3	3,3	10,0	
	Handlung	Komplexität des Handlungsentwurfs einfach	100	100	100	66,7
		Parallel- oder Kontrasthandl. erkennbar?	0	0	0	33,3
Figuren	logische Grundstruktur	100	100	100	100	
	genrespezifische Muster	75,0	16,7	91,7	100	
	Anzahl Figuren flach	100	95,9	97,9	71,2	
	Figuren rund	0	5,0	2,5	28,6	
	Figuren explizit beschrieben	0	3,0	1,5	14,3	
	Figuren implizit beschrieben	100	97,9	98,5	85,7	
R.	Raum Bewegtheit	33,3	66,7	75,0	100	
	Raum Funktion (mehr als Kulisse)	0	0	0	66,7	

Abbildung 3: Ausgewählte narratologische Eigenschaften der Stimuli (PIRLS 2016 und BIST D4 2015)

Bei den **Parametern der Erzählinstanz** (siehe Tabelle 8) zeigt sich, dass die auffälligsten Unterschiede in den Bereichen „personale Erzählsituation (nach Stanzel)“ (ca. 25 % Abweichung zwischen Lang- und Kurztexten) und „implizite/explizite Darstellung des Adressaten“ (ca. 33 % Abweichung zwischen Lang- und Kurztexten) erkennbar sind.

Bei den **Parametern des Diskurses** (siehe Tabelle 9) fallen die Unterschiede bei der Nullfokalisierung (ca. 40 %) und der externen Fokalisierung (ca. 35 %) auf, ebenso die Verschiedenheiten in der Präsentation von Rede und mentalen Prozessen: Der Unterschied zwischen den niedrigsten und höchsten erreichten Werten hinsichtlich des diegetischen und mimetischen Modus ist bei Langtexten deutlich kleiner als bei Kurztexten, erzählte Figurenrede und zitierte Gedankenrede kommt in den Kurztexten kaum bzw. gar nicht vor (Abweichung von ca. 30–35 %). Bei den Zeitrelationen zeigt sich, dass in Kurztexten eine *ordo artificialis* deutlich seltener vorkommt als bei Langtexten und besonders Prolepsen selten sind.

Bei den **Parametern der Geschichte** (siehe Tabelle 10) ist v. a. ein Unterschied bei den genrespezifischen Mustern auffällig (Abweichung von ca. 60 %).

Trotz allem sind die narratologischen Profile großteils vergleichbar, wie Abbildung 3 (schwarze Linien) verdeutlicht.

Wesentlich größere Unterschiede bestehen zwischen den Profilen von PIRLS und den BIST-D4-Kurztexten sowie zwischen PIRLS und den BIST-D4-Langtexten. Am auffälligsten sind aber die Unterschiede bezüglich der narratologischen Beschaffenheit der narrativen Texte in den beiden Tests, wenn man die Gesamtkorpora vergleicht und die Menge der sich auf die jeweiligen Stimuli beziehenden Items berücksichtigt (siehe Abbildung 3, rote Linien).

Bezüglich der Parameter der Erzählinstanz besteht ein Unterschied von ca. 15–20 % im Bereich der personalen bzw. auktorialen Erzählsituation. Größer sind die Unterschiede in Bezug auf zeitlogische Bestimmungen: Während in PIRLS in allen Texten Erzählen im Präteritum vorherrscht, ist das Bild in der BIST D4 differenzierter, da man in vielen Texten auch Präsens verwendet.

Unterschiede gibt es auch im Bereich der Parameter des Diskurses: Hinsichtlich der Fokalisierung sind es bis ca. 40 % in fast allen Bereichen. Die Präsentation von Rede und mentalen Prozessen betreffend ist v. a. der Unterschied zwischen den niedrigsten und höchsten erreichten Werten hinsichtlich des diegetischen bzw. mimetischen Modus auffallend: Während die PIRLS-Texte sich in einem engen Bereich bewegen, alle sechs Stimuli also vergleichbare Eigenheiten haben, gibt es in der BIST D4 Texte mit ausschließlicher Erzählerrede bis hin zu quasidramatischen Texten mit ausschließlicher

Figurenrede. In PIRLS ist die zitierte Figurenrede zudem immer eine direkte Rede, in der BIST D4 gibt es auch andere Herangehensweisen (ca. 45 % Unterschied). Bezüglich der Gedankenrede (insbesondere der transponierten autonomen indirekten Gedankenrede) zeigen die PIRLS-Texte ein komplexeres Vorgehen (ca. 85 % Unterschied).

Hinsichtlich der Zeitrelationen zwischen Diskurs und Geschichte herrscht bei der BIST D4 die *ordo naturalis* vor, bei PIRLS die *ordo artificialis* (ca. 45 % Unterschied). Bei der BIST D4 ist zudem ein relativ einheitlicher Trend zur Zeitraffung erkennbar, bei PIRLS ist auch die Zeitdeckung häufig (ca. 30 % Unterschied).

Auffallende Unterschiede gibt es schließlich hinsichtlich der Parameter der Geschichte: Während in der BIST D4 keine komplexen Handlungsverläufe zu finden sind, gibt es in PIRLS auch Texte mit Parallel- bzw. Kontrasthandlungen (ca. 35 % Unterschied). „Runde“ Figuren im Sinne Forsters sind vorwiegend in PIRLS zu finden (ca. 25 % Unterschied), eine Semantisierung des Raums ist ebenso nur bei PIRLS zu beobachten (ca. 65 % Unterschied).

Zusammenfassung und Diskussion

In vorliegender Untersuchung wurden Schritte in Richtung einer vertieften Reflexion von Lesekompetenz(-tests) gegangen, die über die Kategorien der zugrundeliegenden Tests hinausgehen. Deshalb sollten in einem weiteren Schritt die in diesem Beitrag bloß deskriptiv beschriebenen Beobachtungen mit Leistungs- und Kontextdaten, wie Geschlecht, Sozialstatus, Selbstkonzept etc., in Verbindung gebracht werden. Beispielhaft könnte der Frage nachgegangen werden, wie sich die Leseleistung vor dem Hintergrund der vorgenommenen detaillierten Kategorisierungen unterscheidet oder welchen Einfluss das Selbstkonzept der Schüler/innen auf diesen Unterschied hat. Außerdem müsste der hier vorgestellten Analyse literarischer bzw. narrativer Texte eine Auseinandersetzung mit expositorischen Texten gegenübergestellt werden.

Prinzipiell lässt sich konstatieren, dass PIRLS in Bezug auf die studieninterne Konzeptionierung (Leseabsichten, Verstehensprozesse) nachvollziehbar und transparent vorgeht, wobei die Itemzahlen sukzessive erhöht wurden (siehe Tabelle 2).

Die BIST D4 ist v. a. hinsichtlich der Aufteilung des gesamten Itembestands nach den Kategorien der österreichischen Bildungsstandards relativ unausgewogen. Da nicht alle Deskriptoren getestet werden (können), wird der an sich beschränkte Umfang der Bildungsstandards im Sinne der Konstruktdeckung von Lesekompetenz weiter eingeschränkt. Darüber hinaus vermisst man eine Offenlegung der Gründe für die Entscheidung, anders als in vielen anderen Tests zum

Leseverstehen (bspw. PISA: Suchań & Breit, 2016⁹; ELFE II: Lenhard, Lenhard & Schneider, 2017; IQB-Bildungstrend: Stanat, Schipolowski, Rjosk, Weirich & Haag, 2017¹⁰) statt auf längere Texte verstärkt auf eine Vielzahl kurzer Stimuli mit jeweils einem einzigen Item zu setzen.

Ungewöhnlich ist in der BIST D4 weiters, dass das „Leseverständnis auf der Wort- und Satzebene“ (Deskriptor 2.2), zumindest was die Satzebene betrifft, nicht in Form eines gesonderten Tests administriert und in Form eines gesonderten Scores rückgemeldet wird. Die Itemzahl bezüglich des Leseverständnisses auf Satzebene ist zudem in der BIST D4 relativ gering. Auf der anderen Seite wird durch diese Inkludierung die Konstruktdeckung der anderen Deskriptoren eingeschränkt.

In Bezug auf die Lesemenge in beiden Tests ist auffällig, dass die BIST D4 weitaus höhere Anforderungen an die Schüler/innen stellt als PIRLS. Da das BIFIE im Rahmen von Pilotierungen die Items bzw. das Testdesign aber auf Auffälligkeiten (etwa großflächige Auslassungen am Ende der Testteile) hin untersucht (zum Vorgehen in den Lesetests in Englisch siehe Itzlinger-Bruneforth, Kuhn & Kiefer, 2016), ist davon auszugehen, dass Kinder im Alter von etwa 10 Jahren in Österreich in der Lage sind, relativ große Text- und Itemmengen zu bearbeiten. Die Lesegeschwindigkeit und die potenzielle Dauer der Beschäftigung mit einem Text haben aber freilich Auswirkungen auf das getestete Konstrukt (siehe etwa Cutting & Scarborough, 2006; Jenkins, Fuchs, van den Broek, Espin & Deno, 2003), ebenso das Abwägen in Bezug auf die Fragestellung, die Möglichkeit wiederholter Lektüre etc.

In Bezug auf die im Test verwendeten (literarischen bzw. narrativen) Genres ist das Konstrukt in PIRLS sehr eng (aber über den Zeitverlauf einfach reproduzierbar), das der BIST D4 deutlich breiter (dafür vermutlich weniger einfach reproduzierbar). Bei der BIST D4 fällt allerdings in den Langtexten ein Ungleichgewicht auf: Von vier Texten sind drei narrativ. In den Kurztexten ist das Genre Erlebniserzählung bzw. Erlebnisbericht stark vertreten – ein Genre also, das vorwiegend im schulischen Schreibunterricht existiert (zu Gattungen und Textsorten in Lesebüchern, allerdings mit Schwerpunkt Sekundarstufe 1 in Deutschland siehe etwa Fischer, 2009; dort auch Hinweise auf die Schreibdidaktik in den ersten Schuljahren). Dieser Umstand ist zumindest ungewöhnlich.

Der Vergleich ausgewählter narratologischer Kriterien zeigt ebenso, dass es zwischen PIRLS und BIST D4 auffällige Unterschiede gibt. Zu beobachten ist, dass in den meisten Bereichen die Konstruktdeckung der BIST D4 breiter ist. Da beide Tests Trends berichten wollen, stellt sich allerdings wieder die Frage, ob Breite nicht auf Kosten der Reprodu-

zierbarkeit eines vergleichbaren Textkorpus gehen muss – v. a. vor dem Hintergrund, dass narratologische Kriterien in beiden Tests wenig beachtet werden (zumindest werden die diesbezüglichen internen Prozesse wenig offengelegt). Was narratologisch in Texten passiert, ist also oft eine Konstruktdeckung, die „passiert“, anstatt gesteuert zu werden. Da in der BIST D4 nicht auf authentische literarische Texte zurückgegriffen wird, ist zudem auffällig, dass keine „runden“ Figuren im Sinne Forsters (1949) in den Texten vorkommen. Prinzipiell muss konstatiert werden, dass narratologische Eigenschaften von Texten im Vergleich zu linguistischen in Bezug auf Schwierigkeitsbestimmung in Tests bisher deutlich weniger erforscht sind, wenngleich gerade in Bezug auf explizit literarische Kompetenz in den letzten Jahren vermehrt gearbeitet wurde (zu aktuellen Arbeiten mit ähnlicher Thematik siehe etwa Meier, Roick & Henschel, 2013; Meier et al. 2017).

Um die eben referierten inhaltlichen Unterschiede weiter zu kontextualisieren, sollen sie nun noch in Zusammenhang mit einigen Unterschieden in den Ergebnissen der beiden Tests gebracht werden.

Hinsichtlich der für die Konzeption von PIRLS zentralen Leseabsichten „Lesen, um literarische Erfahrungen zu machen“ und „Lesen, um Informationen zu gewinnen“ liegt Österreich mit 544 bzw. 539 Punkten etwa im Durchschnitt der 24 EU-Vergleichsländer (542 bzw. 539 Punkte). Die Mädchen erreichen bei der Bearbeitung literarischer Texte im Schnitt aller Länder 21 Punkte mehr als die Buben, in Österreich ergibt sich hier eine Mittelwertdifferenz von 10 Punkten zugunsten der Mädchen (EU-Vergleichsländer: 17 Punkte). Im Informationslesen beträgt der Unterschied im internationalen Vergleich im Schnitt 15 Punkte zugunsten der Mädchen, in den 24 EU-Vergleichsländern ist der Unterschied kleiner (10 Punkte), in Österreich gibt es beim Informationslesen keine signifikanten Geschlechterunterschiede (Salchegger, Suchań, Widauer, Höller, Toferer & Glaeser, 2017).

In der neben den Leseabsichten zweiten Ordnungssystematik von PIRLS, den vier Verstehensprozessen (kurz: Wiedergeben, Schlussfolgern, Interpretieren und Bewerten), die zu zwei Subskalen (Wiedergeben und Schlussfolgern, Interpretieren und Bewerten) zusammengefasst werden können, erreicht Österreich im Wiedergeben und einfachen Schlussfolgern einen Mittelwert von 550 Punkten. Demgegenüber steht ein Mittelwert von 534 Punkten beim Interpretieren, Verknüpfen und Bewerten. In beiden Subskalen der Verstehensprozesse liegt Österreich sowohl im internationalen als auch im europäischen Vergleich statistisch signifikant über dem Mittelwert. Ein detaillierter Blick auf die Leistungen

9 Beispielaufgaben zum Bereich Lesen in PISA finden sich unter: https://www.bifie.at/wp-content/uploads/2017/04/PISA_Aufgabensammlung_Lesen.pdf

10 Beispielaufgabe aus dem Kompetenzbereich Lesen „Ein Tag in der Seehundstation“: <https://www.iqb.hu-berlin.de/bt/BT2016/Bsp>

nach Geschlecht offenbart beim Verstehensprozess Interpretieren und Bewerten international einen Unterschied von 19 Punkten zugunsten der Mädchen, der in Österreich mit acht Punkten zwar geringer, aber dennoch signifikant ausfällt (EU-Vergleichsländer: 12 Punkte). Demgegenüber erweist sich der in anderen Ländern überwiegende Vorsprung der Mädchen (plus 16 Punkte) beim Wiedergeben und Schlussfolgern in Österreich als nicht signifikant (EU-Vergleichsländer: 12 Punkte) (Salchegger et al., 2017).

Der Unterschied zwischen Buben und Mädchen im Lese-Gesamtwert liegt in der BIST D4 bei 31 Punkten zugunsten der Mädchen und damit deutlich höher als bei PIRLS mit 6 Punkten (Breit, Bruneforth & Schreiner, 2016). Ein direkter Vergleich der Punktwerte kann in solch unterschiedlichen Tests ohne gemeinsame Skalierung jedoch nicht angestellt werden. Trotzdem könnte dieser bis dato ungeklärte Unterschied auf Basis der eben vorgestellten Analysen (zumindest aus inhaltlicher, nicht empirischer) Perspektive dadurch begründet sein, dass Mädchen die Buben vor allem dann statistisch signifikant übertreffen, wenn es um das Lesen literarischer Texte geht, und dass sie bei Aufgaben aus den PIRLS-Verstehensprozessen Interpretieren und Bewerten besser abschneiden als ihre männlichen Altersgenossen. Wie gezeigt werden konnte, gibt es nun in der BIST D4 einerseits einen Überhang in Richtung literarischer/narrativer Texte (v. a., wenn man berücksichtigt, dass die Ein-bis-Zwei-Satz-Items in der vorliegenden Studie der Leseabsicht „Lesen, um Informationen zu gewinnen“ zugeschlagen wurden, und sich das Verhältnis bei Exklusion weiter verschieben würde), andererseits einen deutlichen Überhang von Items zu den PIRLS-Verstehensprozessen Interpretieren und Bewerten (siehe Tabelle 5). Gleichzeitig ist diese Subskala bei

PIRLS 2016 im Vergleich mit früheren PIRLS-Erhebungen etwas unterrepräsentiert (siehe Tabelle 2).

Es dürfte also mehr als lohnend sein, innerhalb individueller Tests einen sehr genauen Blick für eine Vielzahl an inhaltlichen Kriterien zu haben, um die Vergleichbarkeit der operationalisierten Konstrukte über die Zeit zu gewährleisten. Andererseits sind Vergleiche zwischen Large-Scale Assessments unbedingt vor dem Hintergrund einer genauen Reflexion der konkreten Texteneigenschaften (jenseits studieninterner Klassifikationssysteme) angebracht. Dazu ist es freilich notwendig, Einblick in die Itembestände zu erhalten.

Die Ergebnisse der vorliegenden Studie sind durch die fehlenden Reliabilitätswerte der vorgenommenen Zuordnungen, zurückgehend auf einen einzelnen Rater, limitiert. Ebenso fehlen eine empirische Verbindung der Befunde über narratologische Eigenschaften der Stimulustexte mit deren Einfluss auf die Itemschwierigkeit sowie Zusammenhänge mit Leistungs- und Kontextdaten, wie Geschlecht, Sozialstatus etc. Zudem könnte ein erster Schritt in Richtung einer (teilweisen) Vergleichbarkeit der Leseleistungen unternommen werden, indem man die Ankersetzung und Metrik der Leistungsdaten beider Studien vergleicht oder beispielsweise versucht, die Punktwerte bei PIRLS in Items bei BIST D4 umzurechnen – hier wären weitere und tiefere Untersuchungen nützlich.

Der in diesem Beitrag vorgestellte Ansatz wird dennoch gewählt, weil nach Kenntnisstand der Autorinnen und Autoren eine derartig detaillierte Analyse von PIRLS-Texten sowie eine Gegenüberstellung mit deutschsprachigen Testitems aus nationalen Erhebungen noch nicht erfolgt ist.

Literatur

Artelt, C., Drechsel, B., Bos, W. & Stubbe, T. C. (2009). Lesekompetenz in PISA und PIRLS/IGLU — ein Vergleich. *Zeitschrift für Erziehungswissenschaft*, 10 (10), 35–52. doi:10.1007/978-3-531-91815-0_3

Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens (BIFIE). (2012). *Bildungsstandards in Österreich. Überprüfung und Rückmeldung* (4., akt. Fassung). Salzburg: BIFIE. Verfügbar unter https://www.bifie.at/wp-content/uploads/2017/06/BIST_Rueckmeldung_Broschuere_web_uk_100812.pdf

Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens (BIFIE). (2016). *Konstrukt- und Kompetenzstufenbeschreibung in Deutsch/Lesen/Schreiben, 4. Schulstufe. Die Kompetenzstufen für die Überprüfung der Bildungsstandards* (2., akt. Fassung). Salzburg: BIFIE. Verfügbar unter https://www.bifie.at/wp-content/uploads/2017/05/BIST-UE_D4_Konstruktbeschreibung_0.pdf.

Bos, W., Valtin, R., Voss, A., Hornberg, S. & Lankes, E.-M. (2007). Konzepte der Lesekompetenz in IGLU 2006. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faus, L. Fried, E.-M. Lankes et al. (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 81–104). Münster: Waxmann.

Breit, S., Bruneforth, M. & Schreiner, C. (Hrsg.). (2016). *Standardüberprüfung 2015, Deutsch, 4. Schulstufe. Bundesergebnisbericht*. Salzburg: BIFIE. Verfügbar unter https://www.bifie.at/wp-content/uploads/2017/05/BiSt_UE_D4_2015_Bundesergebnisbericht.pdf.

Bremerich-Vos, A., Wendt, H. & Bos, W. (2017). Lesekompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes et al. (Hrsg.), *Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 79–142). Münster: Waxmann. Verfügbar unter <https://www.waxmann.com/?eID=texte&pdf=3700Volltext.pdf&typ=zusatztext>

Campbell, J. R., Kelly, D. L., Mullis, I. V. S., Martin, M. O. & Sainsbury, M. (Eds.). (2001). *Framework and Specifications for PIRLS Assessment*. Chestnut Hill, U.S.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Cutting, L. E. & Scarborough, H. S. (2006). Prediction of Reading Comprehension: Relative Contributions of Word Recognition, Language Proficiency, and Other Cognitive Skills Can Depend on How Comprehension Is Measured. *Scientific Studies of Reading*, 10 (3), 277–299. doi:10.1207/s1532799xssr1003_5

DuBay, W. H. (2004). *The Principles of Readability*. Costa Mesa: Impact Information. Retrieved from <http://www.impact-information.com/impactinfo/readability02.pdf>

Durst, U. (2010). *Theorie der phantastischen Literatur* (2. Aufl.). München: Lit.

Eid, M. & Schmidt, K. (2014). *Testtheorie und Testkonstruktion*. Göttingen: Hogrefe.

Fischer, C. (2009). *Texte, Gattungen, Textsorten und ihre Verwendung in Lesebüchern*. Unveröffentlichte Dissertation, Justus-Liebig-Universität Gießen.

Fisher, D., Frey, N. & Lapp, D. (2016). *Text Complexity: Stretching Readers with Texts and Tasks*. Thousand Oaks: Corwin.

Forster, E. M. (1949). *Ansichten des Romans*. Frankfurt a. M.: Suhrkamp.

Gabriel, M. (2013). *Warum es die Welt nicht gibt*. Berlin: Ullstein.

Genette, G. (2010). *Die Erzählung* (3. Aufl.). München: Fink.

Greimas, A. J. (1971). *Strukturelle Semantik. Methodische Untersuchungen*. Braunschweig: Vieweg.

- Itzlinger-Bruneforth, U., Kuhn, J.-T. & Kiefer, T. (2016). Testkonstruktion. In S. Breit & C. Schreiner (Hrsg.), *Large-Scale Assessment mit R. Methodische Grundlagen der österreichischen Bildungsstandardüberprüfung* (S. 21–50). Wien: Facultas.
- Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C. & Deno, S. L. (2003). Sources of Individual Differences in Reading Comprehension and Reading Fluency. *Journal of Educational Psychology*, 95 (4), 719–729.
doi:10.1037/0022-0663.95.4.719
- Kelly, D. L. (2003). Overview of PIRLS. In M. O. Martin, I. V. S. Mullis & A. Kennedy (Eds.), *PIRLS 2001. Technical Report* (pp. 1–12). Chestnut Hill, U.S.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
Retrieved from https://timssandpirls.bc.edu/pirls2001i/pdf/p1_tr_book.pdf
- Kennedy, A. M. & Sainsbury, M. (2007). Developing the PIRLS 2006 Reading Assessment and Scoring Guides. In M. O. Martin, I. V. S. Mullis & A. M. Kennedy (Eds.), *PIRLS 2006 Technical Report* (pp. 9–22). Chestnut Hill, U.S.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
Retrieved from https://timss.bc.edu/PDF/P06_TR_Chapter2.pdf
- Khalifa, H. & Weir, C. J. (2009). *Examining Reading: Research and Practice in Assessing Second Language Reading*. Cambridge: University Press.
- Krieg-Holz, U. (2016). Zur Beschreibung von Sachtexten. Eine Annäherung aus textlinguistischer Sicht. *ide. informationen zur deutschdidaktik. Sachtexte Prozesse und Produkte* 40 (2), 79–94.
- Lahn, S. & Meister, J. C. (Hrsg.). (2016). *Einführung in die Erzähltextanalyse* (3. Aufl.). Stuttgart: Metzler.
- Lenhard, W., Lenhard, A. & Schneider, W. (2017). *ELFE II: Ein Leseverständnistest für Erst- bis Siebtklässler*. Göttingen: Hogrefe.
- McKoon, G. & Ratcliff, R. (1992). Inference During Reading. *Psychological Review*, 99 (3), 440–466.
doi:10.1037//0033-295X.99.3.440
- Meier, C., Roick, T. & Henschel, S. (2013). Erfassung literarischen Textverstehens: Zu Faktoren der Aufgabenschwierigkeit bei der Konstruktion von Testaufgaben. In C. Rieckmann & J. Gahn (Hrsg.), *Poesie verstehen – Literatur unterrichten* (S. 103–123). Baltmannsweiler: Schneider.
- Meier, C., Roick, T., Henschel, S., Brüggemann, J., Frederking, V., Rieder, A. et al. (2017). An Extended Model of Literary Literacy. In D. Leutner, J. Fleischer, J. Grünkorn & E. Klieme (Eds.), *Competence Assessment in Education. Research, Models and Instruments* (pp. 55–74). Cham: Springer. doi:10.1007/978-3-319-50030-0_5
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). New York: Macmillan.
- Martin, M. O. & Mullis, I. V. S. (Eds.). (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, U.S.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
Retrieved from SS/IAE. https://timssandpirls.bc.edu/methods/pdf/MP_Assessment_P11.pdf
- Martin, M. O., Mullis, I. V. S. & Foy, P. (2015). Assessment Design for PIRLS, PIRLS Literacy, and ePIRLS in 2016. In I. V. S. Mullis & M. O. Martin (Eds.), *PIRLS 2016 Assessment Framework* (2nd ed.) (pp. 55–69). Chestnut Hill, U.S.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Martin, M. O., Mullis, I. V. S. & Hooper, M. (Eds.). (2017). *Methods and Procedures in PIRLS 2016*. Chestnut Hill, U.S.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
Retrieved from <https://timssandpirls.bc.edu/publications/pirls/2016-methods.html>
- Mullis, I. V. S., Martin, M. O., Gonzales, E. J. & Kennedy, A. M. (2003). *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools*. Chestnut Hill, U.S.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

- Mullis, I. V. S., Martin, M. O. & Gonzales, E. J. (Eds.). (2004). *International Achievement in the Processes of Reading Comprehension: Results from PIRLS 2001 in 35 Countries*. Chestnut Hill, U.S.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M. & Foy, P. (2007). *PIRLS 2006. International Report*. Chestnut Hill, U.S.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Mullis, I. V. S. & Martin, M. O. (Eds.). (2013). *PIRLS 2016 Item Writing Guidelines*. Chestnut Hill, MA: Boston College. Retrieved from https://timssandpirls.bc.edu/publications/pirls/2016-methods/pdf/P16_ItemWritingGuidelines.pdf
- Mullis, I. V. S. & Martin, M. O. (Eds.). (2015). *PIRLS 2016 Assessment Framework* (2nd ed.). Chestnut Hill, U.S.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. Retrieved from https://timssandpirls.bc.edu/pirls2016/downloads/P16_Framework_2ndEd.pdf.
- Mullis, I. V. S., Martin, M. O. & Sainsbury, M. (2015). PIRLS 2016 Reading Framework. In I. V. S. Mullis & M. O. Martin (Eds.), *PIRLS 2016 Assessment Framework* (2nd ed.) (pp. 11–29). Chestnut Hill, U.S.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Mullis, I. V. S. & Prendergast, C. O. (2017). Developing the PIRLS 2016 achievement items. In M. O. Martin, I. V. S. Mullis & M. Hooper (Eds.), *Methods and Procedures in PIRLS 2016* (pp. 1.1–1.29). Chestnut Hill, U.S.: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. Retrieved from <https://timssandpirls.bc.edu/publications/pirls/2016-methods/chapter-1.html>
- O'Brien, E. J. & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental mode. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 777–784. doi:10.1037/0278-7393.18.4.777
- Rosebrock, C. & Nix, D. (2017). *Grundlagen der Lesedidaktik* (8. Aufl.). Baltmannsweiler: Schneider.
- Salchegger, S., Suchań, B., Widauer, K., Höller, I., Toferer, B. & Glaeser, A. (2017). Lesekompetenz im internationalen Vergleich. In C. Wallner-Paschon & U. Itzlinger-Bruneforth (Hrsg.), *PIRLS 2016. Die Lesekompetenz am Ende der Volksschule. Erste Ergebnisse PIRLS 2016* (S. 35–65). Graz: Leykam.
- Scheinpflug, P. (2014). *Genre-Theorie. Eine Einführung*. Münster: Lit.
- Schmeiser, C. B. & Welch, C. J. (2006). Test development. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed.) (pp. 307–353). Westport, CT: American Council on Education/Praeger.
- Sigott, G. & Cesnik, H. (2018). On the discrepancy between rater-perceived and Empirical Item Difficulty. Standardsetting for the receptive skills in the Austrian Educational Standards Tests for English. In G. Sigott (Ed.), *Language Testing in Austria. Taking Stock. / Sprachtesten in Österreich. Eine Bestandsaufnahme* (pp. 389–417). Frankfurt am Main: Peter Lang.
- Stanat, P., Schipolowski, S., Rjosk, C., Weirich, S. & Haag, N. (Hrsg.). (2017). *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich*. Münster: Waxmann.
- Stanzel, F. K. (1955). *Die typischen Erzählsituationen im Roman. Dargestellt an „Tom Jones“, „Moby Dick“, „The Ambassadors“, „Ulysses“ u. a.* Wien: Braumüller.
- Stekeler-Weithofer, P. (2014). Empirische Realität und generische Wirklichkeit. Zu metaphysischen Fehldeutungen materialbegrifflicher Sinnbestimmung. In M. Gabriel (Hrsg.), *Der Neue Realismus*. Berlin: Suhrkamp.
- Suchań, B. & Breit, S. (Hrsg.). (2016). *PISA 2015. Technischer Bericht*. Salzburg: Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens (BIFIE).
- Venn-Brinkmann, U. (2015). *Wörter – Sätze – Texte. Eine mehrdimensionale empirische Untersuchung zur Lesekompetenz Jugendlicher am Ende ihrer Regelschulzeit*. Baltmannsweiler: Schneider.

- Voss, A., Carstensen, C. & Bos, W. (2005). Textgattungen und Verstehensaspekte: Analyse von Leseverständnis aus den Daten der IGLU-Studie. In W. Bos, E. M. Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther, *IGLU: Vertiefende Analysen zu Leseverständnis, Rahmenbedingungen und Zusatzstudien* (S. 1–36). Münster: Waxmann.
- Wallner-Paschon, C. (2007). Testinstrumente. In G. Haider & B. Suchań (Hrsg.), *PIRLS 2006. Internationaler Vergleich von Schülerleistungen. Technischer Bericht. Lesen in der Grundschule* (S. 23–42). Salzburg: Projektzentrum für Vergleichende Bildungsforschung (ZVB).
Verfügbar unter https://www.bifie.at/wp-content/uploads/2017/05/PIRLS-2006_technischer-bericht_2007-11-28.pdf
- Wallner-Paschon, C., Grafendorfer, A., Widauer, K. & Suchań, B. (2012). Die getesteten Kompetenzbereiche. In B. Suchań & C. Schreiner (Hrsg.), *PIRLS & TIMSS 2011. Die Kompetenzen in Lesen, Mathematik und Naturwissenschaft am Ende der Volksschule. Technischer Bericht* (S. 11–25). Salzburg: Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens (BIFIE).
Verfügbar unter https://www.bifie.at/wp-content/uploads/2017/10/PIRLS_TIMSS_2011_Technischer-Bericht.pdf
- Wallner-Paschon, C. & Salchegger, S. (2017). Grundkonzeption, Ziele und Organisation von PIRLS. In C. Wallner-Paschon & U. Itzlinger-Bruneforth (Hrsg.), *PIRLS 2016. Technischer Bericht* (S. 3–8). Salzburg: Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens (BIFIE).
Verfügbar unter https://www.bifie.at/wp-content/uploads/2018/04/PIRLS-2016_TechnischerBericht_final.pdf
- Wallner-Paschon, C. & Widauer, K. (2017). PIRLS 2016 – eine Einführung in die Studie. In C. Wallner-Paschon & U. Itzlinger-Bruneforth (Hrsg.), *PIRLS 2016. Die Lesekompetenz am Ende der Volksschule. Erste Ergebnisse* (S. 7–34). Graz: Leykam.
Verfügbar unter https://www.bifie.at/wp-content/uploads/2017/12/PIRLS2016_ErsteErgebnisse_final_web.pdf
- West, R. & Stanovich, K. (2000). Automatic contextual facilitation in readers of three ages. In K. Stanovich (Ed.), *Progress in understanding reading* (pp. 13–20). New York: Guilford. doi:10.2307/1128240
- Widauer, K. & Wallner-Paschon, C. (2017). Entwicklung und Aufbau der Testinstrumente und Kontextfragebögen. In C. Wallner-Paschon & U. Itzlinger-Bruneforth (Hrsg.), *PIRLS 2016. Technischer Bericht* (S. 9–22). Salzburg: Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens (BIFIE).
Verfügbar unter https://www.bifie.at/wp-content/uploads/2018/04/PIRLS-2016_TechnischerBericht_final.pdf
- Wiesner, C., Pacher, K., George, A. C., Breit, S. & Schreiner, C. (2017). *Allgemeine Informationen zur IKM*. Salzburg: Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens (BIFIE).
Verfügbar unter <https://www.bifie.at/allgemeine-informationen-zur-ikm>
- Zamarian, M. & Heydari, P. (2012). Readability of Texts: State of the Art. *Theory and Practice in Language Studies*, 2, 43–53. doi:10.4304/tpis.2.1.01-05
- Zwaan, R. & Singer, M. (2003). Text comprehension. In A. Graesser, M. Gernsbacher & S. Goldman (Eds.), *Handbook of discourse processes* (pp. 83–122). Mahwah: Erlbaum. doi:10.4324/9781410607348

