

Schoor, Cornelia; Hahnel, Carolin; Artelt, Cordula; Reimann, Daniel; Kroehne, Ulf; Goldhammer, Frank
**Entwicklung und Skalierung eines Tests zur Erfassung des Verständnisses
multipler Dokumente von Studierenden**

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:

formally and content revised edition of the original source in:

Diagnostica 66 (2020) 2, S. 123-135



Bitte verwenden Sie in der Quellenangabe folgende URN oder DOI /

Please use the following URN or DOI for reference:

urn:nbn:de:0111-pedocs-218434

10.25656/01:21843

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-218434>

<https://doi.org/10.25656/01:21843>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz:
<http://creativecommons.org/licenses/by-nc/4.0/deed.de> - Sie dürfen das
Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich
machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes
anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm
festgelegten Weise nennen und das Werk bzw. den Inhalt nicht für
kommerzielle Zwecke verwenden.

Mit der Verwendung dieses Dokuments erkennen Sie die
Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License:
<http://creativecommons.org/licenses/by-nc/4.0/deed.en> - You may copy,
distribute and render this document accessible, make adaptations of this work
or its contents accessible to the public as long as you attribute the work in the
manner specified by the author or licensor. You are not allowed to make
commercial use of the work, provided that the work or its contents are not
used for commercial purposes.

By using this particular document, you accept the above-stated conditions of
use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Akzeptierte Manuskriptfassung (nach peer review) des folgenden Artikels:

Schoor, C., Hahnel, C., Artelt, C., Reimann, D., Kröhne, U. & Goldhammer, F. (2020). Entwicklung und Skalierung eines Tests zur Erfassung des Verständnisses multipler Dokumente von Studierenden. *Diagnostica*, 66, 123-135. <https://doi.org/10.1026/0012-1924/a000231>

© Hogrefe Verlag, Göttingen 2020

Diese Artikelfassung entspricht nicht vollständig dem in der Zeitschrift veröffentlichten Artikel. Dies ist nicht die Originalversion des Artikels und kann daher nicht zur Zitierung herangezogen werden.

Die akzeptierte Manuskriptfassung unterliegt der Creative Commons License CC-BY-NC.

Diagnostica

Entwicklung und Skalierung eines Tests zur Erfassung des Verständnisses multipler Dokumente von Studierenden Developing and Scaling a Test of Multiple Document Comprehension in University Students --Manuskript-Entwurf--

Manuskriptnummer:	DIA-D-17-00031R1
Vollständiger Titel:	Entwicklung und Skalierung eines Tests zur Erfassung des Verständnisses multipler Dokumente von Studierenden Developing and Scaling a Test of Multiple Document Comprehension in University Students
Artikeltyp:	Originalarbeit
Schlüsselwörter:	Verständnis multipler Dokumente; Studierende; Test; Multiple Document Literacy multiple document comprehension; university students; assessment; multiple document literacy
Korrespond. Autor:	Cornelia Schoor, Dr. Otto-Friedrich-Universität Bamberg Bamberg, GERMANY
Korrespondierender Autor, Zweitinformationen:	
Korrespondierender Autor, Institution:	Otto-Friedrich-Universität Bamberg
Korrespondierender Autor, zweite Institution:	
Erstautor:	Cornelia Schoor, Dr.
Erstautor, Zweitinformationen:	
Reihenfolge der Autoren:	Cornelia Schoor, Dr. Carolin Hahnel, Dr. Cordula Artelt, Prof. Dr. Daniel Reimann Ulf Kröhne, Dr. Frank Goldhammer, Prof. Dr.
Reihenfolge 'Zweite Informationen' von Autoren:	
Zusammenfassung:	<p>Das Verständnis multipler Dokumente (Multiple Document Comprehension, MDC) wird als Fähigkeit verstanden, aus verschiedenen Informationsquellen eine integrierte Repräsentation eines inhaltlichen Gegenstandsbereichs zu konstruieren. Als solche ist sie sowohl für die erfolgreiche Bewältigung eines Studiums als auch für gesellschaftliche Partizipation eine wichtige Kompetenz. Bisher gibt es jedoch kein etabliertes Diagnostikum in diesem Bereich. Um diese Lücke zu schließen, wurde ein Test entwickelt, der vier zentrale kognitive Anforderungen von MDC abdeckt und auf Basis der Daten von 310 Studierenden sozial- und geisteswissenschaftlicher Fächer überprüft wurde. Die im MDC-Test gemessene Kompetenz erwies sich als eindimensional. Der MDC-Testwert wies theoriekonforme Zusammenhänge mit der Abiturnote, dem Studienabschnitt und der Leistung in einer Essay-Aufgabe auf. Insgesamt liefern die Ergebnisse empirische Belege dafür, dass der Testwert aus dem MDC-Test die fächerübergreifende Fähigkeit von Studierenden wiedergibt, multiple Dokumente zu verstehen.</p> <p>Multiple document comprehension (MDC) is defined as the ability to construct an integrated representation based on different sources of information on a particular topic. It is an important competence for both the successful accomplishment of university studies and participation in societal discussions. Yet, there is no established</p>

assessment instrument for MDC. Therefore, we have developed a test covering four theory-based cognitive requirements of multiple document comprehension. Based on data of 310 university students of social sciences and humanities, the MDC test proved to be a unidimensional measure. Furthermore, the test score was related to the final school exam grade, the study level (bachelor/master), and the performance in an essay task. All in all, the empirical results suggest that the score of the MDC test can be interpreted as the generic competence of university students to understand multiple documents.

1 Um den Anforderungen im Studium und im späteren Beruf gerecht zu werden, müssen
2 Studierende aller Fächer in der Lage sein, sich selbständig in verschiedene Themenbereiche
3 einzuarbeiten und mit wesentlichen Kernaspekten zielgerichtet auseinanderzusetzen.
4 Insbesondere im wissenschaftlichen Kontext der Hochschule werden Studierende dabei mit
5 mehreren Quellen und somit (Text-)Dokumenten konfrontiert (z. B. als Ergebnis einer
6 Datenbank- oder Internetrecherche), aus denen sie relevante Informationen herausfiltern und
7 miteinander integrieren müssen (v. a. wenn Verfasseransichten sich widersprechen). Diese
8 Anforderungen gehen über das Verstehen einzelner Texte hinaus, da Informationen aus
9 verschiedenen Quellen zueinander in Bezug gesetzt und hierfür gegebenenfalls Quellen
10 (z. B. hinsichtlich ihrer Glaubwürdigkeit) bewertet werden müssen.
11
12
13
14
15
16
17
18
19
20
21
22
23

24 Nach den Bildungsstandards für das Abitur (Kultusministerkonferenz, 2012) sollten
25 bereits Schulabsolventen in der Lage sein – hier bezogen auf das Fach Deutsch –
26 „vielschichtige, voraussetzungsreiche Sachtexte, darunter auch wissenschaftliche und
27 berufsbezogene Fachtexte aus unterschiedlichen Domänen [für den eigenen
28 Erkenntnisprozess nutzen zu können und] Gehalt, Kontext und Wirkungsabsicht
29 pragmatischer Texte [zu erschließen, zu analysieren und zu beurteilen]“
30 (Kultusministerkonferenz, 2012, S. 19). Aus den allgemeinen Darstellungen des
31 Qualifikationsprofils von Hochschulabsolventinnen und Hochschulabsolventen geht hervor,
32 dass diese fähig sein sollen, „relevante Informationen, insbesondere in ihrem
33 Studienprogramm zu sammeln, zu bewerten und zu interpretieren [sowie] daraus
34 wissenschaftlich fundierte Urteile abzuleiten, die gesellschaftliche, wissenschaftliche und
35 ethische Erkenntnisse berücksichtigen“ (Arbeitskreis Deutscher Qualifikationsrahmen, 2011,
36 S. 16) sowie als Master-Absolventinnen und -Absolventen „Wissen zu integrieren und mit
37 Komplexität umzugehen [und] selbständig sich neues Wissen und Können anzueignen“
38 (Arbeitskreis Deutscher Qualifikationsrahmen, 2011, S. 17).
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Die beiden Anforderungsprofile verweisen auf das Sammeln, Bewerten und Integrieren
2 von Informationen, die im Wissenschaftskontext meistens aus verschiedenen Quellen
3 stammen. Wir bezeichnen diese Fähigkeit als *Multiple Document Comprehension* (MDC)
4 und definieren sie als die Fähigkeit, aus verschiedenen Informationsquellen eine integrierte
5 Repräsentation eines bestimmten inhaltlichen Gegenstandsbereichs oder Sachverhalts zu
6 konstruieren (vgl. Anmarkrud, Bråten & Strømsø, 2014). Es kann daher von einer
7 Kompetenz im Sinne einer kontextspezifischen kognitiven Leistungsdisposition (Weinert,
8 1999, 2001) gesprochen werden.
9

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Obwohl MDC eine Kompetenz ist, die Studierende unabhängig vom Studienfach für ihr Studium benötigen und spätestens im Verlauf des Studiums entwickeln sollten, gibt es bisher kein Instrument, das diese Kompetenz bei Studierenden angemessen und verschiedene theoretische Facetten abdeckend erfasst. Mit dem Ziel, diese Lücke zu schließen, wurde ein computerbasierter Test zur Erfassung des Verständnisses multipler Dokumente für Studierende entwickelt. Die dem Test zugrunde liegenden multiplen Dokumente stellen dabei im Grundsatz glaubwürdige Texte dar und behandeln Inhalte, bezüglich deren Studierende über wenig inhaltliches Vorwissen verfügen. Diese Anforderungssituation ist unseres Erachtens typisch für universitäres Lernen mit vorgegebenen Texten. Im Test nicht abgebildet werden damit Kompetenzen, die sich auf die eigenständige Suche und Auswahl von Dokumenten beziehen und die den Umgang mit Inhalten umfassen, die den eigenen Vorstellungen widersprechen. Das Grundprinzip der Testentwicklung sowie die empirische Prüfung der intendierten Testwertinterpretation (d. h. Konstruktinterpretation) sind Gegenstand des vorliegenden Artikels.

1 Verstehen multipler Dokumente

Der Umgang mit mehreren Dokumenten stellt eine Vielzahl von Anforderungen. So müssen beispielsweise ...

1. ... verschiedene Perspektiven auf den gleichen Sachverhalt als solche wahrgenommen und in Relation gesetzt (vgl. Britt, Perfetti, Sandak & Rouet, 1999; Britt & Rouet, 2012),
2. ... Informationen über Dokumente hinweg integriert (vgl. Britt & Sommer, 2004; Cerdán & Vidal-Abarca, 2008; Gil, Bråten, Vidal-Abarca & Strømsø, 2010b),
3. ... widersprüchliche Aussagen erkannt und adäquat gehandhabt (vgl. Bråten, Ferguson, Strømsø & Anmarkrud, 2014; Keck, Kammerer & Starauschek, 2015; Maier & Richter, 2013; Stadtler & Bromme, 2014),
4. ... Verzerrungen durch Eigenschaften der Quelle (wie z. B. Intentionen des Autors, intendierte Leserschaft) erkannt und für eigene Entscheidungen gewichtet (vgl. Braasch, Rouet, Vibert & Britt, 2012; Bråten, Salmerón & Strømsø, 2016; Britt & Aglinskas, 2002; Kammerer & Gerjets, 2014; Kammerer, Kalbfell & Gerjets, 2016) sowie
5. ... Aussagen aus verschiedenen Quellen miteinander verglichen beziehungsweise abglichen werden (vgl. Rouet, Favart, Britt & Perfetti, 1997; Wineburg, 1991).

Die genannten Anforderungen sind zwar typisch für den Umgang mit mehreren Dokumenten, können aber auch beim Lesen innerhalb eines Dokuments auftreten (z. B. wenn in einem Lehrbuchtext verschiedene theoretische Perspektiven auf den gleichen Sachverhalt dargestellt werden). Prinzipiell ist MDC daher nicht formal durch das Vorhandensein multipler Dokumente definiert, sondern durch die für multiple Dokumente typischen Anforderungen. Diese treten also nicht notwendigerweise bei multiplen

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Dokumenten, wohl aber bei multiplen Quellen auf. Unter einem *Dokument* verstehen wir eine strukturelle Einheit, die Inhalte vermittelt (z. B. ein Zeitschriftenartikel, eine Webseite, ein Lehrbuchkapitel). Mit *Quelle* hingegen bezeichnen wir die Urheberin oder den Urheber von Inhalten. Beispielsweise können verschiedene Autorinnen oder Autoren (Quellen) in Zeitschriftenartikeln (Dokumenten) unterschiedliche Modelle (Inhalte) zu einem thematischen Bereich vertreten. Auch ein Lehrbuchkapitel (Dokument) kann solche Modelle (Inhalte) der jeweiligen Autorinnen oder Autoren (Quellen) aufgreifen und beschreiben. In diesem Fall hat das Lehrbuchkapitel natürlich auch eine eigene Quelle (die Kapitelautorin oder den Kapitelautor). Im Gegensatz zu multiplen Dokumenten kann von einem gut geschriebenen Einzeltext jedoch erwartet werden, dass die Leserin oder der Leser in den auftretenden Anforderungen unterstützt wird (z. B. sollten im Lehrbuchtext die Unterschiede und Gemeinsamkeiten verschiedener theoretischer Perspektiven explizit herausgearbeitet werden; vgl. Goldman & Scardamalia, 2013a). Unter Quelleninformationen verstehen wir im Folgenden (Meta-)Informationen über die Quelle (z. B. Referenz, Autorennamen und Publikationsjahr; vgl. Perfetti, Rouet & Britt, 1999).

1.1 Theoretische Ansätze: Documents Model Framework und MDC-Strategien

41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Im Documents Model Framework (z. B. Britt & Rouet, 2012) wird postuliert, dass über die beim Lesen einzelner Texte entstehenden kognitiven Repräsentationen (z. B. nach Kintsch, 1998: wörtliche Repräsentation, Textbasis, Situationsmodell) hinaus idealerweise ein sogenanntes *Documents Model* aufgebaut werden sollte. Dieses ist eine Kombination zweier mentaler Modelle: des *integrierten Situationsmodells* und des *Intertext Model*. Das integrierte Situationsmodell enthält die Integration der Inhalte aus mehreren Dokumenten. Metainformationen über die einzelnen Dokumente sowie ihre Beziehungen untereinander werden im Intertext Model repräsentiert. Unter Metainformationen sind dabei nach Perfetti

1 et al. (1999) beispielsweise Informationen über die Autorin oder den Autor, die Form und
2 die rhetorischen Ziele des Dokuments sowie Informationen über den (kulturellen, zeitlichen
3 etc.) Hintergrund zu verstehen. Zusätzlich werden zentrale Informationen aus dem
4 integrierten Situationsmodell mit der Information verbunden, aus welchem Dokument sie
5 stammen (z. B. Perfetti et al., 1999). Weniger gute alternative Repräsentationen von
6 multiplen Dokumenten schließen die getrennte Repräsentation der Dokumente (*separate*
7 *representation model*), die Vermischung aller Informationen ohne Verbindung zu
8 Quelleninformationen (*mush model*) oder die Repräsentation jeder Information mit
9 zugehöriger Quelle ein (*tag-all model*), die höchstens von Personen mit großer Expertise
10 leistbar ist (Britt et al., 1999). Für diese verschiedenen Repräsentationsformen leiteten Britt
11 et al. (1999) jeweils spezifische Vorhersagen hinsichtlich der Erinnerung an Inhalte und
12 Quellen ab. In einer Studie zu einer nicht näher beschriebenen Stichprobe legten sie 80
13 „students“ (ob es sich hierbei um Schülerinnen und Schüler oder Studierende handelte,
14 wurde nicht erläutert) jeweils zwei Texte zum US-amerikanischen Erwerb des Panama-
15 Kanals sowie Tests zur Wiedererkennung von Inhalten und Quellen vor. Auf der Basis dieser
16 Daten kommen Britt et al. (1999) zu dem Schluss, dass das Documents Model die
17 wahrscheinlichste der genannten Repräsentationsformen ist.

18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Jenseits der kognitiven Repräsentation analysierte Wineburg (1991) auf der Ebene von Strategien, inwiefern sich Expertinnen und Experten von Novizinnen und Novizen im Umgang mit multiplen Dokumenten unterscheiden. Im Gegensatz zu Novizinnen und Novizen vergleichen Expertinnen und Experten häufiger die Informationen aus mehreren Texten (*Corroboration*), setzen Informationen häufiger in Bezug zu ihrem generellen Vorwissen beziehungsweise zu dem, was sie über den Kontext der Entstehung der Dokumente wissen (*Contextualization*) und beziehen häufiger Informationen über die Quelle in ihre Überlegungen mit ein (*Sourcing*).

1.2 Besonderheiten von MDC im Hochschulkontext

Ein Schwerpunkt bisheriger Forschung zu MDC betrifft die Nutzung und Bewertung von Quelleninformationen (z. B. Braasch & Bråten, 2017; Bråten et al., 2016; Britt & Aglinskas, 2002; Kammerer & Gerjets, 2014; Paul, Macedo-Rouet, Rouet & Stadtler, 2017; Scharrer & Salmerón, 2016; Strømsø, Bråten, Britt & Ferguson, 2013). Diese Komponente von MDC spielt vor allem bei der Informationssuche im Internet eine zentrale Rolle, wo Dokumente verschiedenster Provenienz und Qualität auffindbar sind. Im universitären Kontext hingegen begegnen Studierende multiplen Dokumenten hauptsächlich nach Literatursuchen in wissenschaftlichen Datenbanken oder durch Literaturvorgaben von Dozierenden. Die Bewertung der Glaubwürdigkeit von Quellen spielt hierbei eher eine untergeordnete Rolle. Die Kompetenz zur Suche und Auswahl von Dokumenten sehen wir als eine eigenständige, von MDC abgrenzbare Kompetenz an (Informationskompetenz, vgl. Deutscher Bibliotheksverband, 2009; Homann, 2000; Lau, 2006). Im entwickelten Test liegt der Fokus auf dem Umgang mit – nicht auf dem Finden von – multiplen Dokumenten.

Auch wenn multiple Dokumente in den unterschiedlichen Fachdisziplinen unterschiedliche Charakteristika haben, so lassen sich doch Gemeinsamkeiten über Fachdisziplinen hinweg feststellen (Goldman et al., 2016), die rechtfertigen, MDC als fächerübergreifende Kompetenz zu betrachten. Der entwickelte Test soll diese Gemeinsamkeiten unabhängig vom Studienfach und vom Vorwissen erfassen. Weiterhin kann davon ausgegangen werden, dass MDC als Kompetenz trainier- und damit veränderbar ist (Britt & Aglinskas, 2002). Auch wenn ein Studium kein explizites Training darstellt und der Umgang mit multiplen Dokumenten in den einzelnen Fächern vermutlich unterschiedlich stark explizit thematisiert wird, gibt es Hinweise darauf, dass sich MDC als Kompetenz im

1 Verlauf eines Studiums positiv entwickelt (vgl. für Unterschiede zwischen Erstsemestern
2 und [Post-] Doktoranden in Sourcing: Mühlen, Richter, Schmid, Schmidt & Berthold, 2016).
3
4
5
6

7 **1.3 Bisherige Ansätze zur Erfassung des Verständnisses multipler Dokumente**

8

9 Da die Untersuchung des Textverständnisses im Umgang mit multiplen Dokumenten ein
10 vergleichsweise junges Forschungsgebiet ist (vgl. z. B. Maier & Richter, 2014), gibt es
11 bislang kaum Ansätze zur standardisierten und auf psychometrischen Modellen aufbauenden
12 Kompetenzmessung. In bisherigen Studien wurde das Verständnis multipler Dokumente
13 häufig in Form von Essays erfasst, die von den Versuchspersonen zum Textmaterial verfasst
14 wurden (z. B. Britt & Aglinskas, 2002; Kammerer & Gerjets, 2014; Rouet et al., 1997;
15 Stadtler, Scharrer, Skodzik & Bromme, 2014). Diese Methode ist – zumindest in der
16 bisherigen Form, bei der die Essays von menschlichen Beurteilern kodiert werden müssen –
17 sehr zeitaufwändig. Zudem werden neben den primär interessierenden rezeptiven
18 Sprachkompetenzen (MDC) auch produktive Fähigkeiten (Schreibproduktion) erfasst,
19 wodurch Verzerrungen in der spezifischen Messung von MDC als separate Kompetenz nicht
20 ausgeschlossen werden können.
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38

39 Ein weiterer Ansatz zur Erfassung des Verständnisses multipler Texte besteht in der
40 Anwendung von intertextuellen Inferenz-Verifikationsaufgaben (z. B. Braasch, Bräten,
41 Strømsø & Anmarkrud, 2014; Salmerón, Gil, Bräten & Strømsø, 2010; vgl. auch Maier &
42 Richter, 2013; Schmalhofer & Glavanov, 1986). Bei diesem Aufgabentyp wird das
43 integrierte Situationsmodell durch die Beurteilung des Wahrheitsgehalts von Aussagen
44 operationalisiert, für deren Bewertung Inferenzen über mehrere Texte hinweg gezogen
45 werden müssen. Das Aufgabenformat erfordert eine einfache Ja-Nein-Entscheidung, die sich
46 ökonomisch und objektiv auswerten lässt. Mit diesem Ansatz wird jedoch nur ein Teilaspekt
47 des Verständnisses multipler Dokumente (das integrierte Situationsmodell, nicht jedoch das
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Intertext Model oder das Documents Model in seiner Gesamtheit) abgebildet. Jenseits von
2 Essays oder intertextuellen Inferenz-Verifikationsaufgaben ist ein umfassender, spezifisch
3 die rezeptive Kompetenz des Verständnisses multipler Dokumente erfassender Ansatz
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

2 Rahmenkonzeption eines Tests zur Erfassung von Multiple Document Comprehension

Der Aufgabenentwicklung für einen Test zur Erfassung von MDC wurden die theoretischen Ansätze des Documents Model Framework (z. B. Britt & Rouet, 2012) sowie die von Wineburg (1991) postulierten Strategien zugrunde gelegt. Neben dieser theoretischen Einbettung war die Überlegung leitend, dass MDC-Items nicht ausschließlich aufgrund von Informationen eines Dokuments beziehungsweise einer Quelle lösbar sein sollten, sondern der Verwendung von Informationen aus mindestens zwei Dokumenten beziehungsweise Quellen bedürfen. Dabei wurde von Dokumenten mit je einer Quelle ausgegangen, so dass die Verwendung mehrerer Quellen die Verwendung mehrerer Dokumente implizierte. Vor diesem Hintergrund wurden zunächst vier zentrale kognitive Anforderungen von MDC abgeleitet, auf denen die Itementwicklung basierte. Diese Anforderungen werden im Folgenden näher dargestellt. Beispielitems sind dem elektronischen Supplement 1 (ESM 1) zu entnehmen.

2.1 Kognitive Anforderungen der Items

Anforderung 1: Inhalte über Texte hinweg vergleichen

Items, die diesem Anforderungsbereich zugeordnet werden, erfordern die Strategie der Corroboration (Wineburg, 1991), wobei diese Strategie auf einzelne Aussagen oder einfache Inferenzen anzuwenden ist. Zur korrekten Aufgabenlösung müssen Textaussagen zur

1 Fragestellung über mehrere Texte hinweg verglichen werden. Sie müssen daher entweder im
2 Gedächtnis verfügbar sein oder in den Texten (wieder-)gefunden werden – in letzterem Fall
3
4 ist eine entsprechende Repräsentation der Texte nötig oder die Texte müssen erneut gelesen
5 werden. Schwierigkeitsbestimmend kann hier wirken, ob die für die Lösung des Items
6 benötigten Inhalte (a) wörtlich in den Texten zu finden oder Inferenzen nötig sind, (b) über
7 Texte hinweg in gleichen oder unterschiedlichen Wortlauten vorliegen, (c) an zu
8 erwartenden beziehungsweise plausiblen Stellen in den Texten zu finden sind
9 (beispielsweise der Chronologie von Ereignissen folgend) oder (d) mit einer größeren
10 Wahrscheinlichkeit aus dem Gedächtnis abrufbar sind (z. B. für das Leseziel relevante
11 Informationen: Kaakinen & Hyönä, 2008; McCrudden & Schraw, 2007; oder wiederholte
12 Informationen).

23 **Anforderung 2: Inhalte über Texte hinweg miteinander verknüpfen**

24
25
26
27
28
29 Aufgaben mit dieser Anforderung erfordern ein integriertes Situationsmodell (z. B.
30 Britt & Rouet, 2012). Es ist jedoch nicht notwendig zu wissen, aus welchem Dokument die
31 jeweilige Information stammt, oder Quellen miteinander in Beziehung zu setzen oder zu
32 bewerten. Zur korrekten Aufgabenlösung müssen daher Inhalte aus verschiedenen
33 Dokumenten miteinander verknüpft werden. Bei leichteren Aufgaben sollten diese
34 Verknüpfungen nur additiv geleistet werden; bei schwierigeren Aufgaben wäre eine Inferenz
35 auf Basis von Informationen aus mehreren Texten zu ziehen.

36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
Wie bei Items der Anforderung 1 sollten demnach für Items mit der Anforderung 2 die
Merkmale (a) – (d) schwierigkeitsbestimmend sein. Zusätzlich sollte ebenfalls
schwierigkeitsbestimmend sein, (e) ob die Verknüpfung der Informationen rein additiv oder
durch das Ziehen einer Inferenz geschieht.

Anforderung 3: Quellen einschätzen und vergleichen (source-source links)

Für die erfolgreiche Bearbeitung von Aufgaben dieses Anforderungsbereichs ist ein Intertext Model (z. B. Britt & Rouet, 2012) notwendig (bzw. es ist spätestens zur Lösung einer entsprechenden Aufgabe aufzubauen). Quellenbezogene Einschätzungen müssen über Dokumente hinweg miteinander in Bezug gesetzt werden. Daher erfordert eine korrekte Aufgabenlösung die Einschätzung und Bewertung der Dokumentenquelle (z. B. hinsichtlich der Intention oder Eigenschaften der Autorin oder des Autors oder hinsichtlich der intendierten Zielgruppe). Anders als bei Items mit der Anforderung 1 oder 2 ist für Items mit der Anforderung 3 der Grad schwierigkeitsbestimmend, (f) indem Aufgaben allein anhand der Quelleninformation lösbar sind oder Inferenzen aus Quelleninformationen oder Texten gezogen werden müssen.

Anforderung 4: Inhalte als von Quellen generiert repräsentieren und Quelle-Inhalt-Kombinationen vergleichen (source-content links vergleichen)

Aufgaben mit dieser Anforderung sind theoretisch gesehen am anspruchsvollsten. Es müssen mit Quelleninformationen verknüpfte Inhalte aus verschiedenen Dokumenten zueinander in Relation gesetzt werden. Für diese Aufgaben ist es notwendig, Inhalte als von bestimmten Quellen generiert zu repräsentieren beziehungsweise Quelle-Inhalt-Kombinationen mit einander zu vergleichen. Das heißt, es muss ein Documents Model vorliegen, da eine ausschließliche Berücksichtigung der Quellen (Anforderung 3) oder Verknüpfung von Inhalten (Anforderung 2) allein unzureichend sind. Zwar dürften für die Unterschiede in der Schwierigkeit dieser Aufgaben wie für Aufgaben der Anforderung 1 und 2 die Merkmale (a) – (d) eine Rolle spielen. Bei Aufgaben der Anforderung 4 ist jedoch vor allem (g) der Umfang an Informationen schwierigkeitsbestimmend, die beachtet werden müssen, um den Vergleich vornehmen zu können.

2.2 Der entwickelte Test

Obwohl MDC als fächerübergreifende Kompetenz Studierender betrachtet wird, können typische Anforderungen in verschiedenen Inhaltsdomänen beziehungsweise Disziplinen oder Fächern unterschiedliche Ausprägungen annehmen (vgl. Goldman et al., 2016). Um in diesem Sinn der Breite des intendierten Konstrukts Rechnung zu tragen und um mögliche Vorwissenseffekte zu reduzieren, sollten in einem Test zur Erfassung von MDC Texte aus verschiedenen Inhaltsdomänen eingesetzt werden. Es wurden daher sechs Units (d. h. Kombinationen aus Texten zum gleichen Thema sowie dazugehörige Items) zu verschiedenen Themenbereichen entwickelt. Hierzu wurden zwei oder drei Texte pro Unit entwickelt, über die auch kurze (z. T. fiktive) Quelleninformationen bereitgestellt wurden. Der Test (d. h. Texte und Items) wurde auf Deutsch konstruiert und eingesetzt.

Für die Items wurde entweder ein Multiple-Choice-Format (eins aus vier) oder das Format eines Verifikationstests (d. h. Zustimmung oder Ablehnung einzelner Aussagen) gewählt. Zusätzlich zu Items der bisher vorgestellten Anforderungstypen wurden Essay-Items konzipiert, die eine integrierende Zusammenfassung der Texte erfordern und damit das integrierte Situationsmodell (z. B. Britt & Rouet, 2012) operationalisieren. Da Essay-Aufgaben nicht nur die hier interessierende rezeptive Sprachkompetenz (MDC) erfassen, sondern auch produktive Komponenten beinhalten (Sprachproduktion), wurden Essay-Items nicht in die Definition des MDC-Konstrukts einbezogen, sondern als Validierungsvariable verwendet. Dementsprechend wurden in zwei Units Essay-Items vorgelegt. Der resultierende Test dauerte somit nicht länger als ca. 1–1.5 Stunden.

Eine Übersicht über die entwickelten Units einschließlich der Eigenschaften der Texte und der Items und ihren kognitiven Anforderungen gibt das ESM 2. Beispielitems zu jedem Anforderungsbereich und Itemantwortformat können dem ESM 1 entnommen werden.

1 Jede Unit begann zunächst mit einer Übersichtsseite, die sowohl die Anzahl der Texte
2 und Items als auch ein Leseziel (vgl. Gil, Bråten, Vidal-Abarca & Strømsø, 2010a; Stadler
3 et al., 2014) vorgab. In Units mit Essay-Item wurde das Leseziel so formuliert, dass es dem
4 späteren Essay-Item entsprach. In Units ohne Essay-Item wurden analoge Leseziele
5 formuliert.
6
7
8
9
10

11 Der Test wurde computerbasiert mit dem CBA ItemBuilder (Rölke, 2012) umgesetzt. In
12
13
14
15
16
17
18
19
20
21
22

23 **2.3 Hypothesen**

24 Da MDC als eindimensionales Konstrukt verstanden wird und dementsprechend das Ziel
25 die Entwicklung eines eindimensionalen Tests von MDC war, wurden zur Bildung des
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

MDC-Testwerts die entlang der Rahmenkonzeption konstruierten Items auf ihre Passung zu
einem eindimensionalen Rasch-Modell sowie auf *Differential Item Functioning* (DIF) hin
überprüft und ausgewählt. Aufgrund dieser Optimierung im Hinblick auf einen
eindimensionalen Test wurde auf explizite Hypothesen zur Dimensionalität verzichtet. Der
auf der Grundlage der verbliebenen Items gebildete Testwert (operationalisiert als
Personenparameterschätzung) soll die Ausprägung der (eindimensionalen) MDC-Kompetenz
von Studierenden im Sinne unserer Definition widerspiegeln. Vor dem beschriebenen
Hintergrund zu MDC und entsprechend der *Standards for Educational and Psychological
Testing* (American Educational Research Association, American Psychological
Association & National Council on Measurement in Education, 2014) soll diese
Konstruktinterpretation durch die folgenden Validierungshypothesen geprüft werden.

Da von beginnenden Bachelorstudierenden angenommen werden kann, dass (a) sie
wissenschaftliche Fachtexte für den eigenen Erkenntnisprozess nutzen können

1 (Kultusministerkonferenz, 2012) und sich diesbezügliche Fähigkeitsunterschiede in der
2 Abiturnote widerspiegeln, sowie dass (b) sich ihre Fähigkeiten, Wissen aus verschiedenen
3 Quellen zu integrieren, im Laufe des Studiums entwickeln (Arbeitskreis Deutscher
4 Qualifikationsrahmen, 2011) wurde erwartet, dass der MDC-Testwert ...
5
6
7
8

9 H1. positiv mit der Abiturnote korreliert,
10

11 H2. höher mit der Abiturnote von Bachelorstudierenden korreliert ist als mit der
12 Abiturnote von Masterstudierenden und
13
14

15 H3. für Masterstudierende höher ausfällt als für Bachelorstudierende.
16
17

18 H4. Da die im MDC-Test integrierten Essay-Aufgaben sowohl rezeptive als auch
19 produktive Fähigkeiten erfassen, wurde erwartet, dass die Leistung in Essay-Aufgaben mit
20 dem MDC-Testwert zwar positiv, aber in niedriger bis moderater Höhe zusammenhängt.
21
22
23
24

25 Darüber hinaus werden im MDC-Test Items mit unterschiedlichen kognitiven
26 Anforderungen vorgegeben, die sich zum Teil auf Teilkomponenten des Documents Model,
27 zum Teil auf das gesamte Documents Model bezogen. Explorativ wurden daher
28 Unterschiede in der Lösungswahrscheinlichkeit der Items in Abhängigkeit von ihrem
29 Anforderungstyp analysiert.
30
31
32
33
34
35
36
37
38
39
40

41 **3 Methode**

42 **3.1 Stichprobe**

43 An der Studie haben 310 Studierende zweier deutscher Universitäten zwischen 18 und 34
44 Jahren ($M = 21.4$, $SD = 2.72$, 79 % weiblich) teilgenommen. Von ihnen waren 201 (65 %) im
45 ersten Bachelor-Semester und 90 Studierende (29 %) im ersten Master-Semester
46 verschiedener sozial- und geisteswissenschaftlicher Studiengänge eingeschrieben (s. ESM
47 4). Die Studierenden erhielten für die Teilnahme 20 €. Zudem wurde unter allen
48 Teilnehmenden ein Gutschein für einen Tabletcomputer verlost.
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

3.2 Ablauf und Design

Nach einer kurzen mündlichen Einführung durch eine Testleiterin oder einen Testleiter und der schriftlichen Einwilligung in die Untersuchungsteilnahme arbeiteten die Teilnehmenden ausschließlich am Computer. Zunächst wurden demografische Angaben und eine Reihe weiterer Merkmale erfasst, die zur weiterführenden Beantwortung MDC-betreffender Forschungsfragen vorgelegt wurden (z. B. Zielorientierungen, epistemische Überzeugungen). Nach einer Einführung in die Funktionalitäten des Tests (Tutorial) bearbeiteten die Teilnehmenden jeweils drei aus sechs MDC-Units ohne zeitliche Beschränkungen. Um sicherzustellen, dass jede Unit und jedes Paar von Units in gleich vielen Kombinationen vorgelegt wurden, wurde ein Multi-Matrix-Design in Form eines Balanced-Incomplete-Block-Designs mittels des R-Pakets `crossdes` (Sailer, 2013) generiert und umgesetzt, wodurch sich 10 verschiedene Unit-Kombinationen ergaben. Um zusätzlich Positionseffekte auszubalancieren, wurde die Reihenfolge der Units rotiert, so dass insgesamt 60 verschiedene Testhefte verwendet wurden. Im Anschluss an jede Unit wurden den Teilnehmenden weitere Items vorgelegt, die sich auf die vorherige Unit bezogen (z. B. zur Erfassung mentaler Anstrengung) und sie bekamen die Gelegenheit, eine kurze Pause zu machen. Nach dem Ende der Testung erhielten die Teilnehmenden ihr Incentive und wurden verabschiedet. Die Erhebung dauerte insgesamt etwa zwei Stunden.

3.3 Scoring der MDC-Items und der Essay-Items

Die Antworten der Multiple-Choice- als auch der Verifikationsitems wurden dichotom kodiert (1 = richtig, 0 = falsch). Fehlende Werte, die durch das Überspringen (*omitted*) oder Nicht-Bearbeiten (*not reached*) von Items entstehen können, traten in der vorliegenden Studie kaum auf (insgesamt in 0.5 % der vorgelegten Items) und wurden daher für die

1 Schätzung so behandelt, als seien die Items nicht administriert worden (Pohl, Gräfe & Rose,
2 2014; Rose, Von Davier & Xu, 2010).
3

4 Für die Bewertung der Essay-Items wurde auf der Basis einer Musterlösung eine Liste
5 von Informationen erstellt, die in der Lösung genannt werden konnten (für eine optimale
6 Lösung mussten nicht alle dieser möglichen Informationen genannt werden, da z. T. für eine
7 optimale Lösung nur nötig war, die eine *oder* die andere Information zu nennen). Die
8 Lösungen wurden daraufhin bewertet, ob die jeweilige Information genannt wurde (1 Punkt)
9 oder nicht (0 Punkte). Als Score für den Essay wurde die Summe der Punkte gebildet. Alle
10 Lösungen wurden von zwei geschulten Kodierenden bewertet. Entsprechend der Bewertung
11 von Landis und Koch (1977) sind die Übereinstimmungen zwischen den beiden
12 Kodierenden (berechnet auf der Ebene der Einzelinformationen) als fast perfekt (Cohens
13 $\kappa = .86$ in der Unit „Nichts“) beziehungsweise substantiell (Cohens $\kappa = .78$ in der Unit
14 „Universum“) einzustufen.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

34 **3.4 Datenanalyse**

35 Die psychometrischen Analysen im Rahmen der Testentwicklung erfolgten auf Basis des
36 Rasch-Modells. Die Analysen wurden mit R 3.3.2 (R Core Team, 2016) und dem R-Paket
37 TAM (Robitzsch, Kiefer & Wu, 2017) vorgenommen. Als Kriterien für die
38 (eindimensionale) Rasch-Konformität wurden dabei deskriptiv Infit und Outfit der Items
39 anhand gängiger Schwellenwerte (0.8–1.2, s. Wright & Linacre, 1994) sowie die
40 Trennschärfe als punktbiseriale Korrelation des Items mit der im Test gemessenen Fähigkeit
41 angelegt. Zusätzlich wurde in einer visuellen Inspektion die erwartete itemcharakteristische
42 Kurve mit der beobachteten, nicht-parametrischen itemcharakteristischen Kurve im Hinblick
43 auf Wölbungen, Nicht-Monotonie und unerwartete Asymptoten verglichen (Douglas &
44 Cohen, 2001). Abschließend wurden der globale Modellfit anhand von χ^2 -Tests auf bedingte
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Abhängigkeit (Chen & Thissen, 1997) sowie die Annahme lokaler Unabhängigkeit anhand
2 von Q3-Statistiken überprüft.
3

4
5 Im Rahmen von Dimensionalitätsanalysen wurden mehrdimensionale Rasch-Modelle
6 unter Verwendung der in TAM implementierten Quasi-Monte-Carlo-Integration (10 000
7 Knoten) geschätzt und anhand der Informationskriterien AIC und BIC und eines Likelihood-
8 Ratio-Tests mit dem eindimensionalen Modell verglichen.
9

10
11
12
13
14 Zusätzlich wurden DIF-Analysen mit dem R-Paket difR (Magis, Béland, Tuerlinckx &
15 De Boeck, 2010) vorgenommen. Als Gruppierungsvariablen wurden neben dem Studienlevel
16 (Bachelor vs. Master), für das Effekte im Rahmen der Hypothesen geprüft werden sollten,
17 die Variablen Geschlecht und Standort gewählt, für die ebenfalls nicht-MDC-spezifische
18 Unterschiede in den Lösungswahrscheinlichkeiten der Items vorliegen könnten (Geschlecht:
19 vgl. die Diskussion um Geschlechtsunterschiede in der Lesekompetenz, z. B. Schwabe,
20 McElvany & Trendtel, 2015; Standort: z. B. aufgrund universitätsspezifischer
21 Eingangsvoraussetzungen oder Ausbildungsstrukturen). Auf der Grundlage der verbliebenen
22 Items wurden als Testwert (Personenparameter) *weighted likelihood estimates* (WLE-Scores;
23 Warm, 1989) geschätzt. Um die Robustheit dieses Vorgehens zu überprüfen, wurden
24 zusätzlich confirmatorische Faktorenanalysen (mit dem MLR-Schätzer, vgl. Muthén,
25 Muthén & Asparouhov, 2015) vorgenommen, deren Ergebnisse ESM 5 entnommen werden
26 können.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45
46 Zur Prüfung der Validierungshypothesen wurden für die Zusammenhangshypothesen
47 (H1, H2, H4) Korrelationsanalysen durchgeführt. Die Unterschiedshypothese (H3) wurde im
48 Skalierungsmodell durch eine Regression des Personenparameters auf den
49 Gruppierungsfaktor überprüft. Die explorativen Analysen zu den anforderungsspezifischen
50 Lösungswahrscheinlichkeiten wurden mittels eines generalisierten linearen gemischten
51 Modells (GLMM) durchgeführt, in dem die Lösungswahrscheinlichkeit der Items durch
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 ihren Anforderungstyp vorhergesagt wurde. Im Modell wurden zudem ein festes Interzept
2 sowie Zufalls-Interzepte für Personen und Items eingeschlossen (R-Syntax: $\text{score} \sim 1 + 1|ID$)
3 $+ (1|item_name) + itemtyp$; R-Paket lme4; Bates, Mächler, Bolker & Walker, 2015). Die
4 Referenzkategorie war Anforderungstyp 1, so dass das feste Interzept den Logit der mittleren
5 Lösungswahrscheinlichkeit von Items des Anforderungstyps 1 repräsentiert, während die
6 Regressionskoeffizienten beschreiben, wie sich dieser Logit verändert, wenn sich der
7 Anforderungstyp verändert (Typ 2–4). Um Unterschiede in der Lösungswahrscheinlichkeit
8 in Abhängigkeit vom Anforderungstyp zu analysieren, wurden a priori vier Kontraste
9 spezifiziert, in denen jeweils die Lösungswahrscheinlichkeit von Items eines
10 Anforderungstyps mit der Lösungswahrscheinlichkeit der Items aller anderen
11 Anforderungstypen verglichen wurde.

28 **4 Ergebnisse**

29 **4.1 Skalierung, DIF-Analysen und Dimensionalitätsanalysen**

30 Von den entwickelten 174 Items wurden 66 Items ausgeschlossen, da ihre Itemkennwerte
31 (d. h. Infit und Outfit, punktbiseriale Korrelation, Merkmale der itemcharakteristischen
32 Kurve, vgl. Abschnitt 3.4) auf Antwortmuster hindeuteten, die unter Gültigkeit des Rasch-
33 Modells nicht zu erwarten wären. Die DIF-Analysen zeigten Nachteile für Männer auf vielen
34 Items der Unit „Verzeihen“, so dass diese Unit komplett ausgeschlossen wurde. Aufgrund
35 von DIF wurden weitere 23 Items in den verbleibenden Units ausgeschlossen. Die
36 verbliebenen 67 Items (s. a. Tabelle 2 in ESM 2) decken weiterhin alle kognitiven
37 Anforderungen gleichmäßig auf die verschiedenen Units verteilt ab und spiegeln daher die
38 MDC-Kompetenzen von Studierenden im Sinne unserer Definition wider. Die 67 Items
39 wiesen in einer abschließenden Skalierung der ausgewählten Items sehr gute Infit- (zwischen
40 0.91 und 1.10) und Outfit-Werte auf (zwischen 0.74 und 1.14). Die Korrelationen der Items
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 mit dem Testwert lagen zwischen .16 und .47 (für die deskriptiven Kennwerte aller Items s.
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

mit dem Testwert lagen zwischen .16 und .47 (für die deskriptiven Kennwerte aller Items s. ESM 6). Der auf Grundlage dieser Items geschätzte Testwert besaß eine akzeptable Expected A Posteriori (EAP) - beziehungsweise WLE-Reliabilität von .69 beziehungsweise .67. χ^2 -Tests auf bedingte Abhängigkeit (Chen & Thissen, 1997) für alle Itempaare mit einer Adjustierung der p -Werte nach Holm zeigten einen akzeptablen globalen Modellfit an (max $\chi^2 = 17.05$, $p_{\text{holm}} = .081$). Zur Überprüfung der Annahme lokaler Unabhängigkeit wurden Q3-Statistiken überprüft ($|\text{Cut-off Kriterium}| > .2$; Chen & Thissen, 1997). Zwar zeigten 13 % aller Itempaare einen Wert über dem Cut-off, der Mittelwert der Q3-Werte war jedoch geringfügig negativ (-0.02), was unter lokaler Unabhängigkeit zu erwarten ist.

Für die verbliebenen Items wurden zusätzlich zum eindimensionalen Modell mehrere Vergleichsmodelle geschätzt (s. ESM 7). Ein (a) vierdimensionales Modell, in dem die vier Anforderungstypen der Items als Dimensionen angenommen wurden, ein (b) fünfdimensionales Modell, in dem Units als Dimensionen angenommen wurden sowie ein (c) Rasch-Testlet-Modell (Wang & Wilson, 2005) spezifiziert, in dem die Testlets den Units entsprechen (d. h. ein Generalfaktor und fünf Subdimensionen). Basierend auf den Informationskriterien AIC und BIC sowie auf den Ergebnissen von Likelihood-Ratio-Tests zum Modellvergleich bildete keines der drei alternativen Modelle die Datenstruktur besser ab als das eindimensionale Modell (s. Tabelle 1). Keines der Vergleichsmodelle wies einen kleineren AIC oder BIC als das eindimensionale Modell auf, die Likelihood-Ratio-Tests waren nicht signifikant.

Dimensionsanalysen mit allen 174 entwickelten Items ergaben für das vierbeziehungsweise sechsdimensionale Modell (Units als Dimensionen, unter Einbezug aller Items ist die sechste Unit „Verzeihen“ weiterhin enthalten) zwar einen signifikanten Likelihood-Ratio-Test, der auf eine bessere Passung der mehrdimensionalen Modelle als des eindimensionalen Modells hindeutet (s. Tabelle 1). Jedoch sprechen in beiden Fällen die

1 Informationskriterien (AIC, BIC) nicht dafür, dass das eindimensionale Modell deutlich
2 schlechter passt. Daher scheint gerechtfertigt, das Rasch-Modell beizubehalten. Das Rasch-
3 Testlet-Modell für alle Items zeigte keine bessere Modellpassung als das eindimensionale
4 Modell (s. Tabelle 1).
5
6
7
8
9

10 Aus diesen Ergebnissen kann geschlossen werden, dass die ausgewählten Items keine
11 Mehrdimensionalität aufgrund der Anforderungstypen oder der Units abbilden und damit die
12 Annahme der Eindimensionalität in Einklang steht mit der gefundenen Faktorstruktur.
13
14
15
16
17
18

19 **4.2 Prüfung der Validierungshypothesen**

21 Die Abiturnote korrelierte in der Gesamtstichprobe signifikant mit dem MDC-Testwert
22 ($r = -.44, p < .001$), das heißt mit einer besseren (kleineren) Abiturnote ging ein besserer
23 (größerer) MDC-Testwert einher (H1). Der Zusammenhang war für Bachelorstudierende
24 ($r = -.44, p < .001$) zwar deskriptiv größer als für Masterstudierende ($r = -.35, p < .001$), der
25 Unterschied (H2) war jedoch nicht statistisch signifikant ($z = -0.83, p = .41$).
26
27 Masterstudierende erreichten signifikant höhere MDC-Testwerte als Bachelorstudierende
28 (H3, $\beta_{\text{unstandardisiert}} = 0.29, SE = 0.03, \beta_{\text{standardisiert}} = 0.22, t_{295} = 9.46, p < .001$, s. Abbildung 1).
29 Dies galt auch unter Kontrolle von Abiturnote und Studienfach ($\beta_{\text{unstandardisiert}} = 0.17,$
30 $SE = 0.03, \beta_{\text{standardisiert}} = 0.15, t_{290} = 6.06, p < .001$). In Bezug auf den Zusammenhang des
31 MDC-Testwerts mit der Leistung in den Essay-Aufgaben (H4) zeigten sich Korrelationen
32 von $r = .32 (p < .001, \text{Unit „Universum“})$ und $r = .44 (p < .001, \text{Unit „Nichts“})$.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

48 Die explorativen Analysen zu den anforderungsspezifischen
49 Lösungswahrscheinlichkeiten (s. Kontraste in Tabelle 2, Abbildung 2) zeigten eine
50 signifikant höhere Lösungswahrscheinlichkeit für Items des Anforderungstyps 1 (Inhalte
51 über Texte hinweg vergleichen) im Vergleich zu den Items aller anderen Anforderungstypen
52 (wobei im Einzelvergleich Items des Anforderungstyps 3 nicht signifikant niedrigere
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Lösungswahrscheinlichkeiten aufwiesen als Items des Anforderungstyps 1, wie aus den
2 Regressionskoeffizienten in Tabelle 2 ersichtlich). Zudem findet sich eine signifikant
3 niedrigere Lösungswahrscheinlichkeit für Items des Anforderungstyps 4 (Inhalte als von
4 Quellen generiert repräsentieren und Quelle-Inhalt-Kombinationen vergleichen) im
5 Vergleich zu den Items aller anderen Anforderungstypen.
6
7
8
9
10

11 **5 Diskussion**

12 Das Ziel des vorliegenden Beitrags war die Entwicklung und Skalierung eines Tests zur
13 Erfassung des Verständnisses multipler Dokumente durch Studierende. Darunter verstehen
14 wir die fächerübergreifende Fähigkeit von Studierenden, aus verschiedenen
15 Informationsquellen eine integrierte Repräsentation eines bestimmten inhaltlichen
16 Gegenstandsbereichs oder Sachverhalts zu konstruieren. Zur Entwicklung des Tests wurden
17 vier kognitive Anforderungen aus der bestehenden Literatur abgeleitet, auf deren Basis die
18 Testitems entwickelt wurden. Die empirischen Ergebnisse zeigen, dass die getroffene
19 Auswahl von Items ein eindimensionales Fähigkeitskonstrukt abbildet. Weder die kognitiven
20 Anforderungen noch die Einteilung in Units stellen – nach der Auswahl von Items mit der
21 Zielstellung eines eindimensionalen Konstrukts – verschiedene Dimensionen dar. Die
22 Reliabilität ist mit .69 (EAP) beziehungsweise .67 (WLE) nicht wünschenswert hoch,
23 allerdings noch akzeptabel. Globale Fit-Indizes sind zufriedenstellend, wenn auch nicht
24 durchgehend optimal.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

46 Die Interpretation des aus dem finalen MDC-Test konstruierten Testwertes wurde
47 anhand von vier Hypothesen überprüft. Die Hypothesen konnten größtenteils gestützt
48 werden. So fand sich die erwartete Korrelation des MDC-Testwerts mit der Abiturnote (H1),
49 die allerdings nur deskriptiv wie erwartet (H2) für Bachelorstudierende höher ausfiel als für
50 Masterstudierende, was sich jedoch inferenzstatistisch nicht absichern ließ. Die Abiturnote
51 hatte auch auf dem Masterniveau noch einen deutlichen Zusammenhang zum MDC-
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Testwert. Zu beachten ist bei diesem Ergebnis, dass Master-Studierende im ersten Semester
2 getestet wurden. Interessant wäre noch die weitere Entwicklung im Studienverlauf bis zum
3 Ende des Masterstudiums. Es ist zu erwarten, dass der Einfluss der Abiturnote absinkt.
4 Dennoch wird er wahrscheinlich nicht gänzlich verschwinden, da in der Abiturnote
5 vermutlich auch kognitive Grundfähigkeiten abgebildet sind, die förderlich für den Umgang
6 mit multiplen Dokumenten sein können. Beispielsweise wäre zu erwarten, dass Personen mit
7 einer größeren Arbeitsgedächtniskapazität die MDC-Aufgaben leichter bearbeiten können
8 und entsprechend besser abschneiden als Personen mit einer kleineren
9 Arbeitsgedächtniskapazität (vgl. Braasch et al., 2014). Der erwartete Niveau-Unterschied
10 zwischen Bachelor- und Masterstudierenden (H3) konnte ebenfalls empirisch bestätigt
11 werden. Insgesamt betrachtet zeigen diese Ergebnisse, dass der Testwert aus dem
12 entwickelten MDC-Test eine Fähigkeit abbildet, die sowohl mit Schulleistung als auch mit
13 der Entwicklung von Kompetenzen im Rahmen eines Studiums zusammenhängt.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

31 Die moderaten Korrelationen des MDC-Testwerts mit den Bewertungen der Essays
32 zeigen ferner, dass der MDC-Testwert tatsächlich eine andere Kompetenz abbildet als mit
33 den Essays erfasst wurde. Diese moderaten Korrelationen waren erwartet worden (H4), da in
34 Essays nicht nur rezeptive Kompetenzen, wie im vorliegenden MDC-Test, sondern auch
35 produktive Kompetenzen (Schreiben) einfließen. Diese Ergebnisse machen daher auch
36 deutlich, wie wichtig eine getrennte Erfassung der rezeptiven Anteile von MDC ist.
37
38
39
40
41
42
43
44
45

46 Innerhalb des MDC-Tests zeigten sich außerdem Differenzierungen hinsichtlich der
47 Schwierigkeit von Items mit unterschiedlichen kognitiven Anforderungen. So waren Items
48 mit der Anforderung, Inhalte über Texte hinweg zu vergleichen (Anforderungstyp 1),
49 signifikant leichter als alle anderen Items (im Direktvergleich allerdings nicht signifikant
50 von Items des Anforderungstyps 3 unterschiedlich) und Items mit der Anforderung, Inhalte
51 als von Quellen generiert zu repräsentieren und Quelle-Inhalt-Kombinationen zu vergleichen
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 (Anforderungstyp 4), signifikant schwerer als alle anderen Items. Dieser Befund ist durchaus
2 plausibel: Bei Items des Anforderungstyps 1 (Inhalte über Texte hinweg vergleichen)
3
4 mussten nur einfache Inhalte verglichen werden, während sich der Vergleich bei Items des
5 Anforderungstyps 4 (source-content links vergleichen) auf Quelle-Inhalt-Kombinationen
6
7 bezieht. Items des Anforderungstyps 2 hingegen erfordern eine Verknüpfung von Inhalten
8
9 zusätzlich zum Finden dieser. Dass Items des Anforderungstyps 1 leichter als alle anderen
10
11 Items sind, lässt sich daher gut nachvollziehen. Items des Anforderungstyps 4 wiederum
12
13 erfordern ein vollständiges Documents Model, das heißt die Inhalte müssen miteinander in
14
15 Beziehung gesetzt werden, während gleichzeitig ihre Quelle mit repräsentiert wird. Items des
16
17 Anforderungstyps 2 beziehungsweise 3 erfordern dagegen jeweils Teilkomponenten des
18
19 Documents Model (integriertes Situationsmodell bzw. Intertext Model), das heißt es ist
20
21 entweder nur eine Verknüpfung von Inhalten (bei Items des Anforderungstyp 2) oder nur
22
23 eine Einschätzung und Repräsentation der Quellen (bei Items des Anforderungstyps 3)
24
25 erforderlich. Damit lässt sich erklären, dass Items des Anforderungstyps 4 schwieriger waren
26
27 als alle anderen. Diese Befunde sollten jedoch mit anderen Materialien repliziert werden, da
28
29 nicht ausgeschlossen werden kann, dass es sich bei den vorliegenden Befunden um Effekte
30
31 handelt, die mit den spezifischen Fragen und Themen zusammenhängen, die im
32
33 vorliegenden Test gewählt wurden.

34
35
36
37
38
39
40
41
42
43
44 Der entwickelte Test zielt im Wesentlichen auf die Erfassung des Verständnisses
45
46 multipler Dokumente, wie es bei geringem oder nicht vorhandenem Vorwissen mit
47
48 weitgehend glaubwürdigen Dokumenten benötigt wird. Dies ist die Situation, mit der
49
50 Studierende im Hochschulkontext wohl oft konfrontiert werden, wenn sie neuen Themen
51
52 begegnen und Literatur aus wissenschaftlichen Datenbanken oder von ihren Dozierenden
53
54 erhalten. Daher wurden größtenteils fiktive Themen sowie Texte gewählt, deren
55
56 Glaubwürdigkeit wenig variierte. Durch diese geringe Varianz in der Glaubwürdigkeit der
57
58
59
60
61
62
63
64
65

1 Texte besteht diesbezüglich eine gute Vergleichbarkeit mit realen Situationen im Studium,
2 wo multiple Texte meist auch nur geringe Unterschiede in der Glaubwürdigkeit aufweisen.
3
4 Ein weiterer Vorteil dieser Wahl von Themen und Texten besteht darin, dass Vorwissens-
5 und Voreinstellungseffekte (vgl. Richter & Maier, 2017) so gering wie möglich gehalten
6 werden können. Im Umkehrschluss kann der Test daher aber nur ansatzweise erfassen, was
7 im alltäglichen Umgang mit multiplen Dokumenten (z. B. nach einer Internetrecherche zu
8 politischen oder gesundheitsbezogenen Fragestellungen) häufig benötigt wird: das
9 Zurückstellen eigener Voreinstellungen und das Verarbeiten auch
10 überzeugungsinkonsistenter Informationen (vgl. Richter & Maier, 2017) sowie die
11 Einschätzung der Glaubwürdigkeit von Quellen (vgl. Goldman & Scardamalia, 2013b). Um
12 diese Aspekte des Verständnisses multipler Dokumente, die bewusst nicht in die aktuelle
13 Testentwicklung aufgenommen wurden, zu erfassen, bedarf es weiterer beziehungsweise
14 anderer Testverfahren.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

31 In der vorliegenden Studie ist das Verhältnis von MDC zu (klassischer) Lesekompetenz
32 noch ungeklärt. Es ist zu erwarten, dass hier eine deutliche Korrelation zu finden ist, wenn
33 auch MDC und Lesekompetenz konzeptuell zu unterscheiden sind. Lesekompetenz ist eine
34 notwendige Voraussetzung für MDC. MDC geht aber in verschiedenen Punkten über
35 klassische Lesekompetenz hinaus, da multiple Dokumente zueinander in Bezug gesetzt
36 werden müssen. Dies erfordert auch den Einsatz von Strategien, die für MDC im Vergleich
37 zu klassischen Lesestrategien spezifisch sind (z. B. der notwendige Vergleich von
38 Informationen oder der wesentlich stärkere Bezug zu Quellen, vgl. Wineburg, 1991).
39 Während auch in klassischen Lesekompetenztests teilweise Quelleneinschätzungen erfragt
40 werden (z. B. Gehrer, Zimmermann, Artelt & Weinert, 2013), ist in MDC neben der reinen
41 Quelleneinschätzung einen Bezug zwischen Quelle und Inhalt (source-content Link)
42 herzustellen beziehungsweise aufrecht zu erhalten. Neben Einstellungen (z. B. epistemische
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Überzeugungen) und (Strategie-) Wissen über den adäquaten Umgang mit multiplen
2 Dokumenten dürfte daher auch die Arbeitsgedächtniskapazität für MDC eine Rolle spielen.
3
4 Diese Hypothesen müssen in Folgestudien weiter untersucht werden.
5
6

7 Die Stichprobe der vorliegenden Studie war im Hinblick auf das Studienfach zu
8 heterogen, um verlässliche Aussagen über studienfachspezifische Unterschiede im MDC-
9 Testwert zwischen den Studierenden treffen zu können. Die Fach(un)abhängigkeit des
10 MDC-Testwerts sollte daher in einer Folgestudie mit größeren Teilstichproben pro
11 Studiengang überprüft werden.
12
13
14
15
16
17
18

19 Auch wenn die vorliegenden Ergebnisse noch an einer unabhängigen Stichprobe (die
20 auch weitere Fächergruppen, wie beispielsweise Natur- und Ingenieurwissenschaften
21 einschließen sollte) überprüft werden müssen, liegt mit dem MDC-Test ein Instrument vor,
22 mit dem theoretisch fundiert und hinreichend differenzierend die Fähigkeit von Studierenden
23 gemessen werden kann, angemessen mit multiplen Dokumenten umzugehen. Der MDC-Test
24 geht insofern über bisherige Messansätze hinaus, als dass nicht nur auf einen Aspekt
25 fokussiert wird (z. B. Sourcing oder Integration), sondern MDC durch mehrere theoretisch
26 hergeleitete kognitive Anforderungen abgebildet wird. Die vorliegenden Ergebnisse
27 sprechen dafür, dass der MDC-Test die fächerübergreifende Kompetenz von Studierenden,
28 aus verschiedenen Informationsquellen eine integrierte Repräsentation eines bestimmten
29 inhaltlichen Gegenstandsbereichs oder Sachverhalts zu konstruieren, objektiv, reliabel und
30 valide misst. Weitere Validitätshinweise – wie beispielsweise die Untersuchung der oben
31 angedeuteten Zusammenhänge mit individuellen Fähigkeiten des Arbeitsgedächtnisses –
32 sind jedoch wünschenswert.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Elektronische Supplemente

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

ESM 1. Beispielaufgaben für die verschiedenen Anforderungstypen und Itemformate

ESM 2. Überblick über die entwickelten Units einschließlich Eigenschaften der Texte
und Anzahl der Items pro Anforderung und Unit

ESM 3. Beschreibung der computerbasierten Umsetzung des Tests

ESM 4. Häufigkeit der Studienfächer der Studienteilnehmerinnen und -teilnehmer

ESM 5. Ergebnisse der konfirmatorischen Faktorenanalysen

ESM 6. Deskriptive Kennwerte der Items

ESM 7. Ergebnisse der Dimensionsanalysen

Literaturverzeichnis

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anmarkrud, Ø., Bråten, I. & Strømsø, H. I. (2014). Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences*, 30, 64–76.
<https://doi.org/10.1016/j.lindif.2013.01.007>
- Arbeitskreis Deutscher Qualifikationsrahmen (2011). *Deutscher Qualifikationsrahmen für lebenslanges Lernen*. Verfügbar unter http://www.dqr.de/media/content/Der_Deutsche_Qualifikationsrahmen_fue_lebenslanges_Lernen.pdf
- Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Braasch, J. L. G. & Bråten, I. (2017). The discrepancy-induced source comprehension (D-ISC) model: Basic assumptions and preliminary evidence. *Educational Psychologist*, 52(3), 167–181. <https://doi.org/10.1080/00461520.2017.1323219>
- Braasch, J. L. G., Bråten, I., Strømsø, H. I. & Anmarkrud, Ø. (2014). Incremental theories of intelligence predict multiple document comprehension. *Learning and Individual Differences*, 31, 11–20. <https://doi.org/10.1016/j.lindif.2013.12.012>
- Braasch, J. L. G., Rouet, J.-F., Vibert, N. & Britt, M. A. (2012). Readers' use of source information in text comprehension. *Memory & Cognition*, 40(3), 450–465.
<https://doi.org/10.3758/s13421-011-0160-6>

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- Bråten, I., Ferguson, L. E., Strømsø, H. I. & Anmarkrud, Ø. (2014). Students working with multiple conflicting documents on a scientific issue: Relations between epistemic cognition while reading and sourcing and argumentation in essays. *British Journal of Educational Psychology*, *84*(1), 58–85. <https://doi.org/10.1111/bjep.12005>
- Bråten, I., Salmerón, L. & Strømsø, H. I. (2016). Who said that? Investigating the Plausibility-Induced Source Focusing assumption with Norwegian undergraduate readers. *Contemporary Educational Psychology*, *46*, 253–262. <https://doi.org/10.1016/j.cedpsych.2016.07.004>
- Britt, M. A. & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, *20*(4), 485–522. https://doi.org/10.1207/s1532690xci2004_2
- Britt, M. A., Perfetti, C. A., Sandak, R. & Rouet, J.-F. (1999). Content integration and source separation in learning from multiple texts. In S. R. Goldman, A. C. Graesser & P. Van den Broek (Eds.), *Narrative, comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 209–233). Mahwah, NJ: Erlbaum.
- Britt, M. A. & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. In J. R. Kirby & M. J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes* (pp. 276–314). New York: Cambridge University Press.
- Britt, M. A. & Sommer, J. (2004). Facilitating textual integration with macro-structure focusing tasks. *Reading Psychology*, *25*(4), 313–339. <https://doi.org/10.1080/02702710490522658>
- Cerdán, R. & Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology*, *100*(1), 209–222. <https://doi.org/10.1037/0022-0663.100.1.209>

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- Chen, W.-H. & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265–289. <https://doi.org/10.2307/1165285>
- Deutscher Bibliotheksverband (2009). *Standards der Informationskompetenz für Studierende*. Verfügbar unter www.bibliotheksverband.de/fileadmin/user_upload/Kommissionen/Kom_Dienstleistung/Publikationen/Standards_Infokompetenz_03.07.2009_endg.pdf
- Douglas, J. & Cohen, A. (2001). Nonparametric item response function estimation for assessing parametric model fit. *Applied Psychological Measurement*, 25(3), 234–243. <https://doi.org/10.1177/01466210122032046>
- Gehrer, K., Zimmermann, S., Artelt, C. & Weinert, S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online*, 5(2), 50–79.
- Gil, L., Bråten, I., Vidal-Abarca, E. & Strømsø, H. I. (2010a). Summary versus argument tasks when working with multiple documents: Which is better for whom? *Contemporary Educational Psychology*, 35(3), 157–173. <https://doi.org/10.1016/j.cedpsych.2009.11.002>
- Gil, L., Bråten, I., Vidal-Abarca, E. & Strømsø, H. I. (2010b). Understanding and integrating multiple science texts: Summary tasks are sometimes better than argument tasks. *Reading Psychology*, 31(1), 30–68. <https://doi.org/10.1080/02702710902733600>
- Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C. et al. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, 51(2), 219–246. <https://doi.org/10.1080/00461520.2016.1168741>
- Goldman, S. R. & Scardamalia, M. (2013a). Managing, understanding, applying, and creating knowledge in the information age: Next-generation challenges and

- opportunities. *Cognition and Instruction*, 31(2), 255–269.
<https://doi.org/10.1080/10824669.2013.773217>
- Goldman, S. R. & Scardamalia, M. (2013b). Multiple document comprehension [Special Issue]. *Cognition and Instruction*, 31(2).
- Homann, B. (2000). Das Dynamische Modell der Informationskompetenz (DYMIK) als Grundlage für bibliothekarische Schulungen. In G. Knorz & R. Kuhlen (Hrsg.), *Informationskompetenz - Basiskompetenz in der Informationsgesellschaft. Proceedings des 7. Internationale Symposiums für Informationswissenschaft (ISI 2000), Darmstadt, 8.-10. November 2000* (S. 195–206). Konstanz: UVK Verlag.
- Kaakinen, J. K. & Hyönä, J. (2008). Perspective-driven text comprehension. *Applied Cognitive Psychology*, 22(3), 319–334.
- Kammerer, Y. & Gerjets, P. (2014). Quellenbewertungen und Quellenverweise bei Lesen und Zusammenfassen wissensbezogener Informationen aus multiplen Webseiten. *Unterrichtswissenschaft*, 42(1), 7–23.
- Kammerer, Y., Kalbfell, E. & Gerjets, P. (2016). Is this information source commercially biased? How contradictions between web pages stimulate the consideration of source information. *Discourse Processes*, 53(5–6), 430–456.
<https://doi.org/10.1080/0163853x.2016.1169968>
- Keck, D., Kammerer, Y. & Staraschek, E. (2015). Reading science texts online: Does source information influence the identification of contradictions within texts? *Computers & Education*, 82, 442–449. <https://doi.org/10.1016/j.compedu.2014.12.005>
- Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: Cambridge University Press.
- Kultusministerkonferenz (2012). *Bildungsstandards im Fach Deutsch für die Allgemeine Hochschulreife (Beschluss der Kultusministerkonferenz vom 18.10.2012)*. Verfügbar

- 1 unter www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-
2 Bildungsstandards-Deutsch-Abi.pdf
3
4
5 Landis, J. R. & Koch, G. (1977). The measurement of observer agreement for categorical
6 data. *Biometrics*, 33(1), 159–174.
7
8
9
10 Lau, J. (2006). *Guidelines on information literacy for lifelong learning*. Verfügbar unter
11 archive.ifla.org/VII/s42/pub/IL-Guidelines2006.pdf
12
13
14 Lenhard, W. & Lenhard, A. (2014). *Berechnung des Lesbarkeitsindex LIX nach Björnson*.
15 Verfügbar unter <http://www.psychometrica.de/lix.html>
16
17
18
19 Magis, D., Béland, S., Tuerlinckx, F. & De Boeck, P. (2010). A general framework and an R
20 package for the detection of dichotomous differential item functioning. *Behavior*
21 *Research Methods*, 42(3), 847–862. <https://doi.org/10.3758/brm.42.3.847>
22
23
24
25
26
27 Maier, J. & Richter, T. (2013). Text belief consistency effects in the comprehension of
28 multiple texts with conflicting information. *Cognition and Instruction*, 31(2), 151–175.
29 <https://doi.org/10.1080/07370008.2013.769997>
30
31
32
33
34 Maier, J. & Richter, T. (2014). Verstehen multipler Texte zu kontroversen
35 wissenschaftlichen Themen: Die Rolle der epistemischen Validierung.
36 *Unterrichtswissenschaft*, 42(1), 24–38.
37
38
39
40
41 McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing.
42 *Educational Psychology Review*, 19(2), 113–139.
43
44
45
46 Mühlen, S. v. d., Richter, T., Schmid, S., Schmidt, E. M. & Berthold, K. (2016). The use of
47 source-related strategies in evaluating multiple psychology texts: A student–scientist
48 comparison. *Reading and Writing*, 29(8), 1677–1698. [https://doi.org/10.1007/s11145-](https://doi.org/10.1007/s11145-015-9601-0)
49 [015-9601-0](https://doi.org/10.1007/s11145-015-9601-0)
50
51
52
53
54
55
56 Muthén, B., Muthén, L. & Asparouhov, T. (2015). *Estimator choices with categorical*
57 *outcomes*. Verfügbar unter <https://www.statmodel.com/download/EstimatorChoices.pdf>
58
59
60
61
62
63
64
65

- 1 Paul, J., Macedo-Rouet, M., Rouet, J.-F. & Stadtler, M. (2017). Why attend to source
2 information when reading online? The perspective of ninth grade students from two
3 different countries. *Computers & Education*, *113*, 339–354.
4
5 <https://doi.org/10.1016/j.compedu.2017.05.020>
6
7
8
9
10 Perfetti, C. A., Rouet, J.-F. & Britt, M. A. (1999). Toward a theory of documents
11 representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of*
12 *mental representations during reading* (pp. 99–122). Mahwah, NJ: Lawrence Erlbaum
13 Associates.
14
15
16
17
18
19 Pohl, S., Gräfe, L. & Rose, N. (2014). Dealing with omitted and not-reached items in
20 competence tests. *Educational and Psychological Measurement*, *74*(3), 423–452.
21
22 <https://doi.org/10.1177/0013164413504926>
23
24
25
26
27 R Core Team (2016). *R: A language and environment for statistical computing*. Vienna,
28 Austria: R Foundation for Statistical Computing. Verfügbar unter [https://www.R-](https://www.R-project.org/)
29 [project.org/](https://www.R-project.org/)
30
31
32
33
34 Richter, T. & Maier, J. (2017). Comprehension of multiple documents with conflicting
35 information: A two-step model of validation. *Educational Psychologist*, *52*(3), 148–166.
36
37 <https://doi.org/10.1080/00461520.2017.1322968>
38
39
40
41 Robitzsch, A., Kiefer, T. & Wu, M. (2017). *TAM: Test analysis modules. R package version*
42 *2.3-18*. Verfügbar unter <https://cran.r-project.org/package=TAM>
43
44
45
46 Rölke, H. (2012). The ItemBuilder: A graphical authoring system for complex item
47 development. In T. Bastiaens & G. Marks (Eds.), *Proceedings of World Conference on*
48 *E-Learning in Corporate, Government, Healthcare, and Higher Education 2012* (Vol.
49 2012, pp. 344–353). Chesapeake, VA: AACE.
50
51
52
53
54
55
56 Rose, N., Von Davier, M. & Xu, X. (2010). *Modeling nonignorable missing data with item*
57 *response theory (IRT)*. Princeton: ETS.
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- Rouet, J.-F., Favart, M., Britt, M. A. & Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction*, 15(1), 85–106. <https://doi.org/10.2307/3233756>
- Sailer, M. O. (2013). *crossdes: Construction of Crossover Designs*. R package version 1.1-1. Verfügbar unter <https://CRAN.R-project.org/package=crossdes>
- Salmerón, L., Gil, L., Bråten, I. & Strømsø, H. (2010). Comprehension effects of signalling relationships between documents in search engines. *Computers in Human Behavior*, 26(3), 419–426. <https://doi.org/10.1016/j.chb.2009.11.013>
- Scharrer, L. & Salmerón, L. (2016). Sourcing in the reading process [Special issue]. *Reading and Writing*, 29(8).
- Schmalhofer, F. & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language*, 25(3), 279–294. [https://doi.org/10.1016/0749-596X\(86\)90002-1](https://doi.org/10.1016/0749-596X(86)90002-1)
- Schwabe, F., McElvany, N. & Trendtel, M. (2015). The school age gender gap in reading achievement: Examining the influences of item format and intrinsic reading motivation. *Reading Research Quarterly*, 50(2), 219–232. <https://doi.org/10.1002/rrq.92>
- Stadtler, M. & Bromme, R. (2014). The content–source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In D. N. Rapp & J. L. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 379–402). Cambridge, MA: MIT Press.
- Stadtler, M., Scharrer, L., Skodzik, T. & Bromme, R. (2014). Comprehending multiple documents on scientific controversies: Effects of reading goals and signaling rhetorical

relationships. *Discourse Processes*, 51(1–2), 93–116.

<https://doi.org/10.1080/0163853x.2013.855535>

Strømsø, H. I., Bråten, I., Britt, M. A. & Ferguson, L. E. (2013). Spontaneous sourcing among students reading multiple documents. *Cognition and Instruction*, 31(2), 176–203.

<https://doi.org/10.1080/07370008.2013.769994>

Wang, W.-C. & Wilson, M. (2005). The Rasch testlet model. *Applied Psychological Measurement*, 29(2), 126–149. <https://doi.org/10.1177/0146621604271053>

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427–450. <https://doi.org/10.1007/bf02294627>

Weinert, F. E. (1999). *Definition and selection of competencies - concepts of competence*. Munich, Germany: Max Planck Institute for Psychological Research.

Weinert, F. E. (2001). Concept of competence. A conceptual clarification. In D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* (pp. 45–65). Göttingen: Hogrefe & Huber.

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83(1), 73–87. <https://doi.org/10.1037/0022-0663.83.1.73>

Wright, B. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.

Running Head: TEST ZUM VERSTÄNDNIS MULTIPLER DOKUMENTE FÜR
STUDIERENDE

Entwicklung und Skalierung eines Tests zur Erfassung des Verständnisses multipler
Dokumente von Studierenden

Cornelia Schoor^a, Carolin Hahnel^{b, c}, Cordula Artelt^{a, d}, Daniel Reimann^d, Ulf Kröhne^b &
Frank Goldhammer^{b, c}

^a Otto-Friedrich-Universität Bamberg, Lehrstuhl für Empirische Bildungsforschung,
Markusplatz 3, 96047 Bamberg

^b DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, Rostocker Straße
6, 60323 Frankfurt am Main

^c Zentrum für internationale Vergleichsstudien (ZIB), Frankfurt am Main

^d Leibniz-Institut für Bildungsverläufe e.V., Wilhelmsplatz 3, 96047 Bamberg

Danksagungen

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01PK15008 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren. Zudem danken wir dem anonymen Mitherausgeber und den anonymen Gutachterinnen und Gutachtern für ihre konstruktive Rückmeldung zur vorherigen Fassung des Manuskripts.

Zusammenfassung (150 Wörter)

Das Verständnis multipler Dokumente (Multiple Document Comprehension, MDC) wird als Fähigkeit verstanden, aus verschiedenen Informationsquellen eine integrierte Repräsentation eines inhaltlichen Gegenstandsbereichs zu konstruieren. Als solche ist sie sowohl für die erfolgreiche Bewältigung eines Studiums als auch für gesellschaftliche Partizipation eine wichtige Kompetenz. Bislang gibt es jedoch kein etabliertes Diagnostikum in diesem Bereich. Um diese Lücke zu schließen, wurde ein Test entwickelt, der vier zentrale kognitive Anforderungen von MDC abdeckt und auf Basis der Daten von 310 Studierenden sozial- und geisteswissenschaftlicher Fächer überprüft wurde. Die im MDC-Test gemessene Kompetenz erwies sich als eindimensional. Der MDC-Testwert wies theoriekonforme Zusammenhänge mit der Abiturnote, dem Studienabschnitt und der Leistung in einer Essay-Aufgabe auf. Insgesamt liefern die Ergebnisse empirische Belege dafür, dass der Testwert aus dem MDC-Test die fächerübergreifende Fähigkeit von Studierenden wiedergibt, multiple Dokumente zu verstehen.

Schlüsselwörter (3-5): Verständnis multipler Dokumente, Studierende, Test, Multiple Document Literacy

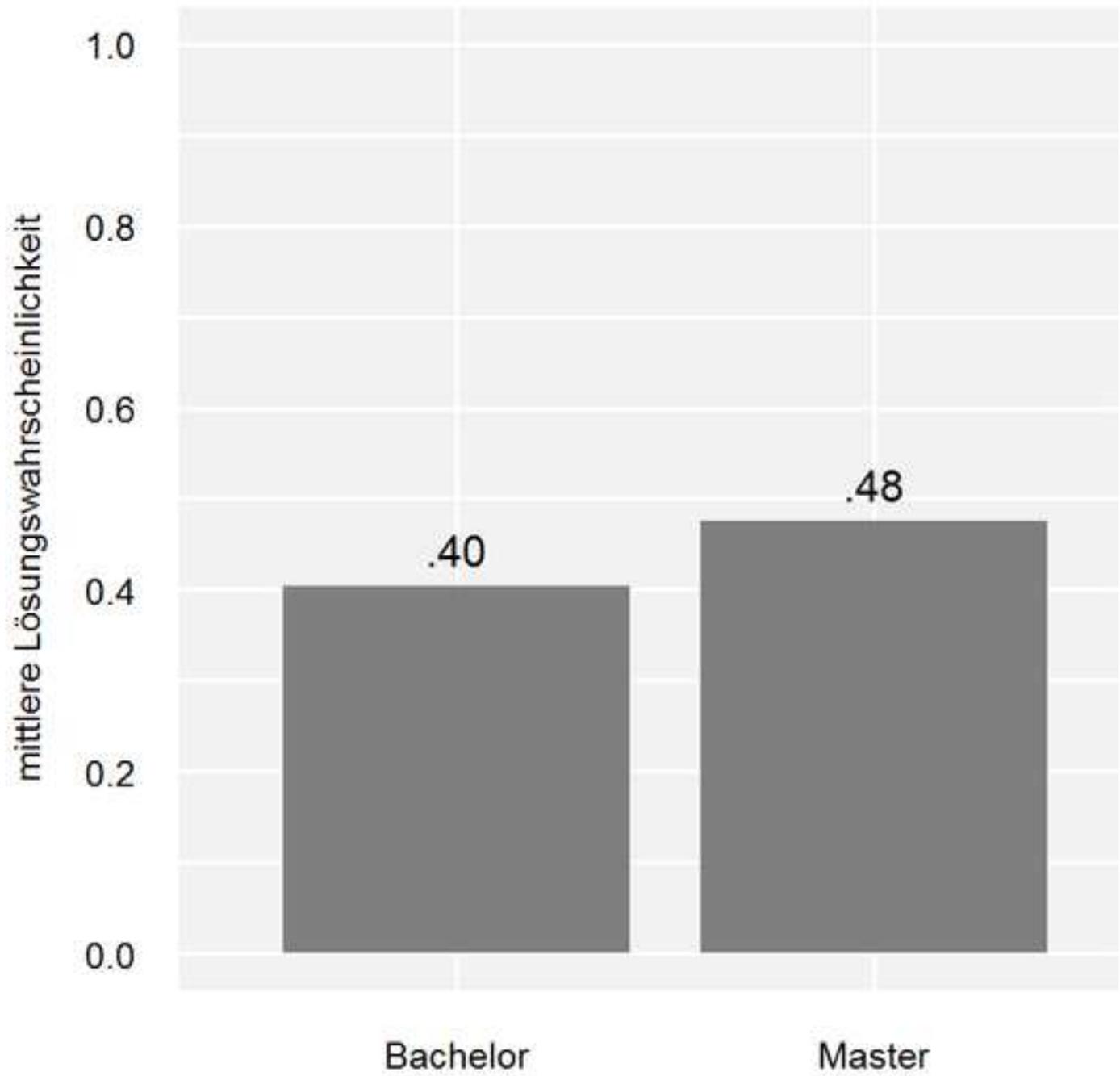
Abstract (150 Wörter)**Developing and Scaling a Test of Multiple Document Comprehension in University****Students**

Multiple document comprehension (MDC) is defined as the ability to construct an integrated representation based on different sources of information on a particular topic. It is an important competence for both the successful accomplishment of university studies and participation in societal discussions. Yet, there is no established assessment instrument for MDC. Therefore, we have developed a test covering four theory-based cognitive requirements of multiple document comprehension. Based on data of 310 university students of social sciences and humanities, the MDC test proved to be a unidimensional measure. Furthermore, the test score was related to the final school exam grade, the study level (bachelor/master), and the performance in an essay task. All in all, the empirical results suggest that the score of the MDC test can be interpreted as the generic competence of university students to understand multiple documents.

Key words (3-5): multiple document comprehension, university students, assessment, multiple document literacy

Fachgebiete: Pädagogische Psychologie

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

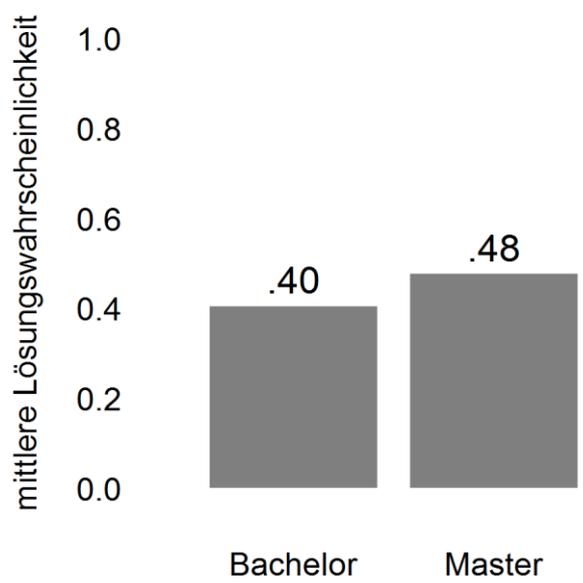
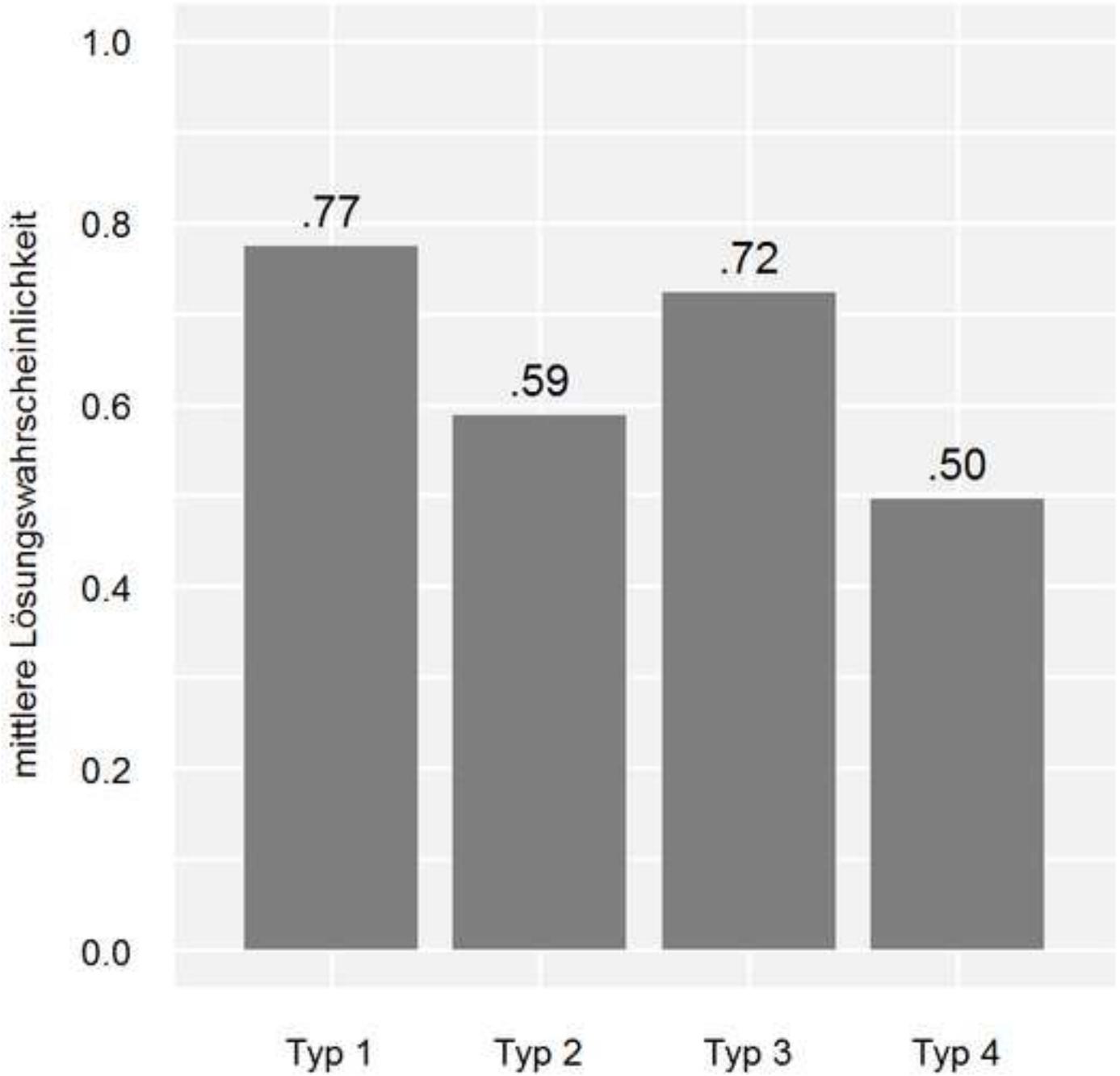


Abbildung 1. Mittlere geschätzte Lösungswahrscheinlichkeiten der Items für Bachelor- vs. Master-Studierende.



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

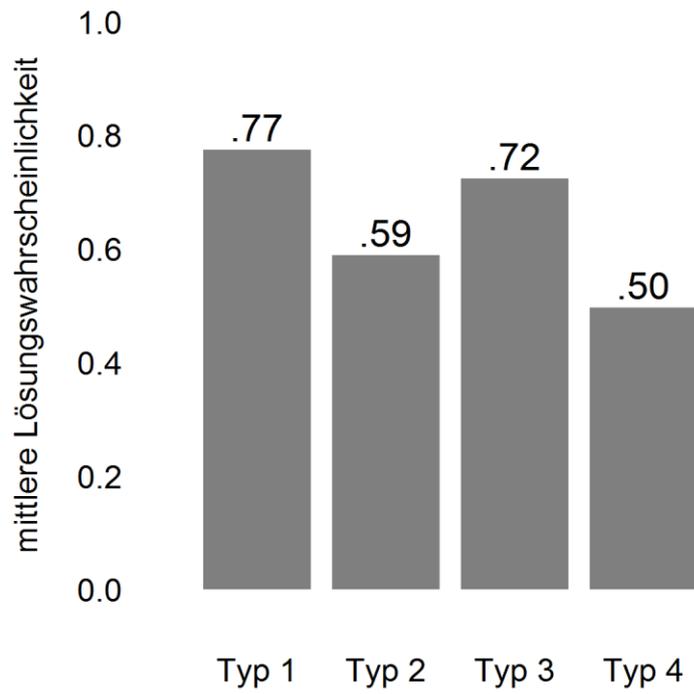


Abbildung 2. Mittlere geschätzte Lösungswahrscheinlichkeit der Items in Abhängigkeit vom Anforderungstyp.

Tabelle 1. Informationskriterien für die geschätzten Modelle sowie Ergebnisse des Modellvergleichs mit dem eindimensionalen Modell

	AIC	BIC	n_{par}	χ^2	df	p
Ausgewählte 67 Items						
eindimensionales Modell	11 687.25	11 941.34	68			
vierdimensionales Modell (Itemanforderungen)	11 697.12	11 984.83	77	8.13	9	.52
fünfdimensionales Modell (Units)	11 697.93	12 004.33	82	17.33	14	.24
Rasch-Testlet-Modell	11 715.97	12 044.79	88	11.28	20	.94
Alle 174 Items						
eindimensionales Modell	29 311.07	29 964.97	175			
vierdimensionales Modell (Itemanforderungen)	29 308.12	29 995.65	184	20.95	9	< .05
sechsdimensionales Modell (Units)	29 311.92	30 040.55	195	39.15	20	< .01
Rasch-Testlet-Modell	29 349.98	30 104.77	202	15.09	27	.97

Anmerkung: n_{par} = Anzahl Parameter.

Tabelle 2. Ergebnisse des GLMM zur Vorhersage der Lösung eines Items durch den Anforderungstyp (Referenzkategorie: Anforderungstyp 1) sowie Kontraste der Lösungswahrscheinlichkeiten

	B (SE)	Varianz
Zufallseffekte		
Person		0.37
Item		0.74
Feste Effekte		
Interzept	1.23 ^{***} (0.21)	
Anforderungstyp 2	-0.87 ^{**} (0.29)	
Anforderungstyp 3	-0.27 (0.30)	
Anforderungstyp 4	-1.24 ^{***} (0.31)	
Geplante Kontraste		
Anforderungstyp 1 vs. 2, 3, 4	0.80 ^{***} (0.24)	
Anforderungstyp 2 vs. 1, 3, 4	-0.37 (0.24)	
Anforderungstyp 3 vs. 1, 2, 4	0.43 ⁺ (0.25)	
Anforderungstyp 4 vs. 1, 2, 3	-0.86 ^{**} (0.26)	

Anmerkungen: ⁺ $p < .10$. ^{**} $p < .01$. ^{***} $p < .001$. AIC = 11 836.7, BIC = 11 880.2, Deviance = 11 824.7, $df_{\text{resid}} = 10\ 331$. GLMM = generalized linear mixed model (generalisiertes lineares gemischtes Modell). B = unstandardisierter Regressionskoeffizient.

Danksagungen

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01PK15008 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren. Zudem danken wir dem anonymen Mitherausgeber und den anonymen Gutachterinnen und Gutachtern für ihre konstruktive Rückmeldung zur vorherigen Fassung des Manuskripts.