

Erfassung des Lesevolumen in Large-Scale Studien: Ein Vergleich von Globalurteil und textspezifischem Urteil

Franziska Maria Locher und Maximilian Pfof

Otto-Friedrich-Universität Bamberg, Lehrstuhl für Empirische Bildungsforschung

Zusammenfassung

In diesem Beitrag werden zwei Möglichkeiten der Erfassung des Lesevolumens, wie sie in Large-Scale-Studien meist verwendet werden, verglichen - zum einen ein differenziertes textsortenspezifisches Urteil und zum anderen ein häufig angewandetes Globalurteil zur Einschätzung des Lesevolumens. Zum Vergleich der beiden Urteile liegen Daten des Nationalen Bildungspanels (NEPS) mit einer Stichprobe von $N = 5\,475$ Schülerinnen und Schülern der fünften Klassenstufe und $N = 16\,425$ Schülerinnen und Schülern der neunten Klassenstufe sowie eine Stichprobe aus PISA 2009 von $N = 4\,979$ Schülerinnen und Schülern im Alter von 15 Jahren vor. Insgesamt zeigt sich, dass im NEPS (insbesondere in Klassenstufe fünf) das Globalurteil das verlässlichere Instrument darstellt. Für die textsortenspezifischen Urteile zeigen sich u. a. deutliche Bodeneffekte und nur geringe Zusammenhänge mit den Kriteriumsmaßen Lesefreude und Lesekompetenz. Im Vergleich der beiden textsorten-spezifischen Urteile in NEPS und PISA zeigt sich ein Vorteil für die in PISA verwendeten Items. Es werden Empfehlungen für den Umgang mit den NEPS-Daten aus den entsprechenden Scientific-Use-Files abgeleitet.

Schlüsselwörter: Lesevolumen, Textsorten, Large-Scale-Panel, NEPS, PISA 2009

Measuring reading volume in Large-Scale-Assessments: A comparison of an overall evaluation and a differentiated evaluation relating different text types

Abstract

In this article, we compared two measures of reading volume, which are common in Large-Scale-Panels: a type of text specific evaluation of students' reading volume as well as an overall evaluation of reading volume. Therefore, we examined a sample of fifth ($N = 5\,475$) and ninth grade ($N = 16\,425$) secondary students from the National Educational Panel Study (NEPS) and a sample of $N = 4\,979$ fifteen-year-old students from the PISA-study 2009. Regarding NEPS-data, results show that the overall evaluation of students reading volume appears to be the more reliable measure. In text type specific measures of reading, floor effects and weak correlations with the criteria variables, reading enjoyment and reading competence, were present. Such weaknesses were less pronounced in items of the PISA study. We discuss the findings and provide recommendations for working with the NEPS scientific use files.

Keywords: reading volume, text types, Large-Scale-Panel, NEPS, PISA 2009

Für das verstehende Lesen von Texten sind eine Reihe komplexer Teilfähigkeiten notwendig, weshalb es für die Entwicklung der Lesekompetenz wichtig ist, regelmäßig zu lesen. Verschiedene Studien konnten hierzu bereits mehrfach zeigen, dass das Lesevolumen positiv mit der Entwicklung der Lesekompetenz zusammenhängt

(Anderson, Wilson & Fielding, 1988; Guthrie, Wigfield, Metsala & Cox, 1999; Mol & Bus, 2011; Pfof, Dörfler & Artelt, 2010, 2013). Über verschiedene Messinstrumente hinweg zeigen sich dabei in der Regel Zusammenhänge zwischen dem Lesevolumen und der Lesekompetenz in

Diese Veröffentlichung wurde unterstützt durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft (Kennzeichen PF 840/2-1).

Post-Print. Diese Artikelfassung entspricht nicht vollständig dem in der Diagnostica, (2019), 65, pp. 26-36 (© 2019 Hogrefe Verlag) veröffentlichten Artikel. Dies ist nicht die Originalversion des Artikels und kann daher nicht zur Zitierung herangezogen werden. Der folgende Link führt zum Originalartikel <https://doi.org/10.1026/0012-1924/a000203>

einer Höhe von $r = .3$ – $.4$. Die Entscheidung für oder gegen ein Messinstrument hängt häufig vom Studiendesign oder den Analysezielen ab. In empirischen Studien mit thematischem Fokus auf dem Leseverhalten ist es nötig, dieses umfangreich zu erfassen (z. B. Guthrie et al., 1999; Spear-Swerling, Brucker & Alfano, 2010). Im Rahmen von Large-Scale-Studien mit multiplen Analysezielen soll hingegen häufig das Lesevolumen als eine Information in einer Reihe von vielen Fragen miterfasst werden. Dabei ist es dann in der Regel unerlässlich, dass die verwendeten Instrumente möglichst zeit- und platzsparend sind. Solche häufig nur oberflächlichen Einschätzungen zum Lesevolumen bieten im Anschluss jedoch oftmals Anlass für, nicht immer empirisch vollständig gerechtfertigte, Kritik. Der vorliegende Beitrag beschäftigt sich daher mit zwei eher ökonomischen Arten von Fragebogenskalen zur Erfassung des Lesevolumens, wie sie oftmals in Large-Scale-Studien zum Einsatz kommen. Zum einen geht es um das häufig verwendete Verfahren des Globalurteils, also einer allgemeinen Einschätzung darüber, wie lange pro Tag durchschnittlich gelesen wird. Zum anderen geht es um ein differenzierteres textsortenspezifisches Urteil, bei dem die Personen Einschätzungen darüber abgeben, wie häufig sie bestimmte Textsorten lesen. Anhand dieser zwei Instrumente soll untersucht werden, ob und wie gut es auch in Studien mit entsprechenden zeitlichen Restriktionen möglich ist, das Lesevolumen differenziert zu erfassen. Beide Urteile sind allerdings nicht voneinander unabhängig. Das Globalurteil ist vielmehr eine implizite Zusammenfassung der gesamten investierten Lesezeit, die für unterschiedliche Textsorten aufgebracht wurde. Mit der internationalen Schulleistungsstudie PISA (Programme for International Student Assessment) und Deutschlands größter Panelstudie zur Erforschung von Bildungsprozessen NEPS (Nationales Bildungspanel) wurden zwei der bekanntesten und für Re-Analysezwecke populärsten Large-Scale-Studien zum Vergleich herangezogen. Analysen, welche etwaige Probleme vorhandener Instrumente und Datensätze darstellen sind grundlegend, um mögliche Fehlinterpretationen in Forschungsarbeiten vorzubeugen. Besonders hinsichtlich der NEPS-Daten, für die über eintausend offizielle nationale und internationale Datennutzer verzeichnet werden können², ist dies wichtig. Darüber hinaus sollen die dargelegten Befunde Forscherinnen und Forscher bei der Gestaltung ihrer Untersuchungsinstrumente unterstützen und somit zukünftig

zu einer effektiveren Nutzung zeitlicher und monetärer Ressourcen in Large-Scale-Forschungen beitragen.³

Theoretischer Hintergrund

Erfassung des Lesevolumens

Bei der Messung des Lesevolumens wurde und wird bislang auf unterschiedliche Instrumente zur Erfassung zurückgegriffen. Bei Verfahren wie dem Titel- oder Autorinnen- und Autoren-Rekognitionstest (TRT & ART) werden die Teilnehmenden nach dem Erkennen von Autorinnen beziehungsweise Autoren oder Buchtiteln gefragt (z. B. Cipielewski & Stanovich, 1992; Spear-Swerling et al., 2010). Durch dieses Vorgehen können Einschränkungen wie sozial erwünschtes Antwortverhalten umgangen werden. Allerdings handelt es sich dabei um eine zeit- und damit kostenintensive Erfassungsmethode. Ferner sind ARTs und TRTs indirekte Messmethoden. Die Tatsache, dass Personen eine Autorin beziehungsweise einen Autor oder Titel kennen, bedeutet also nicht gleichermaßen, dass das Buch auch tatsächlich gelesen wurde (Schroeder, Segbers & Schröter, 2016). Dieser Punkt schränkt die Aussagekraft dieses Instruments erheblich ein. Ebenfalls nicht einfach, wenn auch häufig als Optimalmethode betrachtet, ist die Erfassung des Lesevolumens durch Lesetagebücher (z. B. Anderson et al., 1988). Diese Art des Protokollierens der Leseaktivitäten führt zu einer vermeintlich genauen Schätzung des Lesevolumens. Aufgrund des hohen Aufwandes kann dieses Verfahren aber nur selten und insbesondere meist nur über einen kurzen Zeitraum eingesetzt werden. Dieses Vorgehen bringt somit erhebliche Einschränkungen bezüglich der Generalisierung der Ergebnisse mit sich (Schroeder et al., 2016).

Häufig werden Fragebogenitems, wie sie auch im Rahmen dieses Artikels im Vordergrund stehen, zu einer globalen Einschätzung des Lesevolumens (z. B. Pfost et al., 2010) und zu einer textsortenspezifischen Einschätzung des Lesevolumens (z. B. Pfost et al., 2013) verwendet. Besonders Globalurteile sind in der Erfassung wenig zeitaufwändig und folglich ökonomisch. Sie finden daher oft im Rahmen großer und längsschnittlicher Datenerhebungen Platz. Häufig werden diese Einschätzungen allerdings dafür kritisiert, nur relative Unterschiede der Perso-

² Vgl. NEPS Neuigkeiten Archiv, 27.10.15, URL: <https://www.neps-data.de/de-de/neuigkeiten/2015.aspx>

³ Diese Arbeit nutzt Daten des Nationalen Bildungspanels (NEPS): Startkohorte Klasse fünf (doi:10.5157/NEPS:SC3:5.0.0) und Startkohorte Klasse neun (doi:10.5157/NEPS:SC4:6.0.0). Die Daten des NEPS wurden von 2008 bis 2013 als Teil des Rahmenprogramms zur

Förderung der empirischen Bildungsforschung erhoben, welches vom Bundesministerium für Bildung und Forschung (BMBF) finanziert wurde. Seit 2014 wird NEPS vom Leibniz-Institut für Bildungsverläufe e.V. (LifBi) an der Otto-Friedrich-Universität Bamberg in Kooperation mit einem deutschlandweiten Netzwerk weitergeführt.

nen zu erfassen und aufgrund sozial erwünschten Antwortverhaltens Verzerrungen zu unterliegen (Schroeder et al., 2016). Darüber hinaus kann in diesem Globalurteil nicht in unterschiedliche Textsorten unterschieden werden. Textsortenspezifische Urteile haben im Vergleich zum reinen Globalurteil den Vorteil, dass darin abgestuft werden kann, welche Art der Lektüre wie oft gelesen wird. Dies erlaubt es in späteren Analysen Unterschiede in der Lektürequalität zu berücksichtigen, die bei Globalurteilen verloren gehen würden.

Die Rolle unterschiedlicher Textsorten

Bei der Wahl der Lektüre gibt es eine Vielzahl an Möglichkeiten: von traditionellen Büchern, Zeitungen bis hin zum Internet unterscheiden sich die Medien nicht nur hinsichtlich des Lesezwecks, sondern auch in ihrem Anspruchsniveau (Gehrer & Artelt, 2013; Hayes & Ahrens, 1988). Es liegt daher nahe, dass sich Zusammenhänge der Lesekompetenz und der Lesemotivation mit der Quantität des Lesens unterschiedlich gestalten, je nachdem, auf welcher Textsorte der Fokus liegt. Ergebnisse von Pfof et al. (2013) weisen darauf hin, dass vor allem das Lesen traditioneller Bücher (z. B. Romane) einen positiven Effekt auf die Entwicklung des Leseverstehens und den Wortschatz hat. Einen eher geringen Zusammenhang konnten die Autoren für das Lesen von Zeitungen nachweisen. Lese- und Schreibaktivitäten im Internet hingen zu Teilen sogar negativ mit der Lesekompetenz der Schülerinnen und Schüler zusammen. Bei textsortenspezifischen Betrachtungen konnten McGeown, Osborne, Warhurst, Norgate und Duncan (2016) zeigen, dass auch unter Kontrolle von Geschlecht, Alter, Sozioökonomischem Status und Lesekompetenz die intrinsische Lesemotivation beziehungsweise die lesebezogenen Einstellungen besonders das Lesevolumen von fiktionalen Büchern ($r = .26$), Sachbüchern ($r = .25$) und Comics ($r = .12$) positiv vorhersagen. Signifikante Zusammenhänge lesebezogener Einstellungen mit dem Lesen von Zeitschriften konnten nicht gefunden werden.

Fragestellung und Ziel

Für den Fall, dass die Möglichkeit der Erfassung des Lesevolumens Restriktionen (Kosten, Zeit, usw.) unterliegt, stellen bislang Fragebogenskalen zum Lesevolumen eine akzeptable Lösung dar. Es stellt sich nun die Frage, wie gut es gelingt, das Lesevolumen global zu erfassen oder ob es günstiger ist, zwischen verschiedenen Textsorten zu differenzieren. Ferner stellt sich auf explorativer Ebene die Frage, ob dies für alle Altersgruppen gleich gut gelingt. Das Alter ist ein potenzieller Einflussfaktor, der in diesem Kontext bislang nicht systematisch untersucht

wurde, aufgrund der gegebenen Datengrundlage hier jedoch mitbetrachtet werden kann. In der folgenden Studie werden daher zwei Ausführungen textsortenspezifischer Urteile zum Lesevolumen (aus NEPS und PISA) und das Globalurteil sowie deren Ergebnisse genauer untersucht und verglichen.

Auswahl der Außenkriterien und Testwertinterpretation

Es soll festgestellt werden, ob die Ergebnisse der Messinstrumente hinsichtlich einer Interpretation im Sinne der Erfassung individueller Unterschiede im Lesevolumen valide sind. Hierzu werden Zusammenhänge der Fragebogenergebnisse mit entsprechenden Außenkriterien beziehungsweise theoretisch relevanten Konstrukten empirisch untersucht (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014). Dazu wird zum einen die Lesekompetenz als Außenkriterium herangezogen, da die Beziehung zum Lesevolumen hinreichend belegt wurde und die bekannten Befunde somit als Referenz verwendet werden können (Anderson et al., 1988; Guthrie et al., 1999; Mol & Bus, 2011; Pfof et al., 2010). Zum anderen wird mit der Lesefreude eine Facette der lesebezogenen Einstellungen und damit der intrinsischen Lesemotivation als Kriteriumsvariable gewählt. Diese kann als primäre Determinante für das Lesevolumen angesehen werden (Guthrie et al., 1999; Schaffner, 2009) und stellt somit ein praktisch unmittelbar relevantes Kriterium dar.

Das Ziel der vorliegenden Untersuchung besteht daher nicht primär in einer inhaltlichen Analyse und Interpretation von Zusammenhängen zwischen Lesevolumen und Außenkriterien. Ziel ist vielmehr die Untersuchung einer empirisch gestützten und theoriebasierten Testwertinterpretation anhand vorher definierter Ergebnisse und Annahmen. Das heißt, es soll geprüft werden, ob die Zusammenhänge zwischen Indikatoren des Lesevolumens und Außenkriterien in theoretisch sowie empirisch erwartbarer Höhe ausfallen und damit auf ein funktionierendes Instrument hindeuten. Für Korrelationen, welche nicht den theoretischen Erwartungen entsprechen (z. B. zu geringe oder unplausible Zusammenhänge), nehmen wir implizit an, dass diese Ergebnisse nicht auf das angenommene Konstrukt (hier: das Lesevolumen), sondern auf Unzulänglichkeiten in der Messung zurückzuführen sind (Hartig, Frey & Jude, 2012). Beispielsweise wählen Schroeder et al. (2016) zur Überprüfung der Validität ihres Kinder-Titelreognitionstests (K-TRT) ein vergleichbares Vorgehen und interpretieren Außenkorrelationen zwischen den Kriteriumsvariablen Wortschatz und Lesegeschwindigkeit und dem K-TRT.

Aufgrund der Annahme, dass das Lesevolumen sowohl mit der Lesekompetenz als auch mit der Lesefreude zusammenhängt, und aufgrund bereits genannter Evidenzquellen (z. B. Greaney & Hegarty, 1987; Guthrie et al., 1999; McGeown et al., 2016; Pfost et al., 2010; Pfost et al., 2013; Walberg & Tsai, 1984) erwarten wir zusammenfassend sowohl in PISA als auch in NEPS Folgendes: Bezogen auf das Globalurteil erwarten wir Korrelationen mit der Lesekompetenz in der Höhe von bisher gefundenen Effekten ($r \sim .3-.4$) sowie noch stärkere Korrelationen ($r \geq .4$) mit den lesebezogenen Einstellungen. Im Hinblick auf die textsortenspezifischen Items erwarten wir verhältnismäßig hohe Korrelationen für Bücher ($r \sim .3-.4$) und geringe Korrelationen für Comics ($r \sim .1$) mit der Lesekompetenz. Für die lesebezogenen Einstellungen beziehungsweise die Lesefreude erwarten wir hohe Korrelationen mit dem Lesen von fiktionalen Büchern ($r \sim .5$), geringe bis mittlere Korrelationen mit dem Lesen von Periodika ($r \sim .2$) und noch geringere Zusammenhänge mit dem Lesen von Comics.

Methode

Studiendesign und Stichprobe

NEPS

In der vorliegenden Arbeit werden Daten von Schülerinnen und Schülern der fünften (Startkohorte 3, Welle 1) und neunten Klassenstufe (Startkohorte 4, Welle 1 und 2) aus NEPS verwendet. Schülerinnen und Schüler der Förderschulen werden für die vorliegenden Betrachtungen ausgeschlossen. Insgesamt liegt damit für Startkohorte drei (Klassenstufe fünf) eine Stichprobe von $N = 5\,475$ Schülerinnen und Schülern (48 % weiblich; durchschnittliches Alter: 10.53 Jahre) vor. Für Startkohorte vier (Klassenstufe neun) liegt eine Stichprobe von $N = 16\,425$ Schülerinnen und Schülern (49 % weiblich; durchschnittliches Alter: 14.75 Jahre) vor.

PISA 2009

Der Schwerpunkt der PISA-Studie 2009 liegt auf der Lesekompetenz, weshalb sich diese Daten besonders gut für die Fragestellung der vorliegenden Arbeit eignen. Im Rahmen unserer Untersuchungen wird die Stichprobe deutscher Schülerinnen und Schüler herangezogen. Daher liegt für die Analysen eine Stichprobe von $N = 4\,979$ (49 % weiblich; durchschnittliches Alter: 15.83 Jahre) vor (Organisation for Economic Cooperation and Development, 2012).

Analysen und Instrumente

Für die Überprüfung der Fragestellung wurden Itemverteilungen und Mittelwerte betrachtet sowie Korrelations- und Faktorenanalysen durchgeführt. Aufgrund des Skalenniveaus der Lesevolumenvariablen und der nicht normalverteilten Daten in NEPS wurden Rangkorrelationskoeffizienten nach Kendalls Tau berechnet. Für die Itemverteilungen und Korrelationsanalysen der PISA-Daten wurden die entsprechenden zur Verfügung gestellten Gewichte verwendet. Die Analysen wurden mit Mplus 7 (Muthén & Muthén, 1998-2007) durchgeführt. Im Anschluss an die Korrelationsanalysen wurde eine explorative Faktorenanalyse für ordinale Daten durchgeführt und die Fit-Werte der Modelle verglichen. Dabei wurde mit der *Oblimin*-Methode eine schiefwinklige Rotationsmethode gewählt. Es wurden folgende Variablen für die Analysen herangezogen:

Globalurteil Lesevolumen

NEPS. Das Lesevolumen in NEPS wurde unter Verwendung zweier Items erfasst. Dabei wurde ein globales Urteil abgefragt („Wieviel Zeit verbringst du normalerweise damit, außerhalb der Schule zu lesen?“), um das durchschnittliche Lesevolumen an einem Schultag und an einem schulfreien Tag zu erheben. Auch das Lesen von E-Mails und Zeitschriften sollte dabei berücksichtigt werden. Dies wurde anhand einer fünfstufigen Likert-Skala (1 = *gar nicht außerhalb der Schule*, 2 = *bis zu einer halben Stunde*, 3 = *zwischen einer halben Stunde und einer Stunde*, 4 = *1 bis 2 Stunden*, 5 = *mehr als 2 Stunden*.) beantwortet. Cronbachs α in Klassenstufe fünf betrug $\alpha = .79$ und $\alpha = .88$ in Klassenstufe neun.

PISA 2009. Ähnlich wie auch bei NEPS wurde in PISA das Globalurteil (1 Item: „Wieviel Zeit verbringst du normalerweise damit, zu deinem Vergnügen zu lesen?“) anhand einer fünfstufigen Likert-Skala (1 = *Ich lese nicht zum Vergnügen*; 2 = *Bis zu 30 Minuten täglich*; 3 = *Zwischen einer halben und 1 Stunde täglich*; 4 = *1 bis 2 Stunden täglich*; 5 = *Mehr als 2 Stunden täglich*) erfasst (Hertel, Hochweber, Mildner, Steinert & Jude, 2014).

Differenziertes textsortenspezifisches Urteil

NEPS. In NEPS wurde neben dem Globalurteil das Lesevolumen spezifisch nach Genre und Inhalt (insgesamt 9 Items) erfragt: „Wie oft liest du in deiner Freizeit normalerweise... Krimis? Thriller? Horror- oder Fantasybücher?; Sachbücher?; klassische Literatur?; Comics?“ und „Liest du folgende Zeitungen oder Zeitschriften? Lokale Tageszeitung; Boulevardzeitung (wie BILD, BZ); Jugendseiten in anderen Zeitschriften (wie SZ, FAZ); Ma-

gazine (wie SPIEGEL, FOCUS Schule); andere Jugendzeitschriften“. Die Einschätzungen wurden anhand einer fünfstufigen Likert-Skala (1 = *nie oder selten*, 2 = *mehrmals im Monat*, 3 = *einmal in der Woche*, 4 = *mehrmals in der Woche*, 5 = *täglich*.) für jede Kategorie separat erfasst.

PISA 2009. Auch in PISA wurde das Lesevolumen textsortenspezifisch erfragt („Wie oft liest du zu deinem Vergnügen...“). Für folgende Textsorten fand eine Einschätzung statt: Zeitschriften und Magazine; Comic-Hefte und Comics; Romane, Erzählungen, Geschichten; Sachbücher (z. B. Geschichte, Biografie, Wissenschaft, Technik) und Tageszeitungen (insgesamt 5 Items). Das spezifische Urteil wurde ebenfalls anhand einer fünfstufigen Likert-Skala (1 = *Nie oder fast nie*; 2 = *Ein paarmal im Jahr*; 3 = *Etwa einmal im Monat*; 4 = *Mehrmals im Monat*; 5 = *Mehrmals in der Woche*) erfasst (Hertel et al., 2014).

Lesekompetenz

NEPS. In den Lesekompetenztests fanden sich 31 beziehungsweise 32 Items (5–7 pro Textsorte) im Multiple-Choice-Format als Entscheidungsaufgaben oder als Zuordnungsaufgaben zu den folgenden fünf alltagsnahen Textsorten: (a) Sachtexte, (b) Werbung, Anzeigen und Aufrufe, (c) Anleitungen und Anweisungen, (d) kommentierende Texte, (e) literarische Texte (Gehrer & Artelt, 2013). Jede Textsorte wurde durch einen Text repräsentiert. Dabei gilt es zu beachten, dass der Lesekompetenztest trotz der Textsortenvielfalt konzipiert wurde, um ein eindimensionales Lesekonstrukt zu erfassen. Diese Eindimensionalität wurde geprüft: Alle Dimensionen korrelieren hoch bis sehr hoch miteinander (Haberkorn, Pohl, Hardt & Wiegand, 2012; Pohl, Haberkorn, Hardt & Wiegand, 2012). Die individuelle Personenfähigkeit wird als WLE-Schätzer (Weighted Likelihood Estimation) angegeben. Die Reliabilität des Tests in Klassenstufe fünf (WLE-Reliabilität = .767) ebenso wie in Klassenstufe neun (WLE-Reliabilität = .749) war zufriedenstellend (Haberkorn et al., 2012; Pohl et al., 2012).

PISA 2009. In PISA orientiert sich das Konstrukt der Lesekompetenz am Literacy-Konzept. Es werden daher keine schulspezifischen Fähigkeiten, sondern sogenannte *Life Skills* gemessen. Die PISA-Aufgaben beinhalten dafür unterschiedliche Textformate (kontinuierlich wie z. B. Sachtexte oder Prosa und nichtkontinuierlich wie z. B. Listen oder Tabellen), die außerdem unterschiedliche Zwecke erfüllen (z. B. Anleitung, Beschreibung, etc.) und unterschiedliche Leseanlässe (z. B. beruflich, privat, etc.) beinhalten (Naumann, Artelt, Schneider & Stanat, 2010).

Für die Analysen mit der Lesekompetenz der Schülerinnen und Schüler wurden Plausible Values verwendet. Die Reliabilität der Plausible Values in der Domäne Lesen betrug .86 (Organisation for Economic Cooperation and Development, 2012).

Lesebezogene Einstellungen

NEPS. Zusätzlich wurden die Lust und das Interesse am Lesen erhoben (z. B. „Es macht mir Spaß Bücher zu lesen.“). Die insgesamt fünf positiv formulierten Items wurden anhand einer vierstufigen Likert-Skala (1 = *stimme gar nicht zu* bis 4 = *stimme völlig zu*) erfasst. Cronbachs α in Klassenstufe fünf betrug $\alpha = .88$ und $\alpha = .81$ in Klassenstufe neun.

PISA 2009. Die lesebezogenen Einstellungen in PISA wurden anhand von fünf invers (z. B. „Ich lese nur, wenn ich muss.“) und sechs positiv formulierten Items (z. B. „Lesen ist eines meiner liebsten Hobbys.“) erfasst und anhand einer vierstufigen Likert-Skala (1 = *Stimmt überhaupt nicht* bis 4 = *Stimmt ganz genau*) eingeschätzt (Hertel et al., 2014). Cronbachs α betrug $\alpha = .92$.

Ergebnisse

Ein Überblick über die deskriptiven Kennwerte der textsortenspezifischen Urteile sowie des *Globalurteils* und deren Interkorrelationen mit den Kriteriumsvariablen *Lesekompetenz* und *lesebezogene Einstellungen* ist in Tabelle 1 (Klassenstufe fünf) und in Tabelle 2 (Klassenstufe neun) für NEPS und in Tabelle 3 für PISA 2009 zu sehen.

Kennwertverteilungen und Korrelationsanalysen

NEPS

Bei Betrachtung der Häufigkeitsverteilung zeigt sich eine stark ungleiche Zellenbesetzung bei den textsortenspezifischen Urteilen. Der überwiegend größte Anteil (33–87 %) an Schülerinnen und Schülern gab an, die jeweilig abgefragte Textsorte niemals zu lesen. Ein nur sehr geringer Anteil (1–22 %) gab an, täglich oder über zwei Stunden täglich die entsprechende Textsorte zu lesen. Im Vergleich dazu ist beim *Globalurteil* eine gleichmäßigere Verteilung auf alle Antwortmöglichkeiten festzustellen. Dies zeigt sich auch im Mittelwert, der beim *Globalurteil* etwa in der Mitte des Skalenrange zu finden ist, wohingegen die Mittelwerte der textsortenspezifischen Urteile sehr niedrig ausgeprägt sind und deutlich auf Bodeneffekte hinweisen.

Für die Schülerinnen und Schüler der Klassenstufe fünf kann auch entnommen werden, dass die spezifischen Einschätzungen des Lesevolumens für unterschiedliche Texte nur sehr gering und überwiegend negativ mit der *Lesekompetenz* der Schülerinnen und Schüler der fünften Klassenstufen korrelieren. Dies entspricht nicht den Korrelationen vorheriger Untersuchungen. Ins-

besondere für klassische Literatur ($r = .01$, n. s.) oder Krimis, Thriller, Horror- oder Fantasybücher ($r = .04$) zeigen sich geringe bis gar keine Zusammenhänge. Auch die Korrelationen der Textsortenurteile mit den lesebezogenen Einstellungen sind vielfach nur schwach ausgeprägt. Das *Globalurteil* korreliert, den Erwartungen eher entsprechend, höher mit den *lesebezogenen Einstellungen* ($r = .44$) und der *Lesekompetenz* ($r = .20$).

Tabelle 1. Deskriptive Kennwerte und Kendall-Tau-Korrelationen (NEPS Klassenstufe fünf)

Variablen	<i>N</i>	<i>M</i>	<i>SD</i>	niedrigste ^a	höchste ^b	LK	LE
Lokale Tageszeitung	4 697	1.85	1.32	64 %	8 %	.04*	.14*
Boulevardzeitung	4 592	1.53	1.07	75 %	4 %	-.08*	.06*
Jugendseiten in anderen überregionalen Zeitungen	4 694	1.79	1.23	64 %	6 %	-.08*	.13*
Magazine	4 692	1.96	1.27	54 %	7 %	.01	.19*
andere Jugendzeitschriften	4 748	2.77	1.56	33 %	22 %	-.08*	.10*
Krimis, Thriller, Horror- oder Fantasybücher	4 789	2.66	1.60	40 %	19 %	.04*	.24*
klassische Literatur	4 582	2.20	1.40	48 %	10 %	.01	.24*
Sachbücher	4 480	2.18	1.32	45 %	8 %	-.02	.26*
Comics	4 592	2.76	1.54	33 %	20 %	-.03*	.11*
Globalurteil: freizeitl. Lesevolumen ^c	4 888	2.97	1.19	8 %	10 %	.20*	.44*
Schultag	4 760	2.95	1.22	12 %	13 %	.16*	.38*
schulfreier Tag	4 600	3.02	1.37	15 %	20 %	.22*	.45*

Anmerkungen: Korrelationen basieren auf manifesten Mittelwerten der Kriteriumsvariablen; LK = Lesekompetenz; (WLE-Schätzer = Weighted Likelihood Estimation); LE = Lesebezogene Einstellungen; * $p < .05$.

^a Prozent der Kategorien „nie oder selten“ bzw. Globalurteil: „gar nicht außerhalb der Schule“.

^b Prozent der Kategorien „täglich“ bzw. Globalurteil: „mehr als 2 Stunden“.

^c Mittelwert aus den Items Lesevolumen an Schultagen und schulfreien Tagen.

Tabelle 2. Deskriptive Kennwerte und Kendall-Tau-Korrelationen (NEPS Klassenstufe neun)

Variablen	<i>N</i>	<i>M</i>	<i>SD</i>	niedrigste ^a	höchste ^b	LK	LE
Lokale Tageszeitung	14 113	2.45	1.50	42 %	15 %	.11*	.13*
Boulevardzeitung	14 029	1.75	1.17	63 %	5 %	-.04*	.01
Jugendseiten in anderen überregionalen Zeitungen	14 039	1.39	.89	79 %	2 %	.02*	.12*
Magazine	14 093	1.62	.98	63 %	2 %	.04*	.15*
andere Jugendzeitschriften	14 142	2.19	1.25	40 %	6 %	-.12*	.06*
Krimis, Thriller, Horror- oder Fantasybücher	14 102	2.26	1.47	48 %	12 %	.22*	.44*
klassische Literatur	13 890	1.23	.68	87 %	1 %	.08*	.23*
Sachbücher	13 849	1.44	.85	72 %	1 %	.07*	.23*
Comics	13 960	1.62	1.05	66 %	3 %	-.02*	.07*
Globalurteil: freizeitl. Lesevolumen ^c	14 263	2.98	1.35	23 %	24 %	.24*	.44*
Schultag	14 190	2.91	1.36	20 %	16 %	.21*	.40*
Schulfreier Tag	14 142	3.05	1.49	22 %	25 %	.26*	.46*

Anmerkungen: Korrelationen basieren auf manifesten Mittelwerten der Kriteriumsvariablen; LK = Lesekompetenz (WLE-Schätzer = Weighted Likelihood Estimation); LE = Lesebezogene Einstellungen. * $p < .05$.

^a Prozent der Kategorie „nie oder selten“ bzw. Globalurteil: „gar nicht außerhalb der Schule“.

^b Prozent der Kategorie „täglich“ bzw. Globalurteil: „mehr als 2 Stunden“.

^c Mittelwert aus den Items Lesevolumen an Schultagen und schulfreien Tagen.

Die in Tabelle 2 abgetragenen Korrelationen der Textsortenurteile und der *Lesekompetenz* in Klassenstufe neun sind ebenfalls nur von geringer Stärke und replizieren gut die Befundlage aus der fünften Klassenstufe. Lediglich der Zusammenhang des textsortenspezifischen Urteils zum Lesevolumen von Krimi, Thriller, Horror oder Fantasybüchern hebt sich positiv mit einer Korrelation von .22 ab und reicht fast an die erwartete Effektgröße heran. Dieses Genre korreliert auch am stärksten mit den *lesebezogenen Einstellungen* ($r = .44$). Ansonsten zeigen sich aber auch hier nur mäßige Zusammenhänge mit den Kriteriumsvariablen. Das *Globalurteil* korreliert dagegen wieder deutlich höher mit den Kriteriumsvariablen *lesebezogene Einstellungen* ($r = .44$) und *Lesekompetenz* ($r = .24$).

PISA 2009

Die in Tabelle 3 abgetragenen Itemverteilungen zeigen eine breite Variation bezüglich der Häufigkeit des Lesens der unterschiedlichen Textsorten. Im Gegensatz zu den Items aus der NEPS-Studie sind hier, mit Ausnahme der Kategorie Comics, kaum oder nur geringe Bodeneffekte, das heißt starke Tendenzen in Richtung einer bestimmten Antwortkategorie zu erkennen. Dies spiegelt

sich auch in den Mittelwerten wider. Wie erwartet zeigt sich eine mittelstarke Korrelation zwischen *Lesekompetenz* und dem Lesen von Romanen, Erzählungen und Geschichten ($r = .44$), nicht aber mit dem Lesen von Comicbüchern ($r = -.03$) oder Zeitschriften und Magazinen ($r = .03$). Für das Lesen von Tageszeitungen und Sachbüchern zeigen sich gering positive Zusammenhänge zur *Lesekompetenz* ($r = .10$). Auch das *Globalurteil* hängt in der erwarteten Stärke mit der *Lesekompetenz* ($r = .36$) zusammen. Bei der Überprüfung der Korrelationen mit den *lesebezogenen Einstellungen* zeigen sich beim differenzierten Urteil bezüglich des Lesens von Romanen, Erzählungen und Geschichten und beim *Globalurteil* ähnlich starke Zusammenhänge ($r = .68$ – $.69$). In der Zusammenschau der Ergebnisse von NEPS und PISA können damit folgende zwei Punkte festgehalten werden: Erstens, im Hinblick auf die textspezifischen Urteile zeigen sich vor allem in der PISA-Studie die erwarteten Zusammenhänge zur *Lesekompetenz*. Zweitens, in beiden Studien fallen die Zusammenhänge zur *Lesekompetenz* und zu den *lesebezogenen Einstellungen* für das *Globalurteil* größer und eher in erwartbarer Höhe aus als für die textspezifischen Urteile.

Tabelle 3. Deskriptive Kennwerte und Korrelationen der Textsortenvariablen aus PISA 2009

Variablen	N	M	SD	niedrigste ^a	höchste ^b	LK	LE
Zeitschriften/ Magazine	4 458	3.46	1.19	7 %	21 %	.03*	.09*
Comic-Hefte/ Comics	4 433	1.76	1.15	60 %	5 %	-.03*	.06*
Romane, Erzählungen und Geschichten	4 415	2.70	1.45	29 %	16 %	.44*	.68*
Sachbücher	4 419	2.21	1.21	38 %	5 %	.10*	.20*
Tageszeitungen	4 441	3.57	1.45	16 %	37 %	.10*	.09*
Globalurteil zum Lesevolumen	4 552	2.15	1.24	42 %	6 %	.36*	.69*

Anmerkungen: Korrelationen basieren auf manifesten Mittelwerten der Kriteriumsvariablen. LK = Lesekompetenz (Plausible Values). LE = leistungsbezogene Einstellungen. * $p < .05$.

^a Prozent der Kategorie „nie oder selten“ bzw. Globalurteil: „gar nicht außerhalb der Schule“.

^b Prozent der Kategorie „täglich“ bzw. Globalurteil: „mehr als 2 Stunden“.

Faktorenanalytische Betrachtungen der NEPS-Daten

Aufgrund der erwartungswidrigen Befunde und um zu überprüfen, ob sich die vergleichsweise hohe Anzahl textspezifischer Urteile der NEPS-Studie sinnvoll reduzieren beziehungsweise zusammenfassen lässt und ob sich basierend auf diesen Ergebnissen Zusammenhänge zu den Kriteriumsvariablen zeigen, wurde eine explorative Faktorenanalyse für ordinale Daten durchgeführt. Tabelle 4 gibt einen Überblick über die Modell-Fit-Werte der explorati-

ven Faktorenmodelle. Keines der Modelle zeigt eine absolut zufriedenstellende Modellpassung, wobei sich für das 2- und das 3-Faktorenmodell eine leichte Verbesserung der Fit-Werte verzeichnen lässt. Eine Betrachtung des Eigenwertverlaufs (vgl. Elektronisches Supplement 1) basierend auf der Stichproben-Korrelationsmatrix zeigt, dass die ersten zwei beziehungsweise drei Faktoren den Großteil der Gesamtstreuung in den Daten erklären. Hier zeigt sich also auch eine leichte Tendenz zum 2-Faktorenmodell.

Tabelle 4. Vergleich der Modellfit-Werte der Explorative Faktorenanalyse

	Klasse 5				Klasse 9			
	χ^2	df	RMSEA	CFI	χ^2	df	RMSEA	CFI
1-Faktor	756.96*	27	.08	.88	2318.72*	27	.08	.83
2-Faktoren	385.38*	19	.07	.94	998.49*	19	.06	.93
3-Faktoren	211.15*	12	.07	.97	379.74*	12	.05	.97
4-Faktoren	26.71*	6	.03	1.00	a			

Anmerkungen: a = 4-Faktorenlösung konvergiert nicht. Rotationsmethode: oblimin.

In Tabelle 5 sind daher die Faktorladungen sowie die Interkorrelationen der latenten Faktoren für die 2-Faktorenlösung abgetragen. Die interne Konsistenz des ersten Faktors in Klassenstufe fünf, bestehend aus allen Zeitschriften und Zeitungen, beträgt .68. Die interne Konsistenz des ersten Faktors in Klassenstufe neun, bestehend aus dem Lesen von Tageszeitungen und Boulevardzeitungen, beträgt .50. Faktor zwei in Klassenstufe 5, bestehend aus dem Lesen klassischer Bücher und Sachbücher, zeigt eine interne Konsistenz von .55. Die interne Konsistenz des Faktors 2 in Klassenstufe neun beträgt .53 und beinhaltet alle Lektürearten, außer der Tageszeitung, Boulevardzeitung und Jugendseiten. Das elektronische Supplement 2 zeigt die Ergebnisse der 3-Faktorenlösung, wobei sich im

Sinne der Einfachinterpretation der Faktorladungen keine deutliche Verbesserung zeigt.

Um zu analysieren, ob die mittels der explorativen Faktorenanalyse gebildeten Faktoren mit den Kriteriumsvariablen sowie dem Globalurteil zumindest stärker korrelieren als die individuellen Einzelitems wurden weitere Korrelationsanalysen durchgeführt. Hierzu wurde die 2-Faktorenlösung basierend auf der explorativen Faktorenanalyse verwendet (siehe Elektronisches Supplement 3). In beiden Klassenstufen korrelieren die zwei Faktoren sehr gering (Klasse 9: $r = .06$ und $.08$) und in Klasse fünf teilweise sogar negativ (Klasse 5: $r = .00$ und $-.05$) mit der *Lesekompetenz*. Positive, im Verhältnis zum *Globalurteil* jedoch geringere Korrelationen zeigen sich mit den *lesebezogenen Einstellungen* ($r = .10$ bis $.35$).

Tabelle 5. Matrix des Zwei-Faktoren-Modell in Klassenstufe fünf und neun

Item	Stufe 5		Stufe 9	
	Faktor 1	Faktor 2	Faktor 1	Faktor 2
Lokale Tageszeitung	.670*		.398*	.313*
Boulevardzeitung	.819*		.880*	
Jugendseiten in anderen Zeitungen	.430*	.373*		.510*
Magazine	.404*	.380*		.522*
andere Jugendzeitschriften	.417*			
Krimis, Thriller, Horror- oder Fantasybücher				.492*
klassische Literatur		.691*		.658*
Sachbücher		.657*		.684*
Comics				.366*
Cronbachs α	.68	.55	.50	.53
Korrelation latenter Faktoren:				
Faktor 1		.54*		.31*

Anmerkungen: Rotationsmethode: Oblimin. Kleine Koeffizienten unter .30 wurden in der Tabelle nicht dargestellt. * = $p < .05$.

Diskussion

Insgesamt zeigt sich in den NEPS-Daten, dass die Kriteriumsvariablen durchweg stärker, und damit in dem Ausmaß von bisher gefundenen Ergebnissen mit dem *Globalurteil* als mit den textsortenspezifischen Urteilen oder den

zwei Faktoren zusammenhängen. Jedoch ist anzumerken, dass die Korrelationen der *Lesekompetenz* mit dem *Globalurteil* insgesamt auch schwächer ausfielen als erwartet. Die Werte liegen allerdings nur leicht unter den eingangs definierten Annahmen bezüglich der Testwertinterpretation. Die Zusammenhänge mit der zweiten Kriteriumsvariable, den *lesebezogenen Einstellungen*, entsprechen den

Annahmen. So kann insgesamt gesehen davon ausgegangen werden, dass eine valide Messung beziehungsweise eine valide Testwertinterpretation im Sinne der Erfassung individueller Unterschiede im Lesevolumen für das *Globalurteil* möglich ist. Im Vergleich zum *Globalurteil* zeigen sich bei den textsortenspezifischen Urteilen erwartungswidrige Zusammenhänge. Über alle Textsorten hinweg zeigen sich zumeist negative bis marginale positive Zusammenhänge des Lesevolumens mit der *Lesekompetenz*. Einzige Ausnahme bildet hier in Klasse neun das Lesen der Textsorten Krimis, Thriller, Horror- oder Fantasybücher und lokale Tageszeitungen, welche relativ deutlich positiv mit der Lesekompetenz korrelieren. Auch die Zusammenhänge des textsortenspezifischen Urteils mit den *lesebezogenen Einstellungen* fallen, besonders in Klasse fünf, nicht wie erwartet aus. Die deutlichen Bodeneffekte weisen ebenfalls darauf hin, dass die Antworten der in der NEPS-Studie verwendeten textsortenspezifischen Skala nicht unkritisch zu betrachten und daher nur bedingt aussagekräftig sind. Auch diese Ergebnisse bestärken eher den Einsatz des *Globalurteils*. Ferner konnte die von NEPS intendierte Unterteilung der Textsorten in Buchlesen und Zeitschriften beziehungsweise Zeitungen in einer faktoranalytischen Betrachtung (2-Faktoren-Modell) nicht bestätigt werden. Auch andere Faktorenmodelle zeigten in einer explorativen Faktorenanalyse unzureichende Ergebnisse auf. Stattdessen zeigte sich eine Heterogenität der Items, die z. B. in Form einer geringen Reliabilität der gebildeten Faktoren beziehungsweise Skalen für weitere Analysen problematisch sein kann.

Dies soll allerdings nicht bedeuten, dass eine differenziertere Abfrage nicht grundsätzlich sinnvoll sein kann. Es stellt sich jedoch die Frage, ab welchem Alter, wie differenziert und vor allem welche Textsorten abgefragt werden sollten, um optimal verwertbare Informationen zu erhalten. Die Ergebnisse aus NEPS lassen vermuten, dass es bei Fragebogenskalen zum Lesevolumen, bei welchen stark zwischen den einzelnen Textsorten unterschieden wird, herausfordernd ist für die Befragten diese Kategorien zu differenzieren und damit präzise zu antworten. Im Vergleich der Klassenstufen fünf und neun der NEPS-Daten zeigt sich ein deskriptiver Trend dahingehend, dass mit zunehmendem Alter differenziertere Urteile abgegeben werden können. Im Vergleich der beiden Stichproben 15-Jähriger aus NEPS und PISA zeigt sich jedoch trotzdem, dass die in PISA verwendete und weniger stark ausdifferenzierte Fragebogenskala, eher den theoretischen und empirischen Erwartungen und vorab bestimmten Prämissen zur Testwertinterpretation entspricht. Für die Beantwortung der Fragebogenskalen stehen die Schülerinnen und Schüler vor der kognitiven Herausforderung, zum einen retrospektiv einzuschätzen, wieviel sie

durchschnittlich am Tag lesen. Zum anderen müssen sie aber auch ihre Tätigkeiten genau aufschlüsseln, welcher Textsorte ihr Buch zuzuordnen ist. Bei breiteren Kategorien wie etwa in der PISA-Studie ist eine solche Einschätzung möglicherweise einfacher abzugeben als bei feineren Dimensionen. Im vorliegenden Daten-Beispiel aus NEPS wurden besonders Periodika sehr genau unterschieden, wie beispielsweise Jugendzeitschriften oder Jugendseiten in anderen Zeitschriften. Es dürfte jedoch (insbesondere für jüngere Schülerinnen und Schüler) nicht eindeutig erkennbar sein, wo hier die Grenze liegt. Dies könnte mitunter eine Ursache für die in den NEPS-Stichproben gefundenen Ergebnisse (deutliche Bodeneffekte, geringe Korrelationen mit den Kriteriumsmaßen und eine praktisch kaum psychometrisch sinnvolle Zusammenfassung der einzelnen textspezifischen Items zu Skalen) sein. Es empfiehlt sich daher bei der Erfassung des Lesevolumens eine Reduktion der Dimensionen für eine klarere Differenzierung. Ferner ist zu diskutieren, ob Fünftklässlerinnen und Fünftklässler überhaupt Boulevardzeitungen oder auch lokale Tageszeitungen lesen. In der PISA Studie 2009 wurde, wie in diesem Beitrag gezeigt werden konnte, das Lesevolumen vergleichsweise erfolgreich für spezifische Textsorten erfasst. Eine gelungene Umsetzung des differenzierten Urteils findet sich auch in Studien von Spear-Swerling et al. (2010) oder Pfof et al. (2013). Dabei wurden im Vergleich zu NEPS deutlich weniger und altersangemessenere Facetten unterschieden. Dies bestärkt weiterhin die Vermutung, dass eine differenzierte Erfassung des Lesevolumens dann sinnvoll und valide sein kann, wenn für die Adressaten eine logische Abgrenzung der Textsorten möglich ist und die entsprechenden Antwortskalen im Sinne der Vermeidung von Bodeneffekten dies auch zulassen. Urteile mit hoher Spezifität, insbesondere bei Kindern und Jugendlichen, erscheinen dagegen kritisch in der Anwendung.

Limitationen und Ausblick

Einschränkend für die vorliegende Untersuchung muss angemerkt werden, dass wir keine Möglichkeit hatten, die differenzierten Urteile und das *Globalurteil* mit dem realen Lesevolumen abzugleichen. Für zukünftige Forschung könnten hier als Ergänzung zusätzliche Lesetagebücher dienen. Darüber hinaus wäre es von Vorteil Methoden heranzuziehen, die nicht exklusiv auf einer Selbsteinschätzung beruhen. Aspekte wie soziale Erwünschtheit bei den Kindern und Jugendlichen oder Erinnerungseffekte z. B. dadurch, dass ein Buch einer bestimmten Kategorie seit längerer Zeit gelesen wird und daher eher präsent ist als andere Bücher, könnten dadurch berücksichtigt werden. Ferner wäre es wünschenswert, Zusammenhänge

zwischen Subdimensionen der Lesekompetenz und dem Lesen unterschiedlicher Textsorten zu betrachten.

Die Erfassung des Lesevolumens ist eine Herausforderung, die je nach Möglichkeiten des Studiendesigns qualitativ unterschiedlich gemeistert werden kann. Schlussendlich muss abgewogen werden, welches Instrument hinsichtlich Kosten, Zeitaufwand und Analysezielen geeignet ist. Trotzdem kann die Verwendung von Globalurteilen zum Lesevolumen besonders in zeit- und platzbegrenzten Large-Scale-Studien auf Grundlage der hier durchgeführten Studie gestärkt werden. In der direkten Gegenüberstellung von Globalurteil und textsortenspezifischem Urteil scheint im vorliegenden Fall das in NEPS eingesetzte Globalurteil das geeignetere Messinstrument zur Erfassung individueller Unterschiede im Lesevolumen zu sein. Zumindest im Hinblick auf die jüngere Startkohorte würden wir daher anderen Datennutzerinnen und -nutzern des NEPS-Scientific-Use-Files von einer Nutzung der spezifischen Einschätzungen des Lesevolumens in ihren Analysen abraten.

Jenseits dieser unmittelbaren Handlungsempfehlungen für Nutzerinnen und Nutzern des NEPS-Scientific-Use-File ergeben sich zwei allgemeine Schlussfolgerungen aus den vorgestellten Analysen im Hinblick auf die Erfassung des Lesevolumens. Zum einen sollten die Stufen der Antwortskalen überprüft werden. Für manche Urteile sind gewisse Antwortmöglichkeiten weniger plausibel als für andere. Im konkreten Beispiel von PISA und NEPS verwenden zwar beide Studien eine fünfstufige Skala, jedoch ist jene aus PISA niedrigschwelliger als jene aus NEPS (z. B. PISA: Kategorie 5 = *Mehrmals in der Woche*; NEPS: Kategorie 5 = *täglich*). Sich wirklich täglich mit einer spezifischen Textsorte wie einem Sachbuch oder einer Jugendzeitschrift zu beschäftigen, scheint aber (und darauf weisen auch die Itemverteilungen hin) unrealistisch zu sein. Diese ungleichen Kategorienbeschriftungen könnten mitunter ein Grund für die teilweise deutlichen Unterschiede in den Ergebnissen von PISA und NEPS sein. Obgleich die erste Kategorie (*nie oder selten* bzw. *fast nie*) bei beiden Studien gleich ist, geben in NEPS in Klassenstufe neun ungefähr doppelt so viele Jugendliche an, nie Sachbücher zu lesen. Betrachtet man aber die zweite Kategorie (PISA: *Ein paarmal im Jahr*; NEPS: *Mehrmals im Monat*) können die Bodeneffekte in NEPS auch daraus resultieren, dass der Schritt von Kategorie eins zu zwei zu groß ist und für die Jugendlichen die Antwortkategorie eins zutreffender erscheint. Zum zweiten sollte, wie bereits erwähnt, die Feingliedrigkeit der abgefragten Textsorten diskutiert werden. Eine zu große Differenzierung scheint kontraproduktiv. Was für die weitere Anwendung von Lesevolumenskalen zusammen-

fassend festgehalten werden kann, ist, dass das Globalurteil zur Erfassung des Lesevolumens, bei Limitationen bezüglich Zeit und Umfang, prinzipiell ein akzeptables Instrument darstellt. Ein differenziertes textsortenspezifisches Urteil sollte im Optimalfall jedoch zur Erfassung des Lesevolumens hinzugezogen werden; allerdings nur, wenn diese (a) plausible sowie adäquat bewertbare, das heißt tatsächlich die vorhandene Merkmalsvariation abbildende, Antwortkategorien umfassen und (b) sinnvoll abgrenzbare und altersangemessene Textsorten enthalten. Aufgrund der Betrachtung modellbasierter Itemtrennschärfen in den Faktorenanalysen, der Häufigkeitsverteilungen und im Sinne einer auch für jüngere Personen einfach unterscheidbaren Textsortenauswahl, würden wir beispielsweise eine Unterteilung in Unterhaltungsliteratur (z. B. Romane, Thriller, Krimis, etc.), Sach- und Fachbücher sowie Zeitungen und Zeitschriften empfehlen.

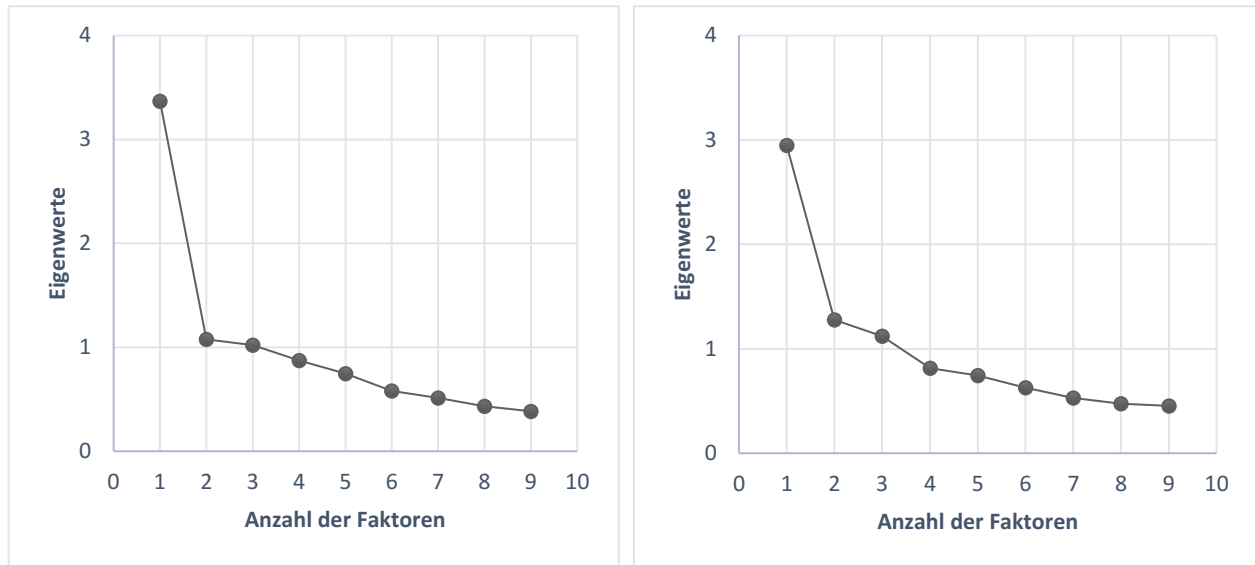
Literatur

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Anderson, R. C., Wilson, P. T. & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly*, 23, 285–303.
- Cipielewski, J. & Stanovich, K. E. (1992). Predicting growth in reading ability from children's exposure to print. *Journal of Experimental Child Psychology*, 54, 74–89. doi: 10.1016/0022-0965(92)90018-2.
- Gehrer, K. & Artelt, C. (2013). Literalität und Bildungslaufbahn: Das Bildungspanel NEPS. In C. Rosebrock & A. Bertschi-Kaufmann (Hrsg.), *Literalität erfassen: bildungspolitisch, kulturell, individuell* (S. 168–187). Weinheim: Beltz Juventa.
- Greaney, V. & Hegarty, M. (1987). Correlates of leisure-time reading. *Journal of Research in Reading*, 10, 3–20. doi: 10.1111/j.1467-9817.1987.tb00278.x.
- Guthrie, J. T., Wigfield, A., Metsala, J. L. & Cox, K. E. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading*, 3, 231–256. doi: 10.1207/s1532799xssr0303_3.
- Haberkorn, K., Pohl, S., Hardt, K. & Wiegand, E. (2012). *NEPS Technical Report for Reading – Scaling Results of Starting Cohort 4 in Ninth Grade (NEPS Working Paper No. 16)*. Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & K. Kelava (Hrsg.), *Testtheorie und*

- Fragebogenkonstruktion* (S. 135–163). Berlin Heidelberg: Springer-Verlag.
- Hayes, D. P. & Ahrens, M. G. (1988). Vocabulary simplification for children: A special case of ‘motherese’?. *Journal of Child Language*, *15*, 395–410. doi: 10.1017/S0305000900012411.
- Hertel, S., Hochweber, J., Mildner, D., Steinert, B. & Jude, N. (2014). *PISA 2009 Skalenhandbuch*. Münster; New York: Waxmann.
- McGeown, S. P., Osborne, C., Warhurst, A., Norgate, R. & Duncan, L. G. (2016). Understanding children's reading activities: Reading motivation, skill and child characteristics as predictors. *Journal of Research in Reading*, *39*, 109–125. doi: 10.1111/1467-9817.12060.
- Mol, S. E. & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*, 267–296. doi: 10.1037/a0021890.
- Muthén, L. K. & Muthén, B. O. (1998–2007). *Mplus user's guide* (5. Aufl.). Los Angeles: Muthén & Muthén.
- Naumann, J., Artelt, C., Schneider, W. & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23–71). Münster/New York/ München/ Berlin: Waxmann.
- Organisation for Economic Cooperation and Development. (2012). *PISA 2009 Technical Report*. PISA, OECD Publishing.
- Pfof, M., Dörfler, T. & Artelt, C. (2010). Der Zusammenhang zwischen außerschulischem Lesen und Lesekompetenz. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *42*, 167–176. doi: 10.1026/0049-8637/a000017.
- Pfof, M., Dörfler, T. & Artelt, C. (2013). Students' extracurricular reading behavior and the development of vocabulary and reading comprehension. *Learning and Individual Differences*, *26*, 89–102. doi: 10.1016/j.lindif.2013.04.008.
- Pohl, S., Haberkorn, K., Hardt, K. & Wiegand, E. (2012). *NEPS Technical Report for Reading – Scaling Results of Starting Cohort 3 in Fifth Grade (NEPS Working Paper No. 15)*. Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.
- Schaffner, E. (2009). Determinanten des Leseverstehens. In W. L. W. Schneider (Hrsg.), *Diagnostik und Förderung des Leseverständnisses* (S. 19–44). Göttingen: Hogrefe.
- Schroeder, S., Segbers, J. & Schröter, P. (2016). Der Kinder-Titelreognitionstest (K-TRT). *Diagnostica*, *62*, 16–30. doi: 10.1026/0012-1924/a000131.
- Spear-Swerling, L., Brucker, P. O. & Alfano, M. P. (2010). Relationships between sixth-graders' reading comprehension and two different measures of print exposure. *Reading and Writing*, *23*, 73–96. doi: 10.1007/s11145-008-9152-8.
- Walberg, H. J. & Tsai, S.-I. (1984). Reading achievement and diminishing returns to time. *Journal of Educational Psychology*, *76*, 442–451. doi: 10.1037/0022-0663.76.3.442.

Elektronische Supplemente (ESM)

ESM 1. Die Abbildung zeigt den Eigenwertverlauf der exploratorischen Faktorenanalyse



Elektronisches Supplement 1. Eigenwertverlauf basierend auf der Stichproben-Korrelationsmatrix der Explorative Faktorenanalyse in Startkohorte 3 (links) und Startkohorte 4 (rechts)

ESM 2. Die Tabelle zeigt die Ergebnisse der 3-Faktorenlösung.

Elektronisches Supplement 2. Matrix des Drei-Faktoren-Modells in Klassenstufe fünf und neun

Item	Stufe 5			Stufe 9		
	Faktor 1	Faktor 2	Faktor 3	Faktor 1	Faktor 2	Faktor 3
Lokale Tageszeitung	.646*			.572*		
Boulevardzeitung (wie BILD, BZ)	.801*			.750*		
Jugendseiten in anderen Zeitungen (wie SZ, FAZ)	.386	.503*		.418*	.394*	
Magazine (wie SPIEGEL, FOCUS Schule)	.402*			.359*	.418*	
andere Jugendzeitschriften (wie Bravo, Popcorn)	.425*					.850*
Krimis, Thriller, Horror- oder Fantasybücher	.315*				.518*	
klassische Literatur			.524*		.652*	
Sachbücher			.805*		.665*	
Comics	.317*				.373*	
Cronbachs α	.59	-	.55	.52	.51	-
Korrelation latenter Faktoren:						
Faktor 1		.36*	.51*		.38*	.18*
Faktor 2			.32			.04*

Anmerkungen: Rotationsmethode: oblimin. Kleine Koeffizienten unter .30 wurden in der Tabelle nicht dargestellt.

Die interne Konsistenz in Klassenstufe fünf wurde berechnet auf Basis folgender Items: Faktor 1: Tageszeitungen, Boulevardzeitungen, Jugendseiten, Magazine und Jugendzeitschriften; Faktor 2: klassische Literatur und Sachbücher. Die interne Konsistenz in Klassenstufe neun wurde berechnet auf Basis folgender Items: Faktor 1: Tageszeitungen; Boulevardzeitung; Faktor 2: Jugendseiten, Magazine, Krimis, Thriller, Horror- oder Fantasybücher, klassische Literatur, Sachbücher und Comics. * = $p < .05$.

ESM 3. Die Tabelle zeigt die Korrelationen mit den gebildeten Lesevolumenfaktoren.

Elektronisches Supplement 3. Kendall-Tau-Interkorrelationen der Skalen und Kriteriumsvariablen von Klassenstufe fünf (unter der Diagonale) und Klassenstufe neun (über der Diagonale)

Variablen	1	2	3	4	5
1. Faktor 1	–	.24*	.06*	.10*	.04*
2. Faktor 2	.27*	–	.08*	.35*	.27*
3. Lesekompetenz	-.05*	.00	–	.23*	.24*
4. Lesebezogene Einstellungen	.18*	.29*	.15*	–	.44*
5. Globalurteil	.12*	.20*	.20*	.44*	–

Anmerkungen: Die geringe Reliabilität der gebildeten Faktoren/Skalen ist einschränkend zu berücksichtigen. Korrelationen basieren auf manifesten Variablen. Die Faktoren in Klassenstufe fünf wurden aus den Mittelwerten folgender Items gebildet: Faktor 1: Tageszeitungen, Boulevardzeitungen, Jugendseiten, Magazine und Jugendzeitschriften; Faktor 2: klassische Literatur und Sachbücher. Die Faktoren in Klassenstufe neun wurde gebildet aus den Mittelwerten folgender Items: Faktor 1: Tageszeitungen; Boulevardzeitung; Faktor 2: Jugendseiten, Magazine, Krimis, Thriller, Horror- oder Fantasybücher, klassische Literatur, Sachbücher und Comics. * $p < .05$.