Kolumnentitel: COMPUTERGESTÜTZTES DIAGNOSTIKUM DES
LESEVERSTEHENS

ELVES: Ein computergestütztes Diagnostikum zur Erfassung der Effizienz von
Teilprozessen des Leseverstehens
Tobias Richter und Nadine van Holt
Universität zu Köln

erschieden in: *Diagnostica*, 51, 169-182.

Zusammenfassung

Lesekompetenz beruht auf der effizienten Bewältigung von Teilprozessen des Leseverstehens, wie sie in hierarchischen Modellen der Textverarbeitung unterschieden werden. ELVES (Effizienz des Leseverstehens bei Erwachsenen nach dem Strategiemodell) ist ein computergestütztes Instrument zur Messung von Lesefähigkeiten bei geübten erwachsenen Lesern/innen. Die 7 Subtests des Instruments erfassen über die simultane Berücksichtigung von Antwortrichtigkeit und Antwortlatenz die Effizienz hierarchieniedriger (mikrostruktureller) und hierarchiehoher (makrostruktureller) Teilprozesse des Lesens nach van Dijk und Kintsch (1983). Vier Untersuchungen mit studentischen Probanden/innen (N=232) dienten einer ersten Konstrukt- und Kriteriumsvalidierung. Anhand der Skaleninterkorrelationen und der Ergebnisse einer konfirmatorischen Faktorenanalyse ließ sich die angenommene Trennung hierarchieniedriger und hierarchiehoher Teilfähigkeiten belegen. Die Subtests des Instruments korrelierten mit der Leistung in klassischen Leseverständnistests aus zwei verschiedenen Inhaltsbereichen und mit Indikatoren der Situationsmodellgüte aus einer naturalistischen Leseaufgabe. Anhand von Testtexten zu verschiedenen Themen konnte gezeigt werden, daß die hierarchiehohen ELVES-Subtests domänenübergreifende Lesefähigkeitskomponenten erfassen, die nicht auf inhaltliches Vorwissen reduzierbar sind.

Schlagnworte: Leistungstest - Leseverständnis - Lesefähigkeit - Strategiemodell

Abstract

Reading ability comprises the efficient mastery of component processes of reading, as distinguished in hierarchical models of text comprehension. ELVES (efficiency of reading comprehension in adult readers according to the strategy model) is a computer-based instrument for the measurement of reading skills in trained adult readers. By combining accuracy and latency of responses, the 7 subtests of the instrument assess the efficiency of lower (microstructural) and higher (macrostructural) processes of reading (van Dijk & Kintsch, 1983). Four studies with student participants ($N=232$) provided evidence for the construct and criterion validity. Scale intercorrelations and confirmatory factor analyses corroborated the distinction of micro- and macrostructural component skills. The instrument's subtests were correlated with performance in traditional reading tests from two content domains, as well as with situation model quality in a naturalistic reading task. A study based on test texts on different topics demonstrated that the higher ELVES subtests capture general reading skills which cannot be reduced to domain-specific knowledge.

Key words: performance test - reading comprehension - reading ability - strategy model

ELVES: Ein computergestütztes Diagnostikum zur Erfassung der Effizienz von
Teilprozessen des Leseverstehens

Die kognitionspsychologische Lesekompetenzforschung der letzten drei Jahrzehnte hat gezeigt, daß die Fähigkeit, schriftlichen Texten Informationen zu entnehmen, auf der Bewältigung einer Reihe funktional getrennter, hierarchisch organisierter Teilprozesse beruht. Auf allen Verarbeitungsebenen - von okulomotorischen Prozessen bis hin zu globalen, wissensbasierten Prozessen auf Textebene - treten interindividuelle Differenzen auf. Nach heutigem Kenntnisstand kommen jedoch bei Lesern/innen ohne Lese-Rechtschreib-Schwächen lediglich Prozesse auf linguistisch beschreibbaren Ebenen für eine ursächliche Erklärung des insgesamt erreichbaren Leseverständnisses in Frage, wobei einzelne Prozeßkomponenten zum Teil additiv, zum Teil kompensatorisch zusammenwirken (vgl. Daneman, 1997; Perfetti, 1994). Eine Vielzahl von Studien hat inzwischen gezeigt, daß selbst bei geübten Lesern/innen mit einem hohen Bildungshintergrund beträchtliche Unterschiede im allgemeinen Lesefähigkeitsniveau, in der Lesegeschwindigkeit sowie in der Bewältigung von Teilprozessen auftreten (z. B. Schüler/innen an Gymnasien oder Studierende; für ein aktuelles Beispiel vgl. Artelt, Stanat, Schneider & Schiefele, 2001; zu interindividuellen Unterschieden in Teilprozessen vgl. Richter & Christmann, 2002).

Untersuchungen im Rahmen der psycholinguistischen und pädagogisch-psychologischen Grundlagenforschung zum Textverstehen greifen in aller Regel auf Studierendestichproben zurück. Für viele Forschungsfragen auch in diesem Bereich wäre es sinnvoll, interindividuelle Differenzen in der Bewältigung von Teilprozessen des Lesens in die Untersuchung einzubeziehen, sei es aus methodischen Gründen (zur Kontrolle von Fehlervarianz) oder aus einem bestimmten theoretischen Interesse. So sind Fortschritte bezüglich der Frage, unter welchen Bedingungen ein mehr oder weniger reichhaltiges Situationsmodell gebildet wird, durch eine Erweiterung der allgemeinpsychologischen durch eine differentiellpsychologische Perspektive zu erwarten (Rinck, 2000). Neben allgemeinen

Aspekten kognitiver Leistungsfähigkeit wie dem visuellen Vorstellungsvermögen oder der Arbeitsgedächtniskapazität spielen hier auch lesespezifische Fähigkeiten eine Rolle.

In diesem Beitrag stellen wir das computergestützte (Forschungs-) Instrument ELVES (Effizienz des Leserstehens bei Erwachsenen nach dem Strategiemodell) vor, das mit seiner Orientierung am Strategiemodell (van Dijk & Kintsch, 1983) einen ersten Schritt zu einer kognitionspsychologisch transparenten, prozeßorientierten Erfassung von Teilfähigkeiten des Lesens darstellen soll. Die grundlegende meßmethodische Idee von ELVES ist die Erfassung der Effizienz von Teilprozessen des Lesens über die Kombination von Antwortgeschwindigkeit und -richtigkeit bei klar definierten Aufgaben. In diesem Beitrag geben wir zunächst einen kurzen Überblick über bislang vorliegende Lesefähigkeits-Diagnostika und skizzieren den kognitionspsychologischen Rahmen von ELVES. Im Anschluß beschreiben wir den Aufbau des Instruments und berichten über erste Ergebnisse zur Konstrukt- und Kriteriumsvalidität der Testwerte, die auf Basis des neuartigen prozeß- und effizienzbezogenen Ansatzes von ELVES ermittelt wurden.

Vorliegende Ansätze zur Erfassung von Lesefähigkeiten bei geübten Lesern/innen

Der Schwerpunkt der bisherigen Lesefähigkeitsdiagnostik bei geübten Lesern/innen lag auf der Entwicklung standardisierter Leseverständnistests für angewandte Zielsetzungen (Flippo & Schumm, 2000). Im angelsächsischen Raum (insbesondere in den USA) haben Leseverständnistests seit langem einen festen Platz in Testbatterien zur Eingangsdagnostik im College-Bereich (z. B. der Verbalteil des Graduate Record Examination Test [GRE], Educational Testing Service, 2001; Nelson-Denny Reading Test, Brown, Fishco & Hanna, 1993). Im Aufbau ähneln sich diese Tests weitgehend: Testpersonen lesen Texte von einer Länge zwischen vier und mehreren hundert Wörtern und beantworten anschließend Mehrfachwahl-Aufgaben (Wiedererkennung von Textaussagen oder Inferenzfragen). Erfäßt werden allgemeine Textverständniswerte, zum Teil auch Maße der Lesegeschwindigkeit. Der oft bemerkenswerten prognostischen Validität solcher Leseverständnistests steht aus

kognitionspsychologischer Sicht eine mangelhafte Konstruktvalidität gegenüber (Daneman, 1997): An der Beantwortung typischer Aufgaben in standardisierten Leseverständnistests ist eine Vielzahl kognitiver Prozesse beteiligt, so daß bei dem ausschließlich produktorientierten Ansatz unklar bleibt, in welchen Teilleistungen sich gute und schlechte Leser/innen eigentlich unterscheiden.

Für den deutschsprachigen Raum liegen zwar Instrumente zur Diagnostik von Leseschwierigkeiten und des Lesenlernens vor (für einen Überblick vgl. Hasselhorn, Schneider & Marx, 2000), für geübte erwachsene Leser/innen gibt es mit dem Subtest "Leseverständnis" aus dem Test für medizinische Studiengänge (TMS, Institut für Test- und Begabungsforschung, 1990) aber lediglich ein einziges Verfahren, das zudem eine recht spezifische selektionsdiagnostische Zielsetzung verfolgt. Der Lesekompetenz-Teil der internationalen PISA-Studie (Zielgruppe: 15jährige Schüler/innen) ist für die Untersuchungen in Deutschland um theoriegeleitet konstruierte Tests erweitert worden (Artelt et al., 2001), die sich wie das hier vorgestellte Instrument an der Theorie von van Dijk und Kintsch (1983) orientieren. Diese Tests sind jedoch als produkt-, nicht als prozeßorientierte Tests konzipiert.

Demnach fehlt bislang ein theoriegeleitet konstruiertes Instrument, das bei erwachsenen Lesern/innen einsetzbar ist, also bei Angehörigen der Population, aus der die Stichproben für psycholinguistische und textverarbeitungspsychologische Experimente in aller Regel gezogen werden. Im Sinne einer hohen Konstruktvalidität und der theoretischen Anschlußfähigkeit muß die Konstruktion eines derartigen Instruments an den zentralen Befunden der kognitionspsychologischen Lesefähigkeitsforschung ansetzen.

Kognitionspsychologische Lesefähigkeitsforschung

Im Mittelpunkt der kognitionspsychologischen Lesefähigkeitsforschung steht die beschreibende Ausdifferenzierung von Teilprozessen des Lesens, die Analyse ihrer kausalen Beziehungen und die Erklärung interindividueller Unterschiede in allgemeinen

Textverständnisleistungen (vgl. Richter & Christmann, 2002). Unabhängig davon, ob zur Erklärung von Lesefähigkeitsunterschieden eher textgeleitete Verarbeitungsprozesse in den Vordergrund gestellt werden (wie in der Theorie der verbalen Effizienz, Perfetti, 1985, oder dem interaktiv-kompensatorischen Modell, Stanovich, 1980) oder ob angenommen wird, daß interindividuelle Differenzen in Lesefähigkeiten auf eine unterschiedlich gute Bewältigung wissensgeleiteter Prozesse zurückzuführen sind (vgl. etwa die Schematheorie, Spiro & Tirre, 1980), werden in der kognitionspsychologischen Lesefähigkeitsforschung vier Grundannahmen weitgehend geteilt:

1. Die Aufgliederung des Lesekompetenz-Konstrukts orientiert sich an hierarchischen Modellen, wie sie für die allgemeinspsychologische Textverarbeitungsforschung kennzeichnend sind (vgl. Johnson-Laird, 1983; Kintsch, 1998; van Dijk & Kintsch, 1983). In derartigen Modellen werden Teilprozesse verschiedenen Verarbeitungsebenen zugeordnet, die von der visuellen Erfassung einzelner Buchstaben über Worterkennung, semantische und syntaktische Integration auf Satzebene bis hin zur Herstellung lokaler und globaler Kohärenz auf Textebene reichen. Die höheren Verarbeitungsebenen schließen die Integration von Textinformation und inhaltlichem Vorwissen über inferentielle Prozesse ein. Dabei hat auch die von van Dijk und Kintsch (1983) eingeführte Unterscheidung zwischen Textbasis als propositionaler Repräsentation des Texts und Situationsmodell als integrierter Repräsentation von Textinformationen und Vorwissen Eingang in die Lesekompetenzforschung gefunden (vgl. z. B. Moravcsik & Kintsch, 1993; Oakhill, 1994). Dasselbe gilt für die Unterscheidung von mikro- und makrostrukturellen Prozessen (vgl. z. B. Bisanz, Das, Varnhagen & Henderson, 1992).

2. Konsens besteht weiterhin darüber, daß zumindest bei Lesern/innen ohne gravierende Leseschwächen ein schlechteres allgemeines Leseverständnis nicht auf Defiziten in kognitiven Leistungsbereichen beruht, die unterhalb einer linguistisch beschreibbaren Verarbeitungsebene anzusetzen sind (wie Okulomotorik, Wahrnehmungsspanne oder

statisches Kurzzeitgedächtnis, vgl. Daneman, 1997). Die wesentlichen Faktoren für Unterschiede in Lesefähigkeiten sind demnach in der Verarbeitung von linguistisch beschreibbarem Material (Wörter, Sätze, Texte) zu suchen.

3. Vor allem für Prozesse auf niedrigeren Ebenen (Worterkennung, lokale Kohärenzbildung), aber auch für bestimmte Prozesse auf höheren Ebenen (z. B. bestimmte Inferenzen) wird angenommen, daß sie in weiten Teilen routinisiert und parallel ablaufen. Der Grad der Routinisierung und damit die Effizienz von Teilprozessen ist jedoch interindividuell variabel. So unterscheiden sich selbst geübte Leser/innen z. B. in der Geschwindigkeit des lexikalischen Zugriffs (Jackson & McClelland, 1979), in der Unterdrückung kontextunangemessener Wortbedeutungen (Gernsbacher & Faust, 1991) oder in der Etablierung thematischer Inferenzen (Long, Oppy & Seely, 1994).

4. Für die Verarbeitung eines schriftlichen Texts stehen nur begrenzte und interindividuell variable kognitive Ressourcen zur Verfügung (Just & Carpenter, 1992). Ein Maß der dynamischen Arbeitsgedächtniskapazität für Text ist die Lesespanne (Daneman & Carpenter, 1980; deutschsprachige Adaptation: Hacker, Veres & Wollenberger, 1994), für die enge Zusammenhänge mit einer Reihe von Teilprozessen nachgewiesen werden konnten (z. B. syntaktische Integration und globale Inferenzen, vgl. Carpenter, Miyake & Just, 1994). Aus der empirisch gut belegten Annahme einer allgemeinen Kapazitätsbeschränkung folgt, daß die Effizienz einzelner Teilprozesse neben der Lesegeschwindigkeit auch die Qualität der Verarbeitung von Textinformationen beeinflussen kann. Je schlechter z. B. der lexikalische Zugriff automatisiert ist, desto weniger Ressourcen stehen für Prozesse auf höheren Verarbeitungsebenen zur Verfügung (Perfetti, 1985). Insbesondere bei geringem inhaltlichen Vorwissen beeinträchtigen ineffiziente hierarchieniedrige Prozesse die Makrostrukturbildung bzw. die Konstruktion von Situationsmodellen.

Konstruktionsprinzipien und Aufbau von ELVES

Die genannten vier theoretischen Eckpunkte der kognitionspsychologischen

Lesekompetenzforschung bilden die konzeptuelle Basis für die Konstruktion von ELVES. Der Aufbau des Instruments stützt sich auf das Strategiemodell von van Dijk und Kintsch (1983), das nach wie vor als das heuristisch fruchtbarste und am breitesten rezipierte hierarchische Modell des Textverstehens gelten kann (vgl. auch Kintsch, 1998). Das Modell bezieht sich auf linguistisch beschreibbare Verarbeitungsebenen (ab Worterkennung aufwärts). Als (computergestützter) geschwindigkeitsbasierter Test (mit simultaner Berücksichtigung von Akkuratheit und Antwortlatenz) soll ELVES die Effizienz von Teilprozessen des Lesens erfassen. Ausgehend von der Annahme, daß die Effizienz hierarchieniedriger Prozesse auch die Qualität höherer Verarbeitungsprozesse beeinflussen kann, werden sowohl mikro- als auch makrostrukturelle Prozesse differenziert erfaßt.

Aufbau des Instruments. Im Strategiemodell (van Dijk & Kintsch, 1983) werden folgende Strategietypen unterschieden, die sich in zwei übergreifende Klassen gliedern: (a) Mikrostrukturelle (hierarchieniedrige) Prozesse, hierzu gehören der Aufbau einer propositionalen Mikrostruktur (propositional strategies) und Prozesse der lokalen Kohärenzbildung (local coherence strategies) sowie b) makrostrukturelle (hierarchiehohe) Prozesse, dazu gehören Makrostrategien (macrostrategies), schematische Strategien zur Bildung von Superstrukturen (schematic strategies) und pragmatische Strategien (rhetorical strategies). Die sieben Subtests von ELVES sind als computergestützte Leistungstests mit unterschiedlichen Aufgabenarten konzipiert, zu deren Bewältigung jeweils bestimmte Teilprozesse (Strategien) zentral sind (für Itembeispiele vgl. Tabelle 1; für eine ausführlichere Dokumentation s. <http://www.allg-psych.uni-koeln.de/elves/index.htm>). Alle Aufgaben sind mit zwei Antworttasten auf einer Standard-Computertastatur zu bearbeiten. Jedem Subtest ist ein Übungslauf vorangestellt. Erhoben werden jeweils Akkuratheit und Antwortlatenz. Der Test ist mit der Versuchssteuerungssoftware Inquisit (Millisecond Software, 1999) programmiert. Die Durchführung des Gesamttests dauert 30-45 Minuten.

Hierarchieniedrige Subtests. Zur Erfassung interindividueller Differenzen in

hierarchieniedrigen Prozessen enthält das Instrument Satzverifikationsaufgaben (1), Sinnhaftigkeitsurteile bezogen auf Sätze (2), Sinnhaftigkeitsurteile bezogen auf Satzfolgen (3) und einen Wortschatztest (4). (1) Der Subtest Satzverifikation besteht aus (je zur Hälfte wahren oder falschen) kurzen begrifflichen Aussagen unterschiedlicher syntaktischer und semantischer Komplexität, die sich auf allgemein geläufige Begriffe beziehen. Systematisch variiert wurden die Länge sowie die Anschaulichkeit und Konkretheit/Abstraktheit der Substantive, die als grammatische Subjekte der Prädikationen fungieren (vgl. Heupst & Hager, 1994). Zur Bewältigung des Subtests Satzverifikation sind in erster Linie Prozesse erforderlich, die zur Kategorie der propositionalen Strategien gehören (lexikalischer Zugriff, Extraktion von Propositionen, semantische und syntaktische Integration von Propositionen auf Satzebene). (2) Die Items des Subtests Sinnhaftigkeitsurteile (Sätze) bestehen wiederum aus unterschiedlich komplexen (sinnvollen oder sinnlosen) Aussagen mit Substantiven an der Subjektstelle, die hinsichtlich Länge, Anschaulichkeit und Konkretheit/Abstraktheit variieren. Auch dieser Subtest bezieht sich in erster Linie auf propositionale Strategien. (3) Im Subtest Sinnhaftigkeitsurteile (Satzfolgen) müssen Paare von Aussagen daraufhin eingeschätzt werden, ob sie eine sinnvolle Folge von Sätzen bilden. Der zweite Satz ist jeweils der Testsatz. Bei einigen Satzfolgen ist die korrekte Bestimmung anaphorischer Relationen für eine korrekte Beurteilung hinreichend (vgl. z. B. Oakhill & Yuill, 1986), für die meisten Satzfolgen muß dagegen eine Brückeninferenz mit Validierung des kausalen Übergangs vorgenommen werden (vgl. Singer, Halldorson, Lear & Andrusiak, 1992). Der Subtest erfordert damit vorwiegend Strategien lokaler Kohärenzbildung (einschließlich lokaler kausaler Inferenzen). (4) Der Subtest Wortschatz zielt mit der Vorgabe (korrekter oder falscher) Bedeutungserklärungen insbesondere auf die Verfügbarkeit seltener und schwieriger Wörter ab, berührt also einen Aspekt des lexikalischen Zugriffs, dessen effiziente Bewältigung sowohl wichtige Voraussetzung als auch wahrscheinliche Folge eines guten allgemeinen Leseverständnisses ist (vgl. z. B. Sternberg, 1987).

Hierarchiehohe Subtests. Die Effizienz hierarchiehoher Prozesse wird mit einem Textverständnis (5), Aufgaben, bei denen Fakten und Meinungen unterschieden werden müssen (6), und über die Identifikation von Textimplikationen (7) erfaßt. (5) Für den Subtest Textverständnis sind zwei einführende Texte zu einem soziologischen Thema zu lesen („Klassen“, modifiziert nach Giddens, 1995). Im Anschluß an die Lektüre jedes Texts werden Aussagen auf Makropropositionsniveau vorgegeben, für die (ohne erneute Zuhilfenahme des Texts) zu entscheiden ist, ob die enthaltene Information im Text vorkam oder nicht. Bei diesem Subtest sind makropropositionale Strategien zentral. (6) Im Subtest Textimplikationen ist nach Lektüre eines populärwissenschaftlichen psychologischen Texts („Gesellschaft und Vererbung“, modifiziert nach Eysenck, 1977) ohne erneute Zuhilfenahme des Texts für einzelne Textaussagen zu entscheiden, ob sie Implikationen des Texts darstellen. Für diesen Subtest sind neben Makrostrategien rhetorische Strategien (einschließlich globaler Inferenzen) erforderlich. (7) Der Subtest Fakten/Meinungen besteht aus Aussagen des Texts von Eysenck (1977), für die zu entscheiden ist, ob es sich um eine Darstellung von Fakten handelt oder ob sie Meinungen des Autors repräsentieren. Die Bearbeitung dieser Aufgaben erfordert die Anwendung rhetorischer Strategien.

Testwerte. Registriert werden Antwortausprägung und Antwortlatenz für die Testsätze, die zu integrierten Testwerten kombiniert werden. Für die Bildung integrierter Testwerte wird die Antwortzeit für jedes Items zunächst an der Anzahl der Silben relativiert, um unterschiedliche Itemlängen näherungsweise auszugleichen. Im Anschluß werden die linksschief verteilten Silben-Antwortzeiten einer Reziprok-Transformation unterzogen, bei der ein Quotient aus einer frei wählbaren Konstanten und dem ursprünglichen Wert gebildet wird (vgl. Lienert & Raatz, 1994, S. 71). Durch diese Transformation wird die Verteilung normalisiert (Unterschiede zwischen kurzen Entscheidungszeiten erhalten ein größeres Gewicht als solche zwischen langen Entscheidungszeiten); außerdem wird die Polung umgekehrt (höhere Werte bedeuten eine schnellere Verarbeitung). Die integrierten Testwerte

für jede Skala werden durch Aufsummierung der reziprok-transformierten Silben-Antwortzeiten für die richtig beantworteten Items gebildet, die um so höher ausfallen, je rascher ein Item bearbeitet wurde. Falsch beantwortete Items werden dagegen mit dem Wert Null kodiert, unabhängig davon, wie schnell sie bearbeitet wurden. Dadurch beinhalten die Skalenwerte sowohl eine Geschwindigkeits- als auch eine Akkuratheitskomponente: Hohe Testwerte, die indikativ für eine effiziente Ausführung der jeweils angezielten Teilprozesse sein sollen, können erreicht werden, wenn viele Items in kurzer Zeit richtig bearbeitet wurden. Während bei den hierarchiehohen Subtests Akkuratheit und Antwortlatenz gleichermaßen zur Variabilität der integrierten Testwerte beitragen sollten, gilt für die hierarchieniedrigen Subtests mit Ausnahme des Wortschatztests, daß die Geschwindigkeitskomponenten im Vordergrund stehen und die Akkuratheitskomponenten in erster Linie Flüchtigkeitsfehler abbilden dürften. Denn mit diesen Subtests wird die Effizienz basaler Teilprozesse des Lesens erfaßt, die von geübten Lesern/innen beherrscht werden, wenn auch mit großen interindividuellen Unterschieden in der Prozeßgeschwindigkeit.

Ausreißer und Verfälschungssicherheit. Aufgrund der Reziproktransformation können Ausreißerwerte, die auf eine ungewöhnlich schnelle Beantwortung zurückzuführen sind, sehr leicht identifiziert werden (z. B. anhand einer graphischen Inspektion der Testwerteverteilung). Dies ist auch deshalb von Bedeutung, weil ansonsten ein hoher integrierter Testwert auch durch beliebige, aber schnelle Antworten erzielt werden könnte (im Prinzip selbst dann, wenn nur ein einziges Item korrekt, dabei aber sehr schnell bearbeitet würde). Neben deutlich erhöhten reziprok transformierten Antwortlatenzen ist eine Testverfälschung durch willkürliche Bearbeitung in aller Regel zusätzlich an einer hohen Fehlerzahl erkennbar; der Erwartungswert für den Anteil richtiger Antworten bei einer unsachgemäßen Bearbeitung liegt aufgrund des dichotomen Antwortformats bei 50%.

Pilotstudie. Eine Vorform des Instruments wurde anhand einer Stichprobe mit 35 studentischen Probanden/innen (Pbn) erprobt. Diese Daten wurden zur Itemselektion und

Verbesserung von Formulierungen einzelner Items bei den hierarchiehohen Subtests genutzt.

Überblick über die Validierungsstudien

Zu Validierungszwecken wurde ELVES Angehörigen der primären Zielpopulation (Universitätsstudierende) vorgelegt, wobei vier Untersuchungen zur Konstrukt- und Kriteriumsvalidität durchgeführt wurden. In Untersuchung 1 standen neben der Ermittlung von deskriptiven Kennwerten und Reliabilitätsschätzern zunächst Aspekte der Konstruktvalidität des Instruments im Sinne des zugrundegelegten textverarbeitungspsychologischen Modells (van Dijk & Kintsch, 1983) im Mittelpunkt. In zwei weiteren Teilstudien wurden Aspekte der Kriteriumsvalidität des Instruments untersucht. Dabei stand zunächst in Untersuchung 2 die Frage im Vordergrund, ob sich mit den entscheidungszeitbasierten Tests in ELVES allgemeine Leseverständnismaße vorhersagen lassen, wie sie mit herkömmlichen Leseverständnistests erfaßt werden. Darüber hinaus wurde in Untersuchung 3 geprüft, ob die Leistungen in naturalistischen Leseaufgaben mit den Testleistungen in den ELVES-Subtests zusammenhängen. In Untersuchung 4 sollten erstens die in Untersuchung 1 erzielten Befunde zu den Reliabilitäten und Skaleninterkorrelationen der ELVES-Subtests repliziert werden. Zweitens wurde anhand einer erweiterten ELVES-Version mit Testtexten zu verschiedenen Themen untersucht, inwieweit die Bearbeitung der hierarchiehohen Subtests vom Ausmaß des inhaltlichen Vorwissens abhängt. Diese Fragestellung ist für die Konstruktvalidität von ELVES von Bedeutung, denn auch mit den hierarchiehohen Subtests sollten vorwissensunabhängige Lesefähigkeitsaspekte erfaßt werden.

Untersuchung 1

Untersuchung 1 diente der Ermittlung von Reliabilitätsschätzern (interne Konsistenzen) für die integrierten Testwerte der einzelnen Subtests. Im Hinblick auf die Konstruktvalidität des Instruments wurden vier Vorhersagen zu Skaleninterkorrelationen

überprüft, die sich aus der Unterscheidung hierarchieniedriger und hierarchiehoher Strategien sensu van Dijk und Kintsch (1983) ergeben. (1) Die hierarchieniedrigen Subtests sollten untereinander hoch korreliert sein, da sie sich auf die Effizienz grundlegender und teilweise überlappender Strategien beziehen. (2) Ähnlich wurde für die hierarchiehohen Subtests erwartet, daß sie ein zu den hierarchieniedrigen Subtests komplementäres Cluster bilden, da sich auch hier die primär angesprochenen Strategien teilweise überlappen (vgl. Tabelle 1). (3) Die Korrelationen der hierarchiehohen Subtests untereinander sollten jedoch insgesamt geringer ausfallen als die Korrelationen der hierarchieniedrigen Subtests untereinander, da die hierarchiehohen Subtests komplexere und damit stärker divergierende Strategien ansprechen. Die Subtests Textverständnis und Textimplikationen sollten dabei untereinander enger zusammenhängen als mit dem Subtest Fakten/Meinungen, da beide auf einem ähnlichen Aufgabentyp beruhen und sich lediglich im Anteil der zur Aufgabenlösung erforderlichen Inferenzen unterscheiden. Der Subtest Fakten/Meinungen bezieht sich demgegenüber auf vergleichsweise spezifische rhetorische Strategien. (4) Schließlich sollten hierarchieniedrige und hierarchiehohe Subtests ebenfalls korreliert sein, da zur Bewältigung der hierarchiehohen Subtests auch die in den hierarchieniedrigen Subtests angesprochenen Strategien erforderlich sind. Diese Korrelationen sollten jedoch nicht höher sein als die Interkorrelationen der hierarchieniedrigen Subtests.

Methode

Stichprobe. An Untersuchung 1 nahmen 158 Studierende (117 Frauen und 41 Männer) der Universitäten Köln und Heidelberg mit einem mittleren Alter von 26.7 Jahren (SD = 6.8) und einer mittleren Semesterzahl von 2.8 (SD = 2.6) teil. In der Mehrzahl handelte es sich um Studierende der Psychologie (92%), die entweder für die Bescheinigung von Versuchspersonenstunden oder für ein geringes Entgelt an der Untersuchung teilnahmen. Die Pbn gaben an, im Mittel 9.2 (SD = 9.5) Stunden pro Woche für das Studium und 8.1 (SD = 8.0) Stunden in ihrer Freizeit zu lesen.

Durchführung. ELVES wurde in Einzel- oder Zweiersitzungen durchgeführt, zum Teil im Rahmen umfangreicherer Untersuchungen (s. Untersuchung 2 und Untersuchung 3).

Ergebnisse und Diskussion

Datenvorbehandlung und Ausreißeranalysen. Für die hier berichteten Analysen wurde die mittlere Silben-Antwortzeit pro Subtest als Konstante in der Reziprok-Transformation herangezogen, was zu einem intuitiv leicht zugänglichen Wertebereich der einzelnen Skalen führt (vgl. Tabelle 2). Die Datensätze von vier Pbn wurden aufgrund offensichtlich instruktionswidriger Bearbeitung einzelner Subtests (unplausibel kurze Antwortlatenzen bei hohen Fehlerzahlen, s. o. Konstruktionsprinzipien und Aufbau von ELVES) von den weiteren Analysen ausgeschlossen. Zusätzlich wurde eine Ausreißeranalyse für die korrekt beantworteten Items der übrigen 154 Datensätze durchgeführt. Als Ausreißer wurden reziprok transformierte Silben-Antwortlatenzen definiert, die mehr als zwei (für jede Person ermittelte) Standardabweichungen vom personenspezifischen Mittelwert abwichen. Nach diesem Kriterium ergab sich nur ein geringer Anteil von Ausreißern pro Subtests (zwischen 0.9 und 3.4 Prozent). Im Sinne einer möglichst konservativen Schätzung der psychometrischen Gütekriterien wurden die Ausreißerwerte in den folgenden Analysen weder ersetzt noch eliminiert.

Deskriptive Kennwerte. Im Unterschied zu den hierarchiehoheren Subtests werden bei der Bearbeitung der hierarchieniedrigen Subtests (mit Ausnahme von Wortschatz) kaum Fehler gemacht (vgl. Tabelle 2); entsprechend gering sind hier die Streuungen der Akkuratheitswerte. In den Antwortlatenzen zeigen sich dagegen für alle Subtests beträchtliche interindividuelle Differenzen. Die internen Konsistenzen (Cronbachs Alpha) der integrierten Testwerte sind als zufriedenstellend bis gut einzuschätzen; lediglich für den Subtest Textverständnis ergibt sich eine nur akzeptable interne Konsistenz ($\alpha=.68$). Die Verteilung der integrierten Testwerte weicht für keinen der Subtests von einer

Normalverteilung ab (alle Kolmogoroff-Smirnoff Z -Werte ≤ 1.0 , alle p -Werte $\geq .25$).

Skaleninterkorrelationen. Die vier hierarchieniedrigen Subtests waren untereinander hoch korreliert (vgl. Tabelle 2; zwischen .62 für Satzverifikation-Wortschatz und .78 für Sinnhaftigkeit (Sätze)-Wortschatz). Die Interkorrelationen der hierarchiehohen Subtests waren insgesamt niedriger; in zwei Fällen ergaben sich jedoch auch hier mittlere bis hohe Zusammenhänge (.58 für Textverständnis-Textimplikationen, .46 für Textverständnis-Fakten/Meinungen). Lediglich Textimplikationen und Fakten/Meinungen korrelierten mit .28 nur moderat. Alle paarweisen Korrelationen zwischen hierarchiehohen und hierarchieniedrigen Subtests waren signifikant, variierten aber zwischen den hierarchiehohen Subtests beträchtlich in ihrer Höhe. Die höchsten Zusammenhänge mit hierarchieniedrigen Subtests ergaben sich für Textverständnis (höchste Korrelation .66 mit Wortschatz), die zweithöchsten für Textimplikationen (höchste Korrelation .56 mit Sinnhaftigkeit [Satzfolgen]) und die niedrigsten für Fakten/Meinungen (höchste Korrelation .35 mit Sinnhaftigkeit [Satzfolgen]). Um zu überprüfen, inwieweit die Korrelationshöhen von den Reliabilitäten der Subtests abhängen, wurden zusätzlich anhand der internen Konsistenzen minderungskorrigierte Korrelationen berechnet. Wie Tabelle 2 zu entnehmen ist, blieb auch bei Berücksichtigung der unterschiedlichen Reliabilitäten der Subtests durch eine doppelte Minderungskorrektur das beschriebene Korrelationsmuster sehr weitgehend erhalten. Lediglich der hierarchiehohe Subtest Textverständnis wies nach der Minderungskorrektur mit drei der vier hierarchieniedrigen Subtests Korrelationen auf, deren Höhe mit den Korrelationen der hierarchieniedrigen Subtests untereinander vergleichbar ist.

Damit folgen die Skaleninterkorrelationen der ELVES-Subtests weitgehend dem nach dem Strategiemodell (van Dijk & Kintsch, 1983) erwarteten Muster. In Anbetracht der teilweise hohen Korrelationen zwischen hierarchieniedrigen und hierarchiehohen Subtests (hier insbesondere dem Subtest Textverständnis) bleibt jedoch zunächst unklar, inwieweit das

Instrument die für das Strategiemodell zentrale Unterscheidung hierarchiehoher und hierarchieniedriger Prozesse abbildet. Zur Überprüfung dieses Validitätsaspekts wurden zusätzlich konfirmatorische Faktorenanalysen durchgeführt.

Konfirmatorische Faktorenanalysen

Modellspezifikation. Im Sinne der differenzierten Meßkonzeption von ELVES wurde ein Hypothesenmodell mit zwei latenten Variablen formuliert, die zum einen die vier hierarchieniedrigen, zum anderen die drei hierarchiehohen Subtests zusammenfassen. Das Hypothesenmodell beinhaltet weiterhin die Annahme, daß hierarchiehohe und hierarchieniedrige Lesefähigkeitskomponenten korreliert sind. Das Zweifaktoren-Hypothesenmodell mit interkorrelierenden Faktoren wurde gegen ein alternatives Modell getestet, das für alle sieben Subtests ein einziges latentes Lesefähigkeits-Konstrukt annimmt. Dieses Einfaktor-Alternativmodell ist äquivalent zu einem Zweifaktoren-Modell, in dem die Korrelation der latenten Variablen auf Eins fixiert ist. Daher stellen Hypothesen- und Alternativmodell geschachtelte Modelle dar (nested models, vgl. Bollen, 1989), die gegeneinander getestet werden können. Die Parameterschätzung erfolgte mit der Maximum-Likelihood-Prozedur von LISREL 8 (Jöreskog & Sörbom, 1996).

Modellbeurteilung. Zur Beurteilung der Modellpassung und des Vergleichs von Hypothesen- und Alternativmodell wurde eine Reihe unterschiedlich fundierter Gütekriterien herangezogen (zur Interpretation vgl. Bollen & Long, 1993; Hu & Bentler, 1995). Sämtliche der betrachteten Gütekriterien weisen auf eine gute Passung des Hypothesenmodells hin, mit $\chi^2(13, \underline{n}=154)=19.45, p>.05$ (NFI=.97, RMSEA=.05, CI_{links}=.00). Demgegenüber zeigt das alternative Einfaktor-Modell eine ungenügende Passung, mit $\chi^2(14, \underline{n}=154)=37.32, p<.001$ (NFI=.94, RMSEA=.10, CI_{links}=.06). Die Güte der Modellanpassung unterscheidet sich signifikant zwischen Zweifaktoren-Hypothesenmodell und Einfaktor-Alternativmodell, mit $\chi^2_{\text{diff}}(1, \underline{n}=154)=17.87, p<.001$ (Δ NFI=.03). Die Faktorladungen in der komplett standardisierten Lösung des Zweifaktoren-Hypothesenmodells sind in Tabelle 2

wiedergegeben. Die Korrelation der beiden latenten Variablen im Hypothesenmodell betrug .84.

Dieses Ergebnismuster spricht insgesamt für die Konstruktvalidität von ELVES und die aus dem Strategiemodell (van Dijk & Kintsch, 1983) abgeleitete Unterscheidung zweier Lesefähigkeitskomponenten. Zwar ist die Korrelation zwischen den beiden latenten Variablen mit .84 als hoch einzuschätzen, was zunächst gegen die diskriminante Validität der ELVES-Subtests spricht. Die bessere Modellpassung des Zweifaktoren- gegenüber dem Einfaktor-Alternativmodell läßt sich grundsätzlich aber als Beleg für eine Trennung hierarchieniedriger und hierarchiehoher Lesefähigkeitskomponenten einordnen. Darüber hinaus schließt das 95%-Konfidenzintervall für den Schätzer der latenten Korrelation nicht den Wert 1 ein, was ein weiteres Argument für eine separate Erfassung der beiden Lesefähigkeitskomponenten darstellt. Eine plausible inhaltliche Interpretation der hohen Faktorinterkorrelation, die sich aus der hierarchischen Konzeption des Strategiemodells ableiten läßt, lautet, daß zur Bearbeitung der hierarchiehohen Subtests die mit den hierarchieniedrigen Subtests erfaßten Lesefähigkeitskomponenten erforderlich sind, während das Umgekehrte nicht gilt. Diese Interpretation läßt sich mit Blick auf die durch die beiden Faktoren jeweils erklärte Indikatorvarianz (variance extracted estimate, vgl. Fornell & Larcker, 1981) erhärten: Der durch den Faktor Hierarchiehohe Lesefähigkeiten erklärte Anteil der Indikatorvarianz fällt mit 48% niedriger aus als der Anteil gemeinsamer Varianz mit dem Faktor Hierarchieniedrige Lesefähigkeiten, der etwa 70% beträgt. Der durch diesen Faktor erklärte Anteil in der Indikatorvarianz beträgt dagegen ebenfalls 70%, ist also nicht geringer als die gemeinsame Varianzanteil der beiden latenten Variablen. Das Verstehen zusammenhängender Texte beruht immer auch - und in nicht unbeträchtlichem Maße - auf Prozessen der Worterkennung, propositionalen Strategien und Strategien der lokalen Kohärenzbildung. Dieser Umstand spiegelt sich in den Parameterschätzungen des Zweifaktoren-Modells wider.

Untersuchung 2

Ausgehend von der Überlegung, daß klassische produktorientierte Leseverständnistests die nach wie vor übliche Methode zur Erfassung des allgemeinen Leseverständnisses darstellen, wurden in Untersuchung 2 neu konstruierte Paper-Pencil-Tests mit Texten aus zwei verschiedenen Inhaltsbereichen eingesetzt, um Hinweise auf die Kriteriumsvalidität von ELVES zu erhalten. Beide Tests beinhalteten je zwei Skalen mit Rekognitionsfragen bezogen auf Detailinformationen sowie Inferenzfragen, in denen vorwiegend zentrale Textinformationen thematisch waren. Erwartet wurden für beide Inhaltsbereiche insgesamt höhere Korrelationen der ELVES-Subtests mit den Inferenzskalen und geringere bzw. fehlende Korrelationen mit den Rekognitionsfragen. Die verwendeten Inferenzfragen zielten auf Textverständnis im engeren Sinne ab; ihre Beantwortung war auf mikro- und makrostrukturelle Prozesse sowie die Konstruktion eines adäquaten Situationsmodells angewiesen und sollte daher die gesamte Bandbreite der mit ELVES angezielten Teilfähigkeiten des Lesens erfordern. Die hier im Sinne von Kontrollvariablen eingesetzten Rekognitionsfragen erfaßten demgegenüber im wesentlichen das Gedächtnis für relativ isolierte Elemente der Oberflächenrepräsentation des Texts und der (mikrostrukturellen) Textbasis und sollten daher - wenn überhaupt - lediglich hierarchieniedrige Strategien in Anspruch nehmen. Dieses Zusammenhangsmuster sollte im wesentlichen erhalten bleiben, wenn das inhaltsspezifische Vorwissen auspartialisiert wird.

Methode

Stichprobe. An Untersuchung 2 nahmen 89 Studierende (70 Frauen, 19 Männer) mit einem mittleren Alter von 26.8 Jahren ($SD = 6.2$) und einer mittleren Semesterzahl von 4.5 ($SD = 3.1$) teil. In der Mehrzahl handelte es sich um Studierende der Psychologie (89%). Die Daten von 50 Pbn gingen auch in die Analysen von Untersuchung 1 ein, die Daten der übrigen 39 Pbn stammen aus der Stichprobe von Untersuchung 4.

Durchführung und Kriteriumsvariablen. Die Pbn bearbeiteten neben ELVES zwei

kurze Paper-Pencil-Tests nach dem Muster klassischer Leseverständnistests zu zwei verschiedenen Inhaltsbereichen: Psychologie ("Signalentdeckungstheorie") und Literaturwissenschaft ("Weimarer Klassik"). Vorab wurde für beide Themen das themenspezifische Vorwissen erhoben, indem die Kenntnis zentraler Begriffe aus den Texten über je 6-8 Items mit offenem Antwortformat erfragt wurde. Die Antworten aus den Vorwissenstests wurden von zwei Kodiererinnen mit 0, 1, oder 2 Punkten bewertet (Interrater-Reliabilität: $\kappa = .70$, Spearman's $\rho = .80$, bestimmt an 3x2 Vorwissenstests) und zu je einer Skala zusammengefaßt. Bei den Leseverständnistests waren unter der Zeitvorgabe von jeweils max. 20 Minuten zwei lehrbuchartige Sachtexte (psychologische Thematik: 821 Wörter; literaturwissenschaftliche Thematik: 1001 Wörter) zu lesen und im Anschluß zwei Arten von Fragen zu bearbeiten: (1) textnah formulierte Aussagen über Detailinformationen, für die zu entscheiden war, ob sie im Text enthalten waren oder nicht (Rekognitionsfragen), und (2) nicht explizit im Text enthaltene Aussagen, für die zu entscheiden war, ob sie Implikationen des Texts darstellen oder nicht (Inferenzfragen). Inferenz- und Rekognitionsfragen wurden zu je einer Skala zusammengefaßt (je 7-8 Items, Anteil richtiger Antworten .46 bis .76, Standardabweichungen .18 bis .26). Die Reihenfolgen von ELVES und Vorwissens-/ Leseverständnistests wurden ebenso wie die Reihenfolgen der Leseverständnistests ausbalanciert.

Ergebnisse und Diskussion

Für die Subtests von ELVES wurden Zusammenhänge mit den Inferenzfragen erwartet, jedoch nur geringe bzw. keine Zusammenhänge mit den zu Kontrollzwecken eingesetzten Rekognitionsfragen. Tabelle 3 läßt sich entnehmen, daß entsprechend dieser Vorhersagen die Mehrzahl der ELVES-Subtests mit den Inferenzfragen zu beiden Testtexten in mittlerer Höhe korreliert war. Diese Korrelationen blieben im Wesentlichen erhalten, wenn der Einfluß des thematischen Vorwissens aus den kriterialen Testwerten auspartialisiert wurde. Die einzige Ausnahme ist der Subtest Fakten/Meinungen, für den sich lediglich ein

geringer Zusammenhang mit den Inferenzfragen zum literaturwissenschaftlichen Testtext ergab, der bei Ausparialisierung des thematischen Vorwissen aus Textverständniswerten nicht mehr signifikant war. Vermutlich waren die spezifischen rhetorischen Strategien, deren Effizienz durch Fakten/Meinungen erfaßt werden soll, für die Bearbeitung der Inferenzfragen eher irrelevant. Die Vorwissensmaße selbst waren mit der Testleistung bei den Inferenzfragen in mittlerer Höhe korreliert (für die psychologische Thematik: $r = .43$, $p < .01$; für die literaturwissenschaftliche Thematik: $r = .36$, $p < .01$). Nur teilweise erwartungskonform waren jedoch die Ergebnisse zu den Rekognitionsfragen als Kriteriumsmaßen. Für die psychologische Thematik zeigten sich wie erwartet nur zwei geringe Korrelationen der ELVES-Subtests mit der Leistung in den Rekognitionsfragen. Beide Korrelationen waren nach Ausparialisierung des thematischen Vorwissens aus den Leistungen in den Rekognitionsfragen nicht mehr signifikant. Bei der literaturwissenschaftlichen Thematik ergaben sich demgegenüber auch für die Rekognitionsfragen Korrelationen mit den ELVES-Subtests (mit Ausnahme von Fakten/Meinungen), und zwar in moderater bis mittlerer Höhe. Hier liegt die Annahme nahe, daß bei der literaturwissenschaftlichen Thematik erwartungswidrig auch die Bearbeitung der Rekognitionsfragen durch die mit ELVES erfaßten Lesefähigkeiten erleichtert wurde.

Untersuchung 3

Mit Untersuchung 3 wurde eine Kriteriumsvalidierung von ELVES anhand einer naturalistischen Leseaufgabe angestrebt. Die Pbn lasen einen umfangreichen Text zu einem psychologischen Thema und verfaßten anschließend einen eigenen Text, der im Hinblick auf Indikatoren der Situationsmodellgüte inhaltsanalytisch ausgewertet wurde. Im Sinne der Meßintention von ELVES sollten sich auf Basis der ELVES-Subtests Aspekte der Verstehensqualität, nicht jedoch Maße der Quantität des Gelernten (ohne Differenzierung nach Qualitätsaspekten) vorhersagen lassen. Zur Überprüfung der Effizienzkonzeption, die der Bildung der integrierten Testwerte zugrundeliegt, wurde schließlich untersucht, ob die

integrierten Testwerte besser zur Vorhersage der Situationsmodellgüte geeignet sind als (a) die Antwortlatenzen (ohne Berücksichtigung der Antwortrichtigkeit) und (b) die Anzahl richtig gelöster Items (ohne Berücksichtigung der Antwortlatenz).

Methode

Stichprobe. An Untersuchung 3 nahmen 94 Psychologiestudierende (71 Frauen, 23 Männer) der Universitäten Köln und Heidelberg teil. Der Altersdurchschnitt lag bei 26 Jahren ($SD = 7.3$), die durchschnittliche Semesterzahl bei 2.0 ($SD = 1.8$). Die Stichprobe stellt eine Substichprobe der Stichprobe aus Untersuchung 1 dar.

Durchführung und Kriteriumsvariablen. Im Rahmen einer Untersuchung zum Lernen mit Text und Hypertext (vgl. Christmann, Groeben, Flender, Naumann & Richter, 1999) bearbeiteten die Pbn zunächst ELVES, gefolgt von einer naturalistischen Leseaufgabe, bei der die Pbn zunächst drei Stunden lang entweder einen Hypertext oder einen inhaltsgleichen Buchtext zum Thema 'Visuelle Wahrnehmung' lasen (zum Textmaterial s. Flender & Christmann, 2000). Im Anschluß sollte in weiteren drei Stunden ohne erneute Zuhilfenahme des Texts ein prüfungsvorbereitendes Skript über Experimente zur visuellen Wahrnehmung verfaßt werden. Die erstellten Skripte wurden in Idea Units (IU) segmentiert (vgl. Ballstaedt & Mandl, 1984) und inhaltsanalytisch ausgewertet. In der vorliegenden Untersuchung standen zwei zentrale Indikatoren der Situationsmodellgüte im Mittelpunkt: (1) Als Indikator für Vorwissensintegration wurde die Häufigkeit inhaltsbezogener IU ermittelt, deren Inhalt über die im Text explizit genannten Informationen hinausgeht (Inferenzen, Problematisierungen von Textinformationen, Metabemerkungen). Der Median der paarweisen Interrater-Reliabilitäten (Cohens κ) lag für diese Variable bei .60 (bestimmt für 7 Rater-Paare anhand von 5 Skripten), die mittlere Häufigkeit bei 3.9 ($SD=7.6$). (2) Als Indikator für die Stringenz (Kohärenz) des gebildeten Situationsmodells wurde die Häufigkeit von IU ermittelt, die die rhetorisch-argumentative Stellung von Informationen im Gesamtzusammenhang anzeigen (rhetorische Explikationen, Signaling, thematische Sätze

und Zielangaben; vgl. Meyer, 1975; Schnotz, 1994). Der Median der Interrater-Reliabilitäten (Cohens κ) betrug hier .54, die mittlere Häufigkeit 13.8 ($\underline{SD}$ =11.5). Im Sinne einer Kontrollvariablen wurde zusätzlich die Gesamtzahl inhaltsbezogener IU ausgezählt ($\underline{M}$ =166.9, $\underline{SD}$ =67.6, Median der Interrater-Reliabilitäten [Cohens κ] .86).

Ergebnisse und Diskussion

In Untersuchung 3 wurden für alle ELVES-Subtests positive Zusammenhänge mit Indikatoren der Situationsmodellgüte aus einer naturalistischen Leseaufgabe erwartet. Dagegen sollten sich mit der zu Kontrollzwecken betrachteten Gesamtzahl inhaltsbezogener IU in den Skripten (als Maß für die reine Quantität des Gelernten) nur geringe und inkonsistente Zusammenhänge zeigen. Insgesamt sollten die integrierten Testwerte eine bessere Vorhersage erlauben als Akkuratheit und Antwortlatenz bei getrennter Betrachtung.

Aus Tabelle 3 (rechte Hälfte) geht hervor, daß jeweils sechs der sieben ELVES-Subtests in mittlerer Höhe mit den Situationsmodellgüte-Indikatoren Vorwissensintegration und Stringenz korrelierten. Im Gegensatz dazu zeigten sich nur für drei Subtests Zusammenhänge mit der Gesamtzahl inhaltsbezogener IU, die numerisch geringer und nicht sämtlich positiv waren. Dieses Korrelationsmuster spricht für die Kriteriumsvalidität der ELVES-Subtests, die als Maße der Effizienz von Teilprozessen des Lesens in erster Linie Verstehensgüte, nicht jedoch die bloße Quantität des Gelernten vorhersagen sollten (wie es etwa für Maße der Lesegeschwindigkeit zu erwarten wäre).

In dieselbe Richtung weist der Befund, daß die Antwortlatenzen für sich genommen mehrheitlich geringere Zusammenhänge mit den Situationsmodellgüte-Indikatoren aufwiesen als die integrierten Testwerte (s. die Korrelationsvergleiche in Tabelle 3). Die mittlere Korrelationshöhe der Antwortlatenzen betrug .22 für den Zusammenhang mit Vorwissensintegration (gegenüber .39 bei den integrierten Testwerten) und .23 für den Zusammenhang mit Stringenz (gegenüber .34 bei den integrierten Testwerten). Im Unterschied zu den integrierten Testwerten war die Mehrzahl der Antwortlatenzen zudem mit

der Gesamtzahl inhaltsbezogener IU in den Skripten korreliert. Die Antwortlatenzen trennen also nicht zwischen Verstehensgüte und bloßer Quantität des Gelernten. Insbesondere bei den drei überwiegend korrekt bearbeiteten hierarchieniedrigen Subtests (Satzverifikation, Sinnhaftigkeit (Sätze) und Sinnhaftigkeit (Satzfolgen)) ist die geringere Prädiktionskraft der Antwortlatenzen allerdings nicht auf die Kombination von Antwortlatenz und Antwortrichtigkeit, sondern auf die den integrierten Testwerten zugrundegelegte Reziproktransformation zurückzuführen, die eine stärker symmetrische Verteilung zur Folge hat. Für die Anzahl richtig beantworteter Items ergaben sich nur in zwei von zwölf Fällen Korrelationen mit Situationsmodellgüte-Indikatoren, die zudem numerisch gering sind. Die mittlere Korrelation mit Vorwissensintegration betrug .12, mit Stringenz .09. Auch hier gilt die Einschränkung, daß die Akkuratheitswerte der drei weitgehend korrekt bearbeiteten hierarchieniedrigen Subtests über eine eingeschränkte Varianz verfügen, was in diesen Fällen zu den geringeren Korrelationen beigetragen haben dürfte. Insgesamt scheint jedoch mit den integrierten Testwerten, die simultan auf Akkuratheit und reziprok transformierte Antwortlatenzen zurückgreifen, ein Zugewinn an diagnostischer Information verknüpft zu sein.

Untersuchung 4

In Untersuchung 4 standen abermals Aspekte der Konstruktvalidität im Mittelpunkt. Dabei wurde erstens geprüft, ob sich die in Untersuchung 1 erzielten Befunde zu den Reliabilitäten und Skaleninterkorrelationen der ELVES-Subtests replizieren lassen. Zweitens wurde anhand einer erweiterten ELVES-Version mit thematisch heterogenen Testtexten die Vorwissensabhängigkeit der hierarchiehohen Subtests untersucht. Das inhaltliche Vorwissen ist einer der wichtigsten leserseitigen Prädiktoren für das Textverständnis und beeinflusst damit auch die Leistung in textbasierten Lesefähigkeitstests (z. B. Johnston, 1984), obwohl es sich nicht um eine lesespezifische Fähigkeitskomponente handelt (vgl. Richter & Christmann, 2002). Wenn mit den hierarchiehohen Subtests von ELVES tatsächlich die

Effizienz der Bewältigung von Teilprozessen des Leseverstehens erfaßt wird, sollten Subtests, die auf denselben Aufgabentypen, aber Texten zu unterschiedlichen Themen beruhen, untereinander substantiell korreliert sein. Diese Zusammenhänge sollten höher sein als die Zusammenhänge mit hierarchiehohen Subtests, die auf andere Teilprozesse des Leseverstehens abzielen. Darüber hinaus sollte das Muster der Korrelationen eines Subtests insbesondere mit den hierarchieniedrigen Subtests davon unabhängig sein, welches Textthema der Messung zugrundeliegt. Diese Vorhersagen zu Zusammenhängen der hierarchiehohen Subtests sollten auch dann gestützt werden können, wenn das jeweils verstehensrelevante inhaltliche Vorwissen statistisch kontrolliert wird.

Methode

Stichprobe. An Untersuchung 4 nahmen 74 Studierende (62 Frauen, 12 Männer) mit einem mittleren Alter von 24.7 Jahren ($SD = 6.4$) und einer mittleren Semesterzahl von 2.2 ($SD = 2.6$) teil. In der Mehrzahl handelte es sich um Studierende der Psychologie (78%).

Untersuchungsmaterial. Zur Untersuchung der Rolle des inhaltlichen Vorwissens kam eine erweiterte Version von ELVES zum Einsatz, bei der für jeden der drei hierarchiehohen Subtests Testtexte zu verschiedenen Themen vorgegeben wurden. Für den Subtest Textverständnis wurden neben dem einführenden soziologischen Text je ein einführender Text zu einem literaturhistorischen Thema („Die literarische Epoche des Realismus“, Länge 260 Wörter, modifiziert nach Wolfrum, 1995) und einem physikalischen Thema („Das erste Bewegungsgesetz“, Länge 265 Wörter, modifiziert nach Ballif & Dibble, 1973) vorgeben. Die Subtests Implikationen und Fakten/Meinungen wurden um einen populärwissenschaftlichen politologischen Text ergänzt („Toleranz als politische Tugend“, Länge 603 Wörter, modifiziert nach Fetscher, 1991). Die Art der Aufgabenstellung und der Aufbau der verwendeten Items unterschieden sich nicht zwischen den thematisch unterschiedlichen Versionen der einzelnen Subtests. Lediglich die Textlängen und die Anzahl der Items (vgl. Tabelle 4) waren bei den neu konstruierten Versionen geringer gehalten, um die

Gesamtbearbeitungsdauer in einem vertretbaren Rahmen zu halten. Zur Erfassung des inhaltlichen Vorwissens wurde für jeden der fünf Testtexte ein Vorwissenstest mit sieben Mehrfachwahlaufgaben konstruiert (eine Bestantwort, vier Distraktoren), mit denen wichtige Begriffe aus den Texten und verstehensrelevantes Hintergrundwissen abgefragt wurde (Spannbreite des Anteils richtiger Antworten von .24 für das Physik-Vorwissen bis .82 für das Politologie-Vorwissen, Standardabweichungen .16 bis .23).

Durchführung. Die Untersuchung wurde als Gruppenuntersuchung mit jeweils 5-7 Pbn pro Termin durchgeführt. Zunächst wurden die Vorwissenstests mit einer Zeitvorgabe von sieben Minuten pro Test vorgegeben. Dann folgte nach einer 15minütigen Pause die Bearbeitung der erweiterten Fassung von ELVES, die im Durchschnitt etwa 60 Minuten in Anspruch nahm.

Ergebnisse

Das erste Ziel von Untersuchung 4 bestand in einer Replikation der Ergebnisse zu Skaleninterkorrelationen und Reliabilitätskennwerten aus Untersuchung 1. Eine zweite Fragestellung beschäftigte sich anhand einer erweiterten ELVES-Version mit der Vorwissensabhängigkeit der hierarchiehoher Subtests. Dabei sollten thematisch heterogene Subtests, die auf die Bewältigung derselben Teilprozesse abzielen, untereinander enger zusammenhängen als mit anderen hierarchiehoher Subtests, die auf die Bewältigung anderer Prozesse abzielen. Zudem sollte das Muster der Skaleninterkorrelationen über Testtexte zu verschiedenen Themen hinweg vergleichbar sein. Beide Vorhersagen sollten sich auch bei Kontrolle des inhaltlichen Vorwissens stützen lassen.

Replikation von Skaleninterkorrelationen und Reliabilitätskennwerten aus Untersuchung 1. Die Skaleninterkorrelationen der Subtests aus der erweiterten ELVES-Fassung sind in Tabelle 4 (linke Dreiecksmatrix) wiedergegeben. Das Zusammenhangsmuster der ELVES-Subtests ähnelt deutlich dem in Untersuchung 1 ermittelten (vgl. Tabelle 1) und entspricht damit den aus dem Strategiemodell ableitbaren

Erwartungen: Untereinander waren die hierarchieniedrigen Subtests hoch, die hierarchiehohe Subtests etwas niedriger korreliert. Auch miteinander waren die hierarchieniedrigen und hierarchiehohe Subtests mehrheitlich korreliert, wobei der Subtest Textverständnis die höchsten und der Subtest Fakten/Meinungen wiederum die niedrigsten Korrelationen mit hierarchieniedrigen Subtests aufwies. In einem Vergleich der Korrelationsmatrizen aus den Untersuchungen 1 und 4 anhand von Multisample-Analysen mit LISREL 8 (Jöreskog & Sörbom, 1996) ergab sich eine sehr gute Modellpassung für ein Modell, das eine identische Korrelationsmatrix für beide Stichproben annimmt, $\chi^2(28, n=229)=27.45, p>.05$ (NFI=.97, RMSEA=.00). Auch die Schätzungen für die internen Konsistenzen der einzelnen Subskalen in Untersuchung 4 (vgl. Tabelle 4, Hauptdiagonale) bewegten sich auf einem zu Untersuchung 1 vergleichbaren Niveau und waren nur in zwei Fällen geringfügig niedriger (bei Sinnhaftigkeit (Sätze) und Textimplikationen, $p < .05$, Überprüfung anhand von z-Tests für Unterschiede in unabhängigen Korrelationen).

Ergebnisse zur Vorwissensabhängigkeit. Entsprechend den Erwartungen waren die hierarchiehohe Subtests, die auf jeweils dieselben Teilprozesse abzielen, untereinander substantiell korreliert, mit einer mittleren Korrelation von .59. Diese Korrelationen waren in der Regel höher als die Korrelationen mit hierarchiehohe Subtests, die auf jeweils andere Teilprozesse abzielen; hier betrug die mittlere Korrelation nur .38. Zur inferenzstatistischen Überprüfung der Hypothese wurden Korrelationsvergleiche für abhängige Korrelationen durchgeführt (vgl. Steiger, 1980). Bei 44 durchgeführten Vergleichen waren in 27 Fällen die Korrelationen zwischen Subtests, die auf dieselben Prozesse abzielen, signifikant höher als die Korrelationen, die auf andere Prozesse abzielen (bei $\alpha = .05$, einseitige Testung). In den übrigen 17 Fällen ergaben sich keine signifikanten Unterschiede. Insbesondere für den Subtest Textimplikationen, der auf dem Testtext zur politologischen Thematik basiert, ergaben sich nicht die erwarteten Korrelationsunterschiede, sondern vergleichsweise hohe Zusammenhänge mit zwei der drei Textverständnis-Subtests. Die Erwartung, daß Subtests,

die auf dieselben Teilprozesse abzielen, ähnliche Korrelationen mit den hierarchieniedrigen Subtests aufweisen, ließ sich dagegen sehr weitgehend stützen: Die drei Textverständnis-Subtests hatten die höchsten, die beiden Fakten/Meinungen-Subtests durchgehend die niedrigsten Zusammenhänge mit den hierarchieniedrigen Subtests. An diesen Verhältnissen ändert sich nichts, wenn Testwerte verwendet werden, die um den Einfluß des inhaltlichen Vorwissens bereinigt sind (vgl. Tabelle 4, rechte Dreiecksmatrix). Die Vorwissensmaße selbst waren in moderater Höhe mit den hierarchiehoher ELVES-Subtests korreliert (von .11 bis .36, mittlere Korrelation .24).

Gesamtdiskussion

Mit ELVES haben wir ein Instrument entwickelt, das eine differenzierte Erfassung der Effizienz von Teilprozessen des Lesens ermöglichen soll, wie sie im Strategiemodell von van Dijk und Kintsch (1983) unterschieden werden. Das Strategiemodell nimmt eine Reihe funktional getrennter hierarchieniedriger (mikrostruktureller) und hierarchiehoher (makrostruktureller) Prozesse an. Die sieben Subtests von ELVES beziehen sich zum Teil auf hierarchieniedrige, zum Teil auf hierarchiehohe Prozesse. Die hier berichteten Studien dienten neben der Ermittlung psychometrischer Kennwerte einer ersten Konstrukt- und Kriteriumsvalidierung des Instruments. Dabei wurden auch Validitätsaspekte der verwendeten integrierten Testwerte untersucht, die mit der Kombination von Antwortlatenz und Antwortrichtigkeit die Effizienz von Teilprozessen des Leseverstehens messen sollen.

Nach unseren Ergebnissen haben sich die integrierten Testwerte zunächst insofern bewährt, als sich auf ihrer Basis reliable Skalenwerte mit günstigen Verteilungseigenschaften bilden ließen. Für die Konstruktvalidität von ELVES spricht, daß sich die nach van Dijk und Kintsch (1983) getroffene Unterscheidung hierarchieniedriger und hierarchiehoher Strategien in den Skaleninterkorrelationen der Subtests abbildet. Ein Modell mit zwei entsprechenden latenten Variablen hat sich in konfirmatorischen Faktorenanalysen gegenüber einem Ein-Faktor-Modell als deutlich überlegen gezeigt. Die hierarchieniedrigen Subtests, die sich auf

grundlegende und teilweise überlappende Prozesse des Leseverstehens beziehen, verfügen dabei über einen höheren Anteil gemeinsamer Varianz als die hierarchiehohen Subtests, mit denen komplexere und stärker divergierende Prozeßkomponenten angezielt sind. Ein beträchtlicher Teil der Ähnlichkeiten zwischen den hierarchiehohen Subtests geht zudem auf die Bewältigung der basalen Strategien zurück, die von den hierarchieniedrigen Subtests erfaßt werden. Dieser Befund mag aus diagnostischer Perspektive auf den ersten Blick ungewöhnlich erscheinen, ist aber gerade vor dem Hintergrund hierarchischer Modelle des Textverstehens plausibel: Lesefähigkeiten, die sich auf das Verstehen zusammenhängender Texte beziehen, lassen sich nicht messen, ohne daß zugleich auch hierarchieniedrige Lesefähigkeitskomponenten mit erfaßt werden. Durch die differenzierte Meßkonzeption von ELVES besteht jedoch prinzipiell die Möglichkeit, Residualwerte für die hierarchiehohen Subtests zu ermitteln, die um den Einfluß hierarchieniedriger Lesefähigkeitswerte bereinigt sind. Der diagnostische Wert eines derartigen „reinen“ Maßes für hierarchiehohe Lesefähigkeitskomponenten ist in weiteren Untersuchungen zu klären.

Belege für die Kriteriumsvalidität des Instruments haben sich in zwei Teilstudien ergeben, in denen als Kriterien die Leistung in klassischen Leseverständnistests aus zwei verschiedenen Inhaltsbereichen und verschiedene Aspekte der Situationsmodellgüte in einer naturalistischen Leseaufgabe erhoben wurden. Für alle ELVES-Subtests fanden sich hier fast durchweg konsistente Zusammenhänge in mittlerer Höhe. Dies ist auch deshalb bemerkenswert, weil die Korrelationen aufgrund der vollkommen unterschiedlichen Konzeption der prozeßorientierten ELVES-Subtests und der verwendeten produktbezogenen Kriterien nicht auf gemeinsame Methodenvarianz attribuierbar sind. Im Sinne der diskriminanten kriterialen Validität ließ sich darüber hinaus zeigen, daß die ELVES-Subtests mit Indikatoren der Verstehensqualität, jedoch nicht oder nur in geringerem Ausmaß mit Maßen der reinen Quantität des Gelernten in Verbindung stehen. Die im Sinne der diskriminanten Validität getroffene Vorhersage, daß sich die ELVES-Subtests nicht zur

Beantwortung schlichter Rekognitionsfragen eignen, ließ sich dagegen nur für eines der beiden Themen stützen, die den kriterialen Leseverständnistests zugrundelagen, für das andere, literaturwissenschaftliche Thema dagegen nicht. Eine mögliche Erklärung dafür könnte sein, daß zur Beantwortung der Rekognitionsfragen für den literaturwissenschaftlichen Testtext nicht nur eine Oberflächenrepräsentation, sondern auch eine propositionale Textbasis hilfreich war. Der Aufbau einer propositionalen Textbasis sollte jedoch durch die von ELVES erfaßten Lesefähigkeiten erleichtert werden. Die integrierten Testwerte, die simultan auf Antwortrichtigkeit und reziprok transformierte Antwortlatenzen zurückgreifen, haben sich im Vergleich zu Antwortrichtigkeit und untransformierter Antwortlatenz allein als insgesamt bessere Prädiktoren von Aspekten der Verstehensgüte erwiesen. Diesen Befund interpretieren wir insgesamt als Indiz für die Validität der zugrundegelegten Effizienzkonzeption.

Ein generelles Problem der Konstruktvalidität, das alle klassischen Lesefähigkeitstests und bei ELVES speziell die hierarchiehohe Subtests betrifft, ist die potentielle Vorwissensabhängigkeit der Testwerte. In einer Untersuchung mit Testtexten aus verschiedenen Inhaltsbereichen und mit Kontrolle des inhaltlichen Vorwissens haben wir erste Belege dafür erbringen können, daß mit den hierarchiehohe Subtests allgemeine Lesefähigkeitskomponenten erfaßt werden, die zudem weitgehend unabhängig vom inhaltlichen Vorwissen sind. Dieser für die Konstruktvalidität von ELVES wichtige Befund parallelisiert die Ergebnisse der Untersuchungen zur Kriteriumsvalidität, bei denen sich das Leseverständnis bei Texten mit sehr unterschiedlichen Themen vorhersagen ließ.

Abgesehen davon, daß ELVES im deutschsprachigen Raum das bislang einzige Lesefähigkeitsdiagnostikum ist, das im Hinblick auf die Zielgruppe geübter erwachsener Leser/innen konstruiert ist, qualifiziert es sich insbesondere als Forschungsinstrument für psycholinguistische und pädagogisch-psychologische Experimente zum Textverstehen. Die Besonderheiten von ELVES gegenüber herkömmlichen Lesefähigkeitstests, wie sie im

angelsächsischen Raum verbreitet sind, liegen in seinem konsequent prozeßorientierten Ansatz und seiner kognitionspsychologisch transparenten Konstruktion. Dadurch ist das Instrument in theoretischer Hinsicht gut in die kognitionspsychologische Forschung zum Textverstehen integrierbar. Eine Einschränkung der hier berichteten Untersuchungen ergibt sich aber daraus, daß die Belege für die Validität des Instruments bislang ausschließlich korrelativer Natur sind. Im Hinblick auf den angezielten Anwendungsbereich des Instruments muß sich ELVES auch in Experimenten zum Textverstehen und zum Lernen aus Texten mit den dort typischerweise verwendeten Online- und Offline-Indikatoren bewähren. Aufgrund der Konzeption der ELVES-Skalen als prozeßorientierte Tests erwarten wir jedoch, daß sich das Instrument in diesem Kontext durchaus bewähren dürfte.

Literatur

Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W.

Schneider, P. Stanat, K. Tillmann & M. Weiß (Hrsg.), PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich (S. 69-137). Opladen:

Leske+Budrich.

Ballif, J. R. & Dibble, W. E. (1973). Anschauliche Physik. Berlin: de Gruyter.

Ballstaedt, S. & Mandl, H. (1984). Zur Beeinflussung der Verstehenstiefe beim Lesen (Forschungsbericht Nr. 27). Tübingen: Universität Tübingen, Deutsches Institut für Fernstudienforschung.

Bisanz, G. L., Das, J. P., Varnhagen, C. K. & Henderson, H. R. (1992). Structural components of reading times and recall for sentences in narratives: Exploring changes with age and reading ability. Journal of Educational Psychology, 84, 103-114.

Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

Bollen, K. A. & Long, J. S. (Eds.) (1993). Testing structural equation models. Newbury Park, CA: Sage.

Brown, J. I., Fishco, V. V. & Hanna, G. S. (1993). Nelson-Denny Reading Test. Itasca, IL: Riverside.

Carpenter, P. A., Miyake, A. & Just, M. A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M. A. Gernsbacher (Ed.), Handbook of psycholinguistics (pp.699-719). San Diego, CA: Academic Press.

Christmann, U., Groeben, N., Flender, J., Naumann, J. & Richter, T. (1999).

Verarbeitungsstrategien von traditionellen (linearen) Buchtexten und zukünftigen (nicht-linearen) Hypertexten. In N. Groeben (Hrsg.), Lesesozialisation in der Mediengesellschaft (S. 175-189). Tübingen: Niemeyer.

Daneman, M. (1997). Individual differences in reading skills. In R. Barr, M. L. Kamil, P. B.

- Mosenthal & P. D. Pearson (Eds.), Handbook of reading research, Vol. 2 (2nd ed., pp. 512-538). Mahwah, NJ: Erlbaum.
- Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, 450-466.
- Educational Testing Service (2001). GRE Graduate Record Examinations: Guide to the use of scores [FTP document]. Princeton, NJ: Educational Testing Service. Available URL: <ftp://ftp.ets.org/pub/gre/992362.pdf> [4. 1. 2002]
- Eysenck, H. J. (1977). Die Zukunft der Psychologie. München: List.
- Fetscher, I. (1991). Toleranz: von der Unentbehrlichkeit einer kleinen Tugend für die Demokratie. Friedrichshafen: Stadt Friedrichshafen.
- Flender, J. & Christmann, U. (2000). Hypertext: prototypische Merkmale und deren Realisierung im Hypertext "Visuelle Wahrnehmung". Medienpsychologie, 12, 94-116.
- Flippo, R. F. & Schumm, J. S. (2000). Reading tests. In R. F. Flippo & D. C. Caverley (Eds.), Handbook of college reading and study strategy research (pp. 403-472). Mahwah, NJ: Erlbaum.
- Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with observable variables and measurement error. Journal of Marketing Research, 18, 39-50.
- Gernsbacher, M. A. & Faust, M. (1991). The mechanism of suppression: A component of general comprehension skill. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 245-262.
- Giddens, A. (1995). Soziologie. Graz: Nausner & Nausner.
- Hacker, W., Veres, T. & Wollenberger, E. (1994). Verarbeitungskapazität für Text: Ergebnisse der Entwicklung eines deutschsprachigen Prüfverfahrens des Arbeitsgedächtnisses. Zeitschrift für Psychologie, 202, 295-320.
- Hasselhorn, M., Schneider, W. & Marx, H. (2000). Früh-, Förder- und Differentialdiagnostik von Leseschwierigkeiten: Eine Einführung. In M. Hasselhorn, W. Schneider & H. Marx (Hrsg.), Tests und Trends 2000 (S. 1-8). Göttingen: Hogrefe.

- Heupst, C. & Hager, W. (1994). Bildhaftigkeit und Konkretheit-Abstraktheit von 100 besonders kurzen und besonders langen Substantiven. In W. Hager & M. Hasselhorn (Hrsg.), Handbuch deutschsprachiger Wortnormen (S. 200-205). Göttingen: Hogrefe.
- Hu, L. T. & Bentler, P. (1995) Evaluating model fit. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 76-99). Thousand Oaks, CA: Sage.
- Institut für Test- und Begabungsforschung (Hrsg.) (1990). Test für medizinische Studiengänge TMS (3. Aufl.). Göttingen: Hogrefe.
- Jöreskog, K. G. & Sörbom, D. (1996). LISREL 8: User's reference guide. Chicago, ILL: Scientific Software.
- Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness. Cambridge: Cambridge University Press.
- Johnston, P. (1984). Prior knowledge and reading comprehension test bias. Reading Research Quarterly, 19, 219-239.
- Just, M. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. Psychological Review, 99, 122-149.
- Kintsch, W. (1998). Comprehension: A paradigm for cognition. Cambridge: Cambridge University Press.
- Lienert, G. A. & Ratz, U. (1994). Testaufbau und Testanalyse. Weinheim: Beltz, Psychologie Verlags Union.
- Long, D. L., Oppy, B. J. & Seely, M. R. (1994). Individual differences in the time course of inferential processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 1456-1470.
- Meyer, B.J.F. (1975). The organization of prose and its effects on memory. Amsterdam: North-Holland.
- Millisecond Software Inc. (1999). Inquisit by Millisecond Software [WWW-document]. Available URL: <http://www.millisecond.com> [14.6.1999]

- Moravcsik, J. E. & Kintsch, W. (1993). Writing quality, reading skills, and domain knowledge as factors in text comprehension. Canadian Journal of Experimental Psychology, 47, 360-374.
- Oakhill, J. V. (1994). Individual differences in children's text comprehension. In M. A. Gernsbacher (Ed.), Handbook of psycholinguistics (pp.821-848). San Diego, CA: Academic Press.
- Oakhill, J. V. & Yuill, N. M. (1986). Pronoun resolution in skilled and less-skilled comprehenders: Effects of memory load and inferential complexity. Language and Speech, 29, 25-37.
- Perfetti, C. A. (1985). Reading ability. New York: Oxford University Press.
- Richter, T. & Christmann, U. (2002). Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In N. Groeben & B. Hurrelmann (Hrsg.), Lesekompetenz: Bedingungen, Dimensionen, Funktionen (S. 25-58). Weinheim: Juventa.
- Rinck, M. (2000). Situationsmodelle und das Verstehen von Erzähltexten: Befunde und Probleme. Psychologische Rundschau, 51, 115-122.
- Schnotz, W. (1994). Aufbau von Wissensstrukturen: Untersuchungen zur Kohärenzbildung beim Wissenserwerb mit Texten. Weinheim: Psychologie Verlags Union.
- Singer, M., Halldorson, M., Lear, J. C. & Andrusiak, P. (1992). Validation of causal bridging inferences in discourse understanding. Journal of Memory and Language, 31, 507-524.
- Spiro, R. J. & Tirre, W. C. (1980). Individual differences in schema utilization during discourse processing. Journal of Educational Psychology, 72, 204-208.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. Psychological Bulletin, 87, 245-251.
- Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. Reading Research Quarterly, 16, 32-71.
- Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. Mc Keown & M. E. Curtis (Eds.), The nature of vocabulary acquisition (pp. 89-105). Hillsdale, NJ: Erlbaum.

van Dijk, T. A. & Kintsch, W. (1983). Strategies of discourse comprehension. New York, NY: Academic Press.

Wolfrum, F. (1995). Zur Epoche Realismus. In Die Deutschen Klassiker [CD-ROM]. München: X Libris.

Autorenhinweis

Tobias Richter, Universität zu Köln; Nadine van Holt, Universität zu Köln

Die Entwicklung des hier vorgestellten Instruments erfolgte im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekts "Wechselwirkungen zwischen Verarbeitungsstrategien von traditionellen (linearen) Buchtexten und zukünftigen (nichtlinearen) Hypertexten" (Leitung: PD Dr. Ursula Christmann und Prof. Dr. Norbert Groeben). Wir danken der DFG für ihre Unterstützung.

Das Instrument steht sowohl als Gesamtest als auch in modularisierter Form mit getrennt einsetzbaren Subtests zur Verfügung. Die Dateien zur Testdarbietung, SPSS-Syntaxfiles zum Einlesen und zur Aufbereitung der Rohdaten sowie detaillierte Hinweise zu Anwendung und Auswertung können über den Erstautor angefordert werden.

Dr. Tobias Richter

Universität zu Köln

Psychologisches Institut, Lehrstuhl Allgemeine Psychologie und Kulturpsychologie

Herbert-Lewin-Straße 2

50931 Köln

E-Mail: tobias.richter@uni-koeln.de

Tabelle 1

Subtests von ELVES mit Beispielitems und Zuordnung zu Prozeßtypen (Strategien bei van Dijk & Kintsch, 1983)

Subtest	Aufgabentyp und Beispielitem	Antwortkategorien	Strategietypen	M_{Worte} (W)	M_{Silben} (W)
Satzverifikation	Beurteilung der Wahrheit begrifflicher Aussagen: <u>Erdbeeren sind eine rote, süß schmeckende Gemüsesorte.</u>	wahr / falsch	propositional	6.6 (4-9)	12.0 (5-16)
Sinnhaftigkeit (Sätze)	Beurteilung der Sinnhaftigkeit von Aussagen: <u>Wenn man verdorbene Speisen isst, verfärbt sich die Haut oft klangvoll.</u>	sinnvoll / sinnlos	propositional, lokale Kohärenzbildung	9.5 (7-12)	19.3 (12-30)
Sinnhaftigkeit (Satzfolgen)	Beurteilung der Sinnhaftigkeit von Satzfolgen (2. Satz ist Testsatz) 1. <u>Die meisten Unfälle auf Autobahnen geschehen im Zusammenhang mit Alkohol.</u> 2. <u>Daher befürworten viele Menschen eine Senkung der Promillegrenze.</u>	sinnvoll / sinnlos	lokale Kohärenzbildung	11.9 (6-17)	21.8 (13-32)
Wortschatz	Beurteilung der Richtigkeit von Bedeutungserklärungen: <u>Der Ausdruck 'Determinante' bedeutet unter anderem 'bestimmender Faktor'.</u>	korrekt / falsch	propositional	10.7 (8-15)	23.1 (19-29)
Textverständnis	Beurteilung von Aussagen (auf Makropropositionsniveau) nach Textlektüre (2 Ausschnitte aus Giddens, 1995, je 250 Worte)	im Text enthalten / nicht enthalten	makropropositional	17.6 (12-22)	33.6 (15-51)
Textimplikationen	Beurteilung möglicher Textimplikationen nach Textlektüre (Ausschnitt aus Eysenck, 1977, 900 Worte)	Implikation / keine Implikation	rhetorisch, makropropositional	10.2 (8-12)	20.8 (18-27)
Fakten/Meinungen	Unterscheiden von Fakten und Meinungen nach Textlektüre (Ausschnitt aus Eysenck, 1977, 900 Worte)	Tatsache/Meinung	rhetorisch, makropropositional	10.7 (7-14)	24.9 (13-26)

Tabelle 2

Untersuchung 1: Skalenkennwerte für Akkuratheit, Silben-Antwortzeiten (aufgeschlüsselt nach Itemtypen und Antworten) und integrierte Testwerte sowie Skaleninterkorrelationen und Faktorladungen im 2-Faktoren-Hypothesenmodell (komplett standardisierte Lösung)

Subtest	Items	Akkuratheit	Latenz/Silbe (ms)	Integrierte Testwerte			Skaleninterkorrelationen unkorrigiert (linke Matrix) / minderungskorrigiert (rechte Matrix)							Ladungen 2-Faktoren- modell		
				<u>M</u> (<u>SD</u>)	<u>M</u> (<u>SD</u>)	<u>M</u> (<u>SD</u>)	<u>Schiefe</u>	α	1	2	3	4	5		6	7
<u>Hierarchieniedrige Subtests</u>																
1 Satzverifikation	15	14.08 (1.03)	265 (94)	18.02 (4.58)	.38	.87	.73**	.77**	.74**	.63**	.46**	.33**		.74		
2 Sinnhaftigkeit (Sätze)	15	14.45 (0.92)	201 (72)	17.43 (5.39)	.82	.93	.65**		.83**	.90**	.78**	.54**	.41**	.89		
3 Sinnhaftigkeit (Satzfolgen)	14	13.41 (0.86)	150 (78)	22.55 (8.46)	1.65	.90	.68**	.76**		.82**	.84**	.62**	.43**	.86		
4 Wortschatz	15	12.06 (1.72)	217 (66)	15.61 (4.79)	.62	.80	.62**	.78**	.70**		.89**	.63**	.59**	.86		
<u>Hierarchiehohe Subtests</u>																
5 Textverständnis	16	11.14 (1.61)	197 (63)	15.63 (4.41)	.37	.68	.48**	.62**	.62**	.66**		.73**	.64**	.87		
6 Textimplikationen	20	12.72 (4.01)	271 (106)	15.95 (7.52)	.61	.91	.41**	.50**	.56**	.54**	.58**		.34**	.68		
7 Fakten/Meinungen	19	12.81 (1.85)	552 (332)	15.63 (7.91)	.45	.75	.27**	.34**	.35**	.28**	.46**	.28**		.48		

Anmerkungen. Akkuratheit: Anzahl richtig beantworteter Items. α : Cronbachs Alpha. $n=154$.

* $p < .05$, ** $p < .01$ (einseitige Testung)

Tabelle 3

Korrelationen und Semipartialkorrelationen (Auspartialisierung von Vorwissen) der ELVES-Subtests mit Kriteriumsvariablen aus Leseverständnistests (Untersuchung 2) sowie der integrierten Testwerte und Latenzen der ELVES-Subtests mit Indikatoren aus der naturalistischen Leseaufgabe (Untersuchung 3)

Subtest	Untersuchung 2: Korrelationen				Untersuchung 2: Semipartialkorr.				Untersuchung 3: Korrelationsvergleiche					
	SDT		Klassik		SDT		Klassik		Integrierte Testwerte			Latenzen (ms)		
	Rekog	Infer	Rekog	Infer	Rekog	Infer	Rekog	Infer	Integration	Stringenz	Anzahl IU	Integration	Stringenz	Anzahl IU
<u>Hierarchieniedrige Subtests</u>														
1 Satzverifikation	.13	.26**	.34**	.34**	.07	.20*	.25**	.26**	.35** ^a	.35** ^a	.23*	-.18*	-.17	-.18*
2 Sinnhaftigkeit (Sätze)	.11	.42**	.47**	.39**	-.02	.31**	.37**	.30**	.46** ^a	.36** ^a	.17	-.26**	-.25**	-.18*
3 Sinnhaftigkeit (Satzfolgen)	.12	.32**	.35**	.39**	.03	.23*	.25**	.31**	.65** ^a	.37** ^a	.12	-.31*	-.20**	-.09*
4 Wortschatz	.12	.40**	.48**	.39**	.00	.30**	.38**	.29**	.37** ^a	.26**	.07	-.24**	-.28*	-.21
<u>Hierarchiehohe Subtests</u>														
5 Textverständnis	.22*	.40**	.31**	.37**	.14	.32**	.22*	.30**	.33**	.37** ^a	.28**	-.23*	-.23*	-.32**
6 Textimplikationen	.24*	.35**	.37**	.42**	.13	.24*	.26**	.34**	.20*	.00	-.23*	-.19	-.25*	-.42**
7 Fakten/Meinungen	-.05	.00	.06	.18*	-.05	.01	.00	.14	.16	.37** ^a	.11	-.10	-.20*	-.26**

Anmerkungen. Untersuchung 2: $n=89$, Untersuchung 3: $n=93$. SDT: Testtext psychologische Thematik (Signalentdeckungstheorie), Klassik: Testtext literaturwissenschaftliche Thematik (Weimarer Klassik). Rekog: Rekognitionsfragen. Infer: Inferenzfragen. Integration: Häufigkeit von Idea Units mit eigenen Vorwissensbeständen, Stringenz: Häufigkeit von Idea Units mit Explikation argumentativ-rhetorischer Relationen oder Signaling, Anzahl IU: Anzahl inhaltsbezogener Idea Units insgesamt. ^a Betrag der Korrelation höher als Betrag der entsprechenden Korrelation für Antwortlatenz.

$p < .05$ (einseitige Testung). * $p < .05$, ** $p < .01$ (einseitige Testung).

Tabelle 4

Skaleninterkorrelationen (linke Dreiecksmatrix), Semipartialkorrelationen unter Auspartialisierung des inhaltlichen Vorwissens (rechte Dreiecksmatrix) und interne Konsistenzen (Hauptdiagonale, Cronbachs α) der Subtests in der erweiterten Fassung von ELVES mit verschiedenen Textthemen für die hierarchiehohe Subtests

		Skaleninterkorrelationen ohne/mit Auspartialisierung von Vorwissen und interne Konsistenzen (Cronbachs α)										
	Items	1	2	3	4	5a	5b	5c	6a	6b	7a	7b
1 Satzverifikation	15	(.86)	—	—	—	.54	.52	.64	.36	.53	.01	.22
2 Sinnhaftigkeit (Sätze)	15	.74	(.92)	—	—	.61	.62	.71	.47	.43	.11	.18
3 Sinnhaftigkeit (Satzfolgen)	14	.79	.70	(.84)	—	.57	.55	.64	.37	.46	.11	.27
4 Wortschatz	15	.75	.86	.71	(.81)	.59	.54	.70	.45	.38	.06	.19
5a Textverständnis: Soziologie	16	.54	.64	.56	.63	(.71)	<u>.49</u>	<u>.63</u>	.41	.44	.27	.39
5b Textverständnis: Literatur	8	.49	.62	.55	.53	<u>.51</u>	(.66)	<u>.62</u>	.28	.45	.20	.29
5c Textverständnis: Physik	8	.64	.69	.65	.68	<u>.62</u>	<u>.59</u>	(.70)	.43	.55	.09	.18
6a Textimplikationen: Psychologie	20	.41	.54	.42	.52	.47	.34	.48	(.82)	<u>.44</u>	.24	.35
6b Textimplikationen: Politologie	14	.58	.52	.54	.47	.46	.51	.56	<u>.51</u>	(.73)	.35	.47
7a Fakten/Meinungen: Psychologie	20	.07	.23	.17	.21	.31	.29	.15	.31	.39	(.81)	<u>.64</u>
7b Fakten/Meinungen: Politologie	16	.25	.26	.31	.28	.38	.32	.16	.43	.48	<u>.70</u>	(.79)

Anmerkung. $n = 73$. Alle Korrelationen $> .05$ signifikant bei $\alpha = .05$. Korrelationen von Subtests, die auf dieselben Teilprozesse abzielen, sind kursiv gesetzt.