

Kopftitel: PROZESSBEZOGENE DIAGNOSTIK VON LESEFÄHIGKEITEN

Prozessbezogene Diagnostik von Lesefähigkeiten bei Grundschulkindern  
**zur Veröffentlichung angenommen in der Zeitschrift für Pädagogische Psychologie**

Tobias Richter und Maj-Britt Isberner

Universität Kassel

Johannes Naumann und Yvonne Kutzner

Deutsches Institut für Internationale

Pädagogische Forschung

Autorenhinweis

Die in diesem Artikel berichtete Forschung wurde durch das Bundesministerium für Bildung und Forschung (BMBF) im Rahmen der Forschungsinitiative *Sprachförderung und Sprachdiagnostik* gefördert (Projekt 01GJ0985). Wir danken David Nitz für seine Unterstützung bei der Programmierung der Tests sowie Lilly Andrich, Claudia Beck, Tatjana DeToia, Wanja von der Felsen, Christian Kirsch, Katrin Klein, Dominik Mathieu, Sarah Rom, Miriam Schwarzenthal, Vanessa Wergin, Barbara Werner und Monika Wieland für ihre Hilfe bei der Erstellung von Stimulusmaterialien und der Datenerhebung.

Eine Veröffentlichung der hier vorgestellten Tests im Hogrefe-Verlag ist vorgesehen. Bis auf weiteres werden Interessentinnen und Interessenten, die ProDi-L für nicht-kommerzielle Forschungsvorhaben nutzen möchten, gebeten, sich an den Erstautor zu wenden.

Tobias Richter

Universität Kassel, Institut für Psychologie

Holländische Str. 36-38

34109 Kassel, Deutschland

E-mail: tobias.richter@uni-kassel.de

Tel.: +49-561 804 3577

Fax: +49-561 804 3586

### Zusammenfassung

Aus kognitionspsychologischer Perspektive beruhen Lesefähigkeiten auf der effizienten Bewältigung von Teilprozessen des Leseverstehens auf Wort-, Satz- und Textebene. In diesem Beitrag stellen wir mit ProDi-L ein neuartiges computergestütztes Diagnostikum vor, das durch die kombinierte Erfassung von Antwortrichtigkeit und Reaktionszeit als Indikatoren für die Zuverlässigkeit und Effizienz einzelner Teilprozesse eine differenzierte prozessbezogene Diagnostik des Leseverstehens bei Grundschulkindern ermöglichen soll. Mittels sechs Subtests sollen zusammenhängende, aber psychometrisch klar trennbare Teilfähigkeiten des Leseverstehens erfasst werden. In einer Querschnittsuntersuchung an 536 Kindern der Klassenstufen 1-4 konnten dieser Annahme entsprechend Belege für die faktorielle Validität von ProDi-L erbracht werden. Die Zusammenhänge der Testwerte von ProDi-L mit kriterialen Lesefähigkeitsmaßen (gemessen mit ELFE 1-6), Lehrerurteilen und sprachfreien Intelligenzmaßen (diskriminante Validität) sprechen außerdem für die Konstrukt- und Kriteriumsvalidität des Instruments.

*Schlagwörter:* Grundschul Kinder, kognitive Prozesse, Lesefähigkeiten, Worterkennung, Satzverstehen, Textverstehen

### Abstract

From a cognitive perspective, reading skills depend on efficient component processes of reading comprehension on the word, sentence and text level. In this article, we present the novel computer-based instrument ProDi-L, which uses both accuracy and reaction time as indicators of reliability and efficiency of each component process, thereby allowing for a differentiated and process-oriented assessment of reading comprehension in primary school children. Six subtests were developed to assess related but psychometrically clearly distinguishable component processes. In line with this assumption, a cross-sectional study with 536 children of grades 1-4 confirmed the factorial validity of ProDi-L. Correlations of ProDi-L scores with external measures of reading comprehension (assessed with ELFE 1-6), teacher ratings, and non-verbal intelligence scores (discriminant validity) also confirmed construct, convergent, and discriminant validity.

*Keywords:* Primary school children, cognitive processes, reading skills, word recognition, sentence comprehension, text comprehension

## 1 Einführung

Lesefähigkeiten sind eine Voraussetzung für die individuelle Teilhabe an der Informations- und Wissensgesellschaft. Die Vermittlung und Förderung von Lesefähigkeiten – einschließlich der Prävention und Intervention bei Leseschwächen – zählen daher zu den wichtigsten Aufgaben des Bildungswesens, wobei der Leseinstruktion in der Grundschule eine zentrale Bedeutung zukommt. Für das theoretische Verständnis von Lesefähigkeiten, ihrer Entwicklung im Laufe der Grundschulzeit und der Ursachen von Leseschwierigkeiten hat sich eine kognitionspsychologische Perspektive als besonders fruchtbar erwiesen, nach der Lesefähigkeiten auf der effizienten Bewältigung von kognitiven Teilprozessen des Leseverstehens beruhen, die von der Wort- über die Satzebene bis hin zum Verstehen zusammenhängender Texte reichen (Richter & Christmann, 2009; Vellutino, Fletcher, Snowling & Scanlon, 2004). Zur Erfassung der Effizienz individueller Unterschiede in Teilprozessen des Leseverstehens werden in der kognitiv orientierten Lesefähigkeitsforschung meist ad hoc konstruierte Verfahren eingesetzt, in denen neben der Antwortrichtigkeit oft auch Reaktionszeiten als diagnostische Größen erfasst werden (z.B. Perfetti, 1999). Die Grundidee bei der Verwendung von Reaktionszeiten als Indikator der Effizienz kognitiver Prozesse besteht darin, dass ein effizienter kognitiver Prozess nicht nur zu einem korrekten Verarbeitungsergebnis führt, sondern auch mit relativ geringer Beanspruchung kognitiver Ressourcen ausgeführt werden kann. In die Konstruktion allgemein einsetzbarer, standardisierter Tests zur Erfassung von Lesefähigkeiten hat ein solches prozessorientiertes Vorgehen dagegen noch kaum Eingang gefunden. Hier dominieren bislang produktorientierte Verfahren, bei denen die Testpersonen längere oder kürzere Texte lesen und im Anschluss mit Verständnisfragen die Qualität der beim Lesen aufgebauten mentalen Repräsentation des Textinhalts geprüft wird (für einen Überblick s. z.B. Lenhard & Schneider, 2009). Lediglich Verfahren, die als Speed-Tests konzipiert sind und den Testpersonen eine Zeitbegrenzung vorgeben (z.B. Würzburger Leise Leseprobe, Küspert & Schneider, 1998; Salzburger Lese-Screening, Mayring & Wimmer, 2003) sowie Leseflüchtigkeits- bzw. -geschwindigkeitsmaße (z.B. CBM Reading Aloud, Deno, Mirkin & Chang, 1982; LGVT 6-12, Schneider, Schlagmüller & Ennemoser, 2007) beziehen zumindest indirekt die Effizienz der am Lesen beteiligten Prozesse mit ein. Allerdings handelt es sich bei

diesen Verfahren in der Regel um globale Screening-Verfahren, die kaum Rückschlüsse auf die Effizienz spezifischer Teilprozesse des Leseverstehens erlauben.<sup>1</sup>

In diesem Beitrag stellen wir mit *ProDi-L (Prozessorientierte Diagnostik von Lesefähigkeiten in der Grundschule)* ein computergestütztes Diagnostikum vor, das eine differenzierte prozessbezogene Diagnostik von Teilprozessen des Leseverstehens ermöglichen soll. Die Zielgruppe des Diagnostikums sind Grundschul Kinder, die im Laufe der Grundschulzeit grundlegende Prozesse des Leseverstehens erlernen und einüben müssen. Entsprechend liegt der Schwerpunkt von ProDi-L auf Prozessen, die bei der Worterkennung und beim Verstehen schriftlicher Sätze und kurzer Texte eine Rolle spielen. Im Folgenden werden wir mit der Konzeption von Lesefähigkeiten als effizienter Bewältigung von Teilprozessen des Leseverstehens zunächst den theoretischen Rahmen skizzieren, der der Konstruktion von ProDi-L zu Grunde liegt. Dann erläutern wir die sechs Subtests von ProDi-L und berichten Daten zur Reliabilität, Konstrukt- und Kriteriumsvalidität des Instruments.

### **1.1 Lesefähigkeiten als effiziente Bewältigung von Teilprozessen des Leseverstehens**

Gemäß der kognitionspsychologischen Perspektive, die bei der Konstruktion von ProDi-L leitend war, beruht Lesefähigkeit auf der effizienten, d.h. zuverlässigen und zugleich

---

<sup>1</sup> In dieser Arbeit werden an zahlreichen Stellen die drei verwandten Begriffe Lesefähigkeit, Leseverstehen und Leseverständnis verwendet. Mit *Lesefähigkeit* ist die Fähigkeit gemeint, geschriebene Wörter, Sätze und ganze Texte zu verstehen. *Leseverstehen* bezieht sich auf den Prozess des Leseverstehens, der sich kognitionspsychologisch betrachtet in kognitive Teilprozesse des Leseverstehens zergliedern lässt, die von der (sub)lexikalischen Ebene über die Satzebene bis hin zur Textebene reichen (vgl. z.B. Richter & Christmann, 2009). Wenn von *Teilfähigkeiten des Leseverstehens* die Rede ist, beziehen sie sich auf die individuelle Bewältigung einzelner Klassen dieser Teilprozesse. Der Begriff *Leseverständnis* schließlich bezieht sich auf das Produkt des Leseverstehens, also Qualität und Umfang der mentalen Repräsentationen, die beim Lesen von Wörtern, Sätzen oder Texten konstruiert wurden.

ressourcenschonenden Bewältigung von kognitiven Teilprozessen auf der Wort-, Satz- und Textebene.

**Teilprozesse auf der Wortebene.** Die visuelle Worterkennung, also die Zuordnung eines Eintrags im mentalen Lexikon zu geschriebenen Wörtern, ist ein elementarer Bestandteil des Leseverstehens. Die Effizienz, mit der Schülerinnen und Schüler die visuelle Worterkennung bewältigen und auf lexikalische Repräsentationen zugreifen können, wird daher oft als eine zentrale Komponente von Lesefähigkeiten angesehen (Hypothese der lexikalischen Qualität, Perfetti & Hart, 2002; interaktiv-kompensatorisches Modell, Stanovich, 1980). Ist die visuelle Worterkennung nicht ausreichend routinisiert, werden kognitive Ressourcen gebunden, die dann für hierarchiehöhere Lese- und Lernstrategien nicht mehr zur Verfügung stehen (Theorie der verbalen Effizienz, Perfetti, 1985; für ein Anwendungsbeispiel vgl. Naumann, Richter, Christmann & Groeben, 2008). Drei Arten von Teilprozessen der visuellen Worterkennung werden üblicherweise unterschieden (vgl. auch das Dual Route Cascaded Model, Coltheart, Rastle, Perry, Langdon & Ziegler, 2001). *Phonologische Rekodierungsprozesse* ordnen einer Graphemfolge eine lautliche (phonologische) Repräsentation zu. Diese Prozesse spielen insbesondere in frühen Phasen des Lesenlernens (alphabetische Strategie, Frith, 1986) und bei der Erkennung seltener Wörter eine entscheidende Rolle. Probleme in der phonologischen (Re-)Kodierung sind daher auch eine wahrscheinliche Ursache der spezifischen Leseschwäche (Stanovich, 1988). Mit zunehmender Vertrautheit mit der Schriftsprache werden dagegen *orthographische Vergleichsprozesse* immer wichtiger, die die Worterkennung durch einen direkten Vergleich der Schreibweise eines Worts mit seiner orthographischen Repräsentation im mentalen Lexikon unterstützen (Cunningham & Stanovich, 1990; orthographische Strategie, Frith, 1986). Insbesondere häufige Wörter und solche mit einer irregulären Schreibweise, die anders ausgesprochen werden als es die sprachspezifischen Graphem-Phonem-Korrespondenzregeln nahe legen (wie z.B. die Fremdwörter *Baby* oder *Friseur* im Deutschen), werden beim Lesen dann (primär) durch einen direkten Vergleich mit einer orthographischen Repräsentation im mentalen Lexikon erkannt (Andrews, 1982; Paap & Noel, 1991). Die Worterkennung beinhaltet schließlich die *Aktivierung von Wortbedeutungen*, die die Basis für das Verständnis von Sätzen und zusammenhängenden Texten darstellt. Defizite in der

Effizienz des Zugriffs auf Wortbedeutungen können Leseschwierigkeiten auch bei solchen Kindern hervorbringen, die über gut routinisierte phonologische Rekodierungsprozesse verfügen (Nation & Snowling, 1998).

**Teilprozesse auf der Satzebene.** Beim Verstehen schriftlicher Sätze müssen zusätzlich zur Worterkennung auch eine Analyse des Satzbaus (syntaktisches Parsing) und eine semantische Integration von Wortbedeutungen zu einem kohärenten Satzsinn geleistet werden. Die Frage, ob das syntaktische Parsing modular, d.h. anhand rein formaler syntaktischer Prinzipien (z.B. minimale Anbindung und späte Grenze, Frazier, 1987) erfolgt oder ob semantische und syntaktische Informationen parallel für den inkrementellen Aufbau einer Interpretation des Satzes genutzt werden (z.B. McRae, Spivey-Knowlton & Tanenhaus, 1998), ist eine klassische und nach wie vor nicht abschließend geklärte Kontroverse der Psycholinguistik (Pickering & van Gompel, 2006). Zudem existieren vielfältige Belege dafür, dass semantische, syntaktische und pragmatische Informationen aus dem Satzkontext auch für die Worterkennung genutzt werden (z.T. in Form einer Vorhersage; Pickering & Garrod, 2007). Damit ist klar, dass Worterkennung und Satzverstehen nicht als sequenzielle Verarbeitungsstufen, sondern als interagierende Prozesse aufzufassen sind (z.B. Stanovich, 1980). Dessen ungeachtet kann kein Zweifel daran bestehen, dass individuelle Unterschiede in der Effizienz syntaktischer wie semantischer Integrationsprozesse eng mit dem allgemein erreichbaren Leseverständnis zusammenhängen. So steigt bei schlechten Leserinnen und Lesern der Verarbeitungsaufwand mit wachsender syntaktischer und semantischer Komplexität von Sätzen besonders stark an (Graesser, Hoffman & Clark, 1980). Außerdem sind gute Leserinnen und Leser besser in der Lage, kontextunangemessene Wortbedeutungen zu unterdrücken (Gernsbacher & Faust, 1991) und den Satzkontext für Vorhersagen zu nutzen (Murray & Burke, 2003).

**Teilprozesse auf der Textebene.** Beim Verstehen zusammenhängender schriftlicher Texte müssen die in den einzelnen Sätzen enthaltenen Informationen zu einer lokal (d.h. zwischen benachbarten Sätzen) und global (d.h. über verschiedene Textteile hinweg) kohärenten Repräsentation des Textinhalts verbunden werden. Dafür müssen neue Informationen auf bereits Gelesenes bezogen und in ein gemeinsames Situationsmodell (mentales Modell) der im Text

beschriebenen Sachverhalte integriert werden (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). Im einfachsten Fall der lokalen Kohärenzbildung sind dabei Sinnrelationen zwischen zwei aufeinander folgenden Sätzen zu rekonstruieren. Dazu müssen Leserinnen und Leser z.B. erkennen, auf welche der bereits zuvor angesprochenen Personen, Objekte oder Ereignisse sich nominale oder pronominale Ausdrücke beziehen (z.B. anaphorische Inferenzen, Garrod & Sanford, 1990), und sie müssen temporale oder kausale Beziehungen zwischen Sätzen nachvollziehen bzw. vorwissensgestützt erschließen (kausale Brückeninferenzen, Singer, 1993). Grundschul Kinder unterscheiden sich in der Effizienz, mit der sie derartige Inferenzen bilden können, und diese individuellen Unterschiede erklären über Lesefähigkeiten auf der Wortebene hinaus einen Teil der Varianz im allgemeinen Leseverständnis (z.B. Oakhill, Cain & Bryant, 2003).

## **1.2 Beschreibung der Subtests von ProDi-L**

ProDi-L enthält drei hierarchieniedrige Subtests, die auf Teilprozesse der visuellen Worterkennung abzielen (*Phonologische Rekodierung, Orthographischer Vergleich und Zugriff auf Wortbedeutungen*) und drei hierarchiehohe Subtests, die auf kognitive Prozesse abzielen, die am Verstehen von Sätzen und Satzfolgen beteiligt sind (*Syntaktische Integration, Semantische Integration, Lokale Kohärenzbildung*). Die Grundidee aller sechs Subtests besteht in der Verwendung von Leseaufgaben, die auf die jeweils angezielten Teilprozesse zugeschnitten sind. Jeder Subtest enthält gut kontrollierte Testitems, in denen als schwierigkeitsgenerierende Merkmale systematisch solche sprachlichen Eigenschaften variiert werden, die aus psycholinguistischer Sicht die jeweils fokalen Teilprozesse erleichtern bzw. erschweren.

Für jedes Testitem werden im Sinne des effizienzbezogenen Messkonzepts sowohl Reaktionszeiten als auch die Antwortrichtigkeit erfasst. Die Reaktionszeiten werden dabei nicht als Rohwerte, sondern in logarithmierter Form verwendet. Die Verteilung von Reaktionszeiten ist typischerweise rechtsschief, mit einem steilen Gipfel im Bereich relativ kurzer Reaktionszeiten und einem langgezogenen Schwanz mit längeren und wenigen sehr langen Reaktionszeiten. Sie lässt sich oft mit einer Lognormalverteilung gut beschreiben (z.B. Ulrich & Miller, 1993). Eine logarithmische Transformation der Reaktionszeiten führt entsprechend zu einer Symmetrisierung der Verteilung, indem der Schwanz der Verteilung gestaucht wird. Ein erwünschter Nebeneffekt

der logarithmischen Transformation besteht darin, dass Ausreißerwerte, d.h. besonders lange Reaktionszeiten bei einzelnen Items, bei der Bildung von Skalenwerten weniger stark gewichtet werden und daher die Skalenwerte bei Vorhandensein von Ausreißerwerten, die sich im Schwanz der Reaktionszeitverteilung „verstecken“, deutlich weniger stark verzerrt werden, als dies bei untransformierten Reaktionszeiten der Fall wäre.

Im Folgenden werden die sechs Subtests von ProDi-L im Einzelnen beschrieben:

**(1) Phonologische Rekodierung.** Die Effizienz phonologischer Rekodierungsprozesse bei der visuellen Worterkennung wird in ProDi-L durch eine Aufgabe gemessen, bei der die phonologische Struktur von paarweise dargebotenen phonotaktisch und orthographisch plausiblen Pseudowörtern, die nicht gegen Regeln der Lautstruktur und Schreibung des Deutschen verstoßen, aber keinen tatsächlichen Wörtern entsprechen, miteinander verglichen wird. Dabei hört die Testperson zunächst ein auditorisch dargebotenes Pseudowort, das als Referenzstimulus dient und daher zuerst dargeboten wird, und sieht unmittelbar darauf ein visuell dargebotenes Pseudowort auf dem Bildschirm. Ihre Aufgabe besteht darin zu entscheiden, ob die phonologische Struktur des visuell dargebotenen Pseudoworts mit dem auditorischen Pseudowort übereinstimmt. Bei der Hälfte der insgesamt 64 Pseudowort-Paare stimmt die phonologische Struktur von auditorisch und visuell dargebotenem Pseudowort überein, bei der anderen Hälfte nicht. Als Stimuli werden Pseudowörter mit 1-4 offenen Silben (ohne Konsonantencluster) verwendet (z.B. *gi*, *banufego*), die durch Permutation von 80 Ausgangssilben gebildet wurden (für ein ähnliches Vorgehen vgl. z.B. Frith, Wimmer & Landerl, 1998; Wimmer, 1996). Durch die Variation der Silbenzahlen sollen Testitems unterschiedlicher Schwierigkeit erzeugt werden. Bei nicht übereinstimmenden Pseudowort-Paaren wird die Schwierigkeit zusätzlich durch eine Variation der Anzahl von Nicht-Übereinstimmungen manipuliert.

**(2) Orthographischer Vergleich.** Die Effizienz orthographischer Vergleichsprozesse (Abruf von orthographischen Repräsentationen und Vergleich mit einem geschriebenen Wort) wird in ProDi-L durch eine lexikalische Entscheidungsaufgabe (Wort/Pseudowort-Unterscheidungen) mit 47 Wörtern und 47 (orthographisch und phonotaktisch plausiblen) Pseudowörtern erfasst. Bei lexikalischen Entscheidungen können sowohl orthographische als auch phonologische Merkmale

von Wörtern genutzt werden (Balota, Cortese, Sergent-Marshall, Spieler & Yap, 2004), wobei das relative Gewicht orthographischer Vergleichs- und phonologischer Rekodierungsprozesse von der Worthäufigkeit und Regularität der Graphem-Phonem-Zuordnung abhängt (Häufigkeits-Regularitäts-Interaktion, Andrews, 1982; Paap & Noel, 1991). Letztlich erfordert die lexikalische Entscheidungsaufgabe jedoch immer den Vergleich einer Graphemfolge mit orthographischen Wortformen (Repräsentationen der Schreibung eines Worts), die im mentalen Lexikon gespeichert sind. Eine zuverlässige und zugleich rasche Unterscheidung von Wörtern und Pseudowörtern ist demnach auf gut routinisierte orthographische Vergleichsprozesse angewiesen. Als schwierigkeitsgenerierende Merkmale der Testitems, die tatsächliche Wörter darstellen, wird in ProDi-L neben der Worthäufigkeit die Anzahl orthographischer Nachbarn variiert, d.h. die Anzahl der gleichlangen Wörter des Deutschen, die sich von dem Testitem in nur einem Buchstaben unterscheiden. Die Anzahl orthographischer Nachbarn gilt als Indikator der orthographischen Redundanz (orthotaktische Regularität, Venezky, 1979) und sollte einen erleichternden Effekt auf orthographische Vergleichsprozesse haben (Andrews, 1989; Balota et al., 2004; Huntsman & Lima, 2002). Bei den Pseudowörtern wird die Wortähnlichkeit variiert. Wortähnliche Pseudowörter wurden dabei meist durch Vertauschung des Onsets (d.h. des Konsonanten am Silbenanfang der ersten Silbe) regulärer Wörter gebildet (z.B. *Bame* gebildet aus *Name*), wortunähnliche durch Austausch der Silben irregulärer Wörter, die nicht den Graphem-Phonem-Korrespondenzregeln des Deutschen entsprechen (z.B. *Chilance* gebildet aus *Chili* und *Balance*). Wörter und Pseudowörter sind im Hinblick auf Länge und Häufigkeit parallelisiert, wobei bei Pseudowörtern die Häufigkeit der Ausgangswörter zu Grunde gelegt wurde. Die Wortmerkmale wurden korpuslinguistischen Datenbanken entnommen (Baayen, Piepenbrock & Gulikers, 1995; Projekt Deutscher Wortschatz, 2007).

Bei Sprachen wie dem Deutschen, die über einen hohen Anteil an regulären Graphem-Phonem-Zuordnungen verfügen, stellt sich das generelle Problem, dass für die Worterkennung relevante orthographische und phonologische Wortmerkmale miteinander konfundiert sind. Eine Möglichkeit, die Effizienz orthographischer Vergleichsprozesse in besonders reiner Form zu erfassen, bieten demgegenüber irreguläre Wörter (z.B. *Chef*, *Jeans*, *Pony*) und Pseudohomophone,

die Pseudowörter sind, aber dieselbe Aussprache wie ein tatsächliches Wort besitzen (z.B. *Tswiebl*, *Saane*, *Wammpier*). Derartige Zeichenfolgen können auf Basis phonologischer Rekodierungsprozesse nicht zuverlässig als Wörter bzw. Pseudowörter erkannt werden. Entsprechend enthält ProDi-L 12 irreguläre Wörter und 14 Pseudohomophone, die bei Bedarf als eine getrennte Skala ausgewertet werden können.

**(3) Zugriff auf Wortbedeutungen.** Die Effizienz des Zugriffs auf Wortbedeutungen wird in ProDi-L durch eine Klassifikationsaufgabe gemessen, bei der die Zugehörigkeit von Wörtern zu einer semantischen Kategorie beurteilt werden muss. Jedes der 32 Testitems besteht aus einem Kategoriennamen (z.B. *Obst*), der auditorisch präsentiert wird, und einem Testwort (z.B. *Apfel*), das unmittelbar darauf visuell auf dem Bildschirm präsentiert wird. Die Aufgabe der Testperson besteht darin zu entscheiden, ob das visuell präsentierte Wort einen Unterbegriff des Konzepts darstellt, das durch den Kategoriennamen angesprochen wird. Dabei entsprechen die Kategoriennamen allgemeinen Alltagskonzepten (auf der Basisebene oder der übergeordneten Ebene, Rosch, 1975), so dass davon auszugehen ist, dass die Aufgabe prinzipiell korrekt beantwortet werden kann, sofern die Bedeutung des Testworts bekannt ist. Bei der Hälfte der Testitems stehen Kategoriennamen und Testwort in einer Oberbegriffs-Unterbegriffs-Relation (wie *Obst-Apfel*), bei der anderen Hälfte ist dies nicht der Fall. Derartige Klassifikationsaufgaben erfordern eine Aktivierung der Wortbedeutung (z.B. Thompson-Schill & Gabrieli, 1999). Als schwierigkeitsgenerierende Merkmale, für die ein Einfluss auf die Leichtigkeit des lexikalischen Zugriffs angenommen werden kann, werden neben der Worthäufigkeit die Vertrautheit der Kategorien für Grundschul Kinder (operationalisiert über dreistufige Lehrerratings, drei unabhängige Rater;  $ICC_{\text{unjust}} = .84$  [Maß für die absolute Übereinstimmung] für die gemittelten Ratings; vgl. Wirtz & Caspar, 2002), bei Kategorienzugehörigkeit die Typizität des Testworts für die Kategorie und bei Nicht-Zugehörigkeit der Grad der semantischen Assoziation von Kategoriennamen und Testwort zweistufig variiert (semantisch assoziiert, z.B. *Schulfächer-Lehrer*, vs. semantisch nicht assoziiert, z.B. *Nahrungsmittel-Bahnhof*). Die Variation der semantischen Assoziation wurde über Cosinus-Werte aus einer Latenten Semantischen Analyse (LSA) nach Übersetzung der Wörter ins Englische und unter Verwendung des Semantischen Raums *General*

*Reading up to 1st Year College* des LSA-Online-Tools der University of Colorado at Boulder (s. <http://lsa.colorado.edu/>) überprüft. Dabei unterschieden sich die acht semantisch assoziierten ( $M = 0.27$ ,  $SD = 0.10$ ) und die acht semantisch nicht assoziierten Wortpaare ( $M = 0.03$ ,  $SD = 0.03$ ) deutlich und statistisch signifikant voneinander ( $t(15) = 6.83$ ,  $p < .001$ ,  $d = 3.91$ ).

**(4) Syntaktische Integration.** Die Effizienz der syntaktischen Integration auf Satzebene wird mit einer Aufgabe erfasst, bei der die Testpersonen die syntaktische Wohlgeformtheit von 40 Sätzen mit unterschiedlicher syntaktischer Komplexität beurteilen sollen. Die Hälfte der Sätze ist syntaktisch wohlgeformt, während die andere Hälfte eine Verletzung (morpho-)syntaktischer Regeln (fehlerhafte Wortstellung, fehlerhafte Tempus- oder Kasusbildung, Verb-Präfix an falscher Stelle; z.B. *Die Autos sind über die Straße fahren*) enthält. Als schwierigkeitsgenerierendes Merkmal wird die syntaktische Komplexität der Testitems variiert (Frazier, 1987; Gibson, 1998).

**(5) Semantische Integration.** Zur Erfassung der Effizienz semantischer Integrationsprozesse auf Satzebene enthält ProDi-L eine Satzverifikationsaufgabe mit 48 Testitems, die jeweils zur Hälfte wahre oder falsche begriffliche Aussagen repräsentieren, eine einfache syntaktische Struktur aufweisen und ausschließlich vertraute Wörter enthalten (z.B. *Schnecken sind schnell*; vgl. den Subtest Satzverifikation aus ELVES, Richter & van Holt, 2005; oder das Salzburger Lesescreening, Mayring & Wimmer, 2003). Diese Aufgabe erfordert die Extraktion eines kohärenten Satzsinn und eine epistemische Validierung der im Satz enthaltenen Aussage, d.h. eine Überprüfung des Wahrheitsgehalts anhand des eigenen Weltwissens (Richter, Schroeder & Wöhrmann, 2009). Als schwierigkeitsgenerierende Merkmale werden die semantische Komplexität (Anzahl von Propositionen, d.h. elementaren Aussageeinheiten, in die sich ein Satz zerlegen lässt; Kintsch & Keenan, 1973) und die Vorhersagbarkeit der Wörter aus dem vorangegangenen Satzkontext variiert (Frisson, Rayner & Pickering, 2005).

**(6) Lokale Kohärenzbildung.** Zur Erfassung von Prozessen der lokalen Kohärenzbildung dient eine Aufgabe, bei der die Testpersonen die Plausibilität von 72 kleinen Geschichten beurteilen sollen, die jeweils aus Folgen von zwei Sätzen bestehen. Die Satzfolgen beschreiben eine plausible oder eine unplausible Konfiguration von Sachverhalten. Zur Bewältigung dieser Aufgabe muss ein Sinnzusammenhang zwischen den beiden Sätzen hergestellt werden, indem ein

Situationsmodell der dargestellten Sachverhalte konstruiert wird, das Informationen aus beiden Sätzen und allgemeines Weltwissen integriert (Zwaan & Radvansky, 1998; vgl. den Subtest *Sinnhaftigkeit Satzfolgen* aus dem Lesefähigkeitsdiagnostikum ELVES, Richter & van Holt, 2005). Damit zielt der Test auf Kohärenzbildungsprozesse ab, die auch über einfache Satzfolgen hinaus essenziell für das Verstehen zusammenhängender Texte sind. In den Testitems sollen mit additiven und kausalen Beziehungen zwei Grundtypen von Kohärenzbeziehungen abgebildet werden (Sanders, Wilbert, Spooren & Noordman, 1992). Eine grundlegende kognitive Operation der Kohärenzherstellung, die bei additiven (z.B. temporalen) Kohärenzbeziehungen zentral ist, stellt die Etablierung referenzieller Kontinuität dar. Um zu entscheiden, ob die Satzfolge *Der Gärtner mäht den Rasen - Danach pflanzt er einen Baum* plausibel ist oder nicht, müssen die Testpersonen beispielsweise den Bezug des Pronomens *er* im zweiten Satz durch eine anaphorische Inferenz bestimmen. Ähnlich erfordert das Verstehen der Satzfolge *Jan und Lena werfen Bälle in einen Korb - Dann zählen sie ihre Punkte* die weltwissensgestützte Inferenz, dass Jan und Lena beim Korbballspiel um Punkte gespielt haben (Haviland & Clark, 1974). Eine anspruchsvollere kognitive Operation der Kohärenzherstellung, die beim Verstehen kausaler Beziehungen angewandt werden muss, ist die Bildung kausaler Inferenzen, mit denen eine Ursache-Wirkungsbeziehung zwischen zwei Sätzen (bzw. den darin ausgedrückten Sachverhalten) etabliert wird (z.B. *Es regnete – Simon spannte den Regenschirm auf*; vgl. Singer, 1993). Neben dem Typus der Kohärenzbeziehung wird im Subtest Lokale Kohärenzbildung variiert, ob die Kohärenzbeziehung explizit signalisiert wird oder nicht. Dies kann z.B. durch den Gebrauch des definiten Artikels geschehen, wenn auf einen im vorangegangenen Satz bereits direkt oder indirekt erwähnten Gegenstand Bezug genommen wird (wie der definite Artikel *das* in *Marie nahm die Picknick-Vorräte aus dem Kofferraum – Das Bier war warm*; Haviland & Clark, 1974). Ein anderes Beispiel ist die Verwendung kausaler Konnektive (z.B. *daher*), um Ursache-Wirkungsbeziehungen zwischen Sätzen zu signalisieren. Schließlich wird zwischen den Items des Subtests Lokale Kohärenzbildung variiert, ob positive oder negative Kohärenzrelationen vorliegen (z.B. *daher* vs. *trotzdem* bei positiven vs. negativen Kausalbeziehungen; vgl. Sanders et al., 1992).

### 1.3 Untersuchungsziele und Hypothesen

In einer korrelativen Untersuchung mit Grundschulkindern der Klassenstufen 1 - 4 wurde die psychometrische Qualität (Reliabilität und Validität) der ProDi-L Subtests überprüft. Im Mittelpunkt stand dabei die Tragfähigkeit des prozessbezogenen Messkonzepts. Dabei wurde erwartet, dass individuelle Unterschiede in den beiden für jeden Subtest erhobenen Variablen, Antwortrichtigkeit und Reaktionszeit, voneinander zumindest partiell unabhängige Aspekte der Effizienz von Teilprozessen des Leseverstehens darstellen, also nur in moderater Höhe miteinander korreliert sind. Im Sinne der faktoriellen Validität des Instruments wurde angenommen, dass die Subtests miteinander zusammenhängende, aber psychometrisch klar trennbare Teilfähigkeiten des Leseverstehens erfassen. Diese Hypothese wurde für die Antwortrichtigkeit und die Reaktionszeit in den hierarchieniedrigen und hierarchiehohen Subtests mit konfirmatorischen Faktorenanalysen überprüft. Dabei wurde auch untersucht, ob sich die Faktorstruktur der mit dem Instrument gemessenen Teilfähigkeiten des Leseverstehens zwischen Mädchen und Jungen unterscheidet (konfigurale und metrische Invarianz, Horn & McArdle, 1992). Die Messinvarianz über Geschlechter hinweg ist eine Voraussetzung dafür, Mittelwerte in den ProDi-L-Testwerten von Mädchen und Jungen zu vergleichen. In Schulleistungsuntersuchungen zu Lesefähigkeiten finden sich meist Geschlechtsunterschiede zugunsten der Mädchen (z.B. in PISA 2009, Naumann, Artelt, Schneider & Stanat, 2010), auch wenn diese Unterschiede in der Grundschule nicht immer groß sind (z.B. IGLU 2006, Hornberg, Valtin, Potthoff, Schwippert & Schulz-Zander, 2007). Vor diesem Hintergrund wurden auch Mittelwertsunterschiede zwischen Mädchen und Jungen in den Testwerten von ProDi-L untersucht.

Zu einer weitergehenden Überprüfung der Konstruktvalidität wurden zunächst Unterschiede der Testwerte von ProDi-L in Abhängigkeit von der Klassenstufe betrachtet. Die Subtests von ProDi-L sollen Teilfähigkeiten des Leseverstehens erfassen, die im Deutschunterricht der Grundschule erworben und zunehmend eingeübt werden. Daher sollten sich in höheren Klassenstufen bessere Testleistungen (niedrigere Fehlerraten und Antwortzeiten) zeigen als in niedrigeren Klassenstufen. Da die Testaufgaben basale Teilprozesse des Leseverstehens erfassen sollten, die am Ende der Grundschulzeit in der Regel relativ zuverlässig ausgeführt werden können, wurde erwartet, dass sich die Abnahme der Fehlerrate mit zunehmender Klassenstufe abschwächt

und einem (niedrigen) asymptotischen Wert annähert. Für die Abnahme der logarithmierten Reaktionszeiten mit höherer Klassenstufe wurde dagegen ein linearer Trend erwartet. Diese Vorhersage lässt sich aus dem allgemeinen Potenzgesetz der Übung (Newell & Rosenbloom, 1981) ableiten: Die Effizienz der jeweils angezielten Teilprozesse sollte auch in höheren Klassen durch Übung noch weiter gesteigert werden können, die erzielte Zeitersparnis (ausgedrückt in der ursprünglichen Metrik) dabei aber immer kleiner werden.

Zur Überprüfung der Konstrukt- und Kriteriumsvalidität wurden schließlich Zusammenhänge der Testwerte von ProDi-L mit Leseverständnismaßen und der Lesegeschwindigkeit (ELFE 1-6, Lenhard & Schneider, 2006) sowie dem Lehrerurteil bestimmt. Hier wurde erwartet, dass die Akkuratheit der ProDi-L-Subtests positiv, die Reaktionszeit dagegen negativ mit den kriterialen Lesefähigkeitsmaßen zusammenhängt. Dagegen sollten die Subtests im Sinne der diskriminanten Validität nur moderat positiv mit sprachfreien Intelligenzmaßen korrelieren. Für das Textverständnis als Kriteriumsvariable wurden für jeden Subtest zusätzlich moderierte Regressionsmodelle geschätzt, in denen neben der Antwortrichtigkeit und der Reaktionszeit auch die Interaktion beider Testwerte einbezogen wurde. Im Sinne des prozessbezogenen Messkonzepts sollten Antwortrichtigkeit und Reaktionszeit inkrementelle Beiträge zur Varianzaufklärung im Textverständnis liefern. Der stärkste Beleg für die Validität von ProDi-L als Instrument zur Erfassung der Effizienz von Teilprozessen des Leseverstehens wäre durch eine Interaktion von Antwortrichtigkeit und Reaktionszeit erbracht, bei der das individuelle Textverständnis einer Person dann besonders hoch ist, wenn sie in den Subtests von ProDi-L sowohl eine hohe Antwortrichtigkeit als auch eine niedrige Reaktionszeit erzielt.

## 2 Methode

### 2.1 Stichprobe

An der Untersuchung nahmen 536 Grundschulkinder aus 10 Grundschulen und 33 Schulklassen in Köln und Frankfurt am Main im Alter zwischen 5 und 11 Jahren teil ( $M=9.19$ ,  $SD=1.12$  Jahre; keine Angaben von 8 Probandinnen und Probanden [Pbn]). Darunter waren 260 Mädchen und 246 Jungen (keine Angaben von 30 Pbn). Sechsendsechzig Pbn (12.3%) waren in der ersten Klasse, 126 (23.5%) in der zweiten Klasse, 169 (31.5%) in der dritten Klasse und 175

(32.6%) in der vierten Klasse. Nach Auskunft der Eltern oder – wenn keine Informationen der Eltern vorlagen (für 16.4% der Pbn) – nach Auskunft der Deutschlehrkraft, war die Erstsprache von 410 Pbn (78.1%) Deutsch (keine Angaben für 11 Pbn). Die Untersuchung wurde gegen Ende des Schuljahrs durchgeführt (im Mai und Juni 2010). Die Teilnahme an der Untersuchung war für die Kinder freiwillig. Nur Kinder, von denen Einverständniserklärungen der Eltern vorlagen, konnten an der Untersuchung teilnehmen. Die Teilnahmequote lag bei 82%.

## 2.2 Versuchsplan und erhobene Variablen

Der Versuchsplan der Validierungsuntersuchung war ein querschnittliches korrelatives Design mit den Subtests von ProDi-L, kriterialen Lesefähigkeitsmaßen und Maßen der sprachfreien Intelligenz.

**Subtests von ProDi-L.** Daten für die drei hierarchieniedrigen Subtests (Phonologische Rekodierung, Orthographischer Vergleich, Zugriff auf Wortbedeutungen) von ProDi-L wurden in allen vier Klassenstufen erhoben. Daten für den Subtest Semantische Integration wurden in der zweiten bis vierten Klasse erhoben. Die Subtests Syntaktische Integration und Lokale Kohärenzbildung wurden in den dritten und vierten Klassen durchgeführt. Aufgrund zeitlicher Restriktionen bei der Testung konnten in manchen Klassen einzelne Subtests nicht durchgeführt werden. Die Größen der Teilstichproben, für die Daten der einzelnen Subtests zur Verfügung stehen, sind in Tabelle 1 wiedergegeben. Die Reaktionszeiten wurden zunächst itemweise logarithmiert (natürlicher Logarithmus). Im Anschluss wurde für jede Person eine Ausreißerbereinigung vorgenommen, indem Reaktionszeiten, die mehr als zwei Standardabweichungen vom personenspezifischen Mittelwert abwichen (insgesamt 4.9% aller Messwerte), durch diesen ersetzt wurden (Windsorizing, Ratcliff, 1993). Da auf diese Weise irreguläre Werte, die aufgrund konsistent instruktionswidrigen Verhaltens zustande kommen, unentdeckt bleiben, wurde zusätzlich geprüft, ob irreguläres Antwortverhalten (immer die gleiche Reaktion) oder konsistent unplausibel kurze Reaktionszeiten ( $< 500$  ms) auftraten. Dies war bei den Pbn durchgängig nicht der Fall.

--- Tabelle 1 hier einfügen ---

**Wort-, Satz- und Textverständnis und Lesegeschwindigkeit.** Als objektive kriteriale Lesefähigkeitsmaße wurden die Subtests Wort-, Satz- und Textverständnis sowie Lesegeschwindigkeit aus ELFE 1-6 (computergestützte Version, Lenhard & Schneider, 2006) eingesetzt.

**Lehrerurteile und Deutschnote.** Die Lehrkräfte, die die betreffende Klasse im Fach Deutsch unterrichteten, wurden gebeten, einen Lehrerfragebogen auszufüllen, bei dem für die einzelnen Schülerinnen und Schüler zusammenfassende Einschätzungen der Leseflüssigkeit (*Wie gut kann der/die Schüler/-in altersangemessene Texte laut vorlesen?*) und des Textverständnisses (*Wie gut versteht der/die Schüler/-in einen altersangemessenen Text?*) abzugeben waren (5-stufige Ratings von 1=*deutlich unterdurchschnittlich* bis 5=*sehr gut*). Außerdem sollten sie ebenfalls mit 5-stufigen Ratings detaillierte Einschätzungen im Hinblick auf die sechs Teilfähigkeiten vornehmen, die in den ProDi-L Subtests abgefragt werden (z.B. *Wie gut kann der/die Schüler/-in im Vergleich zu anderen Schüler(inne)n seiner Klassenstufe beim Lesen seltene Wörter erkennen, die anders geschrieben als ausgesprochen werden (z.B. „Baby“)?*). Da diese sechs Ratings hoch miteinander korreliert waren ( $r > .68$ ), wurden sie zu einer Skala Teilfähigkeiten des Leseverstehens zusammengefasst (Cronbachs  $\alpha = .94$ ). Schließlich wurden die Lehrkräfte in den dritten und vierten Klassen um die Angabe der letzten Deutschnote gebeten.

**Sprachunabhängige Intelligenz.** Zur Ermittlung von Indikatoren der diskriminanten Validität wurden Maße der sprachunabhängigen Intelligenz erhoben. In den Klassen 1 und 2 kamen die Untertests 3 (Klassifikation) und 5 (Matrizen) des Grundintelligenztests Skala 1 (CFT 1, Cattell, Weiß & Osterland, 1997) und in den Klassen 3 und 4 die Untertests 1 (Reihen Fortsetzen) und 3 (Matrizen) des Grundintelligenztests Skala 2 (CFT 20-R, Weiß, 2008) zum Einsatz.

### **2.3 Ablauf der Untersuchung**

Die Untersuchung wurde im Klassenverband im Klassenraum der jeweils untersuchten Schulklassen durchgeführt und dauerte zwei Schulstunden, die auf zwei verschiedene Tage verteilt waren. Am ersten Termin bearbeiteten die Pbn die computergestützten Subtests von ProDi-L an Notebook-Computern (15-Zoll-Bildschirm). Auditive Stimuli und Instruktionen wurden über Kopfhörer dargeboten. Zu Beginn erläuterten geschulte Versuchsleiterinnen und Versuchsleiter den

Pbn, dass sie gemeinsam mit dem Außerirdischen Reli Aufgaben zum Lese- und Hörverstehen bearbeiten würden. Den Kindern wurde gesagt, dass Reli zu Besuch auf der Erde sei, um ihre Sprache zu lernen. Ihnen wurde auch mitgeteilt, dass die Teilnahme freiwillig sei und dass sie Fragen stellen könnten, falls sie eine Aufgabe nicht verstehen sollten. Die Instruktionen zur Aufgabenbearbeitung wurden von Reli in mündlicher Form gegeben (unterstützt durch animierte Videos) und konnten von den Pbn bei Bedarf mehrmals angehört werden. Die genaue Zusammenstellung der Leseverständnistests von ProDi-L unterschied sich zwischen verschiedenen Schulklassen, wobei der Fokus in der ersten und zweiten Klassen auf den hierarchieniedrigen Subtests zur Worterkennung lag (s. Abschnitt *Erhobene Variablen*). Außerdem bearbeiteten die Pbn auch computergestützte Hörverstehenstests, die nicht Gegenstand der hier berichteten Untersuchung sind. Jeder der computergestützten Leseverständnistests von ProDi-L begann mit zwei Übungsdurchgängen (mit einem positiven und einem negativen Beispiel), bei denen die Kinder von Reli Feedback für ihre Antworten erhielten. Die Testitems wurden in zufälliger Reihenfolge präsentiert. Am zweiten Untersuchungstermin wurden Maße zur Konstrukt- und Kriteriumsvalidierung erhoben (ELFE 1-6, CFT 1/20).

Während der Untersuchungstermine in den Schulen wurden die Deutschlehrkräfte der untersuchten Klassen gebeten, einen Lehrerfragebogen auszufüllen, der unter anderem Einschätzungen der Leistungsfähigkeit der einzelnen Schülerinnen und Schüler bei verschiedenen Leseanforderungen erfragte (s. Abschnitt *Erhobene Variablen*). Außerdem wurden die Eltern gebeten, einen Elternfragebogen zu soziodemographischen Informationen auszufüllen. Eine anonymisierte Behandlung und Verarbeitung sämtlicher Daten wurde den Eltern und Lehrkräften zugesichert und durch ein Code-System umgesetzt.

### **3 Ergebnisse und Diskussion**

Im Folgenden berichten wir zunächst Skalenskennwerte und Interkorrelationen der Subtests von ProDi-L. Im Anschluss folgen konfirmatorische Faktorenanalysen, die Aufschlüsse über die faktorielle Validität des Instruments liefern, und Analysen zu Klassenstufen-Unterschieden der mit ProDi-L erhobenen Testwerte. Schließlich berichten wir Ergebnisse zu Zusammenhängen mit weiteren Lesefähigkeitsmaßen, die Aufschlüsse über die Konstrukt- und Kriteriumsvalidität des

Instruments liefern können. Für alle Signifikanztests wurde eine Typ-I-Irrtumswahrscheinlichkeit von .05 angesetzt (zweiseitige Testung). Sofern zu einer einzigen Hypothese bzw. Fragestellung mehrere Mittelwertsunterschiede oder bivariate Korrelationen auf Signifikanz getestet wurden, wurde eine Alpha-Fehler-Adjustierung nach der sequenziellen Holm-Bonferroni-Prozedur vorgenommen (Holm, 1979).

### 3.1 Skalenkennwerte der Subtests

Die Skalenkennwerte der sechs Subtests von ProDi-L sind in Tabelle 2 wiedergegeben. Für die Antwortrichtigkeit der Subtests ergaben sich mit Ausnahme des Subtests Lokale Kohärenzbildung befriedigende bis sehr gute interne Konsistenzen (Cronbachs  $\alpha$  .70-.94) und für die Testwerte, die auf logarithmierten Reaktionszeiten beruhen, sehr gute interne Konsistenzen (Cronbachs  $\alpha \geq .98$ ). Die mittleren Itemtrennschärfen zeigten ein ähnliches Bild. Nach beiden Indikatoren können demnach die Subtests von ProDi-L im Hinblick auf die Antwortrichtigkeit als weitgehend homogen und im Hinblick auf die Reaktionszeiten als sehr homogen gelten.

--- Tabelle 2 hier einfügen ---

Die Items aller sechs Subtests erwiesen sich erwartungsgemäß als eher leicht mit moderat linksschiefen Verteilungen der Antwortrichtigkeit. Diese Verteilungscharakteristika sind mit der Zielvorgabe konsistent, dass mit ProDi-L tatsächlich basale Teilfähigkeiten des Leseverstehens erfasst werden, die Grundschulkinder (insbesondere in höheren Klassen) prinzipiell bewältigen können. Auch die Testwerte, die auf (logarithmierten) Reaktionszeiten beruhen, waren moderat linksschief verteilt, wobei die linksschiefe Verteilung durch wenige sehr schnelle Reaktionszeiten bedingt war.

In den explorativen Analysen von Geschlechtsunterschieden erzielten die Mädchen in fast allen Subtests von ProDi-L (mit Ausnahme der Antwortrichtigkeit beim Subtest orthographischer Vergleich) höhere Werte in der Antwortrichtigkeit und in den Reaktionszeiten (Tabelle 2). Die Unterschiede waren allerdings insgesamt eher gering und in den paarweisen Vergleichen nach Alpha-Fehler-Adjustierung nur in vier Fällen statistisch signifikant. Gleichwohl könnte das konsistente Muster auf einen Geschwindigkeits-Genauigkeits-Zielkonflikt (speed-accuracy trade-

off) auf Gruppenebene hindeuten: Die Jungen erzielen möglicherweise deshalb eine etwas niedrigere Antwortrichtigkeit, weil sie eher als die Mädchen zu einer oberflächlichen Sprachverarbeitung auf der Wort- und Satzebene tendieren (vgl. Hannon & Daneman, 2004). Dieses Befundmuster sollte in weiteren Untersuchungen zunächst repliziert werden, um dann die Interpretation im Sinne eines Geschwindigkeits-Genauigkeits-Zielkonflikts gezielt zu überprüfen. In praktischer Hinsicht könnten sich daraus möglicherweise Hinweise auf konkrete Interventionsmaßnahmen zur Egalisierung von Geschlechtsunterschieden in der Leseleistung ableiten.

### **3.2 Interkorrelationen der Subtests**

Die Antwortrichtigkeiten der einzelnen Subtests waren untereinander in mittlerer Höhe korreliert (Tabelle 2). Dasselbe gilt für die Testwerte, die auf den logarithmierten Reaktionszeiten beruhen. Dieses Korrelationsmuster, das auch bei Auspartialisierung der Klassenstufe erhalten blieb, ist konsistent mit der Annahme, dass die Subtests von ProDi-L miteinander zusammenhängende Teilfähigkeiten des Leseverstehens erfassen.

Die Antwortrichtigkeit und die Reaktionszeiten waren bei den hierarchieniedrigen Subtests überwiegend nicht oder schwach negativ miteinander korreliert. Bei den hierarchiehohen Subtests zeigten sich dagegen moderate bis mittlere positive Korrelationen, die auf einen Speed-Accuracy-Zielkonflikt auf Gruppenebene hindeuten: Bei manchen Kindern ging eine schnellere Bearbeitung der hierarchiehohen Subtests offenbar auf Kosten der Zuverlässigkeit der jeweils angezielten Prozesse. Dies deutet auf eine Tendenz zu einer oberflächlichen Verarbeitung durch die Leserinnen und Leser hin, die den Test inakkurat bearbeitet haben (vgl. Hannon & Daneman, 2004). Diese Tendenz war beim Subtest Lokale Kohärenzbildung besonders ausgeprägt, bei dem die Reaktionszeiten mit der Antwortrichtigkeit aller anderen Subtests positiv korreliert waren.

### **3.3 Faktorielle Validität**

Zur näheren Überprüfung der Dimensionalität und Homogenität der Subtests von ProDi-L wurden getrennt für Antwortrichtigkeit und Reaktionszeit konfirmatorische Faktorenmodelle geschätzt. Für die drei hierarchieniedrigen Subtests (Worterkennung), die für die Klassen 1 bis 4 konzipiert sind, und für die drei hierarchiehohen Subtests (Sätze und Satzfolgen), die in erster Linie

für die Klassen 3 und 4 konzipiert sind, wurden zunächst separate Messmodelle auf Basis von Teilstichproben des jeweils angezielten Altersspektrums geschätzt. Zusätzlich wurde anhand der Daten der Klassenstufen 3 und 4 ein Gesamtmodell mit allen sechs Subtests geschätzt.

Die Konzeption von ProDi-L beruht auf der Annahme psychometrisch trennbarer, aber miteinander zusammenhängender Teilfähigkeiten des Leseverstehens. Zur Überprüfung dieser Annahme wurden Messmodelle mit interkorrelierenden latenten Variablen (Teilfähigkeiten) geschätzt und mit alternativen Modellen verglichen, in denen die Unterscheidung von Paaren von Teilfähigkeiten nivelliert wurde, indem die Korrelationen der latenten Variablen auf 1 fixiert wurden. Jede latente Variable wurde durch vier gleich große Itempäckchen mit Items aus dem jeweiligen Subtest von ProDi-L operationalisiert (vgl. Hau & Marsh, 2004). Bei der Bildung der Itempäckchen wurde darauf geachtet, dass sich objektive Itemmerkmale (z.B. Worthäufigkeit, Silbenzahl, Satzlänge etc.) gleichmäßig auf die Itempäckchen verteilen. In jedem Messmodell wurden die Ladungen der vier Itempäckchen, die derselben latenten Variablen zugeordnet waren, gleichgesetzt. Diese Restriktion entspricht der relativ strikten Homogenitätsannahme, dass die vier Itempäckchen Paralleltests darstellen, die die jeweilige Teilfähigkeit gleich gut erfassen. Die Parameterschätzung erfolgte mit der Maximum-Likelihood-Methode in Lisrel 8 (Jöreskog & Sörbom, 1996) bzw. MPlus 4 (Muthén & Muthén, 1998-2006), die auch bei moderaten Verletzungen der Annahme multivariat normalverteilter beobachtbarer Variablen erwartungstreue und effiziente Schätzungen liefert (West, Finch & Curran, 1995). Zur Kontrolle möglicher Verzerrungen durchgeführte Parameterschätzungen mit der Unweighted-Least-Squares Methode, die nicht die multivariate Normalverteilung der beobachtbaren Variablen voraussetzt, führten zu nahezu identischen Parameterschätzungen und Fit-Statistiken, so dass die Ergebnisse hier nicht gesondert berichtet werden.

--- Abbildung 1 hier einfügen ---

**Messmodelle für die hierarchieniedrigen Subtests (Worterkennung).** Die Parameterschätzungen der hypothetisierten Messmodelle für Antwortrichtigkeit und Reaktionszeit bei den drei hierarchieniedrigen Subtests sind in Abbildung 1a dargestellt. Beide Messmodelle wiesen eine gute Passung zu den Daten auf (Antwortrichtigkeit:  $\chi^2(60)=77.17, p=.07,$

RMSEA=0.03, SRMR=0.00, CFI=1.00; Reaktionszeit:  $\chi^2(60)=113.90, p<.001$ , RMSEA=0.04, SRMR=0.02, CFI=1.00). Die Modellpassung der alternativen Modelle, in denen die Interkorrelationen der latenten Variablen paarweise auf 1 fixiert waren, war demgegenüber deutlich und statistisch signifikant verschlechtert (Antwortrichtigkeit: für alle drei Alternativmodelle  $\Delta\chi^2(1) > 368.94, p < .001$ ; Reaktionszeit: für alle drei Alternativmodelle  $\Delta\chi^2(1) > 1889.94, p<.001$ ). Demnach stützen die Ergebnisse der konfirmatorischen Faktorenanalysen sowohl für die Antwortrichtigkeit als auch für die Reaktionszeiten die Annahme, dass individuelle Unterschiede in phonologischen Rekodierungsprozessen, orthographischen Vergleichsprozessen und im Zugriff auf Wortbedeutungen mit ProDi-L als miteinander zusammenhängende, aber psychometrisch trennbare Teilfähigkeiten der Worterkennung gemessen werden. Auch bei Gleichsetzung der Faktorladungen der vier Itempäckchen der drei latenten Variablen konnte eine gute Modellpassung erzielt werden, was für die Homogenität der hierarchieniedrigen Subtests von ProDi-L spricht.

Angesichts der oft berichteten Unterschiede zwischen Jungen und Mädchen in Leseleistungen (z.B. Naumann et al., 2010) wurde zusätzlich untersucht, ob sich das Messmodell für die hierarchieniedrigen Subtests über die Geschlechter generalisieren lässt. Hierzu wurden zwei geschachtelte Zweigruppenmodelle gegenübergestellt, von denen das liberalere konfigurale Invarianz (gleiche Faktorenstruktur und gleiche Intercepts) annahm und das strengere metrische Invarianz (zusätzlich gleiche Faktorladungen; Horn & McArdle, 1992). Für die Antwortrichtigkeit konnte für das Modell mit konfiguraler Invarianz eine gute Modellpassung gesichert werden ( $\chi^2(120) = 131.85, p = .21, CFI = 1.00, RMSEA = 0.02, SRMR = 0.06$ ). Das Modell mit metrischer Invarianz wies eine nur unwesentlich schlechtere Modellpassung auf ( $\chi^2(123) = 146.88, p = .07, CFI = 0.99, RMSEA = 0.03, SRMR = 0.09$ ; Modellvergleich:  $\Delta\chi^2(3) = 12.92, p < .01$ ). Für die Reaktionszeit fanden sich analoge Ergebnisse, wobei hier kein statistisch signifikanter Unterschied in der Passung der Modelle mit konfiguraler und metrischer Invarianz auftrat (konfigurale Invarianz:  $\chi^2(120) = 144.88, p = .06, CFI = 1.00, RMSEA = 0.03, SRMR = 0.12$ ; metrische Invarianz:  $\chi^2(123) = 144.08, p = .09, CFI = 1.00, RMSEA = 0.03, SRMR = 0.12$ ; Modellvergleich:  $\Delta\chi^2(3) = 0.61, p = .89$ ).

**Messmodelle für die hierarchiehohen Subtests (Sätze und Satzfolgen).** Die

Parameterschätzungen der hypothetisierten Messmodelle für die drei hierarchiehohen Subtests sind in Abbildung 1b dargestellt. Das Messmodell für Antwortrichtigkeit hatte eine gute Passung ( $\chi^2(60)=100.39, p<.001, RMSEA=0.05, SRMR=0.05, CFI=1.00$ ). Die Passung des Messmodells für die Reaktionszeiten war akzeptabel ( $\chi^2(60)=112.44, p<.001, RMSEA=0.06, SRMR=0.02, CFI=.99$ ). Bei den hierarchiehohen Subtests verschlechterte sich die Modellpassung wiederum deutlich, wenn die Interkorrelationen der latenten Variablen paarweise auf 1 fixiert wurden (Antwortrichtigkeit: für alle drei Alternativmodelle  $\Delta\chi^2(1)>95.90, p<.001$ ; Reaktionszeit: für alle drei Alternativmodelle  $\Delta\chi^2(1)>1158.08, p<.001$ ). Demnach scheinen auch die hierarchiehohen Subtests miteinander zusammenhängende, aber psychometrisch unterscheidbare Teilfähigkeiten darzustellen, die mit homogenen Skalen gemessen werden.

Analog zu den Analysen für die hierarchieniedrigen Subtests wurde für die hierarchiehohen Subtests die Messäquivalenz über Geschlecht durch zwei geschachtelte Modelle geprüft, die jeweils konfigurale und metrische Invarianz annahmen. Für die Antwortrichtigkeit ergab sich für beide Modelle eine sehr gute Modellpassung. Die Passung beider Modelle unterschied sich nicht statistisch signifikant voneinander (konfigurale Invarianz:  $\chi^2(120) = 117.97, p = .53, RMSEA = 0.00, SRMR = 0.08, CFI = 1.00$ ; metrische Invarianz:  $\chi^2(123) = 120.12, p = .56, RMSEA = 0.00, SRMR = 0.10, CFI = 1.00$ ; Modellvergleich:  $\Delta\chi^2(3) = 2.48, p = .48$ ). Für die Reaktionszeiten ergab sich eine insgesamt etwas weniger gute, aber immer noch akzeptable Modellpassung. Die Passung der Modelle mit konfiguraler und metrischer Invarianz unterschied sich wiederum nicht statistisch signifikant voneinander (konfigurale Invarianz:  $\chi^2(120) = 166.53, p < .01, CFI = 0.98, RMSEA = 0.05, SRMR = 0.15$ ; metrische Invarianz:  $\chi^2(123) = 172.30, p < .01, CFI = 0.98, RMSEA = 0.05, SRMR = 0.13$ ; Modellvergleich:  $\Delta\chi^2(3) = 6.25, p = .10$ ).

**Messmodelle für den Gesamttest (hierarchieniedrige und –hohe Subtests).** Zur

Prüfung der faktoriellen Validität des Gesamttests wurde anhand der Daten für die dritten und vierten Klassen ( $N = 344$ ) ein sechsfaktorielles Modell geschätzt. Als Indikatoren dienten

dieselben Itempäckchen, die für die Modelle für die hierarchieniedrigen und –hohen Subtests verwendet worden waren. Dieses Modell wies eine gute Passung zu den Daten auf ( $\chi^2(249) = 332.80, p < .01, RMSEA = 0.03, SRMR = 0.06, CFI = 0.99$ ). Mehrgruppenanalysen mit Geschlecht als Gruppierungsvariable ergaben für die Antwortrichtigkeit eine akzeptable bis gute Modellpassung sowohl für das Modell mit konfiguraler als auch für das Modell mit metrischer Invarianz, wobei das Modell mit metrischer Invarianz etwas schlechter an die Daten angepasst war (konfigurale Invarianz:  $\chi^2(510) = 647.05, p < .001, RMSEA = 0.04, SRMR = 0.10, CFI = 0.96$ ; metrische Invarianz:  $\chi^2(516) = 674.20, p < .001, RMSEA = 0.05, SRMR = 0.13, CFI = 0.95$ ; Modellvergleich:  $\Delta\chi^2(6) = 27.14, p < .001$ ). Für die Reaktionszeit ergab sich sowohl für das Modell mit konfiguraler als auch für das Modell mit metrischer Invarianz ebenfalls eine akzeptable Modellpassung. Die Passung beider Modelle unterschied sich nicht statistisch signifikant voneinander (konfigurale Invarianz:  $\chi^2(510) = 804.70, p < .001, CFI = 0.97, TLI = 0.97, RMSEA = 0.06, SRMR = 0.11$ ; metrische Invarianz:  $\chi^2(516) = 813.31, p < .001, CFI = 0.97, TLI = 0.97, RMSEA = 0.06, SRMR = 0.12$ ; Modellvergleich:  $\Delta\chi^2(6) = 8.79, p = .19$ ).

### 3.4 Unterschiede zwischen Klassenstufen

Sowohl die mittleren Reaktionszeiten als auch die mittleren Fehlerraten (1 – Anteil richtiger Antworten) der Subtests von ProDi-L nahmen erwartungsgemäß monoton von Klasse 1 bis Klasse 4 ab (Abbildung 2). Bei den drei hierarchieniedrigen Subtests zeigten sich in allen Fällen statistisch signifikante und starke globale Effekte der Klassenstufe. Bei den Reaktionszeiten waren diese Effekte für den Subtest Orthographischer Vergleich besonders stark ( $F(3,524)=38.70, p<.001, \eta_p^2=.19$ ), für die phonologische Rekodierung ( $F(3,524)=17.27, p<.001, \eta_p^2=.09$ ) und den Zugriff auf Wortbedeutungen ( $F(3,434)=11.84, p<.001, \eta_p^2=.08$ ) dagegen geringer. Ähnlich fanden sich bei den Fehlerraten (Antwortrichtigkeit) die größten Klassenstufenunterschiede für den Subtest Orthographischer Vergleich ( $F(3,524)=109.29, p<.001, \eta_p^2=.39$ ) und weniger stark ausgeprägte Klassenstufenunterschiede für die Subtests Zugriff auf Wortbedeutungen ( $F(3,434)=34.53, p<.001, \eta_p^2=.19$ ) und Phonologische Rekodierung ( $F(3,524)=15.72, p<.001, \eta_p^2=.08$ ).

--- Abbildung 2 hier einfügen ---

Für die drei hierarchiehohen Subtests wurden Daten nur für die Klassenstufen 3 und 4 (Syntaktische Integration und Lokale Kohärenzbildung) bzw. 2 bis 4 (Semantische Integration) erhoben. Hier ergaben sich statistisch signifikante Klassenstufenunterschiede in den Reaktionszeiten für den Subtest Syntaktische Integration ( $F(1,331)=25.95, p<.001, \eta_p^2=.07$ ) und den Subtest Semantische Integration ( $F(2,442)=17.50, p<.001, \eta_p^2=.08$ ). Beim Subtest Lokale Kohärenzbildung zeigte sich nur ein geringer und nicht statistisch signifikanter Unterschied zwischen der dritten und der vierten Klassenstufe ( $F(1,264)=3.73, p=.05, \eta_p^2=.01$ ). Für die Fehlerraten ergaben sich geringe, aber statistisch signifikante Klassenstufenunterschiede in den Subtests Semantische Integration ( $F(2,442)=4.51, p<.001, \eta_p^2=.02$ ) und Lokale Kohärenzbildung ( $F(1,264)=8.89, p<.001, \eta_p^2=.03$ ). Für die Fehlerrate des Subtests Syntaktische Integration zeigten sich keine statistisch signifikanten Unterschiede zwischen der dritten und vierten Klasse ( $F(1,264)=8.89, p<.001, \eta_p^2=.03$ ).

Unter der Voraussetzung, dass die Zusammensetzung der Teilstichproben auf den einzelnen Klassenstufen vergleichbar ist, stützt das Muster der Klassenstufenunterschiede die Annahme, dass mit ProDi-L Teilfähigkeiten des Leseverstehens erfasst werden, die sich über die Grundschulzeit kontinuierlich verbessern. Mit Ausnahme des relativ schweren Subtests Lokale Kohärenzbildung erreichte die Fehlerrate in den vierten Klassen durchgängig einen niedrigen Wert um 10%. Bei allen Subtests mit Ausnahme der Lokalen Kohärenzbildung waren zudem die Unterschiede in den Fehlerraten zwischen dritter und vierter Klasse nur gering und – sofern entsprechende Daten vorlagen – geringer als die Unterschiede zwischen niedrigeren Klassenstufen. Das Muster der Klassenunterschiede in den Testwerten, die auf logarithmierten Reaktionszeiten beruhen, scheint dagegen eher für eine lineare Abnahme mit steigender Klassenstufe zu entsprechen.

Für die drei hierarchieniedrigen Subtests mit Daten aus allen vier Klassenstufen konnte das vorhergesagte Muster der Klassenstufenunterschiede auch inferenzstatistisch überprüft werden. Dazu wurden Regressionsmodelle geschätzt, die neben der Klassenstufe als linearem Prädiktor auch die quadrierte Klassenstufe als Prädiktor enthielten. Dabei wurde die Klassenstufe in zentrierter Form einbezogen (vgl. Cohen, Cohen, West & Aiken, 2003, Kap. 6). Für die Fehlerraten

als Kriteriumsvariablen sollte sich die angenommene Abschwächung der Verbesserung mit höheren Klassenstufen in einem negativen Effekt der Klassenstufe und einem positiven Effekt der quadrierten Klassenstufe niederschlagen. Diese Vorhersage konnte für die Subtests Orthographischer Vergleich und Zugriff auf Wortbedeutungen gestützt werden. Bei diesen beiden Subtests hatte die Klassenstufe einen negativen Effekt (Orthographischer Vergleich:  $B=-0.07$ ,  $SE_B=0.004$ ,  $p<.001$ ,  $\Delta R^2=.11$ ; Zugriff auf Wortbedeutungen:  $B=-0.05$ ,  $SE_B=0.01$ ,  $p<.001$ ,  $\Delta R^2=.07$ ), die quadrierte Klassenstufe aber einen inkrementellen positiven Effekt (Orthographischer Vergleich:  $B=0.02$ ,  $SE_B=0.004$ ,  $p<.001$ ,  $\Delta R^2=.02$ ; Zugriff auf Wortbedeutungen:  $B=0.02$ ,  $SE_B=0.01$ ,  $p<.01$ ,  $\Delta R^2=.02$ ). Dieses Muster deutet auf eine kontinuierliche Verringerung der Fehlerraten von der ersten bis zur vierten Klasse hin, die sich in den höheren Klassenstufen aber abschwächt. Für den Subtest Phonologische Rekodierung zeigte sich dagegen lediglich der erwartete negative Effekt der Klassenstufe als linearem Prädiktor ( $B=-0.04$ ,  $SE_B=0.01$ ,  $p<.001$ ,  $\Delta R^2=.07$ ), während der Effekt der quadrierten Klassenstufe nicht statistisch signifikant war ( $B=0.01$ ,  $SE_B=0.02$ ,  $p=.62$ ).

Mit den logarithmierten Reaktionszeiten als Kriteriumsvariable zeigten sich dagegen bei allen Subtests der erwartete negative Effekt der Klassenstufe (Phonologische Rekodierung:  $B=-0.11$ ,  $SE_B=0.02$ ,  $p<.001$ ,  $\Delta R^2=.07$ ; Orthographischer Vergleich:  $B=-0.22$ ,  $SE_B=0.02$ ,  $p<.001$ ,  $\Delta R^2=.17$ ; Zugriff auf Wortbedeutungen:  $B=-0.15$ ,  $SE_B=0.03$ ,  $p<.001$ ,  $\Delta R^2=.08$ ), aber in keinem Fall ein statistisch signifikanter positiver Effekt der quadrierten Klassenstufe, der eine Abschwächung dieses Effekts in den höheren Klassenstufen anzeigen würde. Stattdessen hatte beim Subtest Zugriff auf Wortbedeutungen die quadrierte Klassenstufe sogar einen negativen Effekt, der auf eine moderate Beschleunigung der Routinisierung dieser Prozesse in den höheren Klassenstufen hindeutet ( $B=-0.05$ ,  $SE_B=0.03$ ,  $p<.05$ ,  $\Delta R^2=.01$ ).

### 3.5 Zusammenhänge mit kriterialen Lesefähigkeitsmaßen

In Tabelle 3 sind Zusammenhänge (Partialkorrelationen nach Auspartialisierung der Klassenstufe) der Subtests von ProDi-L mit kriterialen Lesefähigkeitsmaßen (Urteile der Lehrkraft im Fach Deutsch, Deutschnote und Subtests von ELFE 1-6, Lenhard & Schneider, 2006) als

Indikatoren der konvergenten Validität wiedergegeben. Weiterhin enthält Tabelle 3 Zusammenhänge mit sprachunabhängigen Intelligenzmaßen (Subtests aus CFT 1, Cattell et al., 1997, bzw. CFT 20-R, Weiß, 2008) als Indikatoren der diskriminanten Validität.

--- Tabelle 3 hier einfügen ---

Die Antwortrichtigkeit korrelierte bei allen ProDi-L Subtests positiv und überwiegend in mittlerer Höhe mit den Subtests Wort-, Satz- und Textverständnis aus ELFE 1-6 und den Urteilen der Deutschlehrkraft sowie negativ mit der in Millisekunden gemessenen Lesegeschwindigkeit aus ELFE 1-6 und der Deutschnote. Die Testwerte, die auf logarithmierten Reaktionszeiten beruhen, korrelierten überwiegend negativ und in mittlerer Höhe mit Wort-, Satz und Textverständnis aus ELFE 1-6. Nur beim Subtest Lokale Kohärenzbildung ergaben sich positive Korrelationen. Ähnlich wie die Korrelationen von Antwortrichtigkeit und Reaktionszeit für diesen Subtest legt auch dieser Befund die Interpretation nahe, dass eine schnellere Bearbeitung der Satzfolgen in diesem Subtest oft auf eine oberflächliche Verarbeitung zurückzuführen ist. Mit den Urteilen der Deutschlehrkraft ergaben sich für die Reaktionszeiten nur vereinzelte und eher geringe Korrelationen. Angesichts der deutlichen Zusammenhänge der Reaktionszeiten der ProDi-L Subtests mit ELFE 1-6 spricht dieser Befund allerdings nicht gegen die Validität der ProDi-L-Subtests. Vielmehr scheinen sich die Deutsch-Lehrkräfte bei Ihren diagnostischen Urteilen eher auf Lesefehler als auf Leseflüssigkeit zu stützen. Darüber hinaus hat vermutlich auch die Tatsache, dass diagnostische Urteile von Lehrkräften auch für Schülerleistungen im Fach Deutsch in der Regel eher moderate Zusammenhänge mit tatsächlichen Schülerleistungen aufweisen (im Bereich von .50-.60, z.B. Lorenz & Artelt, 2009; s. auch Karing, Matthäi & Artelt, 2011; Südkamp, Kaiser & Möller, 2012), zu der geringen Höhe der Korrelationen der ProDi-L-Subtests mit Lehrerurteilen beigetragen.

--- Tabelle 4 hier einfügen ---

Zusätzlich zu den bivariaten Korrelationen wurden für jeden der Subtests von ProDi-L multiple Zusammenhänge von Antwortrichtigkeit, logarithmierter Reaktionszeit und der Interaktion beider Variablen mit dem Subtest Textverständnis aus ELFE 1-6 (Lenhard & Schneider, 2006) geschätzt (Tabelle 4). Bei allen Subtests bis auf Lokale Kohärenzbildung ergab sich für die Reaktionszeiten ein negativer und für die Antwortrichtigkeit ein positiver Effekt auf das

Textverständnis. Bei allen Subtests bis auf Lokale Kohärenzbildung und Orthographischer Vergleich zeigte sich außerdem ein statistisch signifikanter Interaktionseffekt dergestalt, dass der Effekt der Antwortrichtigkeit umso positiver ausfiel, je stärker der jeweilige Teilprozess routinisiert war.

#### 4. Gesamtdiskussion und Ausblick

In diesem Beitrag haben wir mit ProDi-L ein neu entwickeltes diagnostisches Instrument zur Erfassung von Teilfähigkeiten des Leseverstehens bei Grundschulkindern vorgestellt, das sich durch ein konsequent prozessbezogenes Messkonzept von gebräuchlichen Lesefähigkeitstests unterscheidet. In einer Querschnittsuntersuchung konnten Belege für die faktorielle Validität des Instruments und seine Konstrukt- und Kriteriumsvalidität erbracht werden. Die prozessorientierten Subtests auf der Ebene von Wörtern, Sätzen und Satzfolgen scheinen zusammenhängende, aber gleichwohl psychometrisch klar trennbare Teilfähigkeiten des Leseverstehens zu erfassen. Damit erweisen sie sich als differenzierter und für eine gezielte Förderdiagnostik vermutlich geeigneter als herkömmliche produktorientierte Leseverständnistests, die auch bei heterogenen Aufgaben in der Regel nur eine globale Lesefähigkeitsdimension erfassen (z.B. Rost, 1989).

Allgemein zeigte sich im Sinne der Erfassung der Effizienz von Teilprozessen des Leseverstehens, dass die Reaktionszeiten der Subtests von ProDi-L als Indikatoren der Ressourcenbeanspruchung gegenüber der Antwortrichtigkeit als Indikator der Zuverlässigkeit zusätzliche Varianzteile im allgemeinen Textverständnis aufklären können. Bei vier Subtests zeigten sich darüber hinaus Interaktionseffekte von Reaktionszeit und Antwortrichtigkeit. Von diesem allgemein sehr konsistenten und schlüssigen Muster weichen lediglich die Ergebnisse zum Subtest Lokale Kohärenzbildung ab, bei dem die interne Konsistenz der Antwortrichtigkeit und Zusammenhänge mit kriterialen Maßen nicht zufriedenstellend ausgefallen und Hinweise auf einen Geschwindigkeits-Genauigkeits-Zielkonflikt aufgetreten sind. Eine mögliche Ursache für diese Probleme könnte sein, dass die Testitems dieses Tests teilweise zu schwer waren und die Kinder überfordert haben (dies gilt insbesondere für Satzfolgen mit negativen Kausalrelationen, wie sie durch konzessive und adversative Konjunktionen wie *trotzdem* oder *obwohl* signalisiert werden). Es ist auch denkbar, dass die Erfassung globaler (satzweiser) Reaktionszeiten, wie sie in den

hierarchiehoher Subtests von ProDi-L umgesetzt wurde, bei komplexen und langen sprachlichen Stimuli wie den Satzfolgen des Subtests Lokale Kohärenzbildung nicht optimal ist. Eine Alternative für eine Erfassung der Effizienz von Teilprozessen des Leseverstehens auf dieser Ebene stellt möglicherweise die Erhebung wortweiser Lesezeiten dar (self-paced reading, vgl. z.B. Schroeder, 2011).

Insgesamt belegen die Ergebnisse aber nachdrücklich den diagnostischen Nutzen des effizienz- und prozessbezogenen Messkonzepts von ProDi-L und der dahinter stehenden kognitionspsychologischen Perspektive auf Lesefähigkeiten. Wenn gute Lesefähigkeiten auf effizienten, d.h. sowohl reliablen als auch schnellen (ressourcenschonenden) Teilprozessen des Leseverstehens beruhen, ist es sinnvoll, sowohl Antwortrichtigkeit als auch Bearbeitungszeiten von klar umrissenen Leseaufgaben zu erfassen, da beide Variablen inkrementelle diagnostische Informationen enthalten. Diese Schlussfolgerung impliziert, dass Maße, die beide Arten von Informationen kombinieren, eine höhere Validität aufweisen als Antwortrichtigkeit und Reaktionszeit alleine. Ähnliche kombinierte Maße sind in jüngerer Zeit in der CBM-Forschung erfolgreich erprobt worden (z.B. Maze accurate response rate, Hale, Henning, Hawkins, Sheeley, Shoemaker, Reynolds & Moch, 2011) und haben sich auch in dem reaktionszeitbasierten Lesefähigkeitsdiagnostikum ELVES für erwachsene Leserinnen und Leser bereits bewährt (Naumann, Richter, Flender, Christmann & Groeben, 2007; Naumann et al., 2008; Richter & van Holt, 2005). In der Tat führt auch bei den ProDi-L-Subtests ein einfacher kombinierter Testwert, der Quotient aus Antwortrichtigkeit und logarithmierter Reaktionszeit pro Subtest, zu Zusammenhängen mit den kriterialen Leseverständnismaßen, die deutlich stärker sind als die Zusammenhänge von Antwortrichtigkeit und Reaktionszeit allein. Die Korrelationen mit ELFE-Textverständnis z.B. sind für das kombinierte Maß mit Ausnahme des Subtests Lokale Kohärenzbildung ( $r = .35$ ) größer als .50, wobei die höchsten Korrelationen mit den hierarchieniedrigen Subtests Orthographischer Vergleich ( $r = .68$ ) und Semantische Integration ( $r = .65$ ) erzielt wurden. Dies deutet darauf hin, dass kombinierte Testwerte für die ProDi-L-Subtests für ein kognitionspsychologisch begründetes Lesefähigkeits-Screening nutzbar gemacht werden könnten. Der Nachteil derartiger kombinierter Testwerte liegt allerdings darin, dass sie nur schwer

interpretierbar sind, da die Art und Weise, wie Reaktionszeit und Antwortrichtigkeit miteinander verrechnet werden, letztlich willkürlich ist. Ein sinnvoller nächster Schritt, der für die ProDi-L-Subtests auch vorgesehen ist, wäre daher hier die modellbasierte Schätzung von Fähigkeitsparametern, die Antwortrichtigkeit und Reaktionszeit integrieren (z.B. Roskam, 1997).

Eine wichtige Frage, die in den hier dargestellten Analysen aus forschungspraktischen Gründen nur am Rande behandelt werden konnte, betrifft die Beziehung der Effizienz von Teilprozessen des Leseverstehens, wie sie mit ProDi-L gemessen werden kann, zu Intelligenz. Unsere Ergebnisse zeigen eher geringe Zusammenhänge von Antwortrichtigkeit und Reaktionszeit mit sprachfreien Intelligenzmaßen, was darauf hindeutet, dass mit ProDi-L tatsächlich lesespezifische kognitive Fähigkeiten erfasst werden. Ungeklärt bleiben damit jedoch noch die Beziehungen zur verbalen Intelligenzmaßen auf der einen und zur allgemeinen kognitiven Leistungsfähigkeit auf der anderen Seite. Was *verbale Intelligenzmaße* angeht, so wären für derartige Maße deutlich höhere Zusammenhänge mit den ProDi-L-Subtests zu erwarten als für die sprachfreien Intelligenzmaße in der vorliegenden Untersuchung. Verbale Intelligenzmaße beinhalten die Verarbeitung von (in der Regel) schriftsprachlichen Informationen (Wörtern, Sätzen und Texten). Zudem stellen die ProDi-L-Subtests Anforderungen, die den Anforderungen verbaler Intelligenzmaße ähneln. So werden z.B. semantische Klassifikationsaufgaben, wie sie in ProDi-L eingesetzt werden, in ähnlicher Form auch zur Messung verbaler Intelligenz eingesetzt. Die Anforderung, Informationen innerhalb und zwischen Sätzen zu integrieren, wie sie in den hierarchiehohen Subtests von ProDi-L gestellt wird, steht auch im Mittelpunkt von Schlussfolgerungsaufgaben mit verbalen Stimuli, wie sie in vielen Intelligenztests vorkommen. Ein bedeutender Unterschied besteht aber darin, dass die ProDi-L-Subtests viel einfacher sind als typische Aufgaben zur Messung verbaler Intelligenz, was sich empirisch in den niedrigen Fehlerraten niederschlägt. Ein Kind, das etwa die Items im ProDi-L-Subtest Zugriff auf Wortbedeutungen zuverlässig lösen kann, scheitert möglicherweise trotzdem an verbalen Schlussfolgerungsaufgaben in einem Intelligenztest. Wenn das Kind allerdings die Items desselben Subtests nicht nur zuverlässig, sondern auch schnell lösen kann, sind schon bessere Voraussetzungen für ein gutes Abschneiden in den Intelligenztestaufgaben gegeben, weil durch den

Zugriff auf Wortbedeutungen nur in geringem Maße kognitive Ressourcen beansprucht werden, so dass diese dann für logisches Schlussfolgern und andere ressourcenintensive Denkprozesse eingesetzt werden können. Aus dieser Perspektive erfasst ProDi-L rezeptiv-(schrift-)sprachliche Fähigkeiten, die in gewisser Weise grundlegender sind als die Fähigkeit zum schlussfolgernden Denken mit sprachlichen Informationen, wie sie in typischen verbalen Intelligenzmaßen erfasst wird.

Eine weitere noch offene Frage betrifft die Beziehung der ProDi-L-Subtests zu *allgemeiner kognitiver Leistungsfähigkeit*, die sich als gemeinsamer Faktor individueller Unterschiede in einer großen Bandbreite von kognitiven Aufgaben auffassen lässt (Spearman's  $g$ ; Spearman, 1927). Es gibt vielfältige empirische Hinweise darauf, dass die allgemeine kognitive Leistungsfähigkeit eng mit der Reaktionszeit in einfachen und komplexen Wahlaufgaben zusammenhängt (für einen Überblick s. z.B. Jensen, 2006, Kap. 9), darunter auch Aufgaben mit sprachlichen Stimuli (z.B. Satzverifikationsaufgaben, Jensen, Larson & Paul, 1988), und es gibt Theorien, die individuelle Unterschiede in allgemeiner kognitiver Leistungsfähigkeit mit individuellen Unterschiede in der Informationsverarbeitungsgeschwindigkeit zu erklären versuchen (z.B. Jensen, 1993; Lehl & Fischer, 1988; Roth, 1964). Vor dem Hintergrund dieser Arbeiten liegt die Annahme nahe, dass ein Teil der Variabilität in den Reaktionszeiten, die mit den ProDi-L-Subtests erhoben werden, auf die allgemeine kognitive Leistungsfähigkeit im Sinne individueller Unterschiede in der Informationsverarbeitungsgeschwindigkeit zurückzuführen ist. Weitere Forschung sollte sich dieser Frage auch deshalb widmen, weil sie möglicherweise Implikationen bezüglich der Trainierbarkeit der Effizienz von Teilprozessen des Leseverstehens hat. Grundsätzlich lassen sich Teilprozesse des Leseverstehens durch Lesepraxis in beträchtlichem Maße routinisieren. Es wäre aber z.B. denkbar, dass die Geschwindigkeit, mit der die Routinisierung fortschreitet, und der insgesamt erreichbare Routinisierungsgrad von der allgemeinen kognitiven Leistungsfähigkeit abhängen.

Zwei der Subtests von ProDi-L (Phonologische Rekodierung und Zugriff auf Wortbedeutungen) beanspruchen sowohl das Lese- und das Hörverstehen, da der jeweilige Referenzstimulus, mit dem das geschriebene Wort oder Pseudowort verglichen wird, auditorisch vorgegeben wird. Dadurch besteht prinzipiell die Gefahr, dass mit diesen beiden ProDi-L-Subtests

Lesefähigkeiten nicht in reiner Form, sondern vermischt mit Fähigkeiten des Hörverstehens erfasst werden. Allerdings wurde bei der Konstruktion der Aufgaben Sorge dafür getragen, dass solche konfundierenden Effekte möglichst minimiert werden. Vor allem wird der auditorische Referenzstimulus immer mit ausreichendem zeitlichen Abstand vor dem fokalen schriftlichen Stimulus präsentiert, auf den sich das Urteil bezieht. Demnach sollten die Testpersonen die Konstruktion der mentalen Repräsentation des Referenzstimulus in der Regel abgeschlossen haben, bevor sie ihr Urteil fällen. Verzerrende Einflüsse sind gleichwohl nicht auszuschließen, weshalb diese Frage in weiteren Untersuchungen empirisch geklärt werden sollte. Dazu kann ein in unserem Projekt neu konstruiertes Instrument zur prozessbezogenen Diagnostik von Teilfähigkeiten des Hörverstehens bei Grundschulkindern (ProDi-H) genutzt werden, das strukturell analog zu ProDi-L aufgebaut ist und sechs Subtests enthält, deren Testitems in relevanten psycholinguistischen Eigenschaften mit den Testitems in ProDi-L parallelisiert sind. Beide Instrumente – ProDi-L für das Leseverstehen und ProDi-H für das Hörverstehen – werden derzeit gemeinsam in einer Längsschnittuntersuchung eingesetzt, in der Grundschulkindern vom Beginn der ersten bis zum Ende der vierten Klasse zu insgesamt fünf Messzeitpunkten getestet werden. Diese Untersuchung soll Aufschlüsse über die Entwicklungssensitivität der Subtests von ProDi-L und ihre prognostische Validität im Hinblick auf Lesefähigkeiten und Schulleistungen am Ende der Grundschulzeit geben. Ein weiteres Ziel ist die Untersuchung von Interdependenzen in der Entwicklungsdynamik von Fähigkeiten des Hör- und Leseverstehens im Laufe der Grundschulzeit.

Viele der am Leseverstehen beteiligten Teilprozesse sind auch für das Verstehen gesprochen-sprachlicher Äußerungen relevant. Gemäß der sogenannten einfachen Sichtweise auf das Lesen (simple view of reading, Gough & Tunmer, 1986) ist Lesefähigkeit identisch mit dem Produkt der Fähigkeit, gesprochene Sprache zu verstehen, und der Fähigkeit zur visuellen Worterkennung (decoding). Aus der Perspektive der meisten Ansätze in der kognitiven und pädagogisch-psychologischen Lesefähigkeitsforschung stellt diese Annahme allerdings eine zu starke Vereinfachung dar. In alltäglichen (Lern-) Situationen unterscheiden sich Lese- und Hörverstehen z.B. durch den Inhalt, die Länge und die sprachliche Komplexität von schriftlich bzw. mündlich kommunizierten Botschaften, den Grad der situativen Einbettung dieser

Botschaften, den Anteil strategischer Verstehensprozesse und die Anforderungen an Arbeitsgedächtnis und Vorwissen. Entsprechend sind Lese- und Hörverstehen in den meisten Untersuchungen zwar substantiell, aber keineswegs perfekt miteinander korreliert (ca. 50% gemeinsame Varianz, vgl. Sticht & James, 1984). Dennoch sind die strukturellen Ähnlichkeiten zwischen Lese- und Hörverstehen so weitgehend, dass eine genauere Analyse der Beziehungen zwischen interindividuellen Unterschieden in der Effizienz der beteiligten Teilprozesse und ihrer Entwicklung sowohl in theoretischer als auch in praktischer Hinsicht sehr lohnenswert scheint. Dabei wäre z.B. denkbar, dass Fähigkeiten des Hörverstehens zu Beginn der Grundschulzeit wichtige Prädiktoren der Entwicklung korrespondierender Teilfähigkeiten des Leseverstehens darstellen, während sich im Laufe der Leseentwicklung die individuellen Profile von Teilfähigkeiten des Lese- und Hörverstehens immer weiter angleichen (vgl. das holistische Modell von Rost & Hartmann, 1992).

Zur weitergehenden Validierung von ProDi-L sind Analysen der Effekte von Itemmerkmalen auf die Itemschwierigkeiten (mittlere Reaktionszeit/Anteil richtiger Antworten) in explanatorischen Item-Response-Modellen vorgesehen. Wenn sich in diesen Analysen zeigen lässt, dass Itemmerkmale, die aus psycholinguistischer Sicht die angezielten Teilprozesse des Leseverstehens erleichtern oder erschweren sollten, tatsächlich Unterschiede in den Itemschwierigkeiten erklären können, wäre dies ein starker Beleg für die Konstruktvalidität der einzelnen Subtests (vgl. Richter & Naumann, 2009). Schließlich verwenden wir ProDi-L dazu, vermittelnde Prozesse der Wirksamkeit von drei verschiedenen Interventionsmethoden (Phonics-, Fluency- und Strategietraining) zur Förderung des Leseverständnisses bei leseschwachen Grundschulkindern zu identifizieren (Müller, Richter, Križan, Hecht & Ennemoser, in Druck). Die hier vorgelegten Ergebnisse stimmen uns zuversichtlich, dass sich ProDi-L auch in diesen Untersuchungen bewähren wird.

### Literatur

- Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory and Cognition*, *10*, 565-575. doi:10.3758/BF03202439
- Andrews, S. (1989). Frequency and neighborhood size effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 802-814. doi:10.1037/0278-7393.15.5.802
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium.
- Balota, D.A., Cortese, M.J., Sergent-Marshall, S.D. & Yap, M.J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283-316. doi:10.1037/0096-3445.133.2.283
- Cattell, R.B., Weiß, R.H. & Osterland, J. (1997). *Grundintelligenztest Skala I* (5., überarb. Aufl.). Göttingen: Hogrefe.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204-256. doi:10.1037/0033-295X.108.1.204
- Cunningham, A.E. & Stanovich, K.E. (1990). Assessing print exposure and orthographic processing skill in children. *Journal of Educational Psychology*, *82*, 733-740. doi:10.1037/0022-0663.82.4.733
- Deno, S. L., Mirkin, P. & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children*, *49*, 36-45.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance, Vol. 12* (pp. 559-586). Hove, UK: Erlbaum.
- Frisson, S., Rayner, K. & Pickering, M.J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 862-877. doi:10.1037/0278-7393.31.5.862
- Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia*, *36*, 69-81.

doi:10.1007/BF02648022

- Frith, U., Wimmer, H. & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, 2, 31-54.  
doi:10.1207/s1532799xssr0201\_2
- Garrod, S.C. & Sanford, A.J. (1990). Referential processes in reading: Focusing on roles and individuals. In G.B. Flores d'Arcais, K. Rayner & D. Balota (Eds.), *Comprehension processes in reading* (pp. 515-533). Hillsdale, NJ: Erlbaum.
- Gernsbacher, M.A. & Faust, M.E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 245-262. doi:10.1037/0278-7393.17.2.245
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-76.
- Gough, P. B. & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6-10.
- Graesser, A.C., Hoffman, N.L. & Clark, L.F. (1980). Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior*, 19, 135-151.
- Hale, A.D., Henning, J.B., Hawkins, R.O., Sheeley, W., Shoemaker, L., Reynolds, J.R. & Moch, C. (2011). Reading assessment methods for middle-school students: An investigation of reading comprehension rate and Maze accurate response rate. *Psychology in the Schools*, 48, 28-26.  
doi: 10.1002/pits.20544
- Hannon, B. & Daneman, B. (2004). Shallow semantic processing of text: An individual-differences account. *Discourse Processes*, 37, 187-204. doi:10.1207/s15326950dp3703\_1
- Hau, K.T. & Marsh, H.W. (2004). The use of item parcels in structural equation modelling: Non-normal data and small sample sizes. *British Journal of Mathematical Statistical Psychology*, 57, 327-351.  
doi:10.1111/j.2044-8317.2004.tb00142.x
- Haviland, S.E. & Clark, H.H. (1974). What's new? Acquiring new information as a process of comprehension. *Journal of Verbal Learning and Verbal Behavior*, 13, 512-521.  
doi:10.1016/S0022-5371(74)80003-4
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of*

*Statistics*, 6, 65–70.

- Horn, J.L. & McArdle, J.J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 117–144. doi: 10.1080/03610739208253916
- Hornberg, S., Valtin, R., Potthoff, B., Schwippert, K. & Schulz-Zander, R. (2009). Lesekompetenzen von Mädchen und Jungen im internationalen Vergleich. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 195-224). Münster: Waxmann.
- Huntsman, L.A. & Lima, S.D. (2002). Orthographic neighbors and visual word recognition. *Journal of Psycholinguistic Research*, 31, 289-306. doi:10.1023/A:1015544213366
- Jensen, A. R. (1993). Why is reaction time correlated with psychometric g? *Current Directions in Psychological Science*, 2, 53-56. DOI: 10.1111/1467-8721.ep10770697
- Jensen, A.R., Larson, J. & Paul, S.M. (1988). Psychometric g and mental processing speed on a semantic verification test. *Personality and Individual Differences*, 9, 243-255.
- Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago, IL: Scientific Software International.
- Kintsch, W. & Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, 5, 257-274. doi:10.1016/0010-0285(73)90036-4
- Karing, C., Matthäi, J. & Artelt, C. (2011). Genauigkeit von Lehrerurteilen über die Lesekompetenz ihrer Schülerinnen und Schüler in der Sekundarstufe I – Eine Frage der Spezifität? *Zeitschrift für Pädagogische Psychologie*, 25, 159-172. doi: 10.1024/1010-0652/a000041
- Küspert, P. & Schneider, W. (1998). *Würzburger Leise Leseprobe*. Göttingen: Hogrefe.
- Lehrl, S. & Fischer, B. (1988). The basic parameters of human information processing: Their role in the determination of intelligence. *Personality and Individual Differences*, 9, 883-896.
- Lenhard, W. & Schneider, W. (2006). *ELFE 1-6: Ein Leseverständnistest für Erst- bis Sechstklässer*. Göttingen, Germany: Hogrefe.
- Lenhard, W. & Schneider, W. (Hrsg.) (2009). *Diagnostik und Förderung des Leseverständnisses*.

Göttingen: Hogrefe.

- Lorenz, C. & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Zeitschrift für Pädagogische Psychologie*, 23, 211-222. doi: 10.1024/1010-0652.23.34.211
- Mayring, H. & Wimmer, H. (2003). *Salzburger Lese-Screening für die Klassenstufen 1-4*. Göttingen: Hogrefe.
- McRae, K., Spivey-Knowlton, M.J. & Tanenhaus, M.K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283-312. doi:10.1006/jmla.1997.2543
- Murray, J.D. & Burke, K.A. (2003). Activation and encoding of predictive inferences: The role of reading skill. *Discourse Processes*, 35, 81-102. doi:10.1207/S15326950DP3502\_1
- Müller, B., Richter, T., Križan, A., Hecht, T., & Ennemoser, M. (in press). Evidenzbasierte Leseförderung: Vorstellung einer Interventionsstudie. *Diskurs Kindheits- und Jugendforschung*.
- Muthén, L.K. & Muthén, B.O. (1998-2006). *MPlus user's guide* (4<sup>th</sup> ed.). Los Angeles, CA: Muthén & Muthén.
- Nation, K. & Snowling, M.J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, 39, 85-101. doi:10.1006/jmla.1998.2564
- Naumann, J., Artelt, C., Schneider, W. & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.). *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23-71). Münster: Waxmann.
- Naumann, J., Richter, T., Christmann, U. & Groeben, N. (2008). Working memory capacity and reading skill moderate the effectiveness of strategy training in learning from hypertext. *Learning and Individual Differences*, 18, 197-213. doi:10.1016/j.lindif.2007.08.007
- Naumann, J., Richter, T., Flender, J., Christmann, U. & Groeben, N. (2007). Signaling in expository hypertexts compensates for deficits in reading skill. *Journal of Educational Psychology*, 99, 791-807. doi: 10.1037/0022-0663.99.4.791
- Newell, A. & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J.R.

- Anderson (Ed), *Cognitive skills and their acquisition* (pp. 1-55) Hillsdale, NJ: Erlbaum.
- Oakhill, J.V., Cain, K. & Bryant, P.E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes*, 18, 443-468.  
doi:10.1080/01690960344000008
- Perfetti, C.A. (1985). *Reading ability*. New York: Academic Press.
- Perfetti, C. A. (1999). Comprehending written language: A blueprint of the reader. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 167-208). Oxford University Press.
- Perfetti, C.A. & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam: John Benjamin.
- Pickering, M.J. & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11, 105-110. doi:0.1016/j.tics.2006.12.002
- Pickering, M.J. & van Gompel, R.P.G. (2006). Syntactic parsing. In M.J. Traxler & M.A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2<sup>nd</sup> ed., pp. 455-503). London: Academic Press.
- Paap, K.R. & Noel, R.W. (1991). Dual route models of print to sound: Still a good horse race. *Psychological Research*, 53, 13-24.
- Projekt Deutscher Wortschatz (2007). *Deutscher Wortschatz – Portal* [Internet-Datenbank]. Verfügbar unter <http://corpora.uni-leipzig.de/> (Zugriff am 20.09.2011)
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510-532. doi: 10.1037/0033-2909.114.3.510
- Richter, T. & Christmann, U. (2009). Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In N. Groeben & B. Hurrelmann (Hrsg.), *Lesekompetenz: Bedingungen, Dimensionen, Funktionen* (3. Aufl., S. 25-58). Weinheim: Juventa.
- Richter, T. & Naumann, J. (2009). Was misst der ELVES-Subtest Satzverifikation? Analysen von Mess- und Itemeigenschaften mit hierarchisch-linearen Modellen. In W. Schneider, H. Marx & W. Lenhard (Hrsg.), *Diagnose und Förderung des Leseverständnisses* (Tests und Trends, N.F. Bd. 7, S. 131-150). Göttingen: Hogrefe.
- Richter, T. & van Holt, N. (2005). ELVES: Ein computergestütztes Diagnostikum zur Erfassung der Effizienz von Teilprozessen des Leseverstehens. *Diagnostica*, 51, 169-182. doi:10.1026/0012-

1924.51.4.169

- Richter, T., Schroeder, S. & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, *96*, 538-558. doi:10.1037/a0014038
- Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*, 532-547.
- Roskam, E.E. (1997). Models for speed and time-limited tests. In W. van der Linden & R.K. Hambleton (Hrsg.), *Handbook of modern item response theory* (S. 187-208). New York: Springer.
- Rost, D.H. (1989). Reading comprehension: Skill or skills? *Journal of Research in Reading*, *12*, 87-113. doi: 10.1111/j.1467-9817.1989.tb00160.x
- Rost, D.H. & Hartmann, A. (1992). Lesen, Hören, Verstehen. *Zeitschrift für Psychologie*, *200*, 345-361.
- Roth, E. (1964). Die Geschwindigkeit der Verarbeitung von Information und ihr Zusammenhang mit Intelligenz. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *11*, 616-622.
- Sanders, T.J.M., Spooren, W.P.M. & Noordman, L.G.M. (1992). Towards a taxonomy of coherence relations. *Discourse Processes*, *15*, 1-35. doi:10.1080/01638539209544800
- Schneider, W., Schlagmüller, M. & Ennemoser, M. (2007). *LGVT 6-12: Lesegeschwindigkeits- und verständnistest für die Klassen 6-12*. Göttingen: Hogrefe.
- Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time on comprehension with and without text availability. *Journal of Educational Psychology*, *103*, 877-896. doi: 10.1037/a0023731
- Singer, M. (1993). Causal bridging inferences: Validating consistent and inconsistent sequences. *Canadian Journal of Experimental Psychology*, *47*, 340-359. doi:10.1037/h0078825
- Spearman, C.E. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.
- Stanovich, K.E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*, 32-71. doi:10.2307/747348
- Stanovich, K.E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological core variable difference model. *Journal of Learning Disabilities*, *21*, 590-604. doi:10.1177/002221948802101003
- Sticht, T. G. & James, J. H. (1984). Listening and reading. In P. D. Pearson, R. Barr, M. L. Kamil & P.

- Mosenthal (Eds.), *Handbook of reading research* (Vol. 1, pp. 293–317). White Plains, NY: Longman.
- Südkamp, A., Kaiser, J. & Möller, J. (2012, March 26). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*. Advance online publication. doi: 10.1037/a0027627
- Thompson-Schill, S.L. & Gabrieli, J.D.E. (1999). Priming of visual and functional knowledge on a semantic classification task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 41-53. doi:10.1037/0278-7393.25.1.41
- Ulrich, R. & Miller, J. (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology*, 37, 513-525.
- van Dijk, T.A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Vellutino, F.R., Fletcher, J.M., Snowling, M.J. & Scanlon, D.M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, 45, 2-40. doi:10.1046/j.0021-9630.2003.00305.x
- Venezky, R. (1979). Orthographic regularities in English words. In P. Kolars, M. Wolstrad & P. Bouma (Eds.), *Processing of visual language* (pp. 283-293). New York: Plenum Press.
- Weiß, R.H. (2008). *Grundintelligenztest Skala 2 Revision (CFT 20-R)*. Göttingen: Hogrefe.
- West, S.G., Finch, J.F. & Curran, P.J. (1995). Structural equation modeling with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 37-55). Thousand Oakes, CA: Sage.
- Wimmer, H. (1996). The nonword reading deficit in developmental dyslexia: Evidence from children learning to read German. *Journal of Experimental Child Psychology*, 61, 80-90. doi:10.1006/jecp.1996.0004
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Zwaan, R.A. & Radvansky, G.A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185. doi:10.1037/0033-2909.123.2.162

Tabelle 1:

*Größe der Teilstichproben (Klasse 1-4) für die Subtests von ProDi-L*

	Klasse 1	Klasse 2	Klasse 3	Klasse 4	Gesamt
Phonologische Rekodierung	66	124	166	172	528
Orthographischer Vergleich	66	124	166	172	528
Zugriff auf Wortbedeutungen	62	103	131	142	438
Syntaktische Integration	—	—	162	171	333
Semantische Integration	—	123	149	170	442
Lokale Kohärenzbildung	—	—	121	145	266

Tabelle 2:

Deskriptive Statistiken, Korrelationen (unterhalb der Hauptdiagonalen), Partialkorrelationen nach Auspartialisierung der Klassenstufe (oberhalb der Hauptdiagonalen) und Reliabilitätsschätzer (Cronbachs  $\alpha$ , auf der Hauptdiagonalen) für die Subtests von ProDi-L (Anteil richtiger Antworten und logarithmisch transformierte Reaktionszeit)

	Skalenkennwerte						Korrelationen der Testwerte											
	<i>M</i> ( <i>SD</i> )	<i>M</i> <sub>Jungen</sub> ( <i>SD</i> )	<i>M</i> <sub>Mädchen</sub> ( <i>SD</i> )	$\bar{r}_{it}$	<i>W</i>	<i>Schiefe</i> ( <i>SE</i> )	1	2	3	4	5	6	7	8	9	10	11	12
<i>Anteil richtiger Antworten</i>																		
1 Phonologische Rekodierung	.83 (.13)	.82 (.13)	.85 (.13)	.25	.27- 1.00	-1.34 (0.11)	(.81)	.51*	.42*	.39*	.41*	.39*	.38*	.12	.14	.05	.09	.22*
2 Orthographischer Vergleich	.83 (.12)	.83 (.12)	.83 (.12)	.32	.42- 1.00	-1.00 (0.11)	.56*	(.94)	.45*	.63*	.67*	.47*	.11	.04	.04	-.01	.16*	.42*
3 Zugriff auf Wortbedeutungen	.84 (.14)	.83 (.14)	.85 (.14)	.25	.33- 1.00	-1.27 (0.12)	.48*	.60*	(.70)	.42*	.48*	.33*	.19*	.13	.34*	.10	.10	.19
4 Syntaktische Integration	.89 (.09)	.88 <sup>a</sup> (.11)	.91 <sup>a</sup> (.08)	.45	.50- 1.00	-1.67 (0.13)	.37*	.53*	.39*	(.91)	.36*	.36*	.22*	-.01	.10	.38*	.40*	.56*
5 Semantische Integration	.91 (.10)	.89 (.12)	.92 (.09)	.26	.44- 1.00	-2.28 (0.12)	.42*	.57*	.46*	.60*	(.74)	.39*	.23*	.23*	.16	.40*	.59*	.65*
6 Lokale Kohärenzbildung	.74 (.18)	.72 <sup>a</sup> (.71)	.77 <sup>a</sup> (.18)	.16	.02- 1.00	-1.33 (0.15)	.38*	.41*	.31*	.36*	.39*	(.55)	.00	-.09	-.06	.05	.10	.32*
<i>Reaktionszeit (log.)</i>																		
7 Phonologische Rekodierung	7.36 (0.38)	7.32 (0.39)	7.40 (0.35)	.65	5.17- 8.23	-1.10 (0.11)	.26*	-.10	.03	.20*	.20*	-.03	(.98)	.65*	.57*	.63*	.49*	.39*
8 Orthographischer Vergleich	7.35 (0.51)	7.28 (0.50)	7.40 (0.49)	.73	4.42- 9.01	-0.55 (0.11)	-.02	-.22*	-.08	.10	.16*	-.12	.69*	(.99)	.64*	.82*	.69*	.40*
9 Zugriff auf Wortbedeutungen	7.28 (0.52)	7.23 (0.52)	7.30 (0.51)	.73	4.36- 8.83	-1.20 (0.12)	.05	-.12	.19*	.10	.12	-.10	.60*	.67*	(.98)	.56*	.52*	.38*
10 Syntaktische Integration	8.15 (0.43)	8.08 <sup>a</sup> (0.44)	8.22 <sup>a</sup> (0.40)	.70	5.55- 9.60	-0.90 (0.13)	.02	-.05	.06	.34*	.36*	-.02	.62*	.78*	.57*	(.98)	.70*	.59*
11 Semantische Integration	8.13 (0.56)	8.04 <sup>a</sup> (0.65)	8.20 <sup>a</sup> (0.47)	.72	4.87- 9.35	-1.94 (0.12)	.02	.02	.02	.32*	.52*	.07	.50*	.69*	.55*	.69*	(.98)	.68*
12 Lokale Kohärenzbildung	7.86 (0.50)	7.79 (0.53)	7.91 (0.46)	.70	5.94- 8.93	-1.69 (0.15)	.20*	.33*	.17	.55*	.63*	.28*	.39*	.39*	.39*	.60*	.67*	(.99)

Anmerkung.  $\bar{r}_{it}$  = mittlere Trennschärfe (geometrisches Mittel) der Items des jeweiligen Tests. *W*: Spannbreite.

<sup>a</sup> Mittelwerte von Jungen ( $161 > n < 244$ ) und Mädchen ( $166 > n < 257$ ) unterscheiden sich paarweise auf einem Signifikanzniveau von  $p < .05$  (zweiseitig, Holm-Bonferroni-Korrektur für 12 simultane Tests; Holm, 1979).

\*  $p < .05$  (zweiseitig, Holm-Bonferroni-Korrektur für 66 simultane Tests)

Tabelle 3:

*Partialkorrelationen der Subtests von ProDi-L mit kriterialen Lesefähigkeitsmaßen und Intelligenz nach Auspartialisierung der Klassenstufe*

	Urteile der Deutsch-Lehrkraft				ELFE 1-6				Intelligenz	
	Texte laut	Text-	Teilfähig-	Deutsch-	Lesese-	Wort-	Satzver-	Textver-	CFT 1 <sup>b</sup>	CFT 20-R <sup>c</sup>
	lesen	verständnis	keiten	note <sup>a</sup>	schwindigkeit	verständnis	ständnis	ständnis		
1 Phonologische Rekodierung (ACC)	.36*	.36*	.39*	-.37	-.39*	.20*	.35*	.44*	.27	.08
2 Orthographischer Vergleich (ACC)	.44*	.46*	.45*	-.60	-.54*	.46*	.61*	.56*	.28	.04
3 Zugriff auf Wortbedeutungen (ACC)	.33*	.38*	.47*	-.40	-.42*	.23*	.33*	.39*	.32	.04
4 Syntaktische Integration (ACC)	.36*	.35*	.33*	-.40	-.47*	.26*	.56*	.44*	–	.02
5 Semantische Integration (ACC)	.24*	.26*	.21*	-.41	-.40*	.23*	.56*	.47*	.24	.01
6 Lokale Kohärenzbildung (ACC)	.34*	.36*	.27*	-.38	-.37*	.34*	.50*	.47*	–	.36*
7 Phonologische Rekodierung (RT)	-.16	-.05	-.07	.01	-.02	-.19	-.01	.03	.02	.11
8 Orthographischer Vergleich (RT)	-.13	-.11	-.15	.19	.29*	-.49*	-.36*	-.22*	-.09	.07
9 Zugriff auf Wortbedeutungen (RT)	.00	-.03	-.03	.10	.20	-.55*	-.44*	-.26*	-.07	-.04
10 Syntaktische Integration (RT)	-.16	-.16	-.13	.06	.26*	-.64*	-.38*	-.22	–	.02
11 Semantische Integration (RT)	-.06	-.03	-.04	-.09	.16	-.53*	-.22*	-.16	.06	.03
12 Lokale Kohärenzbildung (RT)	.15	.10	.01	-.21	-.25	.03	.39*	.41*	–	.05
<i>M (SD)</i>	3.95(0.98)	3.93 (1.02)	4.07 (0.86)	2.19 (0.78)	477 (418)	39.29 (13.68)	16.58 (6.58)	10.60 (15.12)	9.09 (1.80)	10.61 (9.63)
<i>n</i>	415	411	248	297	278	278	270	210	66	180

*Anmerkung.* ACC: Akkuratheit (Anteil richtiger Antworten). RT: Reaktionszeit in ms (log. transformiert). <sup>a</sup> Nur Daten aus den Klassen 3 ( $n=144$ ) und 4 ( $n=144$ ). <sup>b</sup> Mittelwert der Untertests 3 und 5. Nur Daten aus Klasse 2 ( $n=66$ ). <sup>c</sup> Mittelwert der Untertests 1 und 3. Nur Daten aus den Klassen 3 ( $n=92$ ) und 4 ( $n=88$ ).

\*  $p < .05$  (zweiseitig, Holm-Bonferroni-Korrektur für 12 simultane Tests).

Tabelle 4:

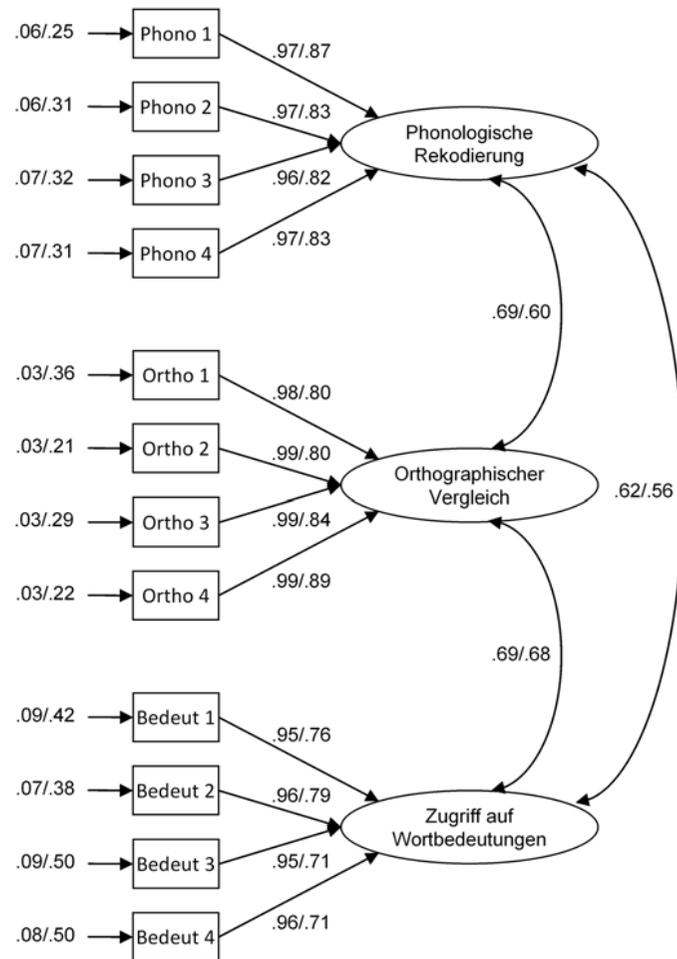
*Moderierte Regressionen mit Antwortrichtigkeit (z-standardisiert) und log. Reaktionszeit (z-standardisiert) der Subtests von ProDi-L als Prädiktoren und Textverständnis (ELFE 1-6) als Kriterium*

	<i>B</i>	<i>SE<sub>B</sub></i>	$\Delta R^2$
<i>Phonologische Rekodierung<sup>a</sup></i>			
Additive Konstante	10.96	0.30	
Antwortrichtigkeit	2.75*	0.31	.24
RT (log.)	-2.14*	0.38	.10
Antwortrichtigkeit X Reaktionszeit	-0.82*	0.25	.03
<i>Orthographischer Vergleich<sup>b</sup></i>			
Additive Konstante	10.63	0.26	
Antwortrichtigkeit	3.11*	0.32	.25
RT (log.)	-1.30*	0.40	.03
Antwortrichtigkeit X Reaktionszeit	0.10	0.25	.00
<i>Zugriff auf Wortbedeutungen<sup>c</sup></i>			
Additive Konstante	10.72	0.29	
Antwortrichtigkeit	2.30*	0.32	.17
RT (log.)	-2.65*	0.42	.13
Antwortrichtigkeit X Reaktionszeit	-0.47*	0.25	.01
<i>Syntaktische Integration<sup>d</sup></i>			
Additive Konstante	12.62	0.40	
Antwortrichtigkeit	2.26*	0.45	.15
RT (log.)	-2.59*	0.51	.15
Antwortrichtigkeit X Reaktionszeit	-0.49 *	0.23	.03
<i>Semantische Integration<sup>e</sup></i>			
Additive Konstante	11.48	0.29	
Antwortrichtigkeit	2.87*	0.37	.18
RT (log.)	-3.30*	0.43	.18
Antwortrichtigkeit X Reaktionszeit	-0.46*	0.20	.02
<i>Lokale Kohärenzbildung<sup>f</sup></i>			
Additive Konstante	13.32	0.50	
Antwortrichtigkeit	2.01*	0.51	.14
RT (log.)	0.94	0.79	.01
Antwortrichtigkeit X Reaktionszeit	-0.41	0.71	.00

*Anmerkung.* RT: Reaktionszeit in ms (logarithmisch transformiert). Antwortrichtigkeit: Anteil richtiger Antworten. Modellgüte:

<sup>a</sup> $R^2=.35$ ,  $F(3,206)=36.89$ ,  $p<.001$ , <sup>b</sup> $R^2=.47$ ,  $F(3,206)=59.86$ ,  $p<.001$ , <sup>c</sup> $R^2=.37$ ,  $F(3,199)=39.28$ ,  $p<.001$ , <sup>d</sup> $R^2=.33$ ,  $F(3,116)=18.88$ ,  $p<.001$ , <sup>e</sup> $R^2=.45$ ,  $F(3,185)=50.29$ ,  $p<.001$ , <sup>f</sup> $R^2=.26$ ,  $F(3,84)=9.85$ ,  $p<.001$ . \*  $p < .05$  (zweiseitig).

a)



b)

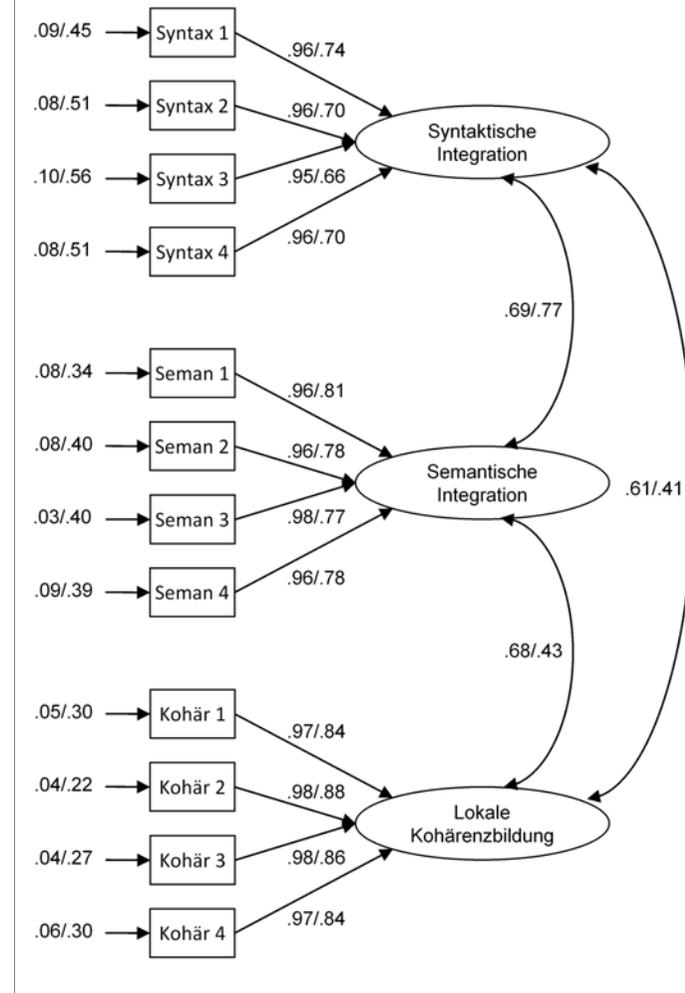


Abbildung 1. Parameterschätzungen des konfirmatorischen Faktormodells (standardisierte Lösung) für (a) die Subtests von ProDi-L zur Erfassung der Effizienz von kognitiven Teilprozessen bei der Worterkennung (Phonologische Rekodierung, Orthographischer Vergleich, Zugriff auf Wortbedeutungen) und (b) die Subtests zur Erfassung der Effizienz von kognitiven Teilprozessen des Leseverstehens auf der Ebene von Sätzen und Satzfolgen (Syntaktische Integration, Semantische Integration, Lokale Kohärenzbildung). Die Faktorladungen der Itempäckchen für jede der drei latenten Variablen sind gleichgesetzt. Der erste der jeweils angegebenen Parameter bezieht sich auf das Modell für Antwortrichtigkeit, der zweite auf das Modell für die logarithmierte Reaktionszeit.

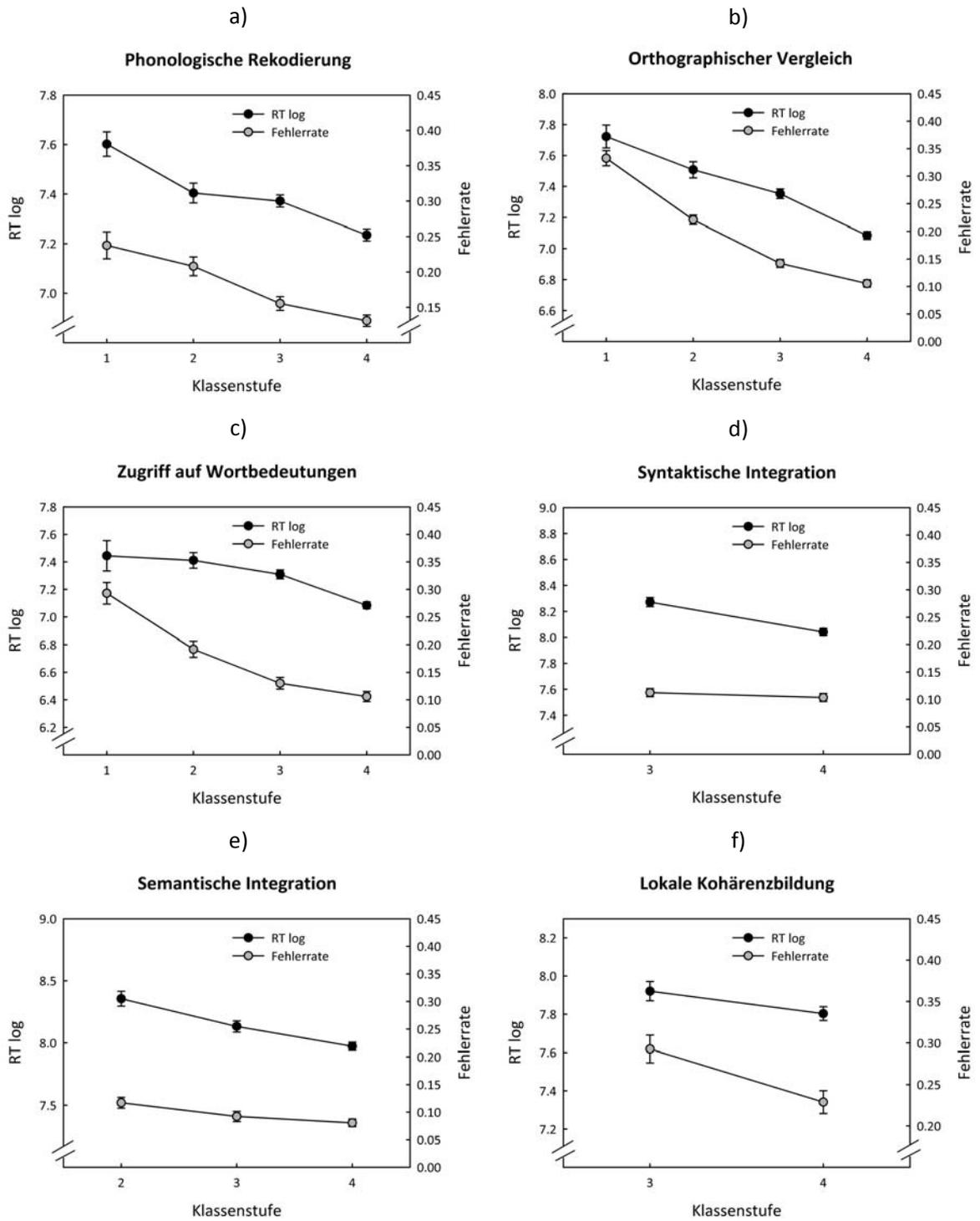


Abbildung 2. Unterschiede zwischen Klassenstufen in den sechs Prodi-L-Subtests. Für jeden Subtest ist die Variable Antwortrichtigkeit als Fehlerrate (1 - Anteil richtiger Antworten) dargestellt. Fehlerbalken repräsentieren den Standardfehler des Mittelwerts.