

Zeitbegrenzungen in Lesetests. Auswirkungen auf das Testkonstrukt, testmethodische Konsequenzen und didakti- sches Potential am Beispiel der VERA-8- Leseaufgaben im Fach Französisch

**Bettina Neugebauer,¹
Rüdiger Grotjahn² und Bernd Tesch³**

The measurement of reading competence has been a major focus of both international and national evaluation studies such as PISA and VERA-8 (eight-grade comparative study in German Schools). However, little research has investigated how task-specific time limits affect reading comprehension. The present study examines the effect of time constraints on French VERA-8 reading comprehension tasks. 36 ninth-graders from various schools in Berlin completed a computer-based French reading comprehension assessment in two versions: One with and the other without time pressure. Other measures were the German d2 attention test, the German LGVT 6-12 test of reading speed and comprehension, and a French C-Test. The results of the study provide us with new insights into the construct of reading comprehension and can be used to develop tests, which are better aligned with a construct comprising the ability to process texts under time-constraints.

1. Einleitung

Lesekompetenz lässt sich anhand einer Vielzahl von Merkmalen charakterisieren. Ein wichtiges Merkmal ist die Lesegeschwindigkeit. Leser können ihren Leseprozess verlangsamen oder beschleunigen. Geübte Leser⁴ passen ihr Zeitmanagement flexibel den Lesezielen an. Sie lesen z.B. deutlich

1 Korrespondenzadresse: Bettina Neugebauer, M.A., Institut zur Qualitätsentwicklung im Bildungswesen (IQB), Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Tel. 030-2093 46570, E-Mail: bettina.neugebauer@iqb.hu-berlin.de

2 Korrespondenzadresse: Prof. Dr. Rüdiger Grotjahn, Ruhr-Universität Bochum, Fakultät für Philologie, Seminar für Sprachlehrforschung, D-44780 Bochum, Tel. 0234-32 28729, E-Mail: ruediger.grotjahn@rub.de

3 Korrespondenzadresse: Dr. Bernd Tesch, Institut zur Qualitätsentwicklung im Bildungswesen (IQB), Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Tel. 030-2093 46531, E-Mail: bernd.tesch@iqb.hu-berlin.de

4 Zur Leseerleichterung wird bei Personenbezeichnungen lediglich die maskuline Form verwendet.

schneller, wenn sie sich lediglich einen Überblick über den Inhalt einer Passage verschaffen wollen, als wenn sie sich möglichst viele Details aus einem Text merken möchten. Hoch kompetente Leser unterscheiden sich damit von weniger kompetenten Lesern nicht nur in der letztendlich erreichten Versteheleistung, sondern auch in der Schnelligkeit, mit der eine Versteheleistung erbracht wird.

Angesichts der Bedeutung zeitabhängiger Verstehensprozesse in authentischen Lesesituationen wird auch in vielen fremdsprachlichen Lesetests bei der Testentwicklung zwischen langsamen und schnellen Leseprozessen und Leseaktivitäten unterschieden und versucht, diesen Unterschied in valider Weise zu operationalisieren. Dies gilt v.a. im Hinblick auf Fremdsprachentests für den akademischen Kontext, wie dem neuen *TOEFL iBT* oder auch dem *Pearson Test of English Academic*, da in vielen akademischen Situationen sowohl das sorgfältige Lesen als auch die Fähigkeit zur schnellen Informationsentnahme wichtige Teilkompetenzen darstellen.

Bevor wir in Abschnitt 4 eine eigene explorative Studie zu den Auswirkungen von Zeitbegrenzungen in Lesetests vorstellen, möchten wir uns zunächst etwas ausführlicher mit der Bedeutung des Faktors Zeit beim Lesen auseinandersetzen.

2. Der Faktor Zeit beim Lesen

2.1 Lesen in der Muttersprache

Innerhalb der Lesetheorie wird u.a. zwischen hermeneutischen, rezeptionsästhetischen und kognitionspsychologischen Ansätzen unterschieden (vgl. Ehlers 2007). Traditionelle hermeneutische Ansätze zielen auf die Rekonstruktion der Autorenintention. Die Rezeptionsästhetik dagegen leitet die Hinwendung zum Leser als konstruktive, sinnbildende Instanz ein (ebd.: 108). Die moderne kognitionspsychologische Leseforschung fokussiert die Verarbeitungsprozesse und die zum Einsatz kommenden Kompetenzen und modelliert Verstehen im weitesten Sinne als Informationsverarbeitung und Informationsentnahme (ebd.: 115). Der wesentliche Vorteil dieser empirischen Forschungsrichtung besteht darin, dass relevante Teilkompetenzen durch spezifische Aufgaben so präzise wie möglich erfasst werden und eine Konfundierung mit anderen Fähigkeitsaspekten so weit wie möglich vermieden wird. Die kognitionspsychologische Lesetheorie stellt die Begriffe Aufmerksamkeitslenkung, Relevanzfilterung, Inferenzbildung und Arbeitsge-

dächtnis in den Mittelpunkt des Interesses. Lesen und lesendes Verstehen werden damit auch in ihrer Zeitlichkeit erfasst.

Das Zeitmanagement beim Lesen kann im Lesekompetenzmodell von Artelt et al. (2007), das in Abb. 1 wiedergegeben ist, im Bereich der Aktivitäten des Lesers verortet werden. Diese Aktivitäten sind komplex und individuell. Ob jemand schnell oder langsam liest, hängt insbesondere in der Fremdsprache neben den jeweiligen Lesezielen sowohl von Faktoren wie der Verfügbarkeit der für das Verständnis des Textes und der Items notwendigen lexikalischen und grammatikalischen Kenntnisse als auch von Persönlichkeitsmerkmalen wie der Konzentrationsfähigkeit ab. Insbesondere mit Blick auf das Lesen in der Fremdsprache besteht hier noch erheblicher Forschungsbedarf. So müssten z.B. die Nutzung des Vorwissens, das Strategiewissen, die Fähigkeit zur Selbstregulation, Motivationen und Einstellungen noch näher untersucht werden. Relativ detaillierte Hinweise zur Textseite des Modells (Beschaffenheit des Textes) finden sich u.a. in den didaktischen Handreichungen zu den Vergleichsarbeiten der achten Jahrgangsstufe (VERA-8).

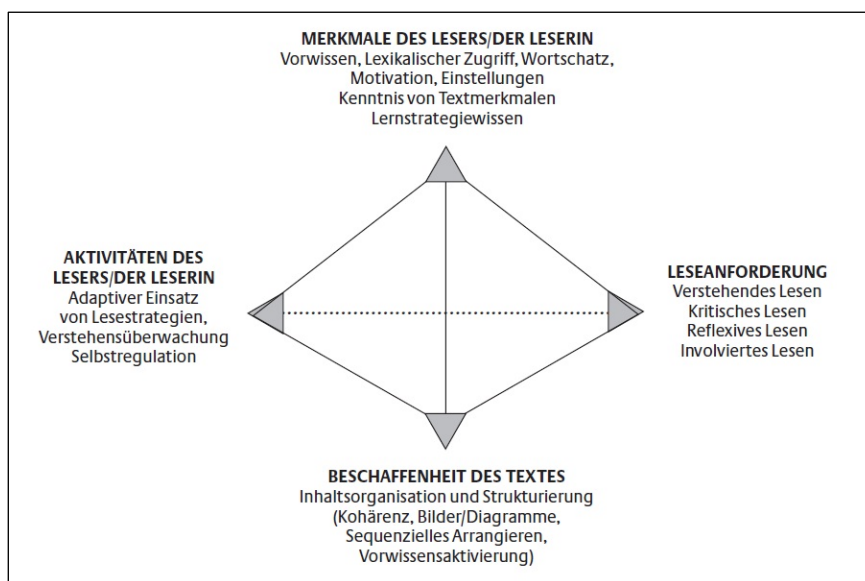


Abbildung 1: Determinanten der Lesekompetenz
(Artelt, McElvany, Christmann, Richter, Groeben, Köster, Schneider, Stanat,
Ostermeier, Schiefele, Valtin, Ring & Saalbach 2007: 12)

Um Kompetenzunterschiede bei schnellen und langsamen Leseaktivitäten valide messen zu können, ist auf der Ebene der Konstruktdefinition möglichst präzise sowie theoretisch und empirisch abgesichert zu beschreiben, was unter schnellen und langsamen Leseaktivitäten zu verstehen ist.

Bereits seit den 1950er Jahren wird bei muttersprachlichen Lesern intensiv zum Thema Lesegeschwindigkeit geforscht, und zahlreiche Trainingsprogramme versprechen durch gezieltes Erlernen von Schnell-Lese-Strategien ein schnelleres Lesen und insgesamt eine Erhöhung der Lesekompetenz. So argumentieren z.B. Musch & Rösler (2011), dass die durchschnittliche Lesegeschwindigkeit von 250 wpm (Wörtern pro Minute) mit Hilfe von Training deutlich gesteigert werden kann. Wood (1960) untersuchte sogar Leserinnen und Leser, die nach Angaben der Autorin ein Lesetempo zwischen 1500 und 6000 wpm erreichen konnten. Andere Autoren zweifeln dies jedoch an, da aus physischen Gründen, genauer der Wortfixierung und der Blickbewegung, keine Geschwindigkeiten von mehr als 900 wpm möglich seien (Spache 1962). Vor dem Hintergrund, dass Schlagzeilen zu Schnelllesern häufig einer empirischen Basis entbehren, untersuchte Carver (1985) 16 Personen, die aufgrund ihres außergewöhnlich hohen Lesetempos aufgefallen waren. Anhand von unterschiedlichen Tests erfasste Carver die Lesegeschwindigkeit sowohl separat als auch im Zusammenhang mit dem Leseverständnis. Dabei identifizierte er u.a. einen Probanden mit einer Lesegeschwindigkeit von 1262 wpm in einem Test ohne Kontrolle des Verständnisses. Bei gleichzeitiger Kontrolle des Leseverständnisses sank die Lesegeschwindigkeit jedoch signifikant und rapide. Carver (1985: 418) schließt insgesamt, dass bei einem Verständnisgrad von mindestens 75% selbst außergewöhnliche Schnellleser keine Werte über 600 wpm erzielen können. Dies zeigt, dass die Messung der Lesegeschwindigkeit wenig aussagekräftig ist, wenn nicht gleichzeitig das Leseverständnis erfasst wird (vgl. auch Musch & Rösler 2011).

In späteren Arbeiten bestimmte Carver dann im Rahmen seiner *Rauding*-Theorie die Abhängigkeit der Lesegeschwindigkeit von der Zielsetzung des Lesens experimentell genauer. Er konnte bei seinen studentischen Versuchspersonen zwischen fünf verschiedenen Leseaktivitäten unterscheiden – von ihm metaphorisch als *reading gears* bezeichnet –, die sich jeweils anhand des Durchschnitts gelesener Wörter pro Minute charakterisieren ließen. Der 5. Gang und damit die schnellste Aktivität ist das überfliegende Scanning einzelner Wörter mit 600 wpm. Der 4. Gang ist das *Skimming* mit durchschnittlich 450 wpm, bei dem über ein Scanning hinaus auch Beziehungen zwischen Wörtern und Wortgruppen hergestellt werden. Der 3., für

normales Lesen typische Gang ist mit 300 wpm das satzübergreifende *Rau- ding*. Die beiden unteren Gänge sind *Learning* mit 200 wpm und *Memori- zing* mit 138 wpm (vgl. Carver 1997: 6; Nix 2011: 84f.).

Auch im Rahmen von PISA 2000 wird der Schnelligkeit der Informati- onsverarbeitung beim Lesen von Texten eine vergleichsweise hohe Bedeu- tung zugewiesen. So betonen z.B. Artelt, Stanat, Schneider & Schiefele (2001) auf der Grundlage ihrer regressionsanalytischen Auswertungen der Daten zur Lesekompetenz im Deutschen bei 15jährigen Schülern, dass ne- ben Variablen wie kognitive Grundfähigkeit, Lernstrategiewissen, inhaltli- ches Vorwissen und Interesse die Fähigkeit zum schnellen Dekodieren eine zentrale Erklärungsvariable für die nachgewiesenen Leistungsunterschiede ist. Zusammenfassend stellen sie hierzu fest:

Eine ebenfalls vorhersagemächtige Variable ist die Decodierfähigkeit der Schülerin- nen und Schüler, die sich in der Schnelligkeit des Erfassens der korrekten Bedeutung von Sätzen eines längeren Textes äußert. Für einen effektiven Umgang mit Texten ist hohe Decodierfähigkeit unter anderem deshalb förderlich, weil durch schnelleres Les- sen Ressourcen für eine tiefere Verarbeitung des Textes zur Verfügung stehen (Artelt et al. 2001: 128).

Zur Messung der Dekodierfähigkeit wurde in PISA 2000 ein Lesege- schwindigkeitstest eingesetzt, bei dem ein Text innerhalb von sieben Minu- ten so schnell wie möglich zu lesen war. In Abständen von ca. drei bis sechs Sätzen waren jeweils drei Wörter eingefügt, aus denen das in den Kontext passende Wort auszuwählen war. Gemessen wurde die Dekodierfähigkeit anhand der Lesegeschwindigkeit gewichtet mit der Zahl der korrekt ausge- wählten Wörter.

Nix (2011: 46f.) charakterisiert dieses weite Verständnis von schneller Dekodierfähigkeit in PISA 2000 folgendermaßen:

Um im Testverfahren passende Wortalternativen nach den jeweiligen Textblöcken auswählen zu können, müssen – zusätzlich zu einer exakten De- und Rekodierung – Satzpropositionen zugeordnet und Bedeutungen über die Satzgrenzen hinweg mitei- nander mental vernetzt werden. Die Lesegeschwindigkeit kann dabei anzeigen, in- wie weit diese Prozesse automatisiert von den Jugendlichen vollzogen werden können. Schüler, die ungenau, unautomatisiert und im Satzzusammenhang disfluent lesen, werden weniger Items in der vorgegebenen Zeit bearbeiten können als Mit- schüler, die die hierarchieniedrigeren Lesefertigkeiten automatisiert haben und flüs- sig im Satzzusammenhang lesen können.

Für Nix (2011) ist damit die Lesegeschwindigkeit ein wichtiger Indikator für den Grad der Automatisierung der Informationsverarbeitung beim ver-

stehenden Lesen von Texten und damit zugleich ein zentraler Indikator für Unterschiede in der Lesekompetenz.⁵

In Ergänzung der Analysen von Artelt et al. (2001) konnten Stanat & Schneider (2004) zudem nachweisen, dass die Dekodierfähigkeiten bei Schülern mit einer geringen Lesekompetenz im Deutschen sowohl bei Jugendlichen ohne als auch mit Migrationshintergrund unterdurchschnittlich ausgeprägt waren, wobei Jugendliche mit Migrationshintergrund jedoch insgesamt deutlich weniger entwickelte Dekodierfähigkeiten zeigten als Jugendliche ohne Migrationshintergrund (ebd.: 267). Auch in dieser clusteranalytischen Studie erwies sich damit die Fähigkeit zum schnellen Erfassen der korrekten Bedeutung auf Wort- und Satzebene als eine wichtige Komponente von Lesekompetenz.

Die Entwicklung von Leseflüssigkeit wird zudem von einer Vielzahl von individuellen Lernervariablen beeinflusst wie z.B. Alter und Bildungsstand (ebd.: 293). Außerdem konnten v.a. Walczyk und Mitarbeiter in einer Reihe von Studien zeigen, dass sowohl Kinder als auch Erwachsene Defizite bei der Worterkennung und beim verbalen Arbeitsgedächtnis z.B. durch Pausieren oder auch erneutes stilles oder lautes Lesen bestimmter Wörter oder Passagen ausgleichen und auf diese Weise zu ähnlichen Ergebnissen in Leseverständnistests kommen konnten wie Leser ohne die entsprechenden Defizite (vgl. z.B. Walczyk, Marsiglia, Johns & Bryan 2004). Dies impliziert zugleich, dass anhand von Leseverständnistests ohne Zeitbegrenzung die Fähigkeit zum Leseverstehen unter zeitlimitierten Bedingungen tendenziell überschätzt wird. Dies spricht dafür, beim Testen des Leseverstehens auch Aufgaben einzusetzen, bei denen ein gewisser Zeitdruck aufgebaut wird (vgl. auch Grabe 2009: 300). Zudem konnten Walczyk, Kelly, Meche & Braud (1999) zeigen, dass der Aufbau milden Zeitdrucks die Achtsamkeit (im Sinne von engl. *mindfulness*) erhöht und auf diese Weise zu einer generell besseren Verstehensleistung führt. Der Aufbau eines deutlichen Zeitdrucks ist allerdings nicht unproblematisch (vgl. Grotjahn 2010). Ein Zeitdruck kann Stress und Testangst produzieren, so dass die "wahren" Fähigkeiten der Testperson überdeckt werden. Insbesondere im Schulkontext sollte der Zeitfaktor deshalb nicht zu stark erhöht werden.

5 Vgl. auch den vielfach nachgewiesenen Zusammenhang zwischen Schnelligkeit der Informationsverarbeitung und Intelligenz. So stellen z.B. Sheppard & Vernon (2008: 535) anhand einer Metaanalyse von 172 Studien fest: "The results indicate that measures of intelligence are significantly correlated with mental speed and that for some measures this relationship shows a trend toward strengthening as the complexity of the speeded tasks increase".

2.2 Lesen in der Fremdsprache

Die bisherigen Ausführungen fokussieren das Lesen in der Muttersprache. Zum Lesen in der Fremdsprache liegt deutlich weniger forschungsbasierte Literatur vor, besonders in Hinblick auf die Leseflüssigkeit im Sinne schnellen Lesens mit ausreichendem Verständnis (vgl. Grabe 2009: 289; Grabe 2010). Zudem spielen beim Lesen in der Fremdsprache auch noch eine Vielzahl weiterer Faktoren eine wichtige Rolle, wie z.B. lexikalische und grammatikalische Kenntnisse in der L2 oder auch die Literalität in der L1 (vgl. z.B. Bernhardt 2011; Grabe 2009; Jung 2009; Van Gelderen, Schoonen, Stoel, De Glopper & Hulstijn 2007). Zudem ist es nicht ungewöhnlich für einen Fremdsprachenleser, eine Sprache gut zu verstehen, sie aber im Vergleich zu einem Muttersprachler nicht hinreichend flüssig lesen zu können (vgl. Grabe 2009: 289f.). Empirische Hinweise hierzu liefert z.B. die Studie von Shaw & McMillion (2008). Die Autoren kamen bei einem Vergleich der englischen Lesekompetenz von Studierenden mit Schwedisch bzw. Englisch als Muttersprache zu folgendem Ergebnis: Wurden in der Studie die Lesetests ohne Zeitdruck administriert, ergaben sich keine signifikanten Mittelwertunterschiede zwischen den beiden Gruppen hinsichtlich der Korrektheit des Verstehens. Wurden die Tests jedoch unter Zeitdruck bearbeitet, zeigten die Muttersprachler (deutlich) höhere Verstehensleistungen als die Nichtmuttersprachler (vgl. hierzu auch die Untersuchung von Fraser 2007 sowie die Ausführungen in Hulstijn 2011b). Shaw & McMillion (2008) weisen in diesem Zusammenhang unter Verweis auf weitere Studien darauf hin, dass tendenziell grundsätzlich langsamer in der L2 gelesen werde als in der L1 – sogar bei hochkompetenten Bilingualen. Die Tatsache, dass in ihrer Untersuchung einige schwedische Probanden sogar höhere Verstehensleistungen als die vergleichbaren britischen Probanden erreicht haben, führen die Autoren ähnlich wie Walczyk et al. (2004) auf die Möglichkeit eines kompensatorischen Effekts der L1-Lesekompetenz zurück:⁶

[...] unlike the weak L1 readers, many of the L2 readers are efficient L1 readers and have functional compensatory behaviors and strategies from their L1 processing that can be used for dealing with various types of reading difficulties in the L2 (Shaw & McMillion 2008: 129).

6 Vgl. die in Abschnitt 2.1 erwähnten Studien von Walczyk und Mitarbeitern (z.B. Walczyk et al. 2004) zu Kompensationsprozessen beim Lesen in der Muttersprache. Vgl. weiterhin Hulstijn (2011a) zum Zusammenhang zwischen Literalität in L1 und Lesen in einer L2 sowie die *Compensatory Theory of Second Language Reading* als theoretische Grundlage der Monografie von Bernhardt (2011).

Für Grabe (2009) ist die Leseflüssigkeit ein zentraler Bestandteil der Lesekompetenz. Der Autor beschreibt drei zentrale Teilfertigkeiten von Leseflüssigkeit: *automaticity* (Automatizität), *accuracy* (Genauigkeit/Korrektheit) und *rate* (Lesegeschwindigkeit), die beim Lesen in der Fremdsprache wichtig sind. *Automaticity* bezieht sich auf schnelle, wenig Verarbeitungskapazität verlangende, unbewusste und damit zugleich auch wenig kontrollierbare Prozesse, *accuracy* dagegen auf eine vollständige und akkurate Worterkennung (vgl. ebd.: 291 sowie auch Grabe 2010; Rasinski, Reutzel, Chard & Linan-Thompson 2011; Segalowitz 2010).

Zum schnellen Lesen können u.a. folgende drei Typen gezählt werden (vgl. Grotjahn & Tesch 2010; Leucht, Retelsdorf, Möller & Köller 2010). Das suchende Lesen umfasst die Lokalisierung von inhaltlichen Informationen zu ausgewählten Fragestellungen, z.B. das Erfassen von Hauptinformationen eines Textes. Ein Spezialfall des suchenden Lesens ist das identifizierende Lesen im Sinne einer schnellen Lokalisierung von bestimmten Zeichenketten, z.B. Jahreszahlen, Namen oder Farben, im Text. Unter orientierendem Lesen versteht man, sich einen möglichst schnellen Überblick über einen Text zu verschaffen, z.B. durch die Erfassung von Überschriften und Grafiken. Um einen Text binnen kürzester Zeit möglichst effektiv bearbeiten zu können, ist zudem metalinguistisches und metakognitives Wissen erforderlich, das es ermöglicht, Lesestrategien bewusst einzusetzen und den Leseprozess zu steuern. Zudem wird der Leseprozess auch von früheren Erfahrungen geleitet, z.B. durch Vorwissen über Themen oder Textsorten (vgl. auch Alderson 2000; Lutjeharms 2007).

Im Hinblick auf die Frage, welche kognitiven und sprachlichen Kompetenzen dem schnellen suchenden Lesen zugrunde liegen, gibt es deutlich divergierende Auffassungen. Während v.a. Guthrie (vgl. Guthrie 1988; Guthrie & Kirsch 1987; Guthrie & Mosenthal 1987) schnelles suchendes Lesen und Leseverstehen als zwei unabhängige Dimensionen textueller Literalität sieht, gehen die Autoren des *TOEFL 2000 Reading Framework* unter Bezug auf Autoren wie Carver (1997) und Perfetti (1994) von einer hierarchischen Beziehung zwischen Prozessen des schnellen suchenden Lesens und komplexeren höheren Verstehensprozessen aus.⁷ Dabei sehen sie für das suchen-

7 Siehe auch Leucht et al. (2010), die zeigen konnten, dass sich selektives und genaues Leseverstehen hinsichtlich der mittleren Aufgabenschwierigkeit unterscheiden: Selektives Leseverstehen im Sinne eines schnellen *scanning* und *search reading* konnte tendenziell eher den Stufen A1 und A2 des Gemeinsamen europäischen Referenzrahmens für Sprachen (GeR, Europarat 2001) und damit der elementaren Sprachverwendung, das langsamere, genaue Lesen eher der kompetenten Sprachverwendung (GeR-Stufen C1 und C2) zugeordnet werden. Allerdings wurde in der Studie die Bearbeitungszeit nicht kontrolliert.

de Lesen die Fähigkeit zur schnellen und effizienten Worterkennung als essentiell an (vgl. Enright, Grabe, Koda, Mosenthal, Mulcahy-Ernt & Schedl 2000; Jamieson, Eignor, Grabe & Kunnan 2008: 71).⁸

Um wirklich schnell und effektiv lesen zu können, muss das Ziel des Lesens (z.B. die Hauptaussage des Textes erfassen) deutlich sein. Die Kenntnis des Ziels ermöglicht die Wahl einer passenden Strategie, wie z.B. den Text zu überfliegen und nach Schlüsselwörtern zu suchen. Entsprechend wird auch im *TOEFL 2000 Reading Framework* und in der Folge im neuen TOEFL iBT das Leseziel als entscheidende Einflussvariable auf den Leseprozess angesehen (vgl. Jamieson et al. 2008: 69-74).

Im Zusammenhang mit den Lesezielen ist die Positionierung der Items in einem Lesetest ein wichtiger Aspekt. Durch eine Voranstellung der Items⁹ vor den Stimulus-Text könnten die Schüler ermutigt werden, eine Entscheidung über die zu verwendende Lesestrategie zu treffen. Sind die Items nicht vorangestellt, haben Schüler mit Testerfahrung (*testwiseness*) unter Umständen den Vorteil, auf Grund ihrer Erfahrung das Item vor dem Stimulus-Text zu lesen und somit die bestmögliche Lesestrategie zu wählen. Eine Voranstellung der Items hingegen würde auch weniger testerfahrenen und unter Umständen auch insgesamt leistungsschwächeren Schülern die Möglichkeit geben, ihren Leseprozess besser zu steuern. Schnelles Lesen bedarf zudem eines gewissen Zeitdrucks (Grotjahn & Tesch 2010) bzw. einer Kontrolle der Darbietungszeit, um Aussagen über den Einsatz der intendierten Lesestrategien treffen zu können (Weir & Khalifa 2008). Insgesamt gilt, dass bei einer deutlich zeitlimitierten Aufgabendarbietung stets zu prüfen ist, ob die Zeitlimitierung das Testkonstrukt in relevanter oder irrelevanter Weise verändert. Konstruktrelevant ist eine Zeitlimitierung dann, wenn die Geschwindigkeit auch bei der Bearbeitung kriterialer Aufgaben außerhalb der Testsituation eine wichtige Rolle spielt (vgl. hierzu bereits Mollenkopf 1960). Bezogen auf Kursabschluss-tests hat u.a. schon Wesman (1960: 272) diese Voraussetzung treffend folgendermaßen formuliert: "Speed is a legitimate element in achievement tests only when the *same kind of speed which the test is measuring* is an objective of the course" (Hervorhebung im Original).

8 Vgl. auch die Untersuchungen von Shiotsu (2010) zu den Komponenten der Lesekompetenz japanischer Englischler. Den besten Erklärungsansatz lieferte dort ein Strukturgleichungsmodell mit zwei latenten Faktoren, und zwar sorgfältige Textverarbeitung sowie Schnelligkeit und Effizienz des lexikalisch-semantischen Zugriffs.

9 Bei der Messung der Fähigkeit zu schnellem, sinnentnehmendem Lesen wird häufig ein Stimulus-Text mit nur einem einzigen Item eingesetzt (vgl. das Beispiel "Vélib" in Abschnitt 6). Um die Möglichkeit eines Stimulus-Textes mit mehreren Items nicht auszuschließen, gebrauchen wir hier und auch an weiteren Textstellen den Plural.

Eindeutig konstruktrelevant dürfte eine Zeitlimitierung z.B. in einem Lesetest für Fluglotsen sein, wenn in der zielsprachlichen Verwendungssituation schnell Anweisungen zu lesen sind und ein inadäquates Leseverstehen zudem mit gravierenden Konsequenzen verbunden ist. Auch im universitären Kontext ist die schnelle lesende Informationsverarbeitung eine wichtige Teilkompetenz und sollte deshalb in kontextvalider und kognitiv valider Form getestet werden. Ähnlich sind auch im Schulkontext Situationen denkbar, in denen der Faktor Geschwindigkeit konstruktrelevant ist, so z.B. wenn man testen möchte, inwieweit Lernende in der Lage sind, sich rasch einen Überblick zu einem bestimmten fremdsprachlichen Themenbereich zu verschaffen.

3. Bearbeitungszeit beim Testen

3.1 *Power-Tests* vs. *Speed-Tests*

Nach der Veröffentlichung der Gesamtstrategie zum Bildungsmonitoring (KMK 2006) haben Tests in Deutschland verstärkt Einzug in die Klassenzimmer gehalten. Ziel ist es, Kompetenzen möglichst umfassend und zugleich möglichst reliabel zu messen. Dies gelingt tendenziell umso besser, je umfangreicher der Test ist, d.h. je mehr Aufgaben eingesetzt werden können. Aus testökonomischer Perspektive sollte jedoch zugleich die Gesamtzeit für die Testung möglichst gering sein. Die Festlegung der Bearbeitungszeit für den Gesamtest oder auch einzelne Aufgaben hat damit nicht nur potentielle Auswirkungen auf die Validität eines Tests, sondern ist auch ein wichtiger Aspekt der Testökonomie (vgl. hierzu auch Kalyuga 2013).

Im Hinblick auf die Bearbeitungszeit ist die Unterscheidung von *Power-Tests* (Niveautests) und *Speed-Tests* (Geschwindigkeitstests) grundlegend (vgl. zum Folgenden die tiefergehenden Ausführungen in Grotjahn 2010: 268-271). Sind die Aufgaben eines Tests so schwierig, dass nur ein Teil korrekt gelöst wird, und ist die Bearbeitungszeit für den Test zugleich so großzügig bemessen, dass (fast) alle Teilnehmer alle Aufgaben bearbeiten können, handelt es sich um einen reinen Niveautest (*Power-Test*). Durch eine Verkürzung der Gesamtbearbeitungszeit kann ein Zeitdruck aufgebaut werden und damit eine Geschwindigkeitskomponente in den jeweiligen Lesetest eingeführt werden. Ist die Geschwindigkeit der Informationsverarbeitung Teil des Testkonstrukts, kann der Aufbau von Zeitdruck, wie bereits erwähnt, die Validität des jeweiligen Lesetests erhöhen. Dabei ist jedoch zu

beachten, dass ein deutlicher Zeitdruck auch zu konstruktirrelevanter Varianz z.B. aufgrund von Testangst führen kann (vgl. auch Bridgeman, McBride & Monaghan 2004).

Ein Niveautest mit einer Zeitkomponente wird im Englischen auch als *Speeded Test* bezeichnet. Die im Abschnitt 4.3.2 beschriebenen zeitbegrenzten VERA-Aufgaben fallen unter diese Kategorie. Zeitbegrenzte Niveautests sind zu unterscheiden von (reinen) *Speed-Tests* (Schnelligkeitstests). Bei reinen *Speed-Tests* werden den Testpersonen sehr einfache Aufgaben vorgelegt, die innerhalb einer bestimmten Zeit gelöst werden müssen und die ohne Vorgabe einer Zeitbeschränkung von den Testpersonen auch alle gelöst werden könnten. Ziel ist die Messung der Schnelligkeit bei der Bearbeitung entsprechender Aufgaben.

Eine oft zitierte frühe Studie zum Einfluss der Bearbeitungszeit auf die Testleistung und das Testkonstrukt ist Lord (1956). Der Autor untersuchte 649 Studierende an der *United States Naval Academy* in Annapolis in den Bereichen Vokabular, räumliches Vorstellungsvermögen und numerisches Schlussfolgern mit Hilfe einer Vielzahl von unterschiedlich stark zeitbegrenzten (experimentellen) Tests. Mit Hilfe von Faktorenanalysen konnten insgesamt vier *Speed-Faktoren* extrahiert werden, die jeweils positiv mit der Studienleistung gemessen anhand von Noten korrelierten. Lord (1956: 49) interpretiert diesen Sachverhalt folgendermaßen:

All correlations between course grades and the four speed factors, with one small exception, were found to be positive, although not large. It is to be concluded that speed of various kinds plays some part in the course grades studies, and that speededness in the admissions examinations is to this extent justified.

3.2 Gesamtdarbietung vs. Einzeldarbietung der Aufgaben

Im Hinblick auf die Operationalisierung des Unterschieds zwischen schnellen und langsamen Leseprozessen ist der Unterschied zwischen Gesamtdarbietung und Einzeldarbietung und die damit verbundenen Möglichkeiten der Steuerung der Bearbeitungszeit von grundlegender Bedeutung (vgl. zum Folgenden Grotjahn 2010). Bei einer *Gesamtdarbietung* werden den Testteilnehmern alle Aufgaben eines Lesetests¹⁰ gleichzeitig dargebo-

10 Wir beziehen uns im Folgenden speziell auf Lesetests. Die Aussagen zu Gesamtdarbietung vs. Einzeldarbietung gelten jedoch prinzipiell auch für andere Testformen wie z.B. C-Tests (vgl. Grotjahn 2010).

ten. Die Teilnehmer können dann selbst entscheiden, in welcher Reihenfolge sie die Aufgaben bearbeiten, wie viel Zeit sie auf jede einzelne Aufgabe verwenden, ob sie bestimmte Aufgaben auslassen und ob sie eventuell zu bereits bearbeiteten Aufgaben zurückkehren. Wird ein Test im Papier-Bleistift-Format administriert, ist die beschriebene Gesamtdarbietung die häufigste Durchführungsform. Insbesondere bei einer computer-basierten Testdurchführung wird die Möglichkeit, zu bereits bearbeiteten Items zurückzukehren, allerdings zuweilen eingeschränkt.

Bei einer zeitlimitierten Gesamtdarbietung kann jedoch nicht kontrolliert werden, ob die Testteilnehmer z.B. bei Aufgaben, die schnelles, suchendes Lesen messen sollen, auch wirklich weniger Zeit aufwenden als bei Aufgaben, die langsames, detailliertes Lesen erfassen sollen. Werden zudem die Items nach dem jeweiligen Lesetext präsentiert, ist z.B. nicht auszuschließen, dass einige Testteilnehmer zuerst die Items und dann den Text lesen, andere Testteilnehmer dagegen zuerst mehr oder minder intensiv den Text und anschließend die Items, wobei in beiden Fällen bei fehlender Zeitkontrolle auch eine mehrfache Textlektüre möglich ist. Von einer validen Messung der Fähigkeit zum schnellen suchenden, identifizierenden oder orientierenden Lesen kann dann kaum noch gesprochen werden.

Bei der sog. Einzeldarbietung wird dagegen jede Aufgabe bzw. jedes Item einzeln mit einer vom Testersteller vorher festgelegten Zeitdauer präsentiert. Ist die Präsentationsdauer knapp gewählt, wird wie bei einer deutlich zeitlimitierten Gesamtdarbietung ein Zeitdruck aufgebaut, diesmal jedoch kontrolliert auf der Ebene der einzelnen Aufgaben bzw. Items. Aus psychometrischer Sicht ist die Einzeldarbietung vorzuziehen, da eine stärkere Steuerung der (zeitabhängigen) Aufgabenlösungsprozesse möglich ist und so die Validität der Aufgaben erhöht werden kann.

Außerdem kann es schwerwiegende psychometrische Konsequenzen haben, wenn bei einer Gesamtdarbietung Aufgaben am Ende eines Tests nicht mehr gelöst und dann wie zumeist üblich als falsch gewertet werden. Hat z.B. ein Testteilnehmer mit einer hohen Lesekompetenz die letzten Aufgaben nur aufgrund eines falschen Zeitmanagements nicht mehr in Angriff genommen, dann wird seine Kompetenz anhand der Gesamtpunktzahl unterschätzt (vgl. hierzu bereits Wesman 1949 sowie die ausführliche Diskussion in Grotjahn 2010). Dieses Problem wird bei einer Einzeldarbietung ebenfalls vermieden.

3.3 Bearbeitungszeit in Fremdsprachentests

In der Sprachtestliteratur wird der Stellenwert des Faktors *Speed* im Hinblick auf die Messung sprachlicher Kompetenzen häufig nur am Rande oder gar nicht thematisiert. So findet sich z.B. in der kürzlich erschienenen Monografie von Fulcher (2010) lediglich an einer Stelle ein kurzer Hinweis, in dem es jedoch in erster Linie um den Faktor *Zeit* als Quelle konstruktirrelevanter Varianz geht: "If test takers are rushing to complete the test it becomes a test of speed, rather than of the intended constructs" (ebd.: 185). Eine explizite Auseinandersetzung mit dem Faktor *Zeit* als Quelle konstruktrelevanter Varianz fehlt dagegen.¹¹

In vielen Lesetests (z.B. in den *Cambridge ESOL-Tests*) oder auch dem *TOEFL iBT* wird schnelles Lesen jedoch als konstruktrelevant erachtet. So weisen Khalifa & Weir (2009) u.a. darauf hin, dass nach dem GeR (Europarat 2001) bereits auf dem B1-Niveau die Lesegeschwindigkeit eine Rolle spiele und dass ab dem B2-Niveau vom Leser erwartet werde "to be able to process text quickly as well as efficiently i.e. expeditious as well as careful reading is expected" (Khalifa & Weir 2009: 62).¹² Allerdings betonen Weir & Khalifa (2008) und Khalifa & Weir (2009) in ihrer Validierungsstudie zu dem *Cambridge ESOL-Tests* selbst, dass die Schnelligkeit nicht kontrolliert werde und dass somit nicht sichergestellt sei, dass die auf der Konstruktebene postulierten Unterschiede z.B. zwischen schnellem selektivem und langsamem detailliertem Lesen durch die Testaufgaben hinreichend valide gemessen werden. Entsprechend stellen sie u.a. in Bezug auf die Items im Teil 3 der B2-Prüfung FCE (*First Certificate in English*), in der schnelles, effizientes Lesen, d.h. *expeditious reading*, getestet werden soll, fest:

Lack of control over the amount of time candidates spend on each part, however, may mean that some candidates are given the opportunity to use careful reading rather than expeditious reading in completing these items (Khalifa & Weir 2009: 65).

Auf die fehlende Kontrolle der von den Testkandidaten gewählten Bearbeitungsprozesse und aufgewendeten Bearbeitungszeit sowie auf die hieraus

11 Etwas ausführlicher wird die Variable *Speed* z.B. in der grundlegenden Monografie von Bachman (1990) behandelt, und zwar als Testmethodenfaktor mit folgenden Facetten: Zeitangaben in der Instruktion, wahrgenommener Zeitdruck beim Input sowie zeitliche Beschränkungen bei der Antwort (ebd.: 119).

12 Vgl. auch die explizite Nennung der Lesegeschwindigkeit in der Kann-Beschreibung für das Niveau 5 (*Good User*) der *Association of Language Testers in Europe* (ALTE): "KANN Texten relevante Informationen entnehmen, die Hauptgedanken eines Textes verstehen und dabei fast so schnell lesen wie ein Muttersprachler" (Europarat 2001: 237).

resultierenden Validitätsprobleme haben u.a. bereits Weir, Yang & Jin (2000) auf der Basis einer Analyse von 10 Lesetests im Bereich *English for Academic Purposes* (EAP) hingewiesen:

Firstly, there appears to be little or no control over the skills/strategies candidates use on current EAP reading tests as questions testing both reading quickly and reading carefully are based on the same passage.

Secondly, there appears to be little or no control over the individual amount of time candidates should spend on each passage. In IELTS [English Language Testing System], TEEP [Test of English for Educational Purposes] and EPTB [English Proficiency Test Battery] times are suggested per passage/section but there is no enforcement of these times. Others simply state the time allowed for the reading test and leave the division of that time to the candidates' discretion (Weir, Yang & Jin 2000: 47).

Die fehlende Kontrolle der Bearbeitungszeiten bedeutet vor dem Hintergrund von Weirs sozio-kognitivem Rahmenmodell, das allen neueren Validierungsstudien zu den *Cambridge ESOL-Tests* zu Grunde liegt, dass die Aufgaben zum schnellen, effizienten Lesen keine hinreichende Kontextvalidität aufweisen und dass auch deren kognitive Validität und Konstruktvalidität nicht hinreichend gesichert ist.

Auch in anderen bekannten Fremdsprachentests wie den Goethe- oder telc-Zertifikaten oder auch universitären Zulassungstests wie TestDaF oder *IELTS* erfolgt keine Steuerung der Bearbeitungszeit auf der Ebene der einzelnen Aufgaben bzw. Items. Auch im Subtest Leseverstehen des neuen *TOEFL iBT* sind seit dem 1.11.2011 nicht mehr einzelne Teile, sondern nur noch der gesamte Subtest zeitbegrenzt.¹³ Der Kandidat kann zudem zu früheren Items zurückgehen. Diese Veränderungen werden als *Timing and Navigation Enhancements* beschrieben (vgl. <http://www.ets.org/s/toefl/newsletter/2011/18643/uk/timing.html>). Mögliche Auswirkungen auf das Testkonstrukt werden nicht thematisiert, obwohl u.a. in den Vorbereitungs-materialien zum neuen *TOEFL iBT* immer wieder auf die Wichtigkeit schnellen Lesens hingewiesen wird und auch im Rahmenmodell für den neuen *TOEFL* die Fähigkeit zur schnellen Informationsentnahme explizit als wichtige Konstruktkomponente genannt wird (vgl. z.B. Enright et al. 2000: 33). Um diese Fähigkeitskomponente direkter als beim alten *TOEFL* zu erfassen, wurden als Teil der prototypischen Aufgabenformate für den neuen *TOEFL* spezifische computer-basierte Aufgaben zur Erfassung der Schnel-

13 Es läuft allerdings eine Uhr mit, so dass der Kandidat weiß, wie viel Zeit er insgesamt noch hat. Auch der neue *Pearson Test of English Academic* hat dieses Design-Merkmal (vgl. Pearson Education 2012: 2).

ligkeit und Effizienz des Leseverstehens entwickelt und erprobt. Enright, Bridgeman, Eignor, Kantor, Mollaun, Nissan, Powers & Schedl (2008: 110) beschreiben diese folgendermaßen:

[...] a scanning task was developed to assess the ability to rapidly recognize words. This task required test takers to rapidly locate specific words within a reading passage. The score on the task was to be the number of words located within a limited amount of time. [...] Basic comprehension requires a reasonably efficient reading rate (Enright et al. 2000), but reading efficiency was not directly assessed on the existing TOEFL. Therefore, a reading-efficiency task to assess this aspect of basic comprehension abilities was also developed. Test takers were allowed a limited amount of time to read a text, and then the text was removed, and three multiple-choice items about the main ideas in the passage were presented.

Dabei gingen die Entwickler der *TOEFL*-Aufgaben u.a. von der Hypothese aus, dass die Scanning- und Leseeffizienz-Aufgaben tendenziell umso leichter werden, je mehr Zeit zum Lesen der Texte zur Verfügung steht. Bei den Erprobungen ergab sich jedoch eine nicht erwartete Beziehung zwischen Lesezeit und Korrektheit des Verstehens. Die Entwicklung dieses Aufgabentyps wurde deshalb nicht weiter verfolgt. Enright et al. (2008: 116) weisen allerdings darauf hin, dass sich bei einer stärkeren Reduzierung der Lesezeiten und bei schwierigeren Items der erwartete Zusammenhang möglicherweise gezeigt hätte. Zur Klärung dieser Frage hätte es allerdings weiterer Untersuchungen bedurft, für die nicht die notwendigen Ressourcen zur Verfügung standen. Wir werden auf den Aspekt einer hinreichenden Reduzierung der Lesezeiten bei der Interpretation der Ergebnisse unserer empirischen Untersuchung zurückkommen.

Die Erprobungen ergaben weiterhin, dass die Testteilnehmer bei der Scanning-Aufgabe nicht genau wussten, was sie tun sollten. Außerdem hatte diese Aufgabe eine relativ geringe Augenscheingültigkeit bei Testteilnehmern und zum Teil auch bei Testspezialisten. Es wurde deshalb entschieden, im Rahmen eines *high-stakes*-Tests wie dem *TOEFL* die Entwicklung dieses Aufgabentyps nicht weiter zu verfolgen (vgl. Enright et al. 2008: 114; 116). Dies bedeutet u.E. allerdings nicht, dass auf die Entwicklung von Aufgaben zur Überprüfung der Fähigkeit zur schnellen Worterkennung grundsätzlich verzichtet werden sollte. Geht man von einem hierarchischen Modell von Lesekompetenz aus, dann ist es aus einer diagnostischen Perspektive und auch im Hinblick auf spezifische Fördermaßnahmen nützlich zu wissen, ob Probleme auf den höheren Ebenen z.B. in erster Linie auf Defiziten im Bereich der schnellen Worterkennung beruhen.

4. Die empirische Studie

4.1 Forschungsfragen

Die Steuerung der zeitlichen Flexibilität erfolgt bei VERA-8 über lesernerne Faktoren wie Aufgabenanzahl, Testblockbesetzung und vorgesehene Bearbeitungszeit pro Leseblock. Eine zeitliche Steuerung auf der Ebene einzelner Aufgaben und Items und der gezielte Aufbau eines Zeitdrucks erfolgt bisher nicht.

Die vorangehenden Ausführungen deuten bereits darauf hin, dass durch eine effektive Zeitkontrolle und eine Voranstellung der Items vor den Stimulus-Text zumindest potentiell die Validität von Aufgaben zur Messung der Fähigkeit zu schnellem, sinnentnehmendem Lesen erhöht werden kann. In der im Folgenden beschriebenen Studie soll untersucht werden, inwiefern die zeitliche Limitierung der Einzelbearbeitungszeit pro Aufgabe die Ergebnisse in einem Französisch-Lesetest verändert. Da es sich um eine erste entsprechende Studie im Rahmen von VERA-8 handelt und zudem die Forschung zum Faktor Zeit bei der Bearbeitung fremdsprachlicher Leseverstehensaufgaben insgesamt eher am Anfang steht, hat diese Studie einen primär explorativen Charakter.

In der Studie werden die Leseaufgaben nacheinander mit und ohne Zeitbeschränkung auf dem Computer bearbeitet. Die Leseintention wird durch die Voranstellung der Instruktionen und der zu bearbeitenden Items gesteuert. Die erste Forschungsfrage lautet demnach:

1) Welchen Einfluss hat die Beschränkung der Bearbeitungszeit der Aufgaben auf die Leseergebnisse beim fremdsprachigen Lesen (Französisch)?

In den vorhergehenden Abschnitten konnte gezeigt werden, dass die Lesegeschwindigkeit im Zusammenhang mit anderen Faktoren betrachtet werden sollte, und zwar insbesondere dem Leseverständnis. Wir vermuten, dass auch die Konzentrationsfähigkeit, d.h. die Fähigkeit, die Aufmerksamkeit (über eine längere Zeitspanne hinweg) auf eine bestimmte Tätigkeit zu fokussieren, einen Einfluss auf die Leseleistung unter Zeitdruck haben kann. Ebenso ist es gerade beim Lesen in einer Fremdsprache wichtig, die allgemeine Sprachkompetenz zu berücksichtigen, und zwar v.a. die Komponenten Wortschatz und Grammatik sowie auch die Literalität in der Muttersprache (vgl. Abschnitt 2.2). Diese Überlegungen führten zu der zweiten Forschungsfrage:

2) Welchen Einfluss haben allgemeine französische Sprachkompetenz, Lesegeschwindigkeit und Leseverstehen in Deutsch sowie die Konzentrationsfähigkeit auf die Leseergebnisse im Französischen?

Vor der Beantwortung der Forschungsfragen gilt es, einige messtheoretische Punkte zu klären.

4.2 Messtheoretische Aspekte computerbasierten Testens

Wie bereits erwähnt, wurden in der vorliegenden Studie die Leseaufgaben von den Teilnehmern in kurzen Zeitabständen mit und ohne Zeitbeschränkung auf dem Computer bearbeitet. Computerbasierte Testverfahren haben eine Reihe von Vorteilen. Zum einen erlauben sie eine weit bessere Kontrolle zeitlicher Abläufe als ein Testleiter. Weitere wichtige Vorteile des computerbasierten Testens sind: (1) kognitive Fähigkeiten, z.B. des Arbeitsgedächtnisses, können präziser erfasst werden, (2) computerbasiertes Testen spart Zeit und Aufwand bei der Testdurchführung und Auswertung, (3) je nach Programmierung sind die Testergebnisse unmittelbar abrufbar und (4) in computerbasierte Tests können Animationen und Audiodateien integriert werden (vgl. Singleton, Thomas & Horne 2000: 162f.). Weiterhin gibt es Hinweise, dass computerbasierte Tests besonders für Sprachtests geeignet sind (Kingston 2009: 22):

Subject did appear to affect comparability, with computer administration appearing to provide a small advantage for English Language Arts and Social Studies test (effect sizes of .11 and .15, respectively), and paper administration appearing to provide a small advantage for Mathematics tests (effect size of -.06).

Auch scheinen Schüler dazu zu neigen, computerbasierte Tests zu bevorzugen. So stellen z.B. Way, Davis & Fitzpatrick (2006: 1) unter Verweis auf mehrere aktuelle Studien fest:

[...] recent surveys indicate that students testing online enjoy their experiences, feel comfortable with taking tests by computer, and tend to prefer it to traditional paper testing [...].

Bereits 1998 hat sich der *Educational Testing Service* (ETS) entschieden, neben der *paper-pencil*-Version auch eine computerbasierte Variante des *Test of English as a Foreign Language* (TOEFL) anzubieten. Diese Entscheidung beruhte unter anderem auf der Untersuchung von Taylor, Jamie-

son, Eigner & Kirsch (1998), die zeigen konnten, dass die Computervertrautheit der Kandidaten bei den meisten *TOEFL*-Aufgaben keinen praktisch bedeutsamen Effekt hatte. Außerdem wurde argumentiert, dass die Anwendung von Technologie im Klassenraum zur Unterstützung der Lehre und Vermittlung von Lerninhalten Schüler zwingt, sich mit dieser vertraut zu machen und sich in rechnergestützten Umgebungen zu bewegen (Enright et al. 2000).

In den letzten Jahren ist eine zunehmende Verwendung von computerbasierten Testverfahren zu beobachten (vgl. z.B. Paek 2005; Scheuermann & Björnsson 2009). Dabei konnte u.a. gezeigt werden, dass *paper-pencil*-Tests und computerbasierte Tests mit jeweils geschlossenen Antwortformaten über Jahrgangsstufen und Inhalte hinweg vergleichbar waren (Kim & Huynh 2007). Das Lesen an Computerbildschirmen ist zudem für die meisten (deutschen) Schüler keine Seltenheit mehr, nicht zuletzt weil zahlreiche Texte digital vorhanden sind und abgerufen werden können (Al-Othman 2003). Insgesamt gesehen kann davon ausgegangen werden, dass Angst und Unsicherheit im Umgang mit Computern unter Schülern heute weniger verbreitet ist als noch vor einigen Jahren (vgl. Khalifa & Weir 2009: 287). Dies ist ein weiteres Argument für den Einsatz von computerbasierten Testverfahren. Hinzu kommt noch, dass sich die technologische Qualität seit den Untersuchungen in den 1980er Jahren und auch die Zugangsmöglichkeiten auf Seiten der Testnutzer deutlich positiv entwickelt haben (vgl. Khalifa & Weir 2009: 285). Studien konnten zudem feststellen, dass computerbasierte Testverfahren mit *paper-pencil*-Tests insgesamt vergleichbar sind, solange mit gleichen Inhalten gleiche Kompetenzen gemessen werden (vgl. u.a. Choi, Kim & Boo 2003; Jones & Maycock 2007; McDonald 2002; Neuman & Baydoun 1998; Sawaki 2001; Schroeders & Wilhelm 2011).

Im Zuge des stetigen Fortschritts durch innovative Technologien sowie in Anbetracht der beschriebenen Vorteile scheinen computerbasierte Tests immer mehr zu bedenkenswerten Alternativen zu *paper-pencil*-Tests zu werden. Sofern die vorhandenen Ressourcen einen Einsatz zulassen, sollte deshalb stets auch die Möglichkeit der Verwendung einer computergestützten Variante geprüft werden.

4.3 Methodische Aspekte

4.3.1 Stichprobe und Durchführung

Die Testungen fanden im November und Dezember 2010 am Institut zur Qualitätsentwicklung im Bildungswesen (IQB) statt. Auf Grund des Einsatzes von computerbasierten Testteilen wurden die Testungen in einem PC-Pool durchgeführt. Durch eine Beschränkung der Sitzplätze konnten pro Testdurchgang maximal zehn Schüler getestet werden. Insgesamt nahmen 37 Schüler aus Berliner Gymnasien und Realschulen an der Testung teil. Der Kontakt zu den Schülern wurde über Lehrer hergestellt, die sich im Schulnetzwerk des Projektes "Französisch Sekundarstufe I" am IQB engagieren. Auf Grund der kaum gegebenen Vergleichbarkeit wurde eine Schülerin französischer Nationalität aus den Analysen ausgeschlossen. Von den verbleibenden 36 Schülern waren 30 (83.3%) weiblichen Geschlechts. Alle Schüler befanden sich zum Testzeitpunkt am Anfang der neunten Jahrgangsstufe. Das Alter der Schüler lag zwischen 14 und 16 Jahren ($M = 14.5$, $SD = 0.7$). Über einen Migrationshintergrund verfügten 22.2% der Schüler. Die Testungen fanden ausschließlich losgelöst vom schulischen Unterricht nachmittags zwischen 14:30 Uhr und 18:30 Uhr statt. Die Testsitzungen wurden von einer erfahrenen Testleiterin vorbereitet und durchgeführt. Technische Fragen wurden im Gegensatz zu inhaltlichen Fragen während der gesamten Testsitzung beantwortet. Eine Testsitzung umfasste circa 90 Minuten – mit einer Pause von fünf Minuten nach den ersten 45 Minuten.

4.3.2 Instrumente

Es wurden folgende Instrumente eingesetzt: der d2-Aufmerksamkeits-Belastungs-Test (Brickenkamp 2002) zur Messung der Konzentrationsfähigkeit, der deutschsprachige Lesegeschwindigkeits- und -verständnis-test für die Klassen 6-12 (Schneider, Schlagmüller & Ennemoser 2007), ein französischer C-Test, VERA-8-Leseverstehensaufgaben (Französisch), die einmal mit Zeitbegrenzung (im Folgenden als *LV-Speed* bezeichnet) und einmal ohne Zeitbegrenzung (im Folgenden als *LV-Power* bezeichnet) dargeboten wurden, sowie ein deutschsprachiger Strategie-Fragebogen.

Wie bereits angedeutet, bieten computerbasierte Tests viele Vorteile. Nichtsdestoweniger mussten wir aus technischen Gründen aber auch *paper-pencil*-Verfahren einsetzen, insbesondere im Fall von Tests, die lediglich als *paper-pencil*-Verfahren normiert waren. Im Fokus der hier beschriebenen Studie stehen die beiden Testteile *LV-Speed* und *LV-Power*. Alle weiteren Instrumente dienen zur Kontrolle von möglichen Einflussfaktoren bezüglich

der Leistung in *LV-Speed* und *LV-Power*. Im Ergebnisteil wird eine Unterscheidung zwischen Leistungstest und Fragebogen vorgenommen. Die Leistungstests beinhalten dabei Tests, bei denen eine kognitive Leistung erforderlich ist, also den Konzentrationstest, den Leseverständnistest, den C-Test und beide Testteile *LV-Speed* und *LV-Power*. Der Fragebogen wird als Einzelaspekt betrachtet. Im Folgenden werden die Instrumente ausführlich beschrieben. Im Anschluss an die Beschreibungen findet sich eine zusammenfassende Tabelle.

4.3.2.1 Deutschsprachige allgemeine Instruktionen

Die Testsitzung begann mit einer allgemeinen Instruktion zur Vorbereitung der Schüler auf die Testung. Die allgemeine Instruktion beinhaltete Informationen zum Ablauf und zum Umgang mit dem Computer. Jeder Testteil wurde zudem durch spezifische Instruktionen eingeleitet. Alle Anweisungen wurden vorgelesen, so dass alle Testteilnehmer dieselben Informationen erhielten.

4.3.2.2 Deutschsprachiger d2-Aufmerksamkeits-Belastungs-Test

Zur Messung der Konzentrationsfähigkeit wurde der d2-Aufmerksamkeits-Belastungs-Test von Brickenkamp (2002) eingesetzt. Der Test bestand aus 14 Testzeilen mit unterschiedlichen Versionen der Buchstaben "d" und "p" mit jeweils einem, zwei, drei oder vier Strichen. Aufgabe der Testpersonen war es, jedes "d" mit zwei Strichen durchzustreichen. Dabei hatten die Schüler pro Testzeile 20 Sekunden Zeit. Die Auswertung erfolgte über die Feststellung richtig und falsch durchgestrichener "d" und der Gesamtbearbeitung einer jeden Testzeile. Mit Hilfe von Normtabellen wurden Standardwerte für die Altersgruppe ermittelt. In die Analysen ging der standardisierte Konzentrationsleistungswert (KL) ein, der einen durch die falsch durchgestrichenen Zeichen bereinigten Summenwert darstellt.

4.3.2.3 Deutschsprachiger Lesegeschwindigkeits- und -verständnistest für die Klassen 6-12 (LGVT 6-12)

Zur Messung der Lesegeschwindigkeit und des Leseverständnisses in der Muttersprache Deutsch wurde der deutschsprachige Lesegeschwindigkeits- und -verständnistest für die Klassen 6-12 (LGVT 6-12) von Schneider, Schlagmüller & Ennemoser (2007) eingesetzt. Die Schüler lasen einen Fließtext mit insgesamt 1727 Wörtern in vier Minuten so weit wie möglich. In den Text waren 23 Lücken integriert, bei denen aus drei Antwortalternativen das am besten passende Wort gewählt werden musste. Jedes korrekt un-

terstrichene Wort ergab zwei Punkte. Ein falsch unterstrichenes Wort wurde mit einem Minuspunkt geahndet. Die daraus resultierenden T-Werte für die Lesegeschwindigkeit sowie das Leseverständnis gingen in die Analysen ein.

4.3.2.4 *Deutschsprachiger Lesestrategie-Fragebogen*

Der eingesetzte Fragebogen bestand aus zwei Skalen, die im Wesentlichen auf den Skalen von Groß (2000) basieren. Er zielte auf Lesestrategien ab (z.B. "Wenn ich beim Lesen französischer Texte Verständnisschwierigkeiten habe, dann versuche ich, die Bedeutung aus dem Zusammenhang zu erschließen") und erfasste diese auf Basis einer vierstufigen Likert-Skala (1 = trifft gar nicht zu, 4 = trifft völlig zu).

4.3.2.5 *Französischsprachige Testteile LV-Speed und LV-Power*

Wie bereits erwähnt, ist die primäre Zielsetzung unserer empirischen Studie herauszufinden, welchen Einfluss der Aufbau von Zeitdruck auf die Lösung von Leseverstehensaufgaben hat und welche weiteren Faktoren hierbei möglicherweise eine Rolle spielen. Dazu wurden allen Schülern dieselben französischsprachigen Aufgaben zuerst unter Zeitdruck (*LV-Speed*) und anschließend ohne Zeitdruck (*LV-Power*) vorgelegt. Angesichts der geringen Zahl an Teilnehmern, die für die Untersuchung gewonnen werden konnten, musste auf die randomisierte Bildung einer Kontrollgruppe, die den Test lediglich unter *Power*-Bedingungen bearbeitet, verzichtet werden. Die Bildung einer solchen Kontrollgruppe hätte erlaubt, die Ergebnisse bei *LV-Power* zusätzlich bezüglich möglicher Übungseffekte zur interpretieren.¹⁴

Der computerbasierte Testteil *LV-Speed* umfasste 14 Testaufgaben mit insgesamt 29 Items. Die Items wurden entwickelt unter Bezug auf die Bildungsstandards Französisch für das vom IQB koordinierte VERA-8-Projekt. Pro Aufgabe wurden zwischen einem und vier Items eingesetzt. Die Aufgaben wurden in Bezug auf das entsprechende Leseziel geprüft, und es wurden ausschließlich Aufgaben für schnelle Lesevorgänge ausgewählt. Zusätzlich wurden zur Steuerung der Leseintention die Items vor den Stimulus-Text gestellt. In einer Vorstudie im November 2010 wurden einer kleinen Gruppe von Schülern ($N = 4$) die Aufgaben vorgelegt und ihre Bearbeitungszeit pro Aufgabe ohne Zeitdruck gemessen. Diese Bearbeitungszeit wurde für die Hauptstudie je nach Aufgabe um circa 20 Sekunden reduziert, um einen an-

14 Die bei einer Bestimmung der Test-Retest-Reliabilität sinnvolle experimentelle Kontrolle der Reihenfolge der Administration von Test und Retest ist in Anbetracht der Fragestellung unserer Untersuchung kontraindiziert, da bei der Reihenfolge *LV-Power* vor *LV-Speed* ein größerer Übungs- und Erinnerungseffekt zu erwarten ist als bei der umgekehrten Reihenfolge.

gemessenen Zeitdruck zu schaffen. Hinsichtlich der Reduzierung der Bearbeitungszeit wurde in dieser Studie der Zeitdruck nicht als fester Faktor gesehen, sondern als variables Zeitlimit in Abhängigkeit von der kognitiven Anforderung, Komplexität und Länge der Aufgabe. Den Schülern standen 20 Minuten zur Bearbeitung der 14 Aufgaben mit insgesamt 29 Items zur Verfügung. Die einzelnen Aufgaben waren wie folgt aufgebaut: Titel – kurze Situierung und Arbeitsanweisung – Item (ggf. Items) – Stimulus-Text. Die Aufgaben wurden den Schülern einzeln dargeboten. Die Schüler sahen jeweils die komplette Aufgabe mit allen Items. Scrollen war an keiner Stelle notwendig. Während der Bearbeitung der einzelnen Aufgaben lief am Bildschirm eine Uhr rückwärts mit. Die Zeitbeschränkung galt jeweils in Bezug auf die gesamte Aufgabe und nicht für einzelne Items. Somit hatten die Schüler stets Kontrolle über die ihnen verbleibende Restzeit. Nach der Bearbeitung einer Aufgabe konnten sie nicht mehr zu dieser Aufgabe zurückgehen.

Die hier näher beschriebene Beispielaufgabe *Vélib* (vgl. Abb. 2) umfasst einen kontinuierlichen Sachtext, eine vorangestellte Instruktion auf Deutsch sowie ein vorangestelltes Dreifachwahl-Item. Der Text gehört zur Textsorte Benutzerinformationen, wie sie im Internet zu finden sind und enthält somit paratextuelle Zusatzinformationen (zwei Grafiken, "Reiter", die im Internet die Navigation unterstützen, Überschriften). Zusätzlich informiert eine Einbettung auf Deutsch, worum es sich bei dem Titel *Vélib* handelt.

Die Voranstellung des Items steuert die Leseintention auf das identifizierende Lesen des Textes im Hinblick auf eine Einzelinformation und zwar die Anzahl erlaubter Nutzungen des Fahrrads pro Zeiteinheit. Sprachlich setzt die Lösung der Aufgabe insbesondere die Kenntnis der Struktur *autant de fois* und logisch-kombinatorisch den Doppelbezug von Anzahl Tage und Anzahl Fahrten voraus. Tabelle 1 gibt Informationen über alle eingesetzten Aufgaben und Items.

Vélib

<< Vélib >> ist ein Fahrradverleihsystem in Paris. Lies dazu den Text und beantworte dann die folgende Frage.

► **Avec les tickets Vélib, tu peux utiliser le vélo ...**

- A** : une fois par jour.
- B** : autant de fois par jour que tu veux.
- C** : autant de fois par mois que tu veux.



Abonnements / tarifs

Abonnement 1 an	Abonnement Courte Durée	Les tarifs
<p>Tickets Vélib' 1 jour (1€) et 7 jours (5€)**(18/06/2007)</p> <p>Les tickets Vélib' 1 jour et 7 jours vous donnent droit à un nombre illimité de trajets pendant toute la durée de votre abonnement.</p> <p>Les 30 premières minutes de chaque trajet sont gratuites. Si vous devez effectuer un trajet supérieur à 30 minutes, le coût du service Vélib' sera débité sur votre compte bancaire en même temps que le coût de votre abonnement, à la fin de sa période de validité. Vous pouvez contracter un abonnement Courte Durée dans n'importe quelle station Vélib' équipée d'une borne.</p> <p>** Dépôt de garantie : dans le cadre de votre abonnement Courte Durée, vous devrez accepter une pré-autorisation de prélèvement (sans encaissement) de 150 € sur votre compte bancaire. En cas de non-respect des Conditions Générales d'Accès et d'Utilisation de Vélib', consultables sur ce site ou en station, et notamment en cas de non-restitution de votre vélo au-delà d'un délai de 24 heures, tout ou partie du montant du dépôt de garantie pourra être encaissé.</p> <p>Lire les Conditions générales d'accès et d'utilisation :</p> <p>CGAU courte durée</p> <p>Retour Abonnements / tarifs</p> <p>Retour Abonnement Courte Durée</p>		

Article paru sur le site www.paris.fr

Abbildung 2: Leseverstehensaufgabe Vélib

Titel der Aufgabe	Zeit in sec.	Wortzahl ¹⁵	Textsorte	Thema	Anzahl und Art der Items
Okapi	60	210	Leserbrief	Freizeit und Sport	1 MC-Item mit 3 Antwortmöglichkeiten
Audrey Tautou	120	258	Biografie	Presse, Rundfunk, Fernsehen	3 Items, sieben Biografieteile in eine Reihenfolge mit drei Lücken bringen
Internet dans le monde	120	269	Alltags-erzählung	elektronische Medien	4 Items, Zuordnung von fünf Ländertexten zu vier Aussagen
Aïcha	120	418	persönlicher Brief	zwischenmenschliche Beziehungen: Liebe, Freundschaft, Familie	1 MC-Item mit 4 Antwortmöglichkeiten
Vélib	45	233	Benutzerinformation	Reisen, Tourismus	1 MC-Item mit 3 Antwortmöglichkeiten
Les clés de l'Actualité	90	249	Zeitungsmeldung	Verkehrsmittel	2 Items, Zuordnung von zwei Artikeln zu vier Titelvorschlägen
Livres pour la jeunesse	120	485	Klappentexte	Literatur	4 Items, Zuordnung von fünf Klappentexten zu den Lesevorlieben von vier Jugendlichen
Titeuf	30	238	Jugendbuchauszug	Schulalltag	1 MC-Item mit 4 Antwortmöglichkeiten
Lucie	45	215	Reisebericht	zwischenmenschliche Beziehungen: Liebe, Freundschaft, Familie	1 MC-Item mit 3 Antwortmöglichkeiten
Téléphone portable	90	351	Alltags-erzählung	elektronische Medien	3 Items, Zuordnung von 4 Aussagen von Jugendlichen zu 3 Zitaten
Le grand-père	60	386	Alltags-erzählung	zwischenmenschliche Beziehungen:	1 MC-Item mit 3 Antwortmöglichkeiten

15 Die Angabe bezieht sich sowohl auf den Stimulus-Text als auch auf die Instruktion und die Items.

				Liebe, Freundschaft, Familie	
Lilian Thuram	90	342	Biografie	Freizeit und Sport	4 Items, 6 Biografie- teile in eine Reihen- folge mit 4 Lücken bringen
Partir seul	90	353	Reise- bericht	Reisen, Tourismus	2 Items, Zuordnung von 3 Aussagen von Jugendlichen zu 2 Zitaten
Pro- gramme d'échange	120	265	Reise- bericht	Sprache und Sprachlernen	1 MC-Item mit 4 Antwort- möglichkeiten

Tabelle 1: Übersicht über eingesetzte Aufgaben und Items¹⁶

4.3.2.6 Französischsprachiger C-Test

Ein C-Test besteht aus mehreren kurzen Texten. Im Gegensatz zum klassischen *Cloze*-Test oder auch vielen anderen Lückentextformaten fehlen beim C-Test keine ganzen Wörter, sondern Wortteile. Für einen klassischen C-Test gilt: Beginnend mit dem zweiten Wort des zweiten Satzes wird in jedem Text bei jedem zweiten Wort die Hälfte getilgt. Die Kandidaten müssen diese Wortteile rekonstruieren. Die Zahl der korrekt rekonstruierten Wörter gilt dann als Maß allgemeiner Sprachkompetenz (Cronjäger, Klappheck, Krätzschar & Walter 2010; Eckes & Grotjahn 2006; Grotjahn 2002, 2011). Die im Hinblick auf das Verstehen französischer Texte wichtigen lexikalischen und morphosyntaktischen Kompetenzen sind dabei zentrale Komponenten des von einem französischen C-Test erfassten Konstrukts. C-Tests sind vor allem dann einsetzbar, wenn eine Sprachstandsfeststellung gewünscht wird, die von Unterricht und individueller Lerngeschichte möglichst unabhängig ist. C-Tests sind hoch objektiv, zumeist hoch reliabel und stellen zudem ein sehr ökonomisches Instrument zur Messung allgemeiner Sprachkompetenz dar. Mittlerweile werden C-Tests in einer Vielzahl von Kontexten verwendet, darunter auch große empirische Bildungsstudien wie z.B. DESI (Deutsch Englisch Schülerleistungen International) oder auch die Hamburger Längsschnittstudien LAU (Aspekte der Lernausgangslage und der Lernentwicklung) und KESS (Kompetenzen und Einstellungen von Schülerinnen und Schülern). Der von uns verwendete klassische C-Test stammt aus dem TESTATT-Projekt (Raatz, Grotjahn, Wockenfuß 2006),

¹⁶ Die Kategorien Textsorte und Thema entsprechen den in VERA-8 üblichen Zuordnungen.

das in erster Linie auf Studierende abzielt. Aus Gründen der besseren Eignung für die von uns anvisierte Zielgruppe wurde ein Text durch einen adaptierten Jugendtext (Phosphore 1993) ersetzt. Insgesamt umfasst der eingesetzte C-Test fünf Texte mit jeweils 20 Lücken, die durch gleich lange, durchgehende Striche gekennzeichnet sind. Für die Bearbeitung jedes Textes standen fünf Minuten zur Verfügung.

4.3.2.7 Französischsprachiger Testteil LV-Power

Der computerbasierte Testteil LV-Power umfasste die bereits im Teil LV-Speed eingesetzten Aufgaben. Im Gegensatz zum ersten Teil arbeiteten die Schüler hier ohne Zeitbegrenzung. Sie hatten demnach die Möglichkeit, ihre Antworten aus LV-Speed zu überprüfen und gegebenenfalls zu korrigieren. Ihre Antworten aus LV-Speed wurden ihnen dabei nicht angezeigt. Die benötigte Zeit der Schüler betrug für LV-Power zwischen zwei und 36 Minuten ($M = 14.58$, $SD = 10.19$). Die Schüler benötigten im Testteil LV-Power demnach im Mittel deutlich weniger Zeit zur Bearbeitung der Items als im Testteil LV-Speed, wo insgesamt 20 Minuten zur Verfügung standen.

Eine Übersicht über die eingesetzten Instrumente findet sich in Tabelle 2. Die Reihenfolge der Instrumente entspricht dem zeitlichen Ablauf der Untersuchung. Im Fall eigener Reliabilitätsberechnungen wurde als Schätzwert der Reliabilität neben dem bekannten α -Koeffizienten von Cronbach auch der weniger bekannte λ_2 -Koeffizient von Guttman angegeben. Letzterer liefert unter bestimmten Bedingungen eine bessere Schätzung der Reliabilität als Cronbachs α (vgl. die Hinweise in Grotjahn 2012) und kann wie α leicht mit Hilfe der Statistik-Software IBM SPSS berechnet werden.¹⁷

Instrumente	vorgesehene Dauer	Modus, Sprache	Anzahl Items	Cronbachs α / Guttmans λ_2
Allgemeine Instruktionen	2 min	mündlich, vorgelesen, deutsch	-	-
d2-Aufmerksamkeits-Belastungs-Test	6 min mit Instruktion	paper-pencil, deutsch	14	0.95 ¹⁸ / -

17 Die statistischen Analysen im vorliegenden Artikel wurden mit Hilfe von IBM SPSS Version 20 durchgeführt.

18 Entspricht den Angaben für die Altersgruppe (13-14 Jahre) aus der Normierung des Tests (Briekenkamp 2002). Die Reliabilität für die Altersgruppe 15-16 Jahre beträgt $\alpha = 0.90$.

Lesegeschwindigkeits- und -verständnistest für die Klassen 6-12 (LGVT 6-12)	7 min mit Instruktion	paper-pencil, deutsch	23	Verständnis: ¹⁹ 0.87 / - Geschwindigkeit: 0.84 / -
Fragebogen	5 min	paper-pencil, deutsch	14	< 0.5 / 0.5 → keine Skalenebildung aus Originalskalen
LV-Speed	20 min	computerbasiert, französisch	29	0.67 / 0.71
Pause	5 min	-	-	-
C-Test Französisch	25 min	paper-pencil, französisch	100 (5 Texte à 20 Lücken)	0.86 / 0.87
LV-Power	25-30 min	computerbasiert, französisch	29	0.80 / 0.82

Tabelle 2: Übersicht über die eingesetzten Instrumente in zeitlicher Reihenfolge

4.4 Ergebnisse

4.4.1 Leistungstests

Zur Beantwortung der Forschungsfragen ist im Vorfeld zu erörtern, welche Schüler als eher langsame oder eher schnelle Leser bezeichnet werden können. Dafür wird der Rohwert der in der vorgegebenen Zeit gelesenen Wörter sowie der korrekt gelösten Lücken errechnet. Zur besseren Vergleichbarkeit liegen dem LGVT-Testhandbuch Tabellen bei, die eine Transformation der Rohwerte in T-Werte unter Berücksichtigung der Schulform und des Alters ermöglichen. Für unsere Stichprobe liegt der resultierende T-Wert im Mittel unter dem normierten T-Wert, der den Vorgaben $M = 50$ und $SD = 10$ entspricht ($M_{\text{Leseverstehen}} = 47.11$, $SD_{\text{Leseverstehen}} = 8.54$, $M_{\text{Lesegeschwindigkeit}} = 50.81$, $SD_{\text{Lesegeschwindigkeit}} = 9.83$). In unsere Analysen gingen die individuellen T-Werte unserer Stichprobe ein. Die reine Lesegeschwindigkeit im LGVT 6-12 (vgl. Abschnitt 4.3.2 Instrumente) variiert zwischen 90 und 309 wpm. Nach Carver (1985) gäbe es in dieser Studie keine außergewöhnlich schnellen Leser. In Bezug auf die Eichstichprobe der 9. Klasse des LGVT 6-

¹⁹ Entspricht den Angaben aus der Normierung des Tests (Schneider, Schlagmüller & Ennemoser 2007).

12 kann man die 36 Probanden jedoch Prozentranggruppen zuordnen. Die entsprechenden Häufigkeitsangaben finden sich in Tabelle 3.

Prozentranggruppe	Lesegeschwindigkeit	Leseverständnis
I (Prozentränge 0 bis 15)	3	7
II (Prozentränge 16 bis 84)	25	28
III (Prozentränge 85 bis 100)	5	1

Tabelle 3: Verteilung der Probanden auf Prozentranggruppen

Die einzige im Leseverständnis in Prozentranggruppe III liegende Person gehört zugleich zu den schnellsten fünf Lesern. Hier könnte demnach von einer ausgeprägten Dekodierfähigkeit gesprochen werden, bei der die Schüler binnen kürzester Zeit einen Text lesen und verstehen können.

Im weiteren Verlauf soll untersucht werden, inwiefern diese Fähigkeit, einen deutschen Text schnell zu lesen und dabei das Gelesene zu verstehen, mit dem Testerfolg im Französischen im Zusammenhang steht.

Im Testteil *LV-Speed* wurden den Schülern computerbasiert Aufgaben vorgelegt, die sie unter Zeitbegrenzung bearbeiten mussten. Von insgesamt 29 Items wurden zwischen zwei und 19 Items ($M = 9.0$) richtig gelöst. Im Testteil *LV-Power* wurden von den Schülern zwischen zwei und 24 Items ($M = 11.1$) richtig gelöst. Da den Schülern in beiden Testteilen die gleichen Aufgaben vorgelegt wurden und sie diese besonders im Testteil *LV-Power* korrigieren konnten, kann man auf den ersten Blick sagen, dass die meisten Schüler diese Chance genutzt haben. Bei 19 Schülern lag die Verbesserung von *LV-Speed* zu *LV-Power* zwischen einem und 14 Punkten, vier Schüler haben sich zwischen beiden Testteilen nicht verbessert, 13 Schüler haben sich zwischen einem und sechs Punkten verschlechtert. Demnach haben sich mit 47% fast die Hälfte aller teilnehmenden Schüler zwischen *LV-Speed* und *LV-Power* nicht verbessert. In manchen Fällen ist es den Schülern jedoch gelungen, bereits im ersten Durchgang die richtige Lösung zu finden, womit eine Verbesserung in *LV-Power* gegenstandslos wurde. Für die weiteren Analysen wurden Summenwerte für beide Testteile gebildet. Eine Reliabilitätsanalyse zeigte dabei suboptimale Ergebnisse für den Testteil *LV-Speed* (Cronbachs $\alpha = 0.50$, $\lambda_2 = 0.58$; für *LV-Power*: Cronbachs $\alpha = 0.86$, $\lambda_2 = 0.88$). Deshalb wurden Items mit negativen Trennschärfen im Testteil *LV-Speed* eliminiert. Diese Items ergaben zudem negative Interkorrelatio-

nen mit den anderen Items.²⁰ In den meisten Fällen betrifft das jene Items, die in einer Aufgabe mit mehreren Items als letztes gelöst wurden. Das zeigt sich auch bei den zahlreichen fehlenden Werten (zwischen 5 und 15 Schüler haben die entsprechenden letzten Items einer Aufgabe nicht beantwortet). Die Reliabilitäten für *LV-Speed* haben sich nach Eliminierung der entsprechenden Items trotz der nunmehr geringeren Itemzahl deutlich verbessert (Cronbachs $\alpha = 0.67$, $\lambda_2 = 0.71$), für *LV-Power* jedoch haben sie sich leicht verschlechtert (Cronbachs $\alpha = 0.80$, $\lambda_2 = 0.82$).²¹ Zu bedenken ist weiterhin, dass die Höhe der Reliabilitäten auch von der Höhe der Varianzen von *LV-Power* ($SD = 4.44$, $M = 8.28$) und *LV-Speed* ($SD = 3.47$, $M = 6.86$) abhängt. Die höhere Varianz von *LV-Power* ist damit eine mögliche Ursache für dessen höhere Reliabilität im Vergleich zu *LV-Speed*. Da mit dem reduzierten Itempool jedoch insgesamt befriedigendere Reliabilitäten erzielt werden konnten, wurden für die weiteren statistischen Analysen die anhand des reduzierten Itempools berechneten Summenwerte verwendet.

Als erstes wurden der Mittelwertunterschied zwischen *LV-Speed* und *LV-Power* auf Signifikanz getestet. Der *t*-Test für abhängige Stichproben ergab $t = -2.37$ ($df = 35$; $p = 0.02$, einseitig) und damit eine hochsignifikante Verbesserung von *LV-Speed* nach *LV-Power*. Allerdings ist dieser Effekt mit $\eta = 0.19$ nur relativ schwach. Betrachtet man die Veränderungen von *LV-Speed* nach *LV-Power* auf der Ebene der einzelnen Items, ergibt sich nur bei fünf der insgesamt 29 Items ein signifikanter Effekt (McNemar-Test).²² Dabei zeigte sich ein knapp signifikanter Effekt zur Verbesserung bei den Items 1 und 2 in *Les clés de l'Actualité* und Item 2 in *Téléphone portable* ($p < 0.05$; einseitig) sowie ein hochsignifikanter Effekt bei Item 2 von *Livres pour la jeunesse* ($p < 0.01$; einseitig). Eine signifikante Tendenz zur Verschlechterung gab es lediglich bei Item 1 in *Lilian Thuram* ($p < 0.01$;

20 Bei allen im vorliegenden Artikel berechneten Korrelationen handelt es sich um Pearson Produkt-Moment-Korrelationen.

21 Zur Vermeidung des Problems der lokalen stochastischen Abhängigkeit der zu einem Lesetext gehörenden Items wurde im Zusammenhang mit der Berechnung der Reliabilitäten auch ein *Testlet Approach* (Behandlung der zum Text gehörenden Items als Itembündel) in Erwägung gezogen (vgl. hierzu auch Grotjahn 2012). Ein solches Vorgehen ist zwar inhaltlich sinnvoll, insbesondere da die Bearbeitungszeitbeschränkung pro Aufgabe und nicht pro Item galt. Da sich das Problem der Itemabhängigkeit allerdings nur bei vier der insgesamt acht Aufgaben stellt, dürfte sich die potentielle Überschätzung der Reliabilitäten aufgrund der Itemabhängigkeiten insgesamt in einer zu vernachlässigenden Größenordnung bewegen. Bei der Berechnung der Reliabilitäten für den C-Test wurde dagegen ein *Testlet Approach* gewählt und die Reliabilität auf der Basis der Summenwerte der fünf Texte berechnet (vgl. Grotjahn 2012).

22 Bei der Interpretation der Signifikanzen ist die Erhöhung des Alpha-Fehlers im Fall multipler Tests zu berücksichtigen.

einseitig). Eine plausible Erklärung, warum ausgerechnet bei diesem Item eine deutlich ausgeprägte Tendenz zur Verschlechterung zu beobachten ist, konnte von uns nicht gefunden werden.

Als nächstes wurde eine Reihe von Korrelationen zwischen den Untersuchungsvariablen berechnet. Die Korrelation der Punktdifferenz, d.h. dem Unterschied zwischen *LV-Speed* und *LV-Power*,²³ mit der Lesegeschwindigkeit im Deutschen ($r = 0.08$) ist minimal und nicht signifikant. Dagegen ist die Korrelation der Punktdifferenz mit dem Leseverständnis im Deutschen mit $r = 0.23$ deutlich höher und zumindest marginal signifikant ($p = 0.09$; einseitig). Dies bedeutet u.a., dass die Verbesserung von *LV-Speed* zu *LV-Power* möglicherweise in Teilen auf den (kompensatorischen) Einsatz muttersprachlicher Verstehenskompetenzen im Fall der *Power*-Bedingung zurückzuführen ist (vgl. auch Abschnitt 2.2). Insgesamt gilt, dass sich die erste Forschungsfrage nicht eindeutig und v.a. nicht ohne Bezug auf die zweite Forschungsfrage beantworten lässt. Es kann zwar eine generelle Verbesserung von *LV-Speed* zu *LV-Power* festgestellt werden, diese scheint jedoch zumindest mit der Lesegeschwindigkeit im Deutschen nicht systematisch in Verbindung zu stehen, möglicherweise jedoch mit den muttersprachlichen Verstehenskompetenzen. Allerdings verbessert sich die Leistung von *LV-Speed* zu *LV-Power* nur relativ geringfügig. Dies lässt sich möglicherweise damit erklären, dass die Schüler durch einen relativ häufigen unterrichtlichen Einsatz von Tests und Übungen gut mit Zeitdruck umgehen können und auch deshalb viele Items bereits in *LV-Speed* korrekt gelöst haben. Zugleich ist jedoch auch denkbar, dass auf Grund des langen Tests und der damit verbundenen hohen kognitiven Belastung eine gewisse Ermüdung und somit nachlässige Bearbeitung der Items in *LV-Power* zu verzeichnen ist und dass v.a. aus diesem Grunde keine grundlegende Verbesserung unter *Power*-Bedingungen erzielt wurde. Weiterhin ist denkbar, dass die Schüler angesichts des *low stakes*-Charakters der Testdurchführung nicht hinreichend motiviert waren.

Tabelle 4 stellt die korrelativen Zusammenhänge zwischen den einzelnen Testinstrumenten genauer dar.²⁴

23 Wir sind uns bewusst, dass die Verwendung von Rohwertdifferenzen zur Messung von Veränderungen mit einer Reihe von Problemen verbunden ist (vgl. Sternberg & Grigorenko 2002: 165-168). In Anbetracht des explorativen Charakters unserer Untersuchung haben wir auf den Einsatz anspruchsvollerer statistischer Verfahren verzichtet.

24 Zur Kontrolle wurden zusätzlich Partialkorrelationen berechnet, die jedoch keine wesentlichen Veränderungen erbrachten.

	Lesever- ständnis	Lesegeschwin- digkeit	C-Test	LV-Speed	LV-Power
Konzentrations- leistung	-0.03	0.30*	0.35*	0.12	0.32*
Leseverständnis		0.50**	0.16	0.16	0.31*
Lesegeschwin- digkeit			0.24	0.29*	0.30*
C-Test				0.57**	0.66**
LV-Speed					0.61**

* signifikant auf dem Niveau 0.05; ** signifikant auf dem Niveau 0.01 (beides einseitig)

Tabelle 4: Korrelation der Leistungstestinstrumente

Die Korrelation der Konzentrationsleistung mit der Lesegeschwindigkeit, dem C-Test sowie LV-Power ist mäßig, jedoch statistisch signifikant. Die Korrelationen deuten darauf hin, dass Schüler, die sich gut konzentrieren können, auch schneller lesen können und zudem auch ein besseres Verständnis französischsprachiger Texte erzielen. Die Korrelation der Konzentrationsleistung mit LV-Speed ist dagegen nicht signifikant. Die Konzentrationsfähigkeit spielt offensichtlich erst dann eine Rolle, wenn die getesteten Schüler die Zeit haben, sich intensiver mit den Aufgaben zu beschäftigen. Dabei ist im Hinblick auf die Korrelation zwischen Konzentrationsleistung und Lesegeschwindigkeit zu berücksichtigen, dass nicht nur im Lesegeschwindigkeitstest, sondern auch in dem zur Messung der Konzentrationsfähigkeit eingesetzten d2-Aufmerksamkeits-Belastungs-Test die Aufgaben unter Zeitdruck bearbeitet werden. Möglicherweise kann die gemeinsame Varianz zumindest teilweise auf diesen Methodenfaktor zurückgeführt werden.²⁵

Die moderate Korrelation zwischen der Lesegeschwindigkeit und LV-Speed deutet darauf hin, dass Schüler, die schnell lesen können, die Aufgaben unter Zeitdruck auch besser bearbeiten können. Schnelle Leser sind hier demnach im Vorteil. Gleiches gilt für den Zusammenhang der Lesegeschwindigkeit mit LV-Power. Das Leseverständnis des deutschen Textes korreliert allerdings nur mit LV-Power signifikant. Die geringe und nicht-signifikante Korrelation mit LV-Speed entspricht dabei den Erwartungen und bedeutet, dass unter Zeitdruck die höheren muttersprachlichen Lesefähigkeiten beim Bearbeiten französischsprachiger Leseaufgaben nicht systematisch genutzt werden können. Der C-Test gilt als Test allgemeiner

25 So argumentiert Bühner (2006: 40) unter Bezug auf Wilhelm & Schulze (2002): "Beispielsweise korrelieren ein zeitbegrenzter Intelligenztest und ein zeitbegrenzter Konzentrationstest miteinander, da beide Tests die gleiche Methode zur Messung unterschiedlicher Fähigkeiten verwenden".

Sprachkompetenz. Somit ist zu erwarten, dass Schüler mit guten Leistungen im französischen C-Test auch gute Leistungen in den beiden französischsprachigen Testteilen zum Leseverstehen erbringen (vgl. die Abschnitte 2.2 und 4.3.2). Die Korrelation des C-Tests mit *LV-Power* ist dabei ungleich höher als mit *LV-Speed*. Ohne Zeitdruck kann die allgemeine Sprachkompetenz demnach besser eingesetzt werden.²⁶

Der beschriebene Zusammenhang zwischen C-Test-Leistung und *LV-Speed* bzw. *LV-Power* zeigt sich auch in Abb. 3. In der vorliegenden Studie erreichte kein Schüler die oberen Perzentile mit 61 bis 80 und 81 bis 100 Prozent korrekt gelöster Lücken. Allgemein wird jedoch sichtbar, dass die Schüler, die im französischen C-Test relativ viele Lücken lösen konnten, auch in den beiden französischen Lesetests gut abschnitten. Dies gilt insbesondere für *LV-Power*.

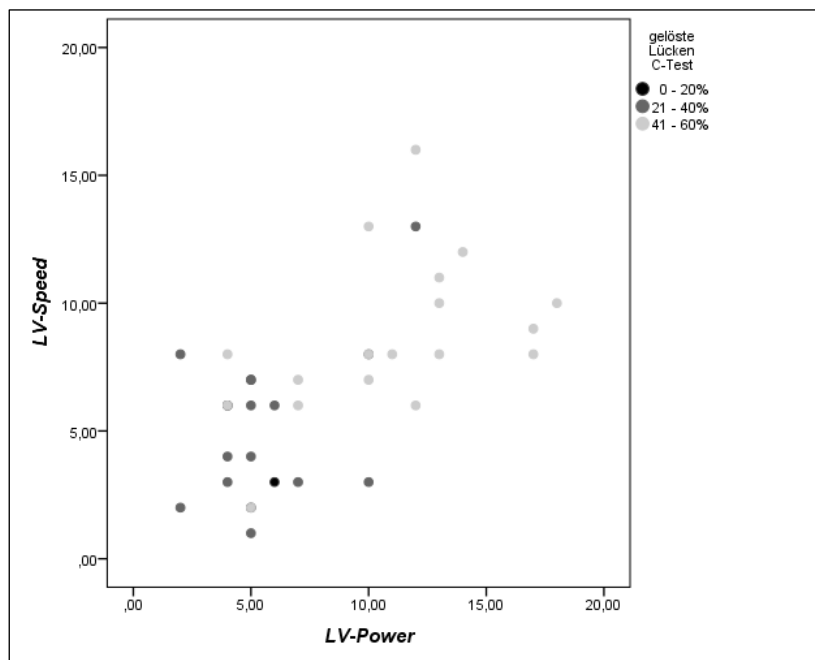


Abbildung 3: Darstellung des Zusammenhangs von C-Test mit *LV-Speed* und *LV-Power*

²⁶ Es ist zu erwarten, dass sich bei einer stärkeren Zeitlimitierung auch des C-Tests eine höhere Korrelation zwischen C-Test-Leistung und *LV-Speed* ergeben hätte (vgl. Grotjahn 2010; Grotjahn, Schlak & Aguado 2010).

Die numerisch relativ hohe Korrelation zwischen *LV-Speed* und *LV-Power* war zu erwarten. Diese Korrelation kann auch als Retest-Reliabilität bei einer teilweisen Variation der Durchführungsbedingungen interpretiert werden. Ein Teil der gemeinsamen Varianz von *LV-Speed* und *LV-Power* dürfte auch auf einem Übungseffekt beruhen (vgl. Sternberg & Grigorenko 2002: 120).²⁷ Bei einer Testwiederholung unter identischen Durchführungsbedingungen würde man allerdings eine deutlich höhere Test-Retest-Korrelation erwarten. So berichten z.B. Grotjahn, Klein-Braley & Raatz (2002: 100) für unterschiedliche C-Tests Test-Retest-Korrelationen im Bereich von 0.70 bis 0.83. Allerdings handelte es sich in allen Fällen um C-Tests mit einer hohen Reliabilität im Sinne innerer Konsistenz, was die Test-Retest-Reliabilität tendenziell erhöht. Im vorliegenden Fall könnte man auch noch Erinnerungseffekte annehmen und würde auch deswegen eine relativ hohe Korrelation erwarten. Die im Vergleich mit Korrelationskoeffizienten aus Testwiederholungen unter identischen Durchführungsbedingungen eher geringe Korrelation von 0.61 zwischen *LV-Speed* und *LV-Power* kann deshalb sicherlich zum Teil auf den unterschiedlichen Zeitdruck bei der Darbietung, aber auch auf Ermüdung im Testteil *LV-Power* zurückgeführt werden (zur Problematik der Interpretation von Korrelationskoeffizienten aus Testdurchführungen mit und ohne Zeitdruck vgl. auch Rindler 1979: 268).

4.4.2 Fragebogen

Wie bereits in Abschnitt 4.3.2 beschrieben, wurde zur Erhebung von Daten zu den Lesestrategien der Schüler ein Fragebogen eingesetzt, der aus zwei Skalen mit vierstufigen Likert-Items bestand und im Wesentlichen auf Groß (2000) basiert. Mit Hilfe einer Faktorenanalyse (Hauptkomponentenanalyse, Varimax) wurden die Befragungsdaten genauer betrachtet, wobei allerdings die Ergebnisse der Faktorenanalyse bei einer Stichprobengröße von lediglich $N = 36$ mit größter Vorsicht zu betrachten sind. Es ergaben sich in unserer Schülerstichprobe zwei Faktoren, die den beiden Skalen von Groß (2000) entsprechen. Eine mittlerweile gängige Faustregel zur Identifikation "guter" Skalen (Schmitt 1996) besagt, dass die Reliabilität im Sinne von Cronbachs α nicht kleiner als 0.5 sein sollte. Für unsere Daten ergab sich für beide Originalskalen eine zum Teil deutlich geringere Reliabilität. Es konnte aber letztendlich aus beiden Skalen eine aus sechs Items beste-

27 Dieser Effekt könnte mit Hilfe einer Kontrollgruppe, die lediglich *LV-Power* bearbeitet, genauer bestimmt werden.

hende Skala "Strategisches Lesen" gebildet werden, deren Reliabilität bei $\alpha = 0.62$ bzw. $\lambda_2 = 0.65$ lag.

Die Korrelationen dieser Skala "Strategisches Lesen" mit den Leistungstestinstrumenten fielen durchgängig nicht statistisch signifikant aus. Es wurden deshalb die ursprünglichen 14 Items aus dem Fragebogen nach Groß (2000) mit den Leistungstests korreliert. Die Darstellung der Ergebnisse erfolgt in Tabelle 5.

Item	Konzentration	Leseverständnis	Lesegeschwindigkeit	C-Test	LV-Speed	LV-Power
01: Lesen nach bestimmtem Ziel	0.26	0.41**	0.42**	0.43**	0.29*	0.14
02: jedes Wort verstehen wollen	0.25	0.01	0.24	0.17	-0.02	-0.14
03: auf zentrale Aussagen konzentrieren	0.13	0.35*	0.28*	0.23	0.30*	0.29*
04: Gedanken über Gesamtbedeutung des Textes machen	-0.04	-0.08	-0.13	0.12	-0.06	-0.02
05: schwierige Stellen später erschließen	-0.16	0.14	0.07	0.38*	0.21	0.30*
06: bei Schwierigkeiten Lesen beenden	0.05	0.05	-0.04	0.21	-0.12	0.16
07: Textteile bei Schwierigkeiten überspringen	-0.20	-0.11	-0.23	0.05	-0.10	-0.18
08: Textstellen bei Problemen mehrfach lesen	-0.24	-0.02	-0.17	0.00	0.04	-0.23
09: Textstelle vor Problemstelle genauer lesen	-0.11	-0.12	-0.27	0.08	-0.06	-0.09
10: bewusst langsamer lesen bei Problemen	0.13	0.13	0.36*	0.07	-0.03	-0.07
11: Bedeutung von Wörtern aus Zusammenhang erschließen	0.02	0.09	-0.09	0.19	-0.02	0.22
12: andere Textstellen für Problemlösung heranziehen	-0.09	0.51**	0.21	0.09	-0.08	0.04
13: Themenwissen einsetzen	-0.18	0.04	-0.02	0.04	0.09	-0.07
14: Bindewörter zum Verständnis nutzen	0.02	-0.03	-0.03	-0.27	-0.33*	-0.33*

* signifikant auf dem Niveau 0.05; ** signifikant auf dem Niveau 0.01 (beides einseitig)

Tabelle 5: Korrelation der Fragebogendaten mit den Leistungstestdaten

Insgesamt zeigten sich bei sechs der 14 Fragebogenitems statistisch signifikante Korrelationen.²⁸ Das Item 1 'Lesen nach bestimmtem Ziel' ("Wenn ich französische Texte lese, überlege ich zuerst, mit welchem Ziel ich den Text lese, z.B. den Text schnell überfliegen, um das Wichtigste zu erkennen oder im Gegenteil besonders langsam und gründlich lesen, um alles genau zu verstehen") korreliert zu 0.41 mit dem Leseverständnis, zu 0.42 mit der Lesegeschwindigkeit, zu 0.43 mit dem C-Test und zu 0.29 mit *LV-Speed*. Schüler, die nach eigener Aussage einen Text mit einem bestimmten Ziel lesen, schneiden somit in den eindeutig lesekompetenzbezogenen Testteilen dieser Studie tendenziell besser ab. Insbesondere in *LV-Speed* und *LV-Power* ist es unabdingbar, die Texte unter einem bestimmten Gesichtspunkt zu bearbeiten; dies gilt noch viel mehr, wenn die Texte unter Zeitdruck bearbeitet werden. Die Strategie, Texte mit einem bestimmten Ziel zu bearbeiten, unterstützt Schüler demnach bei der erfolgreichen Bearbeitung.

'Sich auf zentrale Aussagen zu konzentrieren' (Item 3: "Wenn ich französische Texte lese, konzentriere ich mich nur auf die zentralen Aussagen des Textes") hilft den Schülern unserer Studie beim Leseverständnis ($r = 0.35$), bei der Lesegeschwindigkeit ($r = 0.28$) und im Testteil *LV-Speed* ($r = 0.29$). Besonders in Hinblick auf zeitgesteuerte Aufträge scheint es somit sinnvoll, sich auf das Wesentliche, d.h. die zentralen Aussagen, zu konzentrieren.

Das Fragebogen-Item 5 'schwierige Stellen später erschließen' ("Wenn ich beim Lesen französischer Texte Verständnisschwierigkeiten habe, fahre ich mit dem Lesen fort und hoffe, dass ich die schwierige Stelle später verstehe") korreliert statistisch signifikant mit dem C-Test zu 0.38 und mit *LV-Power* zu 0.30. Besonders im C-Test, in dem Lücken ausgefüllt werden sollen, ist es wichtig, dass die Lösungen der zu rekonstruierenden Lücken aus dem Kontext erschlossen werden können. Dies wird jedoch manchmal erst möglich, wenn man die entsprechende unbekannte Stelle vorerst überspringt. Ebenso wenig ist die statistisch signifikante Korrelation des Items mit *LV-Power* verwunderlich, da die Schüler hier ihre Zeit nutzen konnten, um den Text in Ruhe zu erschließen. Schüler, die auch beim regulären Lesen in französischer Sprache dazu tendieren, schwierige Textstellen vorerst zu überspringen und sie sich später zu erschließen, schneiden auch im C-Test und in *LV-Power* besser ab.

Verwunderlich ist jedoch die schwache positive Korrelation des Items 10 'bewusst langsamer lesen bei Problemen' ("Wenn ich beim Lesen französi-

28 Bei der Interpretation der Signifikanzen ist die Erhöhung des Alpha-Fehlers im Fall multipler Tests zu berücksichtigen.

scher Texte Verständnisschwierigkeiten habe, dann lese ich bewusst langsamer") mit der Lesegeschwindigkeit ($r = 0.36$). Schüler mit einer relativ geringen Lesegeschwindigkeit tendieren somit nach eigener Aussage dazu, bei Verständnisproblemen nicht bewusst langsamer zu lesen. Eventuell ist es für diese Schüler auch nicht notwendig, langsamer zu lesen, da eine hohe Lesegeschwindigkeit in engem Zusammenhang mit dem Leseverständnis gesehen werden kann und somit die Anwendung der mit Item 10 beschriebenen Strategie nicht notwendig wird.

Weiterhin gibt es einen statistisch signifikanten Zusammenhang ($r = 0.51$) zwischen dem Item 12 'andere Textstellen für Problemlösung heranziehen' ("Wenn ich beim Lesen französischer Texte Verständnisschwierigkeiten habe, dann gehe ich zu anderen Stellen im Text, die mir vielleicht beim Lösen meines Problems helfen könnten") und dem Leseverständnistest. Der Leseverständnistest macht es erforderlich, dass passende Wörter im Kontext erschlossen werden. Somit ist eine relativ hohe Korrelation zwischen der genannten Strategie und dem Leseverständnistest zu erwarten.

Das letzte Item, das Korrelationen mit einem Leistungstest zeigt, ist das Item 14 'Bindewörter zum Verständnis nutzen' ("Wenn ich beim Lesen französischer Texte Verständnisschwierigkeiten habe, achte ich meistens auf Wörter wie z.B. *parce que...*, die Sätze oder Satzteile verbinden, um den Text zu verstehen"). Dieses Item korreliert statistisch signifikant ($r = -0.33$) mit *LV-Speed* und auch mit *LV-Power*. Interpretiert hieße das, je weniger Schüler nach eigener Aussage auf Bindewörter achten, desto erfolgreicher sind sie im Test *LV-Speed* und *LV-Power*. Es ist zu vermuten, dass den Schülern bei der Bearbeitung von *LV-Speed* keine Zeit bleibt, sich strategisch auf Bindewörter zu konzentrieren und dass die Korrelation deshalb negativ ausfällt. Eventuell greift hier jedoch auch die Interpretation zu Item 10: Durch den hohen Zusammenhang der Lesegeschwindigkeit mit dem Leseverständnis erreichen Schüler selbst bei schnellem Lesen ein so hohes Verständnis, dass sie auf diese Strategie (Bindewörter finden) nicht mehr zurückgreifen müssen.

5. Diskussion

In der vorliegenden Studie wurde untersucht, welchen Einfluss eine Beschränkung der Bearbeitungszeit auf die Lösung der Aufgaben hat. Zumindest in dieser Studie konnte auf der Ebene der Lösungen (korrekt, inkorrekt) ein Zusammenhang zwischen der limitierten Bearbeitungszeit und der Lö-

sung der Aufgaben in den Testteilen *LV-Speed* und *LV-Power* festgestellt werden. In einem zweiten Schritt wurden die einzelnen Leistungstests betrachtet. Es konnten insbesondere folgende Zusammenhänge gezeigt werden: zwischen Lesegeschwindigkeit und Konzentrationsleistung; zwischen Leseverständnis und *LV-Power*; zwischen *LV-Speed* und *LV-Power*. Des Weiteren wurde ein Zusammenhang zwischen dem C-Test und den beiden Testteilen *LV-Speed* und *LV-Power* nachgewiesen. Bezüglich der Lesestrategien lassen sich einige Zusammenhänge mit den Testleistungen erkennen. Generell konnte gezeigt werden, dass die Anwendung geeigneter Lesestrategien die Bearbeitung der Leistungstests eher begünstigt. Einzige Ausnahme bildet das Fragebogen-Item 'Bindewörter', das mit *LV-Speed* statistisch signifikant negativ korreliert.

Abschließend kann gesagt werden, dass *Speed* als Teilkomponente von Lesekompetenz in unserer Untersuchung vermutlich nicht präzise genug operationalisiert wurde, z.B. wegen des Aufbaus eines unzureichenden Zeitdrucks. Dieser Aspekt müsste (auch in Folgestudien) noch gründlicher erforscht werden.

Bei der Interpretation der Ergebnisse der vorliegenden Studie müssen einige Einschränkungen beachtet werden. Zum einen muss die Möglichkeit eines Testwiederholungseffekts bei der Interpretation der relativ hohen Übereinstimmung zwischen den Ergebnissen in *LV-Speed* und *LV-Power* in Betracht gezogen werden. So ist z.B. zu erwarten, dass einige Testteilnehmer die Aufgaben in *LV-Power* nicht mehr sorgfältig bearbeitet haben, weil sie zu Recht oder auch zu Unrecht glaubten, bereits eine richtige Lösung in *LV-Speed* gefunden zu haben. Auch ein Ermüdungseffekt könnte ein Grund für eine unzureichende Bearbeitung der Aufgaben in *LV-Power* und für die relativ hohe Übereinstimmung in den Lösungen zwischen *LV-Speed* und *LV-Power* sein. Weiterhin könnte eine unzureichende Motivation der Schüler angesichts des *low stakes*-Charakters der Testdurchführung die Bereitschaft zur Revision der Lösungen unter der *Power*-Bedingung beeinflusst haben. Hätte sich dagegen die Leistung von *LV-Speed* zu *LV-Power* deutlich verbessert – und zwar differentiell für die 36 Testteilnehmer – so hätte dies auch zu einer Verringerung der Korrelation zwischen *LV-Speed* und *LV-Power* geführt. Die ausgebliebene grundlegende Verbesserung und die relativ hohe Korrelation von *LV-Speed* zu *LV-Power* sprechen somit für den Einfluss von Variablen wie Testwiederholung, Ermüdung und Motivation.

Des Weiteren ist nicht bekannt, welchen Einfluss das Lesen am Computer hatte. Obwohl man davon ausgehen kann, dass viele Schüler mit dem Lesen an Computern gut vertraut sind (z.B. durch Blogs, soziale Netzwer-

ke), könnte das Lesen längerer, fremdsprachlicher Texte am Computer eher ungewohnt sein. Hieraus könnte ein differentieller Effekt auf das Leseverständnis unter den Bedingungen *Speed* und *Power* resultieren. Inwiefern dies wirklich der Fall war, konnte auf Grund des Fehlens einer Kontrollgruppe nicht erfasst werden.

Schließlich müssen auch der geringe Stichprobenumfang ($N = 36$) und die damit verbundenen Einschränkungen in den Analyseverfahren bei der Interpretation der Ergebnisse berücksichtigt werden. Ein Grund für den geringen Stichprobenumfang ist die explorative Zielsetzung unserer Studie. Außerdem haben wir ursprünglich mit einer höheren Beteiligung gerechnet, konnten jedoch nicht ausreichend Schüler akquirieren. Auf Grund von Zeitplänen und damit verbundenen organisatorischen Aspekten war ein Verschieben der Testungen nicht möglich. Es ist durchaus denkbar, dass bei einer größeren Stichprobe noch weitere Korrelationskoeffizienten statistisch signifikant ausgefallen wären (oder auch einige statistisch signifikante Korrelationen unter die Signifikanzgrenze gefallen wären). Auch im Fall größerer Stichproben mögliche komplexe multivariate Analysen hätten eventuell zu partiell anderen Ergebnissen geführt.²⁹ Mit einer ausreichend großen Stichprobe wäre es auch möglich gewesen, eine Kontrollgruppe zu bilden, die die Aufgaben im Testteil Leseverstehen ausschließlich ohne Zeitbegrenzung, d.h. nur *LV-Power*, bearbeitet.

Eine weitere Einschränkung könnte in der Setzung des Leseziels ohne Berücksichtigung der Leseinteressen gesehen werden. Man könnte davon ausgehen, dass Schüler für sie interessante Texte mit einer höheren Motivation lesen und sich dadurch auch mehr im Test anstrengen. Unser Vorgehen hingegen birgt einige Parallelen zum unterrichtlichen Lesen, bei dem auch häufig für Schüler uninteressante Texte gelesen werden müssen.

Insbesondere führt aber der rückblickend zu schwach gewählte Zeitdruck in *LV-Speed* zu den wenig belastbaren Ergebnissen. Hier wäre zu überlegen, bei Folgestudien den Zeitdruck nicht nur auf Aufgabenebene zu beschränken, sondern die Bearbeitungszeit (auch) auf Itemebene zu begrenzen und diese Begrenzung zu untersuchen. Hierbei müsste die zur Verfügung stehende Zeit jedoch noch sorgfältiger gewählt werden, als es in unserer Studie der Fall war. Diese Erkenntnis und die Ausführungen in Abschnitt 3.3 zur Begrenzung der Bearbeitungszeit in Fremdsprachentests sollten als Hinweis

29 Wie bereits erwähnt, wurden auch Partialkorrelationen berechnet, die jedoch die Ergebnisse der bivariaten Analysen bestätigten.

auf die Notwendigkeit gesehen werden, in künftigen Untersuchungen zur Bearbeitungszeit den Zeitdruck zu erhöhen.

Trotz der Einschränkungen kann diese Studie einige Hinweise und Empfehlungen für den Unterricht liefern. So hat z.B. ein Vergleich der Leseverstehensleistung unter *Speed*- und *Power*-Bedingungen auch ein nicht unerhebliches diagnostisches Potential im Hinblick auf unterrichtliche Interventionsmöglichkeiten. Testet man Leseverstehen nur unter *Power*-Bedingungen, dann ist u.a. denkbar, dass sich Lerner trotz ähnlicher Testwerte deutlich im Grad der Automatisierung ihrer Kompetenzen unterscheiden. Setzt man einen Leseverstehenstest jedoch sowohl unter *Speed*- als auch unter *Power*-Bedingungen ein und erhält bei einem Lerner einen hohen Zuwachs richtiger Lösungen von *LV-Speed* nach *LV-Power*, dann ist dies ein mögliches Indiz, dass die zur Lösung der Aufgaben notwendigen sprachlichen Wissensbestände bei dem entsprechenden Lerner unzureichend prozeduralisiert bzw. automatisiert sind. Solche Lerner müssten dann v.a. bei der Automatisierung ihres Wissens unterstützt werden.

Die Studie liefert auch Hinweise in Bezug auf ein unterrichtliches Training von Lesestrategien. Die Bewusstmachung und Vermittlung von Strategien wird von fachdidaktischer Seite seit jeher gefordert (vgl. den Überblick in Oxford 2011). Gestützt wird diese Forderung durch eine Reihe von empirischen Untersuchungen. So konnten z.B. Souvignier & Mokhlesgerami (2005) zeigen, dass Klassen, in denen Lesestrategien explizit trainiert wurden, eine deutliche Verbesserung im Wissen über Lesestrategien und deren Einsatz zeigten. Auch die Korrelationen zwischen einzelnen Strategien und Leistungstests in unserer Untersuchung deuten darauf hin, dass Schüler, die Lesestrategien bewusst nutzen, bessere Leistungen erzielen. Eine deutliche Steigerung der Lesegeschwindigkeit dürfte ohne begleitende Bewusstmachung der Strategien nicht möglich sein. Im Zusammenhang damit erscheint es sinnvoll, Lerner auf die verschiedenen Ebenen des Strategienzugriffs aufmerksam zu machen: Mikrostrategien der Wort- und Satzdekodierung, Mesostrategien der Lesezielidentifikation und Makrostrategien der Selbstkontrolle im Hinblick auf Leseinteresse und Konzentration.

Ein weiteres Ergebnis der Studie ist der starke Zusammenhang der Aufgabenbearbeitung mit der Konzentrationsfähigkeit. Empfehlenswert scheint somit, das schnelle Lesen unter Berücksichtigung des Leseverständnisses sowie der Konzentrationsfähigkeit in der Klasse zu üben. Beide Aspekte zeigten einen zum Teil relativ hohen Zusammenhang mit der weiteren Leistung in *LV-Speed* und *LV-Power*. Schüler, die sich gut konzentrieren können, haben weniger Schwierigkeiten, eine große Anzahl an Aufgaben auch

unter Zeitlimitierung zu bearbeiten. Dies dürfte unabhängig vom Vorwissen sein und sich somit auch als durch Übung steigerungsfähig erweisen.

Unklar bleibt indes, wie tief das Verstehen tatsächlich geht. Weitere Forschungen mit deutlich stärkerer Zeitbegrenzung müssten zeigen, welche Verarbeitungs- bzw. Verstehentiefe dann noch möglich ist bzw. wo die Konzentrationsfähigkeit nicht mehr ausreicht, um Texte hinreichend tief zu verarbeiten. In diesem Zusammenhang müssten auch die unterschiedlichen Verstehensprozesse bei literarischen im Vergleich zu nicht-literarischen Texten berücksichtigt und näher erforscht werden, um Aussagen über die Mobilisierung spezifischer Wissensressourcen treffen zu können. Welches Genrewissen, welches Textstruktur- und Textmusterwissen, welches Thema- und welches Weltwissen, welche motivationalen Ressourcen werden jeweils aktiviert? Die Vermutung liegt nahe, dass Faktoren für Leseleistungsunterschiede beim literarischen im Vergleich zum nicht-literarischen Lesen nicht ohne weiteres zu isolieren sind und eher auf unterschiedliche Leserpersönlichkeiten zurückgehen. Dies wiederum würde für den Fremdsprachenunterricht bedeuten, dass die genannten Faktoren bzw. Wissensressourcen nicht stillschweigend vorausgesetzt sondern vielmehr thematisiert und bewusst gemacht werden sollten. Die Modellierung einer Text- und Medienkompetenz (z.B. KMK 2012, Schmölzer-Eibinger 2008) beschreitet genau diesen Weg. Als Konsequenz für weitere VERA-Testungen im Fach Französisch kann geschlossen werden, dass Schüler noch stärker auf den Einsatz geeigneter Lesestrategien hingewiesen werden sollten als dies bisher möglicherweise der Fall war. Dies kann explizit durch entsprechende Anweisungen geschehen, eventuell mit Hinweis auf die Beachtung des Zeitbudgets. Ebenso ist es sinnvoll, durch eine Voranstellung der Items das Lesen des Stimulustextes zu steuern. Schließlich ist zu erwägen, ob Aufgaben, die schnelle Lesevorgänge erfordern, d.h. insbesondere das Erfassen der Kernaussage sowie das gezielte Auffinden einer bestimmten Information, zusätzlich zur Voranstellung des oder der Items in einer Leseinheit zusammengefasst werden sollten. Zuweilen haben Schüler in den VERA-Tests ein Zeiteinteilungsproblem: Durch schlechtes Zeitmanagement erreichen sie nicht das Ende des Testheftes. Als Folge wird ihre Lesekompetenz unter Umständen unterschätzt. Eine Zusammenfassung von Aufgaben – z.B. zwei Aufgaben zum globalen Leseverstehen müssen bei freier Einteilung der Zeit innerhalb von fünf Minuten gelöst werden – könnte dieses Problem ein Stück weit umgehen. Durch zwei deutlich getrennte Aufgabensequenzen könnten wenig flexible Leser trennschärfer in ihrer Fähigkeit beurteilt werden, Lesestrategien für bestimmte Leseziele auszuwählen. Zudem hätten sie die Möglichkeit, al-

le Aufgaben im Testheft zu bearbeiten. Es kann davon ausgegangen werden, dass die Schüler fünf Minuten für zwei Aufgaben besser einteilen können als 20 Minuten für 10 bis 12 Aufgaben.

Neben der Voranstellung der Items scheint es jedoch auch sinnvoll, den Faktor Zeit stärker zu kontrollieren (kontextuelle und kognitive Validität). Insbesondere in Hinblick auf die Vorbereitung auf das Abitur und die Studierfähigkeit wird der Aspekt einer schnellen Informationsentnahme immer wichtiger.

Eine weitere Perspektive für VERA zeigt sich jedoch auch im Einsatz von computerbasierten Testverfahren. Dieser Ansatz ist dem IQB nicht neu, und der Wunsch nach Implementierung wurde bereits an das IQB herangetragen. Diese Studie hat gezeigt, dass die Bearbeitung von VERA-Aufgaben am Computer tendenziell möglich ist und zudem noch weitere Möglichkeiten bietet, die im Rahmen unserer Studie nicht weiter ausgeführt wurden (Animationen, eigenständige Steuerung von Hörsequenzen, etc.).³⁰ Über die Vergleichbarkeit von computerbasierten und klassischen *Paper-Pencil*-Verfahren wurde im Abschnitt 4.3 berichtet. Darüber hinaus wäre eine computerbasierte Testung des Lese- und Hörverstehens eine enorme Erleichterung für die beteiligten Lehrkräfte, da ihnen sämtliche Auswertungen abgenommen werden könnten. Einschränkungen sehen wir lediglich bei der technischen Ausstattung von Schulen: Nicht in allen Schulen wird zum jetzigen Zeitpunkt eine computerbasierte Testung möglich sein.

Eingang des revidierten Manuskripts 01.03.2013

Literaturverzeichnis

- Alderson, J. Charles (2000), *Assessing Reading*. Cambridge: Cambridge University Press.
- Al-Othman, Nawal M. (2003), The relationship between online reading rates and performance on proficiency tests. *The Reading Matrix* 3: 3, 120-136.
- Artelt, Cordula; McElvany, Nele; Christmann, Ursula; Richter, Tobias; Groeben, Norbert; Köster, Juliane; Schneider, Wolfgang; Stanat, Petra; Ostermeier, Christian; Schiefele, Ulrich; Valtin, Renate; Ring, Klaus & Saalbach, Henrik (2007), *Förderung von Lesekompetenz – Expertise*. Berlin: BMBF.

30 Vgl. z.B. die Studie von Porsch, Grotjahn & Tesch (2010) zum Hör-Seh-Verstehen, die wie die vorliegende Untersuchung im Rahmen eines Kooperationsprojektes des IQB mit dem Seminar für Sprachlehrforschung der Ruhr-Universität Bochum entstanden ist.

- Artelt, Cordula; Stanat, Petra; Schneider, Wolfgang & Schiefele, Ulrich (2001), Lesekompetenz: Testkonzeption und Ergebnisse. In: Baumert, Jürgen; Klieme, Eckhard; Neubrand, Martin; Prenzel, Manfred; Schiefele, Ulrich; Schneider, Wolfgang; Stanat, Petra; Tillmann, Klaus-Jürgen & Weiß, Manfred (Hrsg.) (2001), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich, 69-137.
- Bachman, Lyle F. (1990), *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bernhardt, Elizabeth B. (2011), *Understanding Advanced Second-Language Reading*. New York: Routledge.
- Brickenkamp, Rolf (2002), *Test d2: Aufmerksamkeits-Belastungs-Test* (9., überarbeitete und neu normierte Auflage). Göttingen: Hogrefe.
- Bridgeman, Brent; McBride, Amanda & Monaghan, William (2004), *Testing and time limits* (R&D Connections 1). Princeton, NJ: Educational Testing Service. [Online: http://www.ets.org/Media/Research/pdf/RD_Connections1.pdf. 15.09.2012]
- Bühner, Markus (2006), *Einführung in die Test- und Fragebogenkonstruktion* (2., aktual. und erw. Aufl.). München: Pearson Studium.
- Carver, Ronald (1985), How good are some of the world's best readers? *Reading Research* 20: 4, 389-419.
- Carver, Ronald P. (1997), Reading for one second, one minute, or one year from the perspective of rauding theory. *Scientific Studies of Reading* 1: 1, 3-43.
- Choi, Inn-Chull; Kim, Kyoung Sung & Boo, Jaeyool (2003), Comparability of a paper-based language test and a computer-based language test. *Language Testing* 20: 3, 295-320.
- Cronjäger, Hanna; Klapheck, Kristine; Krätzschar, Maike & Walter, Oliver (2010), Entwicklung eines C-Tests für Lernanfänger der Sek. I mit Methoden der klassischen und probabilistischen Testtheorie. In: Grotjahn, Rüdiger (Hrsg.) (2010), *Der C-Test: Beiträge aus der aktuellen Forschung/The C-Test: Contributions from Current Research*. Frankfurt am Main: Lang, 71-100.
- Eckes, Thomas & Grotjahn, Rüdiger (2006), A closer look at the construct validity of C-tests. *Language Testing* 23: 3, 290-325.
- Ehlers, Swantje (2007), Lesetheorien, Lesekompetenz und Narrative. In: Bredella, Lothar & Hallet, Wolfgang (Hrsg.) (2007), *Literaturunterricht, Kompetenzen und Bildung*. Trier: Wissenschaftlicher Verlag Trier, 109-124.
- Enright, Mary K.; Bridgeman, Brent; Eignor, Daniel; Kantor, Robert N.; Mollaun, Pamela; Nissan, Susan; Powers, Donald E. & Schedl, Mary (2008), Prototyping new assessment tasks. In: Chapelle, Carol; Enright, Mary K. & Jamieson, Joan M. (Hrsg.) (2008), *Building a Validity Argument for the Test of English as a Foreign LanguageTM*. New York: Routledge, 97-143.
- Enright, Mary K.; Grabe, William; Koda, Keiko; Mosenthal, Peter; Mulcahy-Ernt, Patricia & Schedl, Mary (2000), *TOEFL 2000 Reading Framework: A Working Paper* (TOEFL Research Report MS-17). Princeton, NJ: Educational Testing Service. [Online: <http://www.ets.org/Media/Research/pdf/RM-00-04.pdf>. 15.09.2012].
- Europarat (2001), *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren, beurteilen*. Berlin: Langenscheidt.
- Fraser, Carol A. (2007), Reading rate in L1 Mandarin Chinese and L2 English across five reading tasks. *The Modern Language Journal* 91: 3, 372-394.
- Fulcher, Glenn (2010), *Practical Language Testing*. London: Hodder Education.

- Grabe, William (2009), *Reading in a Second Language. Moving from Theory to Practice*. Cambridge: Cambridge University Press.
- Grabe, William (2010), Fluency in reading: Thirty-five years later. *Reading in a Foreign Language* 22: 1, 71-83 [Online: <http://nflrc.hawaii.edu/rfl/April2010/articles/grabe.pdf>. 01.03.2013].
- Groß, Annette (2000), *Verstehensprozesse beim Lesen fremdsprachlicher Hypertexte: Eine empirische Untersuchung*. Dissertation Bergische Universität Wuppertal [Online: <http://elpub.bib.uni-wuppertal.de/servlets/DerivateServlet/Derivate-278/d040004.pdf>. 01.03.2013].
- Grotjahn, Rüdiger (2002), Konstruktion und Einsatz von C-Tests: Ein Leitfaden für die Praxis. In: Grotjahn, Rüdiger (Hrsg.) (2002), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 4). Bochum: AKS-Verlag, 211-225 [Online: http://homepage.ruhr-uni-bochum.de/ruediger.grotjahn/Grotjahn_KonstruktionC-Test_2002.pdf. 01.03.2013]
- Grotjahn, Rüdiger (2010), Gesamtdarbietung, Einzeldarbietung, Zeitbegrenzung und Zeitdruck: Auswirkungen auf Item- und Testkennwerte und C-Test-Konstrukt. In: Grotjahn, Rüdiger (Hrsg.) (2010), *Der C-Test: Beiträge aus der aktuellen Forschung/The C-Test: Contributions from Current Research*. Frankfurt am Main: Lang, 265-296.
- Grotjahn, Rüdiger (2011), C-Tests – Aspekte der Validität. *Deutsch als Fremdsprache* 48: 3, 131-137.
- Grotjahn, Rüdiger (2012), Theoretical misconceptions and misuse of statistics: A critique of Khodadady and Hashemi (2011) and some general remarks on Cronbach's alpha. *Iranian Journal of Language Testing* 2: 1, 20-27 [Online: <http://ijlt.ir/portal/images/PDF/406-2012-2-1.pdf>. 15.09.2012].
- Grotjahn, Rüdiger; Klein-Braley, Christine & Raatz, Ulrich (2002), C-Tests: An overview. In: Coleman, James A.; Grotjahn, Rüdiger & Raatz, Ulrich (Hrsg.) (2002), *University Language Testing and the C-Test*. Bochum: AKS-Verlag, 93-114.
- Grotjahn, Rüdiger; Schlak, Torsten & Aguado, Karin (2010), S-C-Tests: Messung automatisierter sprachlicher Kompetenzen anhand von C-Tests mit massiver textspezifischer Zeitlimitierung. In: Grotjahn, Rüdiger (Hrsg.) (2010), *Der C-Test: Beiträge aus der aktuellen Forschung/The C-Test: Contributions from Current Research*. Frankfurt am Main: Lang, 297-319.
- Grotjahn, Rüdiger & Tesch, Bernd (2010), Die Messung der Leseverstehenskompetenz im Fach Französisch. In: Porsch, Raphaela; Tesch, Bernd & Köller, Olaf (Hrsg.) (2010), *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I*. Münster: Waxmann, 91-124.
- Guthrie, John T. (1988), Locating information in documents: Examination of a cognitive model. *Reading Research Quarterly* 23: 2, 178-199.
- Guthrie, John T. & Kirsch, Irwin S. (1987), Distinctions between reading comprehension and locating information in text. *Journal of Educational Psychology* 79: 3, 220-227.
- Guthrie, John T. & Mosenthal, Peter (1987), Literacy as multidimensional: Locating information and reading comprehension. *Educational Psychologist* 22: 3-4, 279-297.
- Hulstijn, Jan H. (2011a), Explanations of associations between L1 and L2 literacy skills. In: Schmid, Monika S. & Lowie, Wander (Hrsg.) (2011), *Modeling Bilingualism from Structure to Chaos: In Honor of Kees de Bot*. Amsterdam: Benjamins, 85-111.
- Hulstijn, Jan H. (2011b), Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly* 8: 3, 229-249.

- Jamieson, Joan M.; Eignor, Daniel; Grabe, William & Kunnan, Antony J. (2008), Frameworks for a new TOEFL. In: Chapelle, Carol; Enright, Mary K. & Jamieson, Joan M. (Eds.) (2008), *Building a Validity Argument for the Test of English as a Foreign LanguageTM*. New York: Routledge, 55-95.
- Jones, Neil & Maycock, Louise (2007), The comparability of computer-based and paper-based tests: Goals, approaches, and a review of research. *Research Notes* 27, 11-14 [Online: http://www.cambridgeesol.org/rs_notes/rs_nts27.pdf. 15.09.2012].
- Jung, Jookyoung (2009), Second language reading and the role of grammar. *Columbia University Working Papers in TESOL & Applied Linguistics* 9: 2, 29-48 [Online: <http://journals.tc-library.org/index.php/tesol/article/view/483/298>. 01.03.2013].
- Kalyuga, Slava (2013), Rapid dynamic assessment for learning. In: Mok, Magdalena M. C. (Hrsg.) (2013), *Self-directed Learning Oriented Assessments in the Asia-Pacific*. Dordrecht: Springer, 43-60.
- Khalifa, Hanan & Weir, Cyril J. (2009), *Examining Reading. Research and Practice in Assessing Second Language Reading*. Cambridge: Cambridge University Press.
- Kim, Do-Hong & Huynh, Huynh (2007), Comparability of computer and paper-and-pencil versions of Algebra and Biology assessments. *Journal of Technology, Learning, and Assessment* 6: 4. [Online: <http://napoleon.bc.edu/ojs/index.php/jtla/article/viewFile/1634/1478>. 15.09.2012].
- Kingston, Neal M. (2009), Comparability of computer- and paper-administered multiple-choice tests for K-12 populations: A synthesis. *Applied Measurement in Education* 22: 1, 22-37.
- KMK (2006), *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Unter Mitarbeit des Institut zur Qualitätsentwicklung im Bildungswesen. Bonn: KMK.
- KMK (Hg.) (2012), *Bildungsstandards für die fortgeführte Fremdsprache (Englisch / Französisch) für die Allgemeine Hochschulreife*. Beschluss der Kultusministerkonferenz vom 18.10.2012. [Online: <http://www.kmk.org/dokumentation/veroeffentlichungen-beschluesse/bildung-schule/qualitaetssicherung-in-schulen.html#c7035>. 07.11.2012].
- Leucht, Michael; Retelsdorf, Jan; Möller, Jens & Köller, Olaf (2010), Zur Dimensionalität rezeptiver englischsprachiger Kompetenzen. *Zeitschrift für Pädagogische Psychologie* 24: 2, 123-138.
- Lord, Frederic (1956), A study of speed factors in tests and academic grades. *Psychometrika* 21: 1, 31-50.
- Lutjeharms, Madeline (2007), Fremdsprachliches Leseverstehen. *Französisch heute* 38: 2, 106-120.
- McDonald, Angus S. (2002), The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers & Education* 39, 299-312.
- Mollenkopf, William G. (1960), Time limits and the behavior of test takers. *Educational and Psychological Measurement* 20: 2, 223-230.
- Musch, Jochen & Rösler, Peter (2011), Schnell-Lesen: Was ist die Grenze der menschlichen Lesegeschwindigkeit? In: Dresler, Martin (Hrsg.) (2011), *Kognitive Leistungen. Intelligenz und mentale Fähigkeiten im Spiegel der Neurowissenschaften*. Heidelberg: Spektrum, 89-106.
- Neuman, George & Baydoun, Ramzi (1998), Computerization of paper-and-pencil tests: When are they equivalent? *Applied Psychological Measurement* 22: 1, 71-83.

- Nix, Daniel (2011), *Förderung der Leseflüssigkeit: Theoretische Fundierung und empirische Überprüfung eines kooperativen Lautlese-Verfahrens im Deutschunterricht*. Weinheim: Juventa.
- Oxford, Rebecca L. (2011), Strategies for learning a second or foreign language. *Language Teaching* 44: 2, 167-180.
- Paek, Pamela (2005), *Recent Trends in Comparability Studies* (Pearson Educational Measurement Report, August 2005 [Online: <http://www.pearsonassessments.com/NR/rdonlyres/5FC04F5A-E79D-45FE-8484-07AACAE2DA75/0/TrendsCompStudies.pdf>. 15.09.2012].
- Pearson Education (2012), *PTE Academic Test Taker Handbook*. London: Pearson Education [Online: http://pearsonpte.com/TestMe/Documents/PTEA_Test_Taker_Handbook_EN.pdf. 15.09.2012].
- Perfetti, Charles A. (1994), Psycholinguistics and reading ability. In: Gernsbacher, Morton A. (Ed.) (1994), *Handbook of Psycholinguistics*. New York: Academic Press, 849-894.
- Phosphore 147 (1993), 71, Paris: Bayard Presse.
- Porsch, Raphaela; Grotjahn, Rüdiger & Tesch, Bernd (2010), Hörverstehen und Hörsehverstehen in der Fremdsprache – unterschiedliche Konstrukte? *Zeitschrift für Fremdsprachenforschung* 21: 2, 143-189.
- Raatz, Ulrich; Grotjahn, Rüdiger & Wockenfuß, Verena (2006), Das TESTATT-Projekt: Entwicklung von C-Tests zur Evaluation des Fremdsprachenlernerfolgs. In: Grotjahn, Rüdiger (Hrsg.) (2006), *Der C-Test: Theorie, Empirie, Anwendungen*. Frankfurt am Main: Lang, 85-99.
- Rasinski, Timothy V.; Reutzel, D. Ray; Chard, David & Linan-Thompson, Sylvia (2011), Reading fluency. In: Kamil, Michael L.; Pearson, P. David; Birr Moje, Elizabeth & Afflerbach, Peter (Eds.) (2011), *Handbook of Reading Research: Volume 4*. New York: Routledge, 286-319.
- Rindler, Susan E. (1979), Pitfalls in assessing test speededness. *Journal of Educational Measurement* 16: 4, 261-270.
- Sawaki, Yasuyo (2001), Comparability of conventional and computerized tests of reading in a second language. *Language Learning & Technology* 5: 2, 38-59.
- Scheuermann, Friedrich & Björnsson, Julius (2009), *The Transition to Computer-Based Assessment. New Approaches to Skills Assessment and Implications for Large-scale Testing* (JRC Scientific and Technical Report EUR 23679 EN). Ispra, Italien: Center for Research on Lifelong Learning (CRELL) [Online: <http://crell.jrc.ec.europa.eu/index.php/publications>. 15.09.2012].
- Schmitt, Neal (1996), Uses and abuses of coefficient alpha. *Psychological Assessment* 8: 4, 350-353.
- Schmölzer-Eibinger, Sabine (2008), *Lernen in der Zweitsprache. Grundlagen und Verfahren der Förderung von Textkompetenz in mehrsprachigen Klassen*. Tübingen: Narr.
- Schneider, Wolfgang; Schlagmüller, Matthias & Ennemoser, Marco (2007), *LGVT 6-12. Lesegeschwindigkeits- und -verständnistest für die Klassen 6-12*. Göttingen: Hogrefe.
- Schroeders, Ulrich & Wilhelm, Oliver (2011), Equivalence of reading and listening comprehension across test media. *Educational and Psychological Measurement* 71: 5, 849-869.
- Segalowitz, Norman (2010), *Cognitive Bases of Second Language Fluency*. London: Routledge.

- Shaw, Philip & McMillion, Alan (2008), Proficiency effects and compensation in advanced second-language reading. *Nordic Journal of English Studies* 7: 3, 123-143 [Online: <http://ojs.uu.se/ojs/index.php/njes/article/viewFile/155/153>. 15.09.2012].
- Sheppard, Leah D. & Vernon, Philip A. (2008), Intelligence and speed of information processing: A review of 50 years of research. *Personality and Individual Differences* 44: 3, 535-551.
- Shiotsu, Toshihiko (2010), *Components of L2 Reading: Linguistic and Processing Factors in the Reading Test Performances of Japanese EFL Learners*. Cambridge: Cambridge University Press.
- Singleton, Chris; Thomas, Kevin & Horne, Joanna (2000), Computer-based cognitive assessment and the development of reading. *Journal of Research in Reading* 23: 2, 158-180.
- Souvignier, Elmar & Mokhesgerami, Judith (2005), Implementation eines Programms zur Vermittlung von Lesestrategien im Deutschunterricht. Die Rolle der Lehrenden. *Zeitschrift für Pädagogische Psychologie* 19: 4, 249-261.
- Spache, George D. (1962), Is this a breakthrough in reading? *The Reading Teacher* 15:4, 258-262.
- Stanat, Petra & Schneider, Wolfgang (2004), Schwache Leser unter 15-jährigen Schülerinnen und Schülern in Deutschland: Beschreibung einer Risikogruppe. In: Schiefele, Ulrich; Artelt, Cordula; Schneider, Wolfgang & Stanat, Petra (Hrsg.) (2004), *Struktur, Entwicklung und Förderung von Lesekompetenz: Vertiefende Analysen im Rahmen von PISA-2000*. Wiesbaden: Verlag für Sozialwissenschaften, 243-273.
- Sternberg, Robert J. & Grigorenko, Elena L. (2002), *Dynamic Testing: The Nature and Measurement of Learning Potential*. Cambridge: Cambridge University Press.
- Taylor, Carol; Jamieson, Joan; Eignor, Daniel & Kirsch, Irwin (1998), *The Relationship between Computer Familiarity and Performance on Computer-based TOEFL Test Tasks* (TOEFL Research Report No. 61). Princeton, NJ: Educational Testing Service [Online: <http://www.ets.org/Media/Research/pdf/RR-98-08.pdf>. 15.09.2012].
- Van Gelderen, Amos; Schoonen, Rob; Stoel, Reinoud D.; De Glopper, Kees & Hulstijn, Jan (2007), Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent processes. *Journal of Educational Psychology* 99: 3, 477-491.
- Walczyk, Jeffrey J.; Kelly, Kathryn E.; Meche, Scott D. & Braud, Hillary (1999), Time limitations enhance reading comprehension. *Contemporary Educational Psychology* 24: 2, 156-165.
- Walczyk, Jeffrey J.; Marsiglia, Cheryl S.; Johns, Amanda K. & Bryan, Keli S. (2004), Children's compensations for poorly automated reading skills. *Discourse Processes* 37: 1, 47-66.
- Way, Walter D.; Davis, Lauren L. & Fitzpatrick, Steven (2006), Score comparability of online and paper administrations of Texas assessment of knowledge and skills, Paper presented at the annual meeting of the National Council on Measurement in Education. San Francisco, CA [Online: https://www.pearsonassessments.com/hai/images/tmrs/Score_Comparability_of_Online_and_Paper_Administrations_of_TAKS_03_26_06_final.pdf. 15.09.2012].
- Weir, Cyril J. & Khalifa, Hanan (2008), A cognitive processing approach towards defining reading comprehension. *Cambridge ESOL: Research Notes* 31, 11-16 [Online: http://www.cambridgeesol.org/rs_notes/rs_nts31.pdf. 15.09.2012].

- Weir, Cyril J.; Yang, Huizhong & Jin, Yan (2000), *An Empirical Investigation of the Componentiality of L2 Reading in English for Academic Purposes*. Cambridge: Cambridge University Press.
- Wesman, Alexander G. (1949), Effect of speed on item-test correlation coefficients. *Educational and Psychological Measurement* 9: 1, 51-57.
- Wesman, Alexander G. (1960), Some effects of speed in test use. *Educational and Psychological Measurement* 20: 2, 267-274.
- Wilhelm, Oliver & Schulze, Ralf (2002), The relation of speeded and unspeeded reasoning with mental speed. *Intelligence: A multidisciplinary journal* 30: 6, 537-554.
- Wood, Evelyn N. (1960), A breakthrough in reading. *The Reading Teacher* 14: 2, 115-117.