

## Originalartikel

# Kognitive Verarbeitung von Leseverständnisitems mit und ohne Text

Sascha Schroeder und Simon P. Tiffin-Richards

Max-Planck-Institut für Bildungsforschung, Berlin

**Zusammenfassung:** Multiple-Choice (MC)-Items sind ein häufig eingesetztes Aufgabenformat, um Leseverständnis zu erfassen. Die vorliegende Studie mit 366 Schülerinnen und Schülern repliziert die Befunde von Rost und Sparfeldt (2007), die für das Deutsche gezeigt haben, dass MC-Items allerdings auch ohne Lektüre der zugehörigen Texte beantwortet werden können. Darüber hinaus zeigen wir, dass die Bearbeitungsqualität von MC-Leseverständnisitems mit Textdarbietung sowohl durch die verbale Problemlösefähigkeiten als auch durch lesespezifische Verstehensprozesse beeinflusst wird. Dies spricht dafür, dass MC-Items nicht alleine verbale Intelligenz erfassen, sondern auch die Effizienz lesespezifischer Prozesse. Darüber hinaus weisen die Ergebnisse darauf hin, dass sehr ähnliche Antwortprozesse auch zur Beantwortung von MC-Items ohne Textdarbietung eingesetzt werden. Diese werden aber zunehmend unreliabler erfasst, je mehr Textinformationen eliminiert werden.

**Schlüsselwörter:** Leseverständnis, Konstruktvalidität, Textverarbeitung

### Cognitive Processing of Reading Comprehension Items with and without Text

**Abstract:** Multiple choice (MC)-items are a frequently employed format used to measure reading comprehension. In a study with 366 students we replicate the finding of Rost and Sparfeldt (2007) who found that MC-items can be solved without reading the relevant text in German. In addition, we show that responses to MC reading comprehension items with available text are influenced both by general verbal resources, as well as processes required to generate text coherence. This suggests that MC-items do also measure reading specific comprehension processes. Our results also indicate that similar processes are used to answer MC-items when the relevant text is not available. The measurement of these processes, however, becomes increasingly unreliable when text information is reduced.

**Keywords:** reading comprehension, construct validity, reading processes

## 1 Theoretischer Hintergrund

Leseverständnis valide zu erfassen, ist eine schwierige Aufgabe. Lesen ist eine komplexe kognitive Leistung, die verschiedene Teilprozesse umfasst (Cain, Oakhill & Bryant, 2004). Um die Inhalte eines Textes verstehen zu können, müssen u. a. die einzelnen Wörter dekodiert, syntaktisch korrekt verbunden und semantisch interpretiert werden (Kintsch, 1998). Zentral für das Textverstehen ist die Konstruktion einer bedeutungsbasierten Repräsentation des Textes, des so genannten Situationsmodells (Richter & Christmann, 2002). Hierfür muss ein Text durch zusätzliche Inferenzen und Vorwissen angereichert und zu einer kohärenten Repräsentation verdichtet werden, die – je nach Qualität der Ausführung der einzelnen Teilprozesse – mehr oder weniger elaboriert sein kann (Hannon & Daneman, 2001).

Zur Prüfung der Qualität des Verständnisprodukts werden unterschiedliche Test- und Itemformate verwendet (Pearson & Hamm, 2005), z. B. offene und geschlossene Antwortformate. Ein nach wie vor häufiges und ökonomisches Testformat sind multiple-choice (MC) Leseverständnistests. Hier lesen Personen einen kurzen Text und bearbeiten anschließend eine Reihe von Items. Jedes Item besteht aus einer Frage (z. B. «Warum geht Peter heute nicht in die Schule?»), für die verschiedene Antworten zur Auswahl stehen (z. B. «Weil er krank ist» etc.). Das Antwortverhalten wird dabei von Personen- (Wortschatz etc.), Text- (Sach- vs. Erzähltext etc.) und Itemmerkmalen (Oberflächen- vs. Tiefenitems etc.) beeinflusst (Magliano, Millis, Ozuru & McNamara, 2007).

Ebenfalls ist bekannt, dass MC-Items auch dann gelöst werden können, wenn die zugehörigen Texte gar nicht gelesen wurden. Im englischsprachigen Raum wurde auf

dieses Phänomen bereits in den 1960er Jahren hingewiesen (z. B. Preston, 1964). Die bekannteste Demonstration stammt aber von Katz, Lautenschlager, Blackburn und Harris (1990). In dieser Untersuchung wurden Studierenden MC-Items aus dem *Scholastic Aptitude Test* (SAT) vorgegeben. Eine Gruppe von Studierenden bearbeitete die Items zusammen mit den jeweiligen Texten. Einer zweiten Gruppe wurden die Texte nicht vorgelegt, sondern nur die Items. Die Ergebnisse zeigen, dass MC-Leseverständnisitems auch ohne Textdarbietung überzufällig häufig gelöst werden konnten, auch wenn die Lösungswahrscheinlichkeit nicht so hoch war wie in der mit-Text-Bedingung. Die Testleistung in den beiden Bedingungen korrelierte stark miteinander ( $r = .5$ ), d. h. wer MC-Items mit Text gut lösen kann, kann dies tendenziell auch ohne Text.

Rost und Sparfeldt (2007) haben nachgewiesen, dass dies auch für den deutschen Sprachraum gilt und das Phänomen systematisch aufgearbeitet. In drei Teilstudien bearbeiteten Schülerinnen und Schüler MC-Items in vier Bedingungen: a) in der Originalversion, b) in einer Fassung, in der der Text eliminiert wurde, c) in einer Fassung, in der nur die Fragen eliminiert wurden und d) in einer Fassung, in der sowohl Text als auch Fragen beseitigt wurden. Die Items konnten in allen Bedingungen überzufällig häufig gelöst werden, wobei der Effekt mit der Wichtigkeit der eliminierten Informationsquellen abnahm. Auch hier korrelierte die Testleistung in der informationsreduzierten Version mit der Originalversion ( $r = .2$  bis  $.6$ ) und zwar in Abhängigkeit davon, wie viele Informationen eliminiert wurden. Das gleiche Muster ergab sich für die verbale Intelligenz.

Der Befund, dass MC-Items auch ohne Text gelöst werden können und dass diese Fähigkeit stark mit der verbalen Intelligenz korreliert, ist inzwischen häufig repliziert worden (u. a. Coleman, Lindstrom, Nelson, Lindstrom & Gregg, 2011; Sparfeldt, Kimmel, Löwenkamp, Steingraber & Rost, 2012). Seine Interpretation ist allerdings weiterhin umstritten. In der Debatte lassen sich drei Positionen unterscheiden:

Die erste Position geht davon aus, dass dieses Ergebnismuster die Konstruktvalidität von MC-Items in Frage stellt (z. B. Katz et al., 1990). Es wird argumentiert, dass MC-Items ohne Text nicht durch lesespezifische Verstehensprozesse gelöst werden, sondern durch leseunspezifische, allgemeine Antwortprozesse, z. B. durch Rückgriff auf verfügbares Weltwissen. Da darüber hinaus die Testleistung stark mit der verbalen Intelligenz korreliert, wird davon ausgegangen, dass solche leseunspezifischen Antwortprozesse auch bei der Bearbeitung von MC-Items mit Textdarbietung wirksam sind, diese also ebenfalls nicht das Leseverstehen erfassen. MC-Items mit und ohne Textpräsentation erheben also das *gleiche* Konstrukt, das in beiden Bedingungen *leseunspezifisch* ist.

Eine zweite Position vertritt die Ansicht, dass die berichteten Befunde die Konstruktvalidität von MC-Items nicht in Frage stellen (z. B. Daneman & Hannon, 2001), da MC-Items mit und ohne Text verschiedene Antwort-

prozesse erfassen: Antworten mit Text werden durch lesespezifische Prozesse generiert, ohne Text jedoch durch allgemeine, leseunspezifische Antwortprozesse. Die Korrelation zwischen den beiden Bedingungen wird dadurch erklärt, dass beide Prozesstypen auf den gleichen Ressourcenpool zurückgreifen. Z. B. werden verbale Intelligenz und Vorwissen sowohl beim Lesen eines Textes als auch beim Raten von Antworten benötigt. Dieser Position zufolge werden also in den beiden Darbietungsbedingungen *unterschiedliche* Prozesse erfasst, wobei diese in der mit-Text-Bedingung *lesespezifisch*, in der ohne-Text-Bedingung jedoch *leseunspezifisch* sind.

Die dritte Position (z. B. Freedle & Kostin, 1994) geht davon aus, dass MC-Items mit und ohne Textdarbietung lesespezifische Prozesse erfassen. Es wird argumentiert, dass für die Bearbeitung von MC-Items ohne Text kognitive Fähigkeiten notwendig sind, die auch für das Lesen relevant sind, z. B. müssen immer noch die Fragen und Antwortalternativen dekodiert und sinnstiftende Verbindungen hergestellt werden. Dies alles sind Prozesse, die auch bei der Rezeption eines Textes ausgeführt werden. Diese funktionale Überschneidung ist hinreichend, um die Korrelation in den beiden Testbedingungen zu erklären. MC-Items mit und ohne Textpräsentation erheben also das *gleiche* Konstrukt, das in beiden Bedingungen *lesespezifisch* ist. Allerdings wird es durch die Elimination von Textinformationen zunehmend fehlerhafter erfasst, was sich in einer geringeren Reliabilität der Testscores niederschlagen sollte.

Wie herausgearbeitet, ergeben sich die drei Positionen aus der unterschiedlichen Beantwortung von zwei prinzipiell verschiedenen Teilfragen: 1. Sind diese Prozesse für das Lesen spezifisch oder beruhen sie auf allgemeinen, leseunspezifischen Teilfähigkeiten? und 2. Erfassen MC-Leseitems mit und ohne Textdarbietung ähnliche oder verschiedene kognitive Prozesse und Leistungen?

Das zentrale Problem der Debatte ist, dass sich diese Fragen mit der bislang vorliegenden Datenbasis empirisch nicht beantworten lassen. Alle Positionen beziehen sich in ihrer Argumentation darauf, welche kognitiven Prozesse in der Rezeptionsphase ausgeführt werden. Allerdings wurden bislang aus kognitionspsychologischer Perspektive nur vergleichsweise distale Produktvariablen untersucht, wie z. B. die verbale Intelligenz oder die verbale Arbeitsgedächtniskapazität. Solche Variablen sind jedoch zu unscharf, um zwischen den verschiedenen Prozesshypothesen zu unterscheiden. Da u. E. die Diskussion ohne darüber hinausgehende Daten zur Struktur des Antwortprozesses stagnieren wird, ist ein zentrales Anliegen dieser Studie, weitere Verarbeitungsmaße zu erheben, die verschiedene Teilprozesse des Lesens direkt(er) erfassen.

Hierfür sollten die Ergebnisse der für den deutschen Sprachraum zentralen Studie von Rost und Sparfeldt (2007) zunächst repliziert werden. Schülerinnen und Schüler bearbeiteten denselben Leseverständnis-Test und zwar sowohl in der Originalversion als auch in einer von zwei informationsreduzierten Bedingungen: a) ohne Text

und b) ohne Text und ohne Fragen. Die verbale Intelligenz wurde mit einem Wortschatztest und einem verbalen Analogietest erhoben. Darüber hinaus wurden auch zwei lesespezifische Variablen erfasst, die verschiedene Prozesse abbilden, die für die Generierung eines elaborierten Situationsmodells relevant sind (Richter & Christmann, 2002).

Dadurch sollen sich die beiden Teilfragen besser beantworten lassen:

1. Wenn die erfassten Prozesse spezifisch für das Textverstehen sind, sollten die lesespezifischen Prozessvariablen die Testleistung substanziell beeinflussen. Wenn hingegen die zugrunde liegenden Antwortprozesse leseunspezifisch sind, dann sollten die allgemeinen kognitiven Hintergrundvariablen ausschließlich die Testleistung bestimmen.
2. Wenn die Testleistung mit und ohne Textpräsentation auf ähnlichen Antwortprozessen basiert, sollte sich in den beiden Bedingungen ein qualitativ ähnliches Muster von Außenkorrelationen ergeben. Wenn sie sich jedoch unterscheiden, sollten die Korrelationsmuster grundsätzlich verschieden sein.

Dabei ist selbstverständlich möglich, dass die Beantwortung dieser beiden Fragen für die beiden informationsreduzierten Bedingungen unterschiedlich ausfällt. So könnte es z. B. durchaus sein, dass eine Testfassung, in der nur die Texte eliminiert wurden, noch eine Bearbeitung erlaubt, die mit der normalen Testbearbeitung vergleichbar ist. Wenn hingegen Text und Fragen gleichzeitig eliminiert werden, könnte es sein, dass die Items grundsätzlich anders beantwortet werden müssen.

## 2 Methode

### 2.1 Stichprobe

Das Experiment war Teil der Studie CLAss (Cognitive Language Assessment), die zwei computergestützte Erhebungstermine umfasste. Insgesamt nahmen 463 Jugendliche (54 % weiblich) aus 10 Klassen an 3 Berliner Gymnasien ( $n = 295$ ) und 6 Klassen an 2 Berliner Gesamtschulen ( $n = 168$ ) an der Studie teil. Die Stichprobe umfasst die Jahrgangsstufen 8 ( $n = 78$ ) und 9 ( $n = 385$ ; Alter:  $M = 14.34$ ,  $SD = .81$ ). Von den Schülerinnen und Schülern hatten 28 % einen Migrationshintergrund (erste oder zweite Generation), was ungefähr dem Berliner Durchschnitt entspricht. Die Beziehungen zwischen Variablen unterschieden sich nicht signifikant zwischen den Jugendlichen mit und ohne Migrationshintergrund (alle  $z < 1$ ). 54 Schülerinnen und Schüler nahmen nur an der ersten und 27 nur an der zweiten Testsitzung teil. Für 12 Schülerinnen und Schüler liegen aus technischen Gründen keine Daten vor. Die Fluktuation in der Teilnahme wurde durch Krankheit oder andere zufällige Ereignisse verursacht und unterschied sich nicht zwischen den Experimentalbedingungen,  $\chi^2_{N=451}(1) = 0.25$ ,  $p = .62$ . Vier Jugendliche wurden wegen

fehlender Werte ausgeschlossen. Dadurch ergab sich eine effektive Stichprobe von  $N = 366$ .

### 2.2 Leseverständnis

Für die Untersuchung der Leseverständnisleistung wurden Aufgaben aus dem Leseverständnisstest «Verständiges Lesen 7–9» (Anger, Bargmann & Voigt, 1972) verwendet, der aus zwei Parallelformen mit jeweils sechs Texten besteht, wobei auf jeden Text 7 bis 10 MC-Items mit jeweils 4 Antwortalternativen folgen. Aus den Parallelformen A und B wurden jeweils vier Texte mit 35 MC-Items ausgewählt. Die Texte der Parallelform A wurden in ihrer Originalform belassen (Testbedingung «mit Text») und von allen Schülerinnen und Schülern bearbeitet. In den Texten der Parallelform B wurden hingegen, wie bei Rost und Sparfeldt (2007), verschiedene Elemente eliminiert: Eine Hälfte der Stichprobe (Gruppe A,  $n = 191$ ) bearbeitete diese Texte in der Bedingung «ohne Text», in der nur die Texte entfernt wurden. Die andere Hälfte (Gruppe B,  $n = 175$ ) bearbeitete dieselben Texte in der Bedingung «ohne Text und ohne Fragen», in der zusätzlich noch die Verständnisfragen entfernt wurden. Die Zuteilung auf die Gruppen erfolgte randomisiert innerhalb jeder Klasse durch paarweises Abzählen.

### 2.3 Verbale Intelligenz

Um festzustellen, in welchem Maß allgemeine verbale Fähigkeiten zur Testleistung in den verschiedenen Bedingungen beitragen, wurde die verbale Intelligenz der Schülerinnen und Schüler erhoben. Hierzu gehörten die verbale Problemlösefähigkeit und der Wortschatz. Das Verhältnis zwischen verbaler Intelligenz und Leseleistung ist komplex und reziprok: Einerseits begünstigt die verbale Intelligenz das Leseverstehen, andererseits führen effizientere Leseprozesse zu einer höheren verbalen Intelligenz. Entsprechend wurde empirisch nachgewiesen, dass verbale Intelligenz und das Leseverständnis hoch miteinander korrelieren (ca.  $r = .5$  bis  $.6$ ).

#### 2.3.1 Verbale Problemlösefähigkeit

Die verbale Problemlösefähigkeit ist eine grundlegende kognitive Ressource, die für unterschiedliche Teilprozesse des Lesens wichtig ist. In Anlehnung an Rost und Sparfeldt (2007) wurde sie als «verbal reasoning»,  $g_v$ , operationalisiert und mit dem Subtest «Wortanalogien» aus dem KFT 4–12+R erhoben (Heller & Perleth, 2000). Alle Schülerinnen und Schüler bearbeiteten dabei die Items der 8. Klasse. Deskriptive Kennwerte und interne Konsistenz der Skala finden sich in Tabelle 1. Zusätzlich wurden für diesen Subtest auch die Normwerte (Perzentile) berechnet, wobei sich ein Mittelwert von  $M = 42.87$  ( $SD = 10.15$ )

Tabelle 1

Interkorrelation und deskriptive Statistik der manifesten Indikatorvariablen

		Variable					
		M	SD	$\alpha$	1	2	3
1	Verbales Problemlösen	9.17	3.97	.76	-		
2	Wortschatz	24.45	3.31	.71	.42	-	
3	Lokale Kohärenzbildung	22.89	6.62	.67	.40	.33	-
4	Globale Kohärenzbildung	24.31	4.82	.79	.37	.21	.43

Anmerkung: Alle Korrelationen sind statistisch signifikant größer als 0.

ergab. Da sich keine wesentlichen Unterschiede ergaben, wurden für alle Analysen die Rohwerte verwendet.

### 2.3.2 Wortschatz

Umfang und Qualität des Wortschatzes sind als prototypische Indikatoren der kristallinen Intelligenz eine wichtige Einflussgröße des Leseverständnisses (Cain et al., 2004). Wenn Leserinnen und Leser die Wörter in einem Text nicht kennen oder mit der falschen Bedeutung verbinden, können sie ihn nicht angemessen verstehen. Gleichzeitig ist es ein Indikator des allgemeinen Weltwissens, das häufig als eine Möglichkeit diskutiert wird, weshalb Personen MC-Items auch ohne Text lösen können (z.B. Katz et al., 1990).

Der Umfang des Wortschatzes der Schülerinnen und Schüler wurde mit der Paralleltestform A des Wortschatztests des CFT 20-R (Weiß, 1998) erhoben. Der vergleichsweise hohe Mittelwert (s. Tab. 1) und die etwas geminderte Reliabilität der Skala weisen auf einen leichten Deckeneffekt hin. Dennoch waren die interindividuellen Unterschiede substanziell, und nur in 1.5 % aller Fälle wurde die höchste Punktzahl erreicht.

## 2.4 Lesespezifische Prozessvariablen

Komplementär zur verbalen Intelligenz wurden zwei weitere Variablen erhoben, welche die Effizienz lesespezifischer Teilprozesse direkter erfassen. Dabei wurden solche Prozesse der lokalen und globalen Inferenzgenerierung ausgewählt, die für die Generierung eines kohärenten Situationsmodells zentral sind (Kintsch, 1998).

### 2.4.1 Lokale Kohärenzbildung: Anaphorische Brückeninferenzen

Bei der lokalen Kohärenzbildung werden die Propositionen eines Textes zu größeren Makropropositionen verbunden. Um diese Verbindungen herzustellen, sind häufig Brückeninferenzen notwendig, bei denen Weltwissen aktiviert wird (Richter & Christmann, 2002). Das ist z. B. bei der Auflösung von Pronomen notwendig (z. B. «Paul war

gerne im Freien. Er spazierte jeden Samstag im Park»), bei der Pronomen ein Referent zugewiesen wird. Um zu entscheiden, ob z. B. «Paul» im vorliegenden Fall als Referent plausibel ist, müssen die Sätze miteinander verbunden und auf ihre Konsistenz mit vorliegendem Weltwissen geprüft werden.

Die Effizienz der satzinternen lokalen Kohärenzbildung wurde mit einer pronominalen Auflösungsaufgabe erfasst (vgl. Daneman & Carpenter, 1980). Es wurden 32 Nebensatzkonstruktionen präsentiert (z. B. «Peter half Frank beim Aufräumen, weil er heute krank war.»), in denen sich das Pronomen entweder auf das Subjekt oder das Objekt des vorhergehenden Hauptsatzes beziehen konnte. Die Schülerinnen und Schüler sollten so schnell wie möglich angeben, wer mit dem Pronomen gemeint ist. Die Korrektheit der Antwort wurde dabei durch einen Vergleich der Plausibilität der zugrunde liegenden Brückeninferenz festgelegt (die in separaten Voruntersuchungen ermittelt wurde). Deskriptive Kennwerte finden sich in Tabelle 1.

### 2.4.2 Globale Kohärenzbildung: Wissensaktivierung und -integration

Bei der globalen Kohärenzbildung werden Propositionen über Satzgrenzen miteinander verbunden. Hierfür müssen zuvor gelesene Textinformationen reaktiviert und mit anderen relevanten Wissensbeständen im Langzeitspeicher verbunden werden (Richter & Christmann, 2002). Hannon und Daneman (2001) haben eine Aufgabe entwickelt, welche basale Prozesse erfasst, die für die Aktivierung und Integration von Textinformationen benötigt werden. Diese Aufgabe wurde inzwischen für verschiedene Altersbereiche weiterentwickelt, und es wurde gezeigt, dass sie spezifische Effekte auf die Leseleistung hat, die über die der verbalen Intelligenz und der Arbeitsgedächtniskapazität hinausgehen (Daneman & Hannon, 2001).

Die eingesetzte deutsche Version dieser Aufgaben setzt sich aus zwei Textpassagen zusammen, zu der jeweils 16 Aussagen gehören. Die Textpassagen beinhalteten Informationen über das Verhältnis zwischen realen und nicht-realen Gegenständen und Wesen (z. B. «Ein WEMP ist größer als eine GIRAFFE./Ein TILL ist kleiner als ein AFFE./Ein LORK ist kleiner als ein TILL.»). Die dazu-

Tabelle 2

Mittelwerte, Standardabweichungen und Korrelationen der Subtests der Globalen Kohärenzaufgabe

		Subtests				
		<i>M</i>	<i>SD</i>	1	2	3
1	Textgedächtnis	5.63	1.80	-		
2	Textinferenzen	5.49	1.70	.53	-	
3	Wissensintegration	5.94	1.68	.56	.45	-
4	Wissenszugriff	7.15	1.25	.23	.27	.31

Anmerkung: Alle Korrelationen sind statistisch signifikant größer als 0.

gehörigen Aussagen thematisieren die Verhältnisse zwischen den in der Textpassage beschriebenen Gegenständen oder Wesen sowie nicht im Text vorkommenden realen Gegenständen und Tieren. Verschiedene Items erfassen dabei vier verschiedene Aspekte des Verstehensprozesses: Textgedächtnis (z. B. «Ein WEMP ist größer als eine GIRAFFE» [richtig]), Textinferenzen (z. B. «Ein LORK ist größer als ein TILL» [falsch]), Wissensintegration (z. B. «Eine GIRAFFE ist größer als ein TILL» [richtig]) und Wissenszugriff (z. B. «Eine GIRAFFE ist größer als ein AFFE» [richtig]). Die vier Subskalen korrelierten hinreichend miteinander und wurden zu einer Gesamtskala zusammengefasst (vgl. Tab. 2).

## 2.5 Durchführung

Die Untersuchung wurde in zwei Sitzungen durchgeführt, die ungefähr eine Woche auseinander lagen. In der ersten Sitzung wurden zuerst Aufgaben zum Wortschatz, dann zur Wissensaktivierung und -integration, zu den anaphorischen Brückeninferenzen und schließlich Leseverständnisaufgaben bearbeitet. In der zweiten Sitzung wurden zuerst die verbalen Intelligenzaufgaben und dann die Leseverständnisaufgaben durchgeführt. In jeder Sitzung wurden dabei zwei Texte in der «mit Text»-Bedingung und zwei Texte der informationsreduzierten Bedingung bearbeitet, wobei immer zunächst die vollständigen und dann die informationsreduzierten Fassungen durchgeführt wurden. Dadurch kann die Retest-Reliabilität in den drei Testbedingungen bestimmt werden.

Alle Testverfahren wurden an Laptops durchgeführt. In der «mit Text»-Bedingung wurde zunächst der Text auf einem eigenen Bildschirm angezeigt und konnte ohne Zeitbeschränkung gelesen werden. Nach einem Tastendruck folgten dann nacheinander die einzelnen MC-Leseverständnisisems, die jeweils unter dem Text eingeblendet wurden. Der Text war also während der Beantwortung verfügbar. Nachdem die Schülerinnen und Schüler ihre Antwort abgegeben hatten, wurde das nächste Item eingeblendet etc. Das Vorgehen in der «ohne Text»-Bedingung war prinzipiell identisch, nur dass der Text vor der Item-Beantwortung nicht gelesen wurde und auch in der Antwortphase nicht verfügbar war. In der «ohne Text und ohne Fragen»-Bedingung war das Vorgehen identisch, nur

wurden zusätzlich bei den eingeblendeten Items noch die Fragen eliminiert.

## 2.6 Auswertung

Die Analysen des folgenden Ergebnisteils sind in drei Schritte gegliedert. Zunächst sollen die Ergebnisse von Rost und Sparfeldt (2007) repliziert werden. Hierfür werden die Mittelwerte in den Untersuchungsbedingungen mit einer zweifaktoriellen Varianzanalyse mit Messwiederholung verglichen. Aufgrund des Designs unserer Studie kann die «mit Text»-Bedingung mit den beiden informationsreduzierten Fassungen jeweils messwiederholt verglichen werden. Der Vergleich der «ohne Text»- und «ohne Text und ohne Fragen»-Bedingung basiert hingegen auf einem Zwischengruppenvergleich. Gleichzeitig wird in jeder Untersuchungsbedingung mit einem Einstichproben-*t*-Test überprüft, ob die mittlere Lösungswahrscheinlichkeit signifikant über dem Zufallsniveau (von 25 %) liegt. Wir erwarten dabei ein ähnliches Effektmuster wie bei Rost und Sparfeldt (2007), d. h. die Lösungswahrscheinlichkeit sollte mit der Anzahl der eliminierten Informationen abnehmen, aber in allen Bedingungen sollten die Aufgaben überzufällig häufig gelöst werden können.

Im zweiten Schritt überprüfen wir die Faktorstruktur der verwendeten Außenkriterien mit einer konfirmatorischen Faktoranalyse. Wir erwarten dabei, dass sich Intelligenz- und Prozessvariablen voneinander trennen lassen, aber auch, dass es substantielle Überschneidungen zwischen ihnen gibt, die sich in einer substantiellen Korrelation der beiden latenten Faktoren niederschlagen sollte.

Im letzten Schritt werden die Zusammenhänge zwischen den Indikatorvariablen und der Testleistung untersucht. Hierfür werden zunächst die korrelativen Beziehungen zwischen den beiden Außenfaktoren und den Testscores in den verschiedenen Bedingungen berichtet. Sollten lesespezifische Prozessvariablen die Leseleistung über die Effekte der verbalen Intelligenz hinaus beeinflussen, deutet dies darauf hin, dass die Testleistung auch die prozessspezifische Varianz erfasst, die für das Lesen spezifisch ist. Wir erwarten dabei ein qualitativ ähnliches Muster von Korrelationen in allen Bedingungen, was für die Homogenität der erfassten Prozesse spricht. Abschlie-

ßend sollte die Hypothese identischer Korrelationen explizit überprüft werden, indem auch die korrespondierenden, minderungskorrigierten Korrelationen berechnet wurden. Für die Korrektur wurden die internen Homogenitätswerte und die Reliabilität der latenten Faktoren ( $\Omega$ ) verwendet. Die inferenzstatistische Prüfung erfolgte in beiden Fällen mittels  $z$ -Tests zur Gleichheit von Korrelationen in ungepaarten Stichproben (Cohen, Cohen, West & Aiken, 2003, Kap. 2).

Die Logik, die diesem letzten Analyseschritt zugrunde liegt, ist dabei folgende: Aus theoretischen Gründen ist zu erwarten, dass der Anteil an Fehlervarianz mit dem Ausmaß der Informationsreduktion ansteigt. Dennoch ist die Frage interessant, welche Prozesse von der kriterialen Restvarianz in den verschiedenen Bedingungen noch erfasst werden. Minderungskorrigierte Korrelationen sind eine Form, kontrafaktisch zu prüfen, welche Beziehungen zwischen der Testleistung und Außenkriterien bestehen würden, wenn diese fehlerfrei erfasst worden wären. Eine latente Modellierung der Variablen im Rahmen von Strukturgleichungsmodellen wäre eine alternative Möglichkeit, dieses Beziehungsmuster zu überprüfen. Dies erwies sich aber aus praktischen Gründen als unmöglich, da die entsprechenden Modelle in der am stärksten informationsreduzierten Bedingung (ohne Text/ohne Fragen) wegen des hohen Fehleranteils nicht konvergierten. Minderungskorrigierte Korrelationen sind hier berechnungstechnisch weniger anspruchsvoll und erlauben, die gleiche Frage in den verschiedenen Bedingungen in einem einheitlichen methodischen Rahmen zu prüfen.

Für alle Analysen wurde ein Signifikanzniveau von 5 % (zweiseitig) festgelegt. Da jedoch aufgrund der Stichprobengröße die Signifikanz eines Effekts nur bedingt aussagekräftig ist, werden für alle Analysen Effektstärkemaße berichtet. Trotz der genesteten Struktur der Daten (Schülerinnen und Schüler in Klassen), wurde wegen der geringen Anzahl von Klassen von einer expliziten Modellierung der Mehrebenenstruktur abgesehen.

## 3 Ergebnisse und Diskussion

### 3.1 Replikation der Befunde aus Rost und Sparfeldt (2007)

In einem ersten Schritt wurde überprüft, ob wir mit unseren Daten das Befundmuster von Rost und Sparfeldt (2007) replizieren können. Mittelwerte und Standardabweichungen der Leseleistung in den drei Bedingungen werden in Tabelle 3 berichtet. Eine zweifaktorielle Varianzanalyse mit dem messwiederholten Faktor Textbedingung (Originalversion vs. informationsreduzierte Version) und dem Zwischenpersonen-Faktor Testfassungen (ohne Text vs. ohne Text und ohne Fragen) ergab einen signifikanten Haupteffekt von Textbedingung,  $F(1, 364) = 378.93$ ,  $p < .01$ ,  $\eta^2_G = .250$ : Schülerinnen und Schüler konnten in der «mit Text»-Bedingung mehr MC-Items lösen als in den

informationsreduzierten Bedingungen. Der signifikante Haupteffekt des Faktors Testfassungen,  $F(1, 364) = 7.63$ ,  $p < .01$ ,  $\eta^2_G = .014$ , zusammen mit dem signifikanten Interaktionseffekt zwischen Textbedingung und Testfassungen,  $F(1, 364) = 16.94$ ,  $p < .01$ ,  $\eta^2_G = .02$ , weist zudem darauf hin, dass mehr MC-Items gelöst werden konnten, wenn nur der Text eliminiert wurde.

Die mittlere Leseverständnisleistung der Schülerinnen und Schüler unter den Testbedingungen «mit Text» in der Gruppe A ( $M = 16.80$ ,  $SD = 6.52$ ) und Gruppe B ( $M = 16.83$ ,  $SD = 6.89$ ) waren nicht statistisch signifikant voneinander verschieden,  $t(363) = 0.04$ ,  $p = .96$ ,  $d = 0.01$ , während die mittlere Leseverständnisleistung der Gruppe A unter der Textbedingung «ohne Text» ( $M = 11.97$ ,  $SD = 3.60$ ) und der Gruppe B unter der Textbedingung «ohne Text und ohne Fragen» ( $M = 9.44$ ,  $SD = 2.81$ ) sich statistisch signifikant voneinander unterschieden,  $t(363) = 7.54$ ,  $p < .01$ ,  $d = 0.79$ . Daraus ist zu schließen, dass die Leseverständnisleistung bei zunehmender Textinformationsreduktion abnimmt.

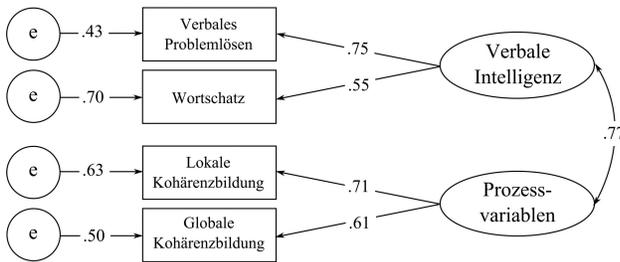
Auch die Korrelationen zwischen der Testleistung in den verschiedenen Bedingungen waren vergleichbar mit der vorherigen Studie (vgl.  $r = .63$  bzw.  $r = .24$  im Vergleich zu  $r = .60$  bzw.  $r = .21$  bei Rost & Sparfeldt, 2007).

Die Ergebnisse replizierten Rost und Sparfeldt (2007) auch dahingehend, dass die erreichte Punktzahl sich der Ratewahrscheinlichkeit annäherte, je mehr Informationsquellen eliminiert wurden. Dennoch blieb die Antwortwahrscheinlichkeit unter der Textbedingung «ohne Text» statistisch signifikant,  $t(191) = 12.28$ ,  $p < .01$ ,  $d = 0.89$ , über dem Zufallsniveau (8.75 Punkte). Auch in der Textbedingung «ohne Text und ohne Fragen» konnten Schülerinnen und Schüler MC-Items noch statistisch signifikant,  $t(176) = 3.15$ ,  $p < .01$ ,  $d = 0.24$ , überzufällig häufig lösen.

### 3.2 Faktorstruktur der Außenkriterien

In nächsten Schritt wenden wir uns der Struktur der Außenkriterien zu, die in der Untersuchung erhoben wurden. Die verschiedenen Variablen korrelierten in mittlerer Höhe miteinander (vgl. Tab. 1). Erwartungsgemäß finden sich die höchsten Korrelationen zwischen den verbalen Intelligenzvariablen (verbale Problemlösefähigkeit und Wortschatz) sowie zwischen den beiden Prozessvariablen (lokale und globale Kohärenzbildung). Das Korrelationsmuster zeigt jedoch ebenfalls, dass es eine enge Verbindung zwischen den beiden Variablentypen gibt.

Die theoretische Unterscheidung zwischen Intelligenz- und Prozessvariablen wurde deshalb mit einer konfirmatorischen Faktorenanalyse explizit überprüft. Das in Abbildung 1 dargestellte Messmodell wies eine zufriedenstellende Passung zu den Daten auf,  $\chi^2(1)_{N=366} = 3.01$ ,  $p = .082$ , CFI = .99, RMSEA = .074, SRMR = .015, die statistisch signifikant besser war als die eines



Anmerkung. Alle dargestellten Effekte sind statistisch signifikant.

Abbildung 1. Parameterschätzung des konfirmatorischen Faktormodells der verbalen Intelligenz (verbales Problemlösen, Wortschatz) und der lesespezifischen Prozessvariablen (lokale und globale Kohärenzbildung).

einfaktoriellen Modells, das davon ausgeht, dass alle vier Variablen nur ein Konstrukt abbilden,  $\Delta\chi^2(1)_{N=366} = 10.471, p < .01$ . Die Reliabilität (*Omega*) der beiden latenten Faktoren war zufriedenstellend,  $\Omega = .60$  bzw.  $\Omega = .63$ . Es wurden zwei Indikatorvariablen gebildet, indem die beiden zugrunde liegenden Testscores zunächst *z*-transformiert und dann gemittelt wurden. Die beiden manifesten Variablen korrelierten statistisch signifikant mit  $r = .46$  miteinander.

### 3.3 Beziehung zwischen dem Leseverständnis und den Intelligenz- und Prozessvariablen

Im nächsten Schritt wurde geprüft, ob und in welchem Maß die beiden Außenkriterien mit der Leseverständnisleistung zusammenhängen. Die erste Spalte von Tabelle 3 zeigt die Korrelationen der Testleistung in der «mit

Text»-Bedingung mit der Intelligenz- und Prozessvariablen. Erwartungsgemäß fällt die Korrelation mit der verbalen Intelligenz am höchsten aus. Die Annahme, dass lesespezifische Prozessvariablen keinen Einfluss auf die Leseverständnisleistung der Schülerinnen und Schüler haben, wird jedoch nicht gestützt. Vielmehr korrelierten die Prozessvariablen in ähnlicher Höhe, aber signifikant niedriger mit der Testleistung wie die verbale Intelligenz. Dies gilt auch, wenn man sich die Semipartialkorrelationen betrachtet, die die Korrelation zwischen der Testleistung und den Intelligenz- und Prozessvariablen angibt, nachdem die jeweils andere aus dieser auspartialisiert wurde (vgl. Tab. 3): Beide Semipartialkorrelationen fallen deutlich niedriger aus als die zugehörigen einfachen Korrelationen, was zeigt, dass es einen deutlichen Überlappungsbereich zwischen den beiden Variablentypen gibt. Darüber hinaus scheinen beide Variablentypen jedoch auch unabhängige, spezifische Beiträge zur Testleistung in den verschiedenen Untersuchungsbedingungen zu haben. Die Effizienz, mit der lesespezifische Prozesse ausgeführt werden, steht also in einem systematischen Zusammenhang mit der Testleistung.

Für die vorliegende Untersuchung ist nun die Frage entscheidend, ob sich das Verhältnis von verbaler Intelligenz und Prozessvariablen ändert, wenn Textinformationen von der Testbearbeitung eliminiert werden. Die Antwort auf diese Frage gibt die mittlere und rechte Spalte von Tabelle 3, in der die Korrelationen der verschiedenen Außenkriterien mit der Testleistung in der «ohne Text»- und der «ohne Text und ohne Fragen»-Bedingung abgetragen sind. Zunächst einmal ist klar zu erkennen, dass die Korrelationen in den beiden informationsreduzierten Bedingungen schrittweise abnehmen, d. h. die Korrelationen sind jeweils statistisch signifikant höher in der «mit Text»-Bedingung als in der «ohne Text»-Bedingung, welche wiederum hö-

Tabelle 3

Mittelwerte, Standardabweichungen und Reliabilitäten der Testleistung in den drei Untersuchungsbedingungen und ihre Korrelationen, Semipartialkorrelationen und minderungskorrigierte Korrelationen mit den verbalen Intelligenz- und Prozessvariablen

Korrelation mit Leseverständnis		Testbedingung		
		mit Text	ohne Text	ohne Text/Fragen
Leseverständnis	<i>M (SD)</i>	16.81 (6.68)	11.97 (3.60)	9.44 (2.80)
	Interne Konsistenz $\alpha$	.84	.49	.17
	Retest-Reliabilität <i>r</i>	.56	.25	.03
Produkt-Moment-Korrelationen	Verbale Intelligenz	<b>.66<sup>a</sup></b>	<b>.47<sup>b</sup></b>	<b>.25<sup>c</sup></b>
	Prozessvariablen	<b>.54<sup>b</sup></b>	<b>.43<sup>b</sup></b>	<b>.23<sup>c</sup></b>
Semipartialkorrelationen	Verbale Intelligenz	<b>.46<sup>a</sup></b>	<b>.29<sup>b</sup></b>	<b>.25<sup>b</sup></b>
	Prozessvariablen	<b>.23<sup>b</sup></b>	<b>.17<sup>b</sup></b>	-.10 <sup>c</sup>
minderungskorrigierte Korrelationen (mit $\alpha$ als Reliabilitätsschätzung)	Verbale Intelligenz	<b>.94<sup>a</sup></b>	<b>.87<sup>b</sup></b>	<b>.79<sup>c</sup></b>
	Prozessvariablen	<b>.74<sup>c</sup></b>	<b>.78<sup>c</sup></b>	<b>.70<sup>d</sup></b>

Anmerkungen: Fett gedruckte Korrelationen sind signifikant größer als 0. Korrelationen mit unterschiedlichen Indizes (<sup>a/b/c/d</sup>) unterscheiden sich statistisch signifikant voneinander.

her sind als in der «ohne Text und ohne Fragen»-Bedingung. Dieser Trend gilt für die verbale Intelligenz und die lesespezifischen Prozessvariablen gleichermaßen und dürfte lediglich eine Konsequenz der geringeren Reliabilitäten in den beiden informationsreduzierten Bedingungen sein (s. 3.4.).

Entscheidend für die Frage, ob in den verschiedenen Informationsbedingungen unterschiedliche Prozesse erfasst werden, ist jedoch vielmehr, ob sich die Muster der Korrelationen qualitativ voneinander unterscheiden. Das Ergebnismuster ist hier nicht vollkommen eindeutig (s. Tab. 3): Einerseits zeigen Intelligenz- und Prozessvariablen das gleiche abfallende Muster über die verschiedenen Untersuchungsbedingungen, und die verbale Intelligenz weist in allen Untersuchungsbedingungen höhere Zusammenhänge auf als die Prozessvariablen. Andererseits sind die Unterschiede in den beiden informationsreduzierten Bedingungen nicht groß genug, um die Signifikanzgrenze zu überschreiten. Ein ähnliches Bild ergibt sich auch, wenn man die Semipartialkorrelationen betrachtet: Ähnlich wie bei den einfachen Korrelationen ist in der «mit Text»-Bedingung der Beitrag der verbalen Intelligenz höher als der der Prozessvariablen, während er in der «ohne Text»-Bedingung ausgewogen ist. Interessant ist, dass in der «ohne Text/ohne Fragen»-Bedingung die Prozessvariablen keinen eigenständigen Beitrag mehr leisten.

### 3.4 Reliabilität und minderungskorrigierte Korrelationen

Wodurch wird das abnehmende Korrelationsmuster in den drei Untersuchungsbedingungen verursacht? Eine naheliegende Hypothese ist, dass in den verschiedenen Testbedingungen der Anteil der Fehlervarianz zunehmend ansteigt, weil immer weniger Kriteriumsvarianz erfasst wird. Wenn diese stimmt, dann sollte die Reliabilität der Testscores schrittweise abnehmen. In Tabelle 3 ist sowohl die interne Konsistenz (Cronbachs  $\alpha$ ) als auch die Retest-Reliabilität in den drei Untersuchungsbedingungen angegeben. Es ist klar zu erkennen, dass die Reliabilität mit dem Anteil der eliminierten Textinformation schrittweise abnimmt. Korrespondierend hierzu nehmen die Beziehungen zu den Außenvariablen ab, allerdings unterscheiden sie sich selbst in der «ohne Text/ohne Fragen»-Bedingung immer noch statistisch signifikant von Null.

Um die Hypothese, dass die Korrelationsmuster in den verschiedenen Bedingungen ähnlich ausfallen, noch klarer zu prüfen, wurden zusätzlich minderungskorrigierte Korrelationen berechnet, welche die unterschiedliche Reliabilität der Testscores berücksichtigten (vgl. Tab. 3, vierte Zeile). Erwartungsgemäß steigen alle Korrelationen durch die Korrektur an, das grundlegende Muster ist jedoch ähnlich wie in den nicht-korrigierten Analysen: Die Korrelationen sinken mit dem Ausmaß der Informationsreduktion leicht ab und die verbale Intelligenz korreliert jeweils höher mit der Testleistung als die Prozessvariablen.

## 4 Zusammenfassung und Schlussfolgerungen

Unsere Ergebnisse replizieren zunächst einmal vollständig die Befunde von Rost und Sparfeldt (2007): Die erreichte Punktzahl nimmt mit der Anzahl der eliminierten Informationen zunehmend ab, liegt aber in der «ohne Text»-Bedingung weit und in der «ohne Text und ohne Fragen»-Bedingung stabil über dem Zufallswert. Auch absolut sind die Ergebnisse der beiden Untersuchungen sehr ähnlich: Zwar fielen die Ergebnisse der Berliner Schülerinnen und Schüler in der «mit Text»-Bedingung insgesamt um 2 Punkte niedriger aus (was angesichts des bekannten Nachteils von Stadt- gegenüber Flächenländern wenig überraschend ist), aber die Ergebnisse in den beiden informationsreduzierten Bedingungen weichen weniger als einen Punkt voneinander ab. Die Korrelationen zwischen der Testleistung in den unterschiedlichen Untersuchungsbedingungen sind ebenfalls identisch. Dass sich die Ergebnisse von Rost und Sparfeldt dermaßen exakt replizieren lassen, ist bemerkenswert und für den Bereich der Pädagogischen Psychologie eher untypisch. Dies spricht für die Stabilität des berichteten Effekts und dafür, dass die Ergebnisse der beiden Studien direkt miteinander vergleichbar sind und sich auch gut auf andere Studien übertragen lassen sollten (Coleman et al., 2010; Katz et al., 1990).

Ähnlich wie bei Rost und Sparfeldt (2007) wurden die verbale Problemlösefähigkeit und der Wortschatz als Indikatoren der verbalen Intelligenz erfasst, die den Schülerinnen und Schülern für ihre Antwort mit den verschiedenen Untersuchungsbedingungen zur Verfügung stehen. Erwartungsgemäß und in Übereinstimmung mit zahlreichen anderen Studien korrelierte die verbale Intelligenz hoch mit der Testleistung in der «mit Text»-Bedingung. Verbale Intelligenz wird jedoch auch als ein wichtiger Prädiktor für potenzielle Rateprozesse in den «ohne Text»-Bedingungen diskutiert (Katz et al., 1990). Dies gilt besonders für den Wortschatz als Indikator des kristallinen Welt- und Vorwissens der Schülerinnen und Schüler. Vor diesem Hintergrund ist erstaunlich, dass die Höhe dieser Korrelation mit der Anzahl der eliminierten Textinformationen abnimmt. Dies ist besonders offensichtlich, wenn man nicht nur die einfachen, sondern auch die minderungskorrigierten Korrelationen betrachtet, die ebenfalls eher ein abnehmendes und auf keinen Fall ein zunehmendes Muster aufweisen.

Ein zentrales Anliegen der vorliegenden Studie war es, zusätzlich zur verbalen Intelligenz der Schülerinnen und Schüler weitere, lesespezifische Prozessvariablen zu erfassen, die sich direkt(er) auf die Effizienz der kognitiven Prozesse beziehen, die für die Etablierung eines kohärenten und stabilen Situationsmodells notwendig sind (Aktivierung von Hintergrundwissen, Generierung von Inferenzen etc., vgl. Kintsch, 1998). Unsere Ergebnisse zeigen, dass solche Prozessvariablen in allen Untersuchungsbedingungen substanziell und über die Effekte der verbalen Intelligenz hinaus die Testleistung beeinflussen. Die Effizienz, mit der die Schülerinnen und Schüler Textinferenzen ge-

nerieren und validieren können, hat also einen eigenständigen Einfluss auf die Testleistung. Gleichzeitig gab es einen erheblichen Überlappungsbereich in der Vorhersageleistung von verbaler Intelligenz und Prozessvariablen. Dieser Befund ist kompatibel mit einem Ressourcen-Prozess-Modell, das davon ausgeht, dass Leserinnen und Leser die ihnen zur Verfügung stehenden verbalen Ressourcen in effizientere Verarbeitungsprozesse investieren, was sich positiv auf die Qualität des Verständnisprodukts auswirkt (vgl. Schroeder, 2011). Wie von Rost und Sparfeldt (2007) herausgearbeitet, ist die verbale Intelligenz in der Tat ein wichtiger Faktor, der für die Bearbeitung von MC-Items von Bedeutung ist. Unsere Ergebnisse zeigen darüber hinaus, dass nicht die verbale Intelligenz einer Person allein entscheidend ist, sondern ob und wie diese im Laufe der Textrezeption genutzt wird.

Die Annahme, dass in den unterschiedlichen Untersuchungsbedingungen kategorial verschiedene Antwortprozesse eingesetzt werden, erhält insgesamt durch unsere Daten nur wenig empirische Unterstützung. Das Muster der bivariaten Korrelationen fiel über die verschiedenen Untersuchungsbedingungen vergleichsweise homogen aus, z. B. zeigten Intelligenz- und Prozessvariablen das gleiche abfallende Muster. Bezüglich der relativen Wichtigkeit der verbalen Intelligenz und der Prozessvariablen ist das Bild weniger einheitlich: In allen Bedingungen fielen die Korrelationen mit dem Intelligenzfaktor höher aus als für die Leseprozesse, allerdings ließen sich diese Unterschiede nicht in allen Bedingungen inferenzstatistisch absichern. Insgesamt legen die Ergebnisse nahe, dass die Testleistung in den verschiedenen Bedingungen von ähnlichen Faktoren beeinflusst wird und sich das Beziehungsmuster nicht fundamental zwischen ihnen unterscheidet.

Wie erwartet werden die Effekte mit der Anzahl der eliminierten Textinformationen insgesamt schwächer, was darauf zurückzuführen ist, dass das erfasste Konstrukt zunehmend unreliabler erfasst wird. Dies ist mit der Vorstellung kompatibel, dass der Anteil von Fehlervarianz in den informationsreduzierten Bedingungen stetig zunimmt, da immer weniger leserelevante Prozesse ausgeführt und stattdessen der Anteil unsystematischen Ratens und anderer strategischer Effekte zunimmt. Werden die Effekte jedoch hierfür bereinigt, dann haben sowohl Intelligenz- als auch Prozessvariablen ähnlich hohe Effekte wie in der «mit Text»-Bedingung. Diese Aussage muss evtl. für die «ohne Text/ohne Fragen»-Bedingung eingeschränkt werden, in der die Effekte insgesamt niedriger ausfielen. Darüber hinaus hatten hier die Prozessvariablen keinen eigenständigen Einfluss auf die Testleistung der Schülerinnen und Schüler. Es könnte also sein, dass sich der Antwortprozess in dieser Bedingung von dem in der «ohne Text»-Bedingung unterscheidet, da hier nicht mehr genügend sprachliche Informationen zur Verfügung stehen, um wirklich noch lesespezifische Varianz erfassen zu können (vgl. auch Rost & Sparfeldt, 2007).

Zusammenfassend sprechen unsere Befunde für die dritte in der Einleitung herausgearbeitete Position, die da-

von ausgeht, dass MC-Items mit und ohne Textpräsentation ein einheitliches Konstrukt erfassen, das selbst dann für den Leseprozess relevante Teilfähigkeiten erfasst, wenn die Texte nicht präsentiert werden. Die relative Homogenität der Effektmuster in den drei Bedingungen widerspricht dabei zunächst der Annahme, dass sich die in den verschiedenen Untersuchungsbedingungen erfassten Fähigkeiten stark voneinander unterscheiden. Insbesondere ist es nicht so, dass lesespezifische Prozessvariablen einen starken Einfluss in der «mit Text»-Bedingungen haben und Intelligenzvariablen in der «ohne Text»-Bedingung (Hannon & Daneman, 2001). Vielmehr ergibt sich ein weitgehend einheitliches Muster, das nahe legt, dass in allen Bedingungen eine komplexe Mischung von verbaler Intelligenz und lesespezifischen Prozessen erfasst wird, die miteinander interagieren und erst in ihrem Zusammenwirken für das Lesen funktional relevant sind (Freedle & Kostin, 1994).

Heißt das, dass es egal ist, ob bei MC-Leseverständnistests die Texte präsentiert und gelesen werden? Sicherlich nicht, und zwar sowohl auf der individuellen als auch auf der messtheoretischen Ebene. Für die individuellen Schülerinnen und Schüler ergibt sich sehr wohl ein entscheidender Vorteil, bestimmte Distraktoren auf der Grundlage der verfügbaren Textinformationen ausschließen zu können (Daneman & Hanon, 2001). Entsprechend steigt ihre Testleistung ja erheblich, sobald sie auf Textinformationen zurückgreifen können. Auf der messtheoretischen Ebene führt die mangelnde Verfügbarkeit von textuellen Kontextinformationen v. a. dazu, dass die Prozesse, die ohne Textdarbietung erfasst werden, zunehmend durch Fehlervarianz kompromittiert werden. Die Wahrscheinlichkeit, leserelevante Prozesse zu erfassen, nimmt deshalb mit der Elimination von Text- und Frageelementen stetig ab, was zu vermeiden ist: Leseverständnistests ohne Textpräsentation erfassen durchaus Lesen – allerdings nicht besonders gut.

## Danksagung

Wir danken Jürgen Baumert für hilfreiche Kommentare zu einer früheren Fassung des Manuskripts und Doreen Burisch, Patricia Ziehn und Julian Berwald für ihre Hilfe bei der Durchführung der Testung.

## Literatur

- Anger, H., Bargmann, R. & Voigt, M. (1972). *Verständiges Lesen* 7–9. Weinheim: Beltz.
- Cain, K., Oakhill, J. & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96, 31–42.
- Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah: Erlbaum.

- Coleman, C., Lindstrom, J., Nelson, J., Lindstrom, W. & Gregg, K. N. (2010). Passageless comprehension on the Nelson-Denny reading test: Well above chance for university students. *Journal of Learning Disabilities, 43*, 244–249.
- Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.
- Daneman, M. & Hannon, B. (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. *Journal of Experimental Psychology: General, 130*, 208–223.
- Freedle, R. & Kostin, I. (1994). Can multiple-choice reading tests be construct-valid? *Psychological Science, 5*, 107–110.
- Hannon, B. & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology, 93*, 103–128.
- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Weinheim: Beltz.
- Katz, S., Lautenschlager, G., Blackburn, A. & Harris, F. (1990). Answering reading comprehension items without passages on the SAT. *Psychological Science, 1*, 122–127.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- Magliano, J. P., Millis, K. K., Ozuru, Y. & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (S. 107–136). Mahwah, NJ: Erlbaum.
- Pearson, P. & Hamm, D. (2005). The assessment of reading comprehension: A review of practices – past, present, and future. In S. Paris & S. Stahl (Eds.), *Children's reading comprehension and assessment* (S. 13–69). Mahwah, NJ: Erlbaum.
- Preston, R. C. (1964). Ability of students to identify correct response before reading. *The Journal of Educational Research, 58*, 181–183.
- Richter, T. & Christmann, U. (2002). Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In N. Groeben & B. Hurrelmann (Hrsg.), *Lesekompetenz: Bedingungen, Dimensionen, Funktionen* (S. 25–58). Weinheim: Juventa.
- Rost, D. H. & Sparfeldt, J. R. (2007). Leseverständnis ohne Lesen? Zur Konstruktvalidität von multiple-choice-Leseverständnisaufgaben. *Zeitschrift für Pädagogische Psychologie, 21*, 305–314.
- Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology, 103*, 877–896.
- Sparfeldt, J., Kimmel, R., Löwenkamp, L., Steingräber, A. & Rost, D. H. (2012). Not read, but nevertheless solved? Three experiments on PIRLS multiple choice reading comprehension test items. *Educational Assessment, 17*, 214–232.
- Weiß, R. H. (1998). *Grundintelligenztest Skala 2 (CFT-20) mit Wortschatztest (WS) und Zahlenfolgentest (ZF)*. Göttingen: Hogrefe.

---

PD Dr. Sascha Schroeder

---

MPFG Reading Education and Development (REaD)  
 Max-Planck-Institut für Bildungsforschung  
 Lentzeallee 94  
 14195 Berlin  
 Deutschland  
 sascha.schroeder@mpib-berlin.mpg.de