Studie 1 – Durchführung der Skalenkonstruktion

9

Im Folgenden wird die Durchführung der Online-Befragungen der ersten Studie zur Skalenkonstruktion dokumentiert. In diesem Rahmen wurden drei Erhebungen durchgeführt, die nachfolgend als Erhebung 1.1, 1.2 und 1.3 bezeichnet werden. Die erste Erhebung 1.1 diente der Identifikation geeigneter Frageitems und der Prüfung der zugrunde liegenden Modellstruktur, die Erhebungen 1.2. und 1.3 der Validierung der in Erhebung 1.1 ermittelten Modellstruktur. Dabei wird zunächst am Beispiel der Erhebung 1.1 detailliert auf den Aufbau und den empirischen Einsatz der Skala in der Befragung eingegangen. Zudem werden einmal ganz allgemein die Auswertungsschritte der Datenanalyse mittels Strukturgleichungsmodellierung nachvollziehbar begründet und dargelegt.

9.1 Erhebung 1.1 der Skalenkonstruktion – Identifikation geeigneter Frageitems und Prüfung der Modellstruktur

Ziel dieser Erhebung war die Identifikation geeigneter Frageitems und eine erste Prüfung der konzeptuellen Modellstruktur. Nachdem die entwickelten Items die

---

¹ Alle Fragebögen der berichteten Studien im Rahmen dieser Arbeit wurden mit der Befragungssoftware des Anbieters SoSci Survey GmbH programmiert (D. J. Leiner, 2019a). Weitere Informationen unter und <https://www.sosciurvey.de/> [zuletzt zugegriffen am 25.04.2020].

Ergänzende Information Die elektronische Version dieses Kapitels enthält Zusatzmaterial, auf das über folgenden Link zugegriffen werden kann https://doi.org/10.1007/978-3-658-36368-0_9.

Experten- und Nutzerinterviews durchlaufen hatten und der Fragepool angepasst und überarbeitet wurde, wurden die Items mit Hilfe der Fragebogensoftware *SoSci Survey*¹ für die Erhebung 1.1 in einen Online-Fragebogen überführt.

9.1.1 Aufbau des Fragebogens

Der Aufbau des Fragebogens sah vor, dass die Befragten zunächst begrüßt und über das Thema und die ungefähre Zielsetzung der Befragung aufgeklärt wurden. Die Befragten wurden eingeladen an einer „Umfrage zur Wahrnehmung und Bewertung der Digitalisierung der Gesellschaft“ teilzunehmen, deren Beantwortung ca. 15 Minuten in Anspruch nehmen sollte. Anschließend folgten die Fragen zur Soziodemografie der Befragten (Alter, Geschlecht, Bildungsniveau und Berufstätigkeit), die zur Quotierung genutzt wurden.

Es folgte dann das Briefing für die Items der Skalenkonstruktion. Ziel des Briefings war es, den Befragten die Zielsetzung und den Befragungsgegenstand zu verdeutlichen, weshalb der mit dem Ausfüllen verbundene Auftrag zusammen mit dem Einleitungstext (siehe Abschnitt 8.4.7) vorgestellt wurde. Das Briefing lautete wie folgt (Hervorhebungen wie im Fragebogen):

Der Gegenstand dieser Studie: die Digitalisierung der Gesellschaft

Die zunehmende Verbreitung von digitalen Technologien wie beispielsweise von Computern und mobilen Endgeräten im privaten Alltag, aber auch in Wirtschaft und öffentlicher Verwaltung führt dazu, dass wir alle laufend Datenspuren hinterlassen, die gespeichert, miteinander kombiniert und ausgewertet werden können. Die hieraus entstehenden großen digitalen Datenbestände werden häufig auch als Big Data bezeichnet.

Im Folgenden möchten wir Sie zu Ihrer **persönlichen Einschätzung zu Entstehung und Verwendung dieser digitalen Datenbestände** befragen. Hierbei gibt es **weder falsche noch richtige Antworten**, uns interessiert Ihre ganz individuelle Meinung.

Bitte lesen Sie daher die folgenden Aussagen aufmerksam durch und **beantworten Sie ehrlich und spontan, inwieweit Sie den Aussagen zustimmen**.

Einige der folgenden Aussagen mögen Ihnen dabei recht ähnlich vorkommen. Dies ist allerdings beabsichtigt und hilft uns Ihre Einschätzungen insgesamt besser zu verstehen, da im Rahmen dieser Studie ein neues Erhebungsinstrument entwickelt werden soll.

Der Fragebogen umfasste 93 Items zur Wahrnehmung digitaler Daten und beinhaltete die 69 zugehörigen Items der Beschreibungsdimensionen von Big Data sowie eine Subskala zu *Technological Solutionism* (TS – Morozov, 2014).² Die insgesamt 69 Items der BDGS-Skala wurden zusammen mit den TS-Items randomisiert und in Frageblöcke zu jeweils zehn Items aufgeteilt. Um die Beantwortung der Frageitems nicht zu eintönig werden zu lassen und Ermüdungseffekten vorzubeugen, wurde nach etwas mehr als der Hälfte der Fragen eine Fragebatterie mit sechs Items zur Nutzung politischer Informationen im Internet während des zum Zeitpunkt der Befragung aktuell laufenden Bundestagswahlkampfes in Deutschland eingeschoben. Dieser Fragenblock hatte keinen Einfluss auf die Reihenfolge der Randomisierung, so dass jedes Item sowohl zufällig vor oder nach dieser Itembatterie angezeigt werden konnte.

Die Rekrutierung für einen Pre-Test des Fragebogens fand über Facebook statt, indem Freunde und Bekannte auf die Befragung hingewiesen wurden. Insgesamt haben 16 Personen den Fragebogen vollständig beantwortet, die im Durchschnitt 11,23 Minuten (SD = 3,45) für die Beantwortung der Fragen brauchten. Es konnten keine Probleme bei der Befragung festgestellt werden. Am Ende des Pre-Test-Fragebogens bestand noch die Möglichkeit für die Befragten, Kommentare zum Fragebogen zu hinterlassen. Einige wenige Kommentare bezogen sich lediglich auf die Länge und Eintönigkeit der Fragebatterien und deren Formulierungen. Nachdem der Online-Fragebogen und die Items der zu konstruierenden Skala die vorgehend ausführlich beschriebene Testphase durchlaufen hatten, konnte der finale Fragebogen ins Feld gehen. Der finale Fragebogen findet sich im Anhang 3.1 im elektronischen Zusatzmaterial.

9.1.2 Feldphase, Datenbereinigung und finale Stichprobe

Die Feldphase der Befragung lief vom 18. bis zum 26. September 2017. Die Rekrutierung der Befragten erfolgte durch die *Respondi AG*, die hierzu auf ihr

---

² In Erhebung 1.1 wurden noch Frageitems aufgenommen, die speziell auf eine Problemlösung durch digitale Daten abzielten und inspiriert waren von Morozovs (2014) Konzept des *Technological Solutionism* (TS). Die 24 Frageitems für TS zielten auf zwei Subdimensionen ab, wobei zwölf Items eine *individuelle Problemlösung* durch digitale Daten betrafen und zwölf Items auf eine *gesamtesellschaftliche Problemlösung* abstellten. Diese Items und die Prüfung ihrer Dimensionierung werden mit Rücksicht auf den Umfang der vorliegenden Arbeit nicht weiter besprochen. Da sie jedoch in der ersten Befragung einen gewissen Teil des Fragebogens einnahmen und daher auch aus Gründen der Vollständigkeit in der Itemliste in der Tabelle 2.C. im Anhang im elektronischen Zusatzmaterial aufgeführt sind, wird an dieser Stelle kurz auf sie hingewiesen.

Online-Access-Panel zugriff, das durch die international gültige Norm ISO 26362 zertifiziert ist.³

Auswahllogik der Stichprobe der ersten Befragung der Skalenkonstruktion

Um ein möglichst ausgeglichenes Sample für die erste Erhebung der Skalenkonstruktion zu erhalten, wurde bei der Rekrutierung eine Quotierung vorgesehen und mit dem Panelanbieter vereinbart. Das Geschlechterverhältnis sollte ausgeglichen sein. Zudem sollten annähernd gleich viele Befragte mit hohem und niedrigem bis mittlerem Bildungsniveau rekrutiert werden. Auch im Hinblick auf das Alter sollten zum einen Befragte ins Sample gelangen, die zu den sogenannten Digital Natives gehören (Prensky, 2001), also zu den Personen zählen, die bereits in einer frühen Lebensphase mit digitalen Medien sozialisiert wurden. Zum anderen sollte eine annähernd gleich große Gruppe von Personen am Fragebogen teilnehmen, die erst in einer späteren Lebensphase regelmäßigen Kontakt mit digitalen Medien hatte und noch Zeiten erlebt hat, in denen es noch keine hohe gesellschaftliche Durchdringung des Internets gab. Hinter diesem Vorgehen stand die Überlegung, dass in späteren Befragungseinsätzen der Skala die Population aus allen Online-Nutzer*innen der Bundesrepublik Deutschland besteht und somit eine Vielzahl an Personen und demografischen Milieus die Skala erfolgreich verstehen und beantworten können muss. Dabei gibt es mit Blick auf die Literatur eine reichhaltige Diskussion, ob und wie zwischen Digital Natives und Digital Immigrants unterschieden werden kann (Helsper & Eynon, 2010; Margaryan et al., 2011). Eine entsprechende Aufteilung schien hier ebenfalls geboten sowie möglich. Die Unterteilung in Digital Natives und Digital Immigrants für Erhebung 1.1 wurde wie folgt vorgenommen: Mit Blick auf Usenet-Gruppen und andere an Privatleute gerichtete Online-Angebote, die zu Beginn der 1980er Jahre Verbreitung fanden (Hauben & Hauben, 1997), wurden alle Befragte, die 1982 oder später geboren wurden, der Gruppe U35 zugewiesen, alle Befragten, die vor 1982 geboren wurden der Gruppe Ü35.

Es wurden 1568 Einladungen verschickt. Insgesamt wurde der Fragebogen 374-mal aufgerufen. Mit der Beantwortung begannen 356 Befragte, wovon wiederum 306 Befragte den Fragebogen beendeten und die letzte Fragebogenseite erreichten. Die Antwortrate lag somit bei 19,5 %.

---

³ Weitere Informationen finden sich unter <https://www.respondi.com/access-panel> [zuletzt zugegriffen am 31. Mai 2019].

Datenbereinigung

Der vollständige Datensatz mit 306 abgeschlossenen Interviews wurde anschließend bereinigt. Bei dieser Bereinigung wurden zwei Ausschlusskriterien angelegt: Zum einen wurde die Beantwortungszeit betrachtet, zum anderen wurde nach offensichtlichen Mustern im Antwortverhalten ‚gescreent‘ (sprich: regelgeleitet geprüft), wobei der Fokus auf dem sogenannten ‚Durchklicken‘ lag. Das bedeutet bspw., dass Befragte immer nacheinander weg die gleiche Antwort geben und somit keine Varianz im Antwortverhalten besteht oder von ‚links oben nach rechts unten‘ klicken, also zwar Varianz entsteht, diese aber keine sinnhafte Bedeutung beinhaltet und somit keine tatsächliche Differenzierung ermöglicht. Die ausführliche Prüfung wird im nachfolgenden Abschnitt dokumentiert. In den weiteren berichteten Erhebungen wird diese strenge Prüfung allein aus forschungsökonomischen Gründen nicht in entsprechendem Maße durchzuhalten sein. Zwar wird bei jeder durchgeführten Erhebung eine Mindestdauer formuliert, die erreicht werden muss, da andernfalls nicht davon ausgegangen werden kann, dass die Beantwortung sinnhaft geleistet wurde. Die intensive Prüfung auf Antwortmuster jedoch ist (a) zeitintensiv und (b) es gibt kein allgemeingültiges Vorgehen bei der Prüfung auf invalide Antwortmuster. Antwortverhalten, das keine Varianz aufweist, lässt sich noch verhältnismäßig leicht erkennen und ist eher unwahrscheinlich. Willkürlich erscheinende Antwortmuster sind jedoch schwieriger zu erkennen und müssen nicht zwingend invalide sein. Eine Möglichkeit, systematisch auf sinngemäße und valide Antwort zu prüfen, sind sogenannte ‚Attention Checks‘, wie sie in Erhebung 1.3 und 2.1 eingesetzt wurden. Da die erste Erhebung zur Skalenkonstruktion nur auf ein kleines Sample zurückgreifen konnte und insbesondere die Differenzierung zwischen den Dimensionen im Fokus stand, wurden hier der erhöhte Aufwand und eine gründliche Plausibilitätsprüfung der Daten betrieben.

Screenout durch zu schnelles Beantworten

Im Durchschnitt benötigten die Befragten rund 8 Minuten und 47 Sekunden ($SD = 4$ Minuten 27,5 Sekunden). Es wurde vor Durchführung der Befragung mit dem finalen Fragebogen getestet, wie lange ein Befragter mindestens brauchen würde, um den Fragebogen zu beenden, wenn er oder sie nur die Itemformulierungen, nicht aber die Begrüßung, Briefing, Einleitungen und Instruktionen lesen und zügig beantworten würde. Der durch diesen Test mit mehreren Probetestern

ermittelte Wert, zu dem auch die Ergebnisse aus dem Pre-Test herangezogen wurden, lag bei genau 3 Minuten. Es wurden daher alle Befragten, die weniger als 3 Minuten gebraucht haben und somit mehr als 1,3 Standardabweichungen nach unten abwichen, aus dem Datensatz ausgeschlossen. Insgesamt 28 Befragte und somit 9,2 % des gesamten Samples benötigten weniger als 3 Minuten für alle Fragen und wurden folglich für die Auswertung ausgeschlossen.

Screenout durch Beantwortung in Antwortmustern

Um sogenannte Durchklicker zu identifizieren, deren Antwortverhalten sich auf fast immer die gleiche Antwort auf alle Frageitems auszeichnet, wurden die Antworthäufigkeiten über alle 93 Items der Skala zur Wahrnehmung digitaler Daten ausgezählt. Da es hier jedoch keine allgemeingültigen Daumenregeln zur Identifizierung gibt, wurde recht liberal festgelegt, dass nicht mehr als 90 % der Antworten auf den gleichen Skalenpunkt entfallen durften, was ab 83 von 93 möglichen Antworten der Fall war. Immerhin ist es möglich, dass bestimmte Befragte eine Tendenz zur Mitte hatten und durchweg „Teils/teils“ angaben oder gar keine Antwort geben konnten und „Kann ich nicht beantworten“ ankreuzten. Daher wurden lediglich Befragte ausgeschlossen, die 83 oder mehr gleiche Antworten gaben, was auf 36 Befragte, also eine ‚Click-Through-Rate‘ von 11,8 % des Samples zutraf.

Da es Befragte gab, die zu beiden Gruppen gehörten, wurden 51 Personen aus dem Auswertungsdatensatz ausgeschlossen, so dass von 306 vollständig ausgefüllten Fragebögen noch 255 Fragebögen in die Auswertung eingingen.

Der Tabelle 9.1 kann nicht nur die demografische Zusammensetzung des finalen Samples entnommen werden, sondern sie enthält auch die demografische Zusammensetzung der ausgeschlossenen Fälle.

Tabelle 9.1 Unterschiede in der Soziodemografie zwischen gültigen und ungültigen Fällen im Sample der Erhebung 1.1

Soziodemografische Variable	Aufgenommene Fälle		Ausgeschlossene Fälle	
	absolut	relativ	absolut	relativ
Geschlecht				
Männlich	128	50.2 %	23	45.1 %
Weiblich	127	49.8 %	28	54.9 %
Alter				
Unter 35 Jahren	113	44.3 %	36	70.6 %
Über 35 Jahren	142	55.7 %	15	29.4 %
Bildung				
Niedrige/mittlere Bildung	112	43.9 %	40	78.4 %
Hohe Bildung	143	56.1 %	11	21.6 %
<i>n</i>	255	83.3 %	51	16.6 %

Hier zeigt sich, dass vor allem jüngere und niedrig bis mittel gebildete Befragte den Fragebogen nicht erwartungsgemäß ausgefüllt haben. Dies führt zu einer leicht ungleichen Verteilung des finalisierten Samples.

Finale Stichprobe

Die finale Stichprobe bestand aus $n = 255$ Befragten, die den Fragebogen vollständig beantwortet haben. Dieses Sample setzt sich dabei wie folgt zusammen: Männer ($n = 128$) und Frauen ($n = 127$) waren fast gleich stark vertreten. Das Durchschnittsalter der Befragten lag bei etwas mehr als 43 Jahren ($M = 43,1$; $SD = 16,6$). Abitur oder einen höheren Abschluss hatten 56,1 % der Befragten, einen niedrigeren Bildungsabschluss hatten 43,9 % der Befragten. Obwohl die Quotierung den Anspruch hatte, für diese drei Variablen annähernd gleich große Gruppen zu rekrutieren, ergab sich wie zuvor dokumentiert durch die Datenbereinigung eine leicht unterschiedliche Gruppengröße. Es sind allerdings für die konfirmatorische Faktorenanalyse keine nachteiligen Auswirkungen zu erwarten, zumal zunächst Items identifiziert werden sollen, die die vermuteten Konstruktdimensionen repräsentativ und reliabel abbilden. Es erfolgt keine Aussage zur Prävalenz der Wahrnehmung digitaler Daten in der Bevölkerung der Bundesrepublik.

9.1.3 Vorgehen bei der Datenanalyse mittels konfirmatorischer Faktorenanalyse durch Strukturgleichungsmodellierung

Im folgenden Abschnitt wird die Datenanalyse ausführlich erörtert, wobei zunächst allgemein auf das Vorgehen bei der konfirmatorischen Faktorenanalyse (‘Confirmatory Factor Analysis’ – kurz: CFA) durch Strukturgleichungsmodellierung eingegangen wird – ein Datenanalyseverfahren, das in allen nachfolgenden Erhebungen angewandt wurde und daher an dieser Stelle grundlegend begründet und erörtert werden soll. Da die Items der einzelnen Dimensionen auf Grundlage der konzeptuellen Modellierung des ersten Teils der Arbeit zum *BDGS* gezielt als manifeste Indikatoren eines zugehörigen latenten Faktors formuliert waren und in ihrem Verhältnis zueinander ausgewogen erstellt wurden, gebot es sich, die vermutete Modellstruktur direkt durch eine konfirmatorische Faktorenanalyse zu prüfen (Bandalos, 2017). Es wurde mithin darauf verzichtet, eine Strukturprüfung ohne vorherige Annahmen durchzuführen wie etwa im Rahmen der explorativen Faktorenanalyse.

Die Auswertung der ersten Erhebung zielte darauf ab, die Items zu identifizieren, die geeignete reflektive Indikatoren sind, die zuvor beschriebenen einzelnen Dimensionen abzubilden und zu erheben. In den nachfolgenden Erhebungen wurden diese Indikatoren dann wiederholt erhoben und die Modellstruktur wurde erneut mittels CFA geprüft. Das datenanalytische Vorgehen unterscheidet sich daher zwischen den Erhebungen kaum: Die Auswertungslogik und die verwendeten Kennwerte bleiben konstant, lediglich in der ersten Erhebung wurde aufgrund der ersten Identifikation geeigneter Indikatoren eine ausführliche Modellrespezifikation vorgenommen.

Für die Datenauswertung der Messmodelle in den insgesamt sieben Erhebungen wurden konfirmatorische Faktorenanalysen durchgeführt, um die jeweiligen reflektiven Messmodelle sowie die Zusammenhänge und die Dimensionalitätszugehörigkeit der beobachteten Messungen zu prüfen. Ziel der Datenanalyse mittels Strukturgleichungsmodellierung ist „die empirische Prüfung der durch das Strukturmodell abgebildeten theoretisch vermuteten Zusammenhänge“ (Weiber & Mülhhaus, 2014, S. 126). Grundannahme der CFA ist, dass das modellierte latente Konstrukt – im vorliegenden Fall somit die jeweilige interessierende Dimension – als exogene Variable einen Einfluss auf die Beantwortung der Fragebogenitems hat, die als Indikatoren für die Dimension herangezogen werden und im jeweiligen Modell als manifeste (d. h. tatsächliche Beobachtungen) endogene Variablen aufgenommen werden. Alle nachfolgenden Auswertungen zur Prüfung wurden mit den Datenanalyseprogrammen SPSS 23 vorbereitet

und mit AMOS 23 (Analysis of Moment Structures) durchgeführt, das einem kovarianzanalytischen Ansatz folgt (Weiber & Mülhhaus, 2014). Die Prüfung der Messinvarianz wurde aus Gründen der Praktikabilität in Teilen mit dem Datenanalyse-Programm R (R Core Team, 2020) und dem Package lavaan durchgeführt (Rosseel, 2012). Für alle Analysen wurde die standardmäßig in AMOS voreingestellte Maximum-Likelihood-Methode zur Modellschätzung gewählt.⁴ Da es jedoch zu fehlenden Daten kam und Modifikationsindizes zur Modellrespezifikation herangezogen werden sollten, die jedoch nur für vollständige Datensets bereitgestellt werden können, wurden Daten teilweise imputiert. Das Vorgehen bei der Modell-Schätzung unter Berücksichtigung der Datenimputation wird nachfolgend erläutert.

Imputation der Daten

Da es vorkam, dass Befragte nicht alle Items beantworteten oder die „weiß nicht“-Option wählten, kam es vor, dass nicht für jeden Fall im Datensatz für alle 93 abgefragten Items Daten vorlagen. Für die Identifikation der auszuwählenden Items sollte jedoch auf die Modifikationsindizes des Programms AMOS zurückgegriffen werden, um Items bzw. deren Fehlervarianzen zu identifizieren, die miteinander oder den latenten Faktoren kovariierten, obwohl das die theoretischen Überlegungen und folglich auch das Messmodell nicht vorsehen. Da AMOS Modifikationsindizes nur ausgeben kann, wenn vollständige Daten vorliegen und die Schätzung nach der Maximum-Likelihood-Methode durchgeführt wird, wurden für die Datenauswertung die fehlenden Daten im Datensatz der Erhebung 1.1 mittels der AMOS-Funktion *Regression Imputation* imputiert. Bei der Imputation ist davon auszugehen, dass die fehlenden Daten „missing at random“ (Kline, 2011, S. 55 f.) sind, also fehlende Angaben nicht systematisch zustande kommen. Insbesondere für die „weiß nicht“-Antworten kann bei fehlenden Werten nicht davon ausgegangen werden, dass diese „missing completely at random“ sind – eine Annahme, die laut Kline (2011, S. 56) bei realen Datensätzen jedoch kaum gegeben sein dürfte und nur schwer zu prüfen ist. Für die Antworten, die bewusst keine Angabe auf der 7-Punkt-Likert-Skala enthalten, kann diese Annahme des zufälligen Fehlens daher nicht gemacht werden. Sie wird für die Imputation jedoch angenommen, um die Modellprüfung mit einem ausreichend

---

⁴ Es wird mithin von einer Normalverteilung der erhobenen Indikatoren ausgegangen. Diese Annahme wird nach visueller Inspektion der Verteilungen der Variablen beibehalten, auch wenn ein Kolmogorov-Smirnov-Test auf Normalverteilung der Fallzahlen zu einem gegenteiligen Ergebnis kommt. Für eine Übersicht über alternative Modell-Schätzverfahren bei der groben Verletzung dieser Annahme siehe die Übersicht von Bandalos (2017, S. 368 ff.).

großen Datensatz zu ermöglichen. Insgesamt lagen von 199 Befragten vollständige Datensätze vor. Im Schnitt fehlten pro Item rund 3,5 % der Werte, die mittels der Regressionsimputation imputiert werden mussten. Lediglich bei zwei Variablen⁵ wurden 18 „weiß nicht“-Antworten gezählt, so dass im Maximalfall knapp 7 % der Angaben für ein Item fehlten.

Das finale Modell wurde aufgrund der Imputation der Daten daher auch noch einmal mittels der Full-Information-Maximum-Likelihood-Methode (FIML) berechnet (siehe Abschnitt 9.1.6), um sich gegen etwaige Probleme abzusichern, die durch die Imputation entstanden sein könnten (Kline, 2011). Die erste Spezifikation sowie die Respezifikationsschritte wurden auf Grundlage des durch Imputation neu entstandenen Datensatzes durchgeführt.

Vorgehen bei der Modellspezifikation und -respezifikation

Der idealtypische Ablauf der Modellspezifikation und -respezifikation ist in Abbildung 9.1 dargestellt. Zunächst wurden alle Frageitems einer Dimension in das jeweilige Messmodell eingefügt. Es wurde in allen Messmodellen zur Skalierung ein Referenzindikator durch ein *Unit Loading Identification Constraint* (ULI; Kline, 2011, S. 127) mit dem Regressionsgewicht = 1 festgelegt.⁶ Bei vollständigen Datensätzen (wie bspw. den Modellen auf Grundlage imputierter Datensätze) wurden die Modelle ohne Mittelwerte (Means) und Regressionskonstanten (auch als Intercepts bezeichnet) berechnet und es wird an entsprechender Stelle gesondert darauf hingewiesen, wenn die Modelle mittels FIML geschätzt wurden.

Abbildung 9.1 Schritte der Modell-(Re)-Spezifikation

⁵ Siehe Items AV01_44 und AV01_85 im Anhang 2.1 im elektronischen Zusatzmaterial.

⁶ Auch für die Störgrößen in Strukturmodellen und die Messfehler in Messmodellen wurden wie üblich ULI Constraints mit dem Regressionsgewicht = 1 festgelegt (Kline, 2011).

Um geeignete Fragebogenitems zu identifizieren und ungeeignete Items auszuschließen, wurden drei unterschiedliche Strategien angewendet, die nachfolgend allgemein beschrieben werden. Dabei werden im Laufe der jeweiligen Respezifikationschritte unzählige kleine und größere Entscheidungen getroffen, die hier im Rahmen der Verschriftlichung des Vorgehens nicht alle bis ins Detail gelistet und nachvollzogen werden können. Es ist bei der Modellspezifizierung abzuwägen zwischen größtmöglicher Transparenz und Sparsamkeit in der Dokumentation. Anhand des zur Verfügung gestellten Datensatzes und der berichteten Modelle können jedoch jederzeit alternative oder (fast-)äquivalente Modelle geprüft werden (*(Near-)Equivalent-Models*, Kline, 2011).

Begutachtung der standardisierten Faktorladungen

Zunächst wurden Items ausgeschlossen, die nur einen geringen Erklärungsbeitrag zum jeweiligen latenten Konstrukt lieferten. Dies war der Fall, wenn die standardisierten Faktorladungen der gemessenen Variable unter einem Wert von .70 lagen, da hier weniger als die Hälfte der Varianz einer Variable durch das zugrunde liegende Konstrukt erklärt werden kann. Werte > .70 der standardisierten Faktorladungen hingegen indizieren Konvergenzvalidität der manifesten Indikatoren eines latenten Konstrukts (Kline, 2011).

Die Verwendung von Modifikationsindizes

Weiterhin wurden mittels der Modifikationsindizes Items identifiziert, deren Fehlervarianzen miteinander kovariierten (Weiber & Mülhhaus, 2014). Das bedeutet konkret, dass diese Items noch durch eine andere zusätzliche latente Variable, die nicht im Modell abgebildet war, beeinflusst wurden, die deren gemeinsam geteilte Varianz zusätzlich zum zufälligen Messfehler erklären kann. Ist dieser Pfad nicht vorhanden, also nicht im Modell abgebildet⁷ und somit nicht freigegeben, ist die Annahme, dass in vorliegendem Fall keine Beziehung zwischen den Fehlervarianzen besteht, weil sich gemeinsame Varianz nur über den modellierten latenten Faktor erklären sollte. Hierauf stellen nun die berechneten Modifikationsindizes ab: "(...) A modification index estimates the amount by which the overall model chi-square statistic, X^2_M , would decrease if a particular fixed-to-zero parameter were freely estimated" (Kline, 2011, S. 217). Die Modifikationsindizes schlagen dann vor, diesen nicht vorhandenen Pfad in das Modell einzufügen; die

---

⁷ Potentiell mögliche Pfade oder Kovarianzen zwischen Variablen im Modell, die jedoch nicht freigegeben oder mit einer Identitätsrestriktion belegt sind, werden bei der Modellschätzung als nicht vorhanden und somit = 0 behandelt.

Fehlerterme können frei variieren. Je größer der Modifikationsindex, desto größer die Verringerung des Chi-Quadrat-Wertes und folglich auch die Stärke des geschätzten Zusammenhangs. Dabei ist zu berücksichtigen, dass dieser Weg der Respezifikation rein empirisch begründet wird und laut Kline (2011) anfällig für statistische Fehlentscheidungen, genauer den Alpha- und Beta-Fehler, ist. Es wird daher nicht empfohlen, die Fehlerterme eines Messmodells miteinander kovariieren zulassen. Dies ist auch messtheoretisch nicht gerechtfertigt, da mit Blick auf die Konzeption des Messmodells von der Uni-Dimensionalität der Indikatoren ausgegangen wird (Hermida, 2015).

Zusätzliche dimensionale Einflüsse am Beispiel begrifflicher Ähnlichkeit

Um den Einfluss dieser zusätzlichen, jedoch unberücksichtigten und nicht zu identifizierenden latenten Variable(n) auszuschließen, wurden die entsprechenden Items aus dem Modell entfernt. Dass bedeutet nicht, dass sie keinen Erklärungsbeitrag für die jeweils abgebildete Dimension lieferten, sondern, dass es vermutlich noch einen anderen nicht zu identifizierenden Einfluss gab, der die Beantwortung durch die Befragten beeinflusste. Nun kann zum einen davon ausgegangen werden, dass möglicherweise ein nicht beabsichtigtes Konstrukt einen Einfluss auf die Beantwortung der in Frage kommenden Variablen hat, die miteinander über das eigentlich zu messende Konstrukt hinaus kovariieren. Es lässt sich nur vermuten, welche zusätzlichen Einflüsse es auf das Antwortverhalten gab. Allerdings kann aufgrund der auffallend häufig vorkommenden unerklärten gemeinsamen Varianz zweier sich in ihrer Formulierung ähnelnden Variablen, die noch nicht durch das gemeinsame latente Konstrukt erklärt wird, zumindest auch davon ausgegangen werden, dass die gewählten Ausdrücke bei der Frageformulierung einen Einfluss haben.

Das Phänomen dieser begrifflichen Ähnlichkeit soll hier anhand dreier Items der Dimension Volume verdeutlicht werden:

- Es ist gut, dass immer mehr digitale Daten ausgewertet werden.
- Es ist gut, so viele digitale Daten wie möglich auszuwerten.
- Es ist erstrebenswert, so viele digitale Daten wie möglich auszuwerten.

Diese drei Items sind Indikatoren einer Dimension. Wenn nun Befragte den Items 2 und 3 auf der 7er-Likert-Skala mit „stimme voll und ganz zu“ den Wert 7 zuwiesen, Item 1 jedoch nur „stimme größtenteils zu“ und somit den Wert 6, ist das für sich genommen unproblematisch. Wenn allerdings über alle Befragten hinweg solch ein Muster auftritt und begriffliche Ähnlichkeit in der Formulierung dazu führt, dass zwei Items nicht nur wegen der gemeinsamen Zugehörigkeit zu einem

gleichen Konstrukt, sondern auch aus speziell diesem Grund ähnlich beantwortet werden, kann dies zu einem systematischen Messfehler führen, der im Modell eben nicht berücksichtigt wird. In diesem Fall kann eines der beiden Items eliminiert werden, unter der Voraussetzung, dass die Faktorladung des beibehaltenen Items gleich oder höher ausfällt, aber mind. größer als .70 ist.

Es wurde beim Ausschluss nach Modifikationsindizes darauf geachtet, zuerst Items zu eliminieren, für die mehrere (hohe) Modifikationsindizes vorlagen und deren addierte Werte größer waren als für andere Items (Kline, 2011). So wurden zunächst Items eliminiert, die auf mehrere latente Konstrukte luden und deren Fehlerterme mit denen anderer Variablen kovariierten.

Konstanthaltung der Itemformulierungen

Bei der Respezifikation der Messmodelle wurde zudem auch auf die jeweiligen Formulierungen der Items einer (Sub-)Dimension geachtet, um eine Vergleichbarkeit der Dimensionen und ihrer Items auch im Hinblick auf den Wortlaut der Fragebogenitems zu gewährleisten. Als Beispiel diene hier die Unterscheidung des Umgangs mit digitalen Daten in Erzeugung, Speicherung und Auswertung der Daten.

Es wurde eingangs vermutet, dass durchaus zwischen diesen Handlungen unterschieden werden kann und auch die Befragten hier Unterschiede machen könnten. Wenn sich allerdings zeigen würde, dass Befragte (in einer oder mehreren der Dimensionen) keinen Unterschied sähen, sobald beide Formulierungen vorhanden sind, dann wäre es nach Ansicht des Autors dennoch sinnvoll, eine der gewählten Formulierungsweisen für die finale Zusammensetzung der Skalenitems beizubehalten. Das bedeutet, entweder nur von Erzeugung oder Speicherung zu sprechen, selbst wenn sich zeigen sollte, dass die Befragten bei der Beantwortung der Items der ersten Erhebung keinen Unterschied machen sollten, um zu vermeiden, dass es in der weiteren Verwendung der Skalenitems in Folgestudien dann nicht doch zu einer möglichen Verwirrung bei manchen Befragten kommt. Diese könnten bspw. nun plötzlich doch vermuten, dass ein Unterschied gemeint sein könnte, wenn plötzlich für eine Dimension von Erzeugung und in der anderen von Speicherung gesprochen würde. Das heißt in der Konsequenz, dass auch Items aus dem Modell entfernt werden konnten, für die möglicherweise keine Anhaltspunkte vorlagen, die nach den beiden zuvor vorgestellten Ausschlusskriterien zu einem Ausschluss geführt hätten. Weitere Ausführungen hierzu finden sich nachfolgend in den Abschnitten zu den jeweiligen Dimensionen.

Zusammenfassend wurden für die Identifikation geeigneter Frageitems die drei zuvor beschriebenen Strategien angewendet: 1) die Betrachtung der standardisierten Faktorladungen, 2) die Verwendung der Modifikationsindizes und 3) die möglichst konstante Begriffsverwendung in der Itemformulierung.

Die Evaluation der CFA-Modelle – notwendige Prüfschritte für die Beurteilung reflektiver Messmodelle

Für die Beurteilung der reflektiven Messmodelle wird ein schrittweises Prüfverfahren angewendet. Zunächst wird die Reliabilität der Indikatoren der reflektiven Messmodelle geprüft und eine allgemeine Plausibilitätsprüfung der geschätzten Lösung vorgenommen, bevor die Anpassungsgüte der spezifizierten CFA-Modelle geprüft wird.

Reliabilität der Indikatoren

Die Reliabilität gibt Auskunft über die Zuverlässigkeit bzw. die Genauigkeit der Messung (Bandalos, 2017). Die Faktorladungen der reflektiven Indikatoren sollten daher $> .7$ sein. Zudem wird für die jeweiligen Indikatoren einer Dimension die tau-äquivalente Reliabilität angegeben werden. Dieses Maß, dass auch unter dem Begriff des Cronbachschen α firmiert, findet in den Sozialwissenschaften weite Verbreitung, ist in der Vergangenheit jedoch zunehmend als ungenau in die Kritik geraten, da es kein geeignetes Maß für die Homogenität und die Ein-dimensionalität der Indikatoren ist (Cho, 2016; Schmitt, 1996). Es wird daher ebenfalls häufig empfohlen, die durchschnittlich extrahierte Varianz (DEV) zu berichten. „Diese gibt an, wie viel Prozent der Streuung des latenten Konstruktes über die Indikatoren durchschnittlich erklärt wird“ (Weiber & Mülhhaus, 2014, S. 151). Als Schwellenwert wird üblicherweise ein DEV-Wert von $\geq .5$ angesehen, da durchschnittlich die Hälfte der Varianz auf den latenten Faktor des gemessenen Konstrukts zurückgeht (Fornell & Larcker, 1981; Segars, 1997; Weiber & Mülhhaus, 2014). Es werden daher nachfolgend die Werte für das standardisierte Cronbachsche α sowie die DEV angegeben. Die Berechnung des Cronbachschen α erfolgte mit SPSS 23 während die Berechnung der DEV manuell durchgeführt wurde.⁸

---

⁸ Die Berechnung erfolgte gemäß Weiber und Mülhhaus (2014, S. 151) mit folgender Formel:

$$\text{DEV}(\xi_j) = \frac{\sum \lambda^2_{ij}}{\sum \lambda^2_{ij} + \sum (1 - \lambda^2_{ij})}$$

Die Plausibilitätsprüfung der geschätzten Lösung

Alle berichteten Modelle wurden zudem zunächst auf Heywood Cases geprüft, d. h., es wurde eine Plausibilitätsprüfung durchgeführt. Heywood Cases liegen bei geschätzten Modellen vor, bei denen die Parametermatrizen nicht ‚positiv definit‘ sind, was sich an negativen Varianzen, Kommunalitäten > 1 und Korrelationen > 1 äußert (Kline, 2011; Weiber & Mühlhaus, 2014). Solche Werte sind unplausibel, weshalb Modelle mit Heywood Cases abgelehnt werden müssen. Es gibt keine Möglichkeiten, Heywood Cases bei einem Datensatz zu heilen, außer den Datensatz zu verändern, bspw., indem man neue zusätzliche Fälle erhebt (West et al., 1995).

Die Anpassungsgüte des geschätzten Modells

Um die Anpassungsgüte des spezifizierten Modells zu prüfen, gibt es eine Vielzahl von Kriterien sowie Kennwerten für die Evaluation des Kausalmodells, wobei Letztere auch als Fit-Maße bezeichnet werden. „Eine hohe Güte eines Kausalmodells (sog. Modell-Fit) ist allgemein dann gegeben, wenn die mit Hilfe der Parameterschätzer berechneten Varianzen und Kovarianzen möglichst gut mit den empirisch gewonnenen Varianzen und Kovarianzen übereinstimmen“ (Weiber & Mühlhaus, 2014, S. 203).

Das wichtigste Gütekriterium ist dabei der Chi-Quadrat-Test, dessen Bedeutung und Gültigkeit in der Literatur jedoch kontrovers diskutiert wird. Bspw. neigt der Test dazu, leicht auf die Stichprobengröße zu reagieren, was laut Kline (2011) bei Samplegrößen von 200 bis 300 Probanden noch unproblematisch ist, bei größeren Stichproben⁹ aber leicht zu einer Ablehnung des Modells führen kann. Während etliche Autoren für eine strenge Beachtung des Chi-Quadrat-Tests argumentieren (Barrett, 2007; Kline, 2011), gibt es auch Quellen, die diese Strenge ablehnen (Bentler & Bonett, 1980) oder zumindest im Gesamtkontext der Modellkonzeption betrachten möchten (Vandenberg, 2006).¹⁰

Um auf die Strenge des Chi-Quadrat-Tests zu reagieren, hat sich zudem etabliert, bei der Evaluation eines Gesamtmodells zur Güteprüfung noch absolute und inkrementelle Fit-Maße heranzuziehen (Holbert & Stephenson, 2002; Weiber & Mühlhaus, 2014). Absolute Fit-Maße prüfen, ob das Modell konsistent mit den Daten ist, ohne ein Vergleichsmodell heranzuziehen. Inkrementelle Fit-Maße hingegen stellen auf einen Vergleich zwischen dem spezifizierten Modell (auch Default-Modell genannt) und einem unabhängigen Modell ab. Das Heranziehen

---

⁹ Hierzu zählen laut Kenny (2015) bereits Stichproben mit mehr als 400 Fällen.

¹⁰ Für eine weitere Übersicht über die Debatte sei an dieser Stelle auch auf das Sonderheft von *Personality and Individual Differences* (2007, 42(5)) verwiesen.

dieser Gütekriterien hat sich in den Sozialwissenschaften etabliert, selbst wenn es keine einheitlichen Standards gibt, welche Fit-Maße herangezogen und welche Cut-off-Kriterien angelegt werden sollten (Holbert & Stephenson, 2002).

Letztendlich empfiehlt es sich *a priori* festzulegen, welche Evaluationskriterien herangezogen werden, um die Anpassungsgüte der geprüften Modelle zu beurteilen. In der vorliegenden Arbeit wird der Modellfit daher durch die Angaben der folgenden Kriterien und Werte geprüft: Für die Evaluation des Gesamtmodells werden der Chi-Quadrat-Test (X^2) und der ‚Root Mean Square Error of Approximation‘ (RMSEA) herangezogen. Während der ermittelte Chi-Quadrat-Test nicht-signifikant sein sollte, ist seine Anwendung voraussetzungsreich und unterliegt, wie zuvor diskutiert, diversen Limitationen (Kline, 2011; Weiber & Mülhhaus, 2014). Mit Blick auf den RMSEA besagt eine geläufige Daumenregel, dass der jeweilige Test-Wert im Idealfall nicht größer als .05 sein sollte (Browne & Cudeck, 1992), wobei zusätzlich das 90 %-Konfidenzintervall (Confidence Interval – CI) angegeben wird. Hier sollte die untere Grenze (LO90) nahe 0 liegen und die obere Grenze (HI90) nicht größer als .08 sein (Kenny, 2015).

Zudem wird nachfolgend der Tucker-Lewis-Index (TLI) als inkrementelles Fit-Maß herangezogen. Der Nachteil des TLI ist, dass dieser nicht-normiert ist, also auch Werte > 1 annehmen kann, wobei ein Wert > 1 für gewöhnlich auf 1.00 abgerundet wird. Weiber und Mülhhaus (2014) geben als Schwellenwert für den TLI Werte von mind. .90 an, laut L. Hu und Bentler (1999) sollte der Wert mind. bei .95 oder besser liegen. Es wird zusätzlich zum TLI auch häufig noch der Wert des Comparative Fit Index (CFI) angegeben (L. Hu & Bentler, 1999; Schreiber et al., 2006). Kenny (2015) rät hiervon jedoch mit folgender Begründung ab: “Because the TLI and CFI are highly correlated only one of the two should be reported. The CFI is reported more often than the TLI, but I think the CFI’s penalty for complexity of just 1 is too low and so I prefer the TLI even though the CFI is reported much more frequently than the TLI” (Kenny, 2015, § 10). Nachfolgend werden daher für alle Modelle der Chi-Quadrat-Test, der RMSEA samt 90 %-Konfidenzintervall sowie der TLI angegeben. Die Werte werden durchgängig wie folgt berichtet, wobei kursiv gesetzte Wörter durch die entsprechenden Werte ersetzt werden: ($X^2(df - \text{Freiheitsgrade}) = \text{Chi-Quadrat-Wert}$, $p < \text{Signifikanzwert}$; RMSEA = RMSEA-Wert [90 %-Konfidenzintervall: LO90-Wert, HI90-Wert]; TLI = TLI-Wert). Bei den Messmodellen der unidimensionalen Faktoren werden zudem noch das Cronbachsche α und die durchschnittlich extrahierte Varianz (DEV) ausgewiesen. Für den Vergleich von hierarchischen Modellen („Model Comparisons“) wird zudem der Chi-Quadrat-Differenz-Test herangezogen (Weiber & Mülhhaus, 2014). Dies

ist bspw. der Fall, wenn man die spezifizierten latenten Faktoren auf Diskriminanzvalidität prüfen möchte (siehe nachfolgender Abschnitt), aber auch bei der Prüfung von Messinvarianz sowie der Restriktion anderweitiger Modellparameter. Für die Evaluation nicht-hierarchischer Modelle, also Modelle, die zwar auf Grundlage der gleichen Indikatoren und Daten spezifiziert werden, jedoch eine gänzlich andere Spezifikation der Kausalpfade haben, wird zudem das Akaike Information Criterion (AIC) herangezogen (Kline, 2011).

Prüfung der Diskriminanzvalidität der BDGS-Dimensionen

Es ist zu diskutieren, inwieweit mit Blick auf die unterschiedenen Faktoren von diskriminanter Validität ausgegangen werden kann, d. h., messen die Indikatoren tatsächlich auch empirisch die konzeptuell unterschiedlichen Dimensionen der BDGS-Skala oder kann empirisch eben nicht unterschieden werden. Letzteres wäre bspw. der Fall, wenn die Befragten nicht zwischen den Indikatoren der einzelnen Dimensionen unterscheiden würden, also die einzelnen Indikatoren so hoch miteinander korrelieren, dass davon ausgegangen werden muss, dass sie die gleiche Ursache teilen (sprich: ein gemeinsames latentes Konstrukt, das ihre jeweilige Ausprägung beeinflusst). Weiber und Mühlhaus (2014) merken hierzu an: „Diskriminanzvalidität als Teilaspekt der Konstruktvalidität liegt vor, wenn sich die Messungen verschiedener Konstrukte signifikant unterscheiden“ (S. 164). Im Rahmen der CFA gibt es zwei Möglichkeiten der Prüfung auf Diskriminanzvalidität: 1) den Chi-Quadrat-Differenztest und 2) das Fornell-Larcker-Kriterium, das dem ‚DEV vs. geteilte Varianz‘-Test zugrunde liegt (Farrell, 2010; Weiber & Mühlhaus, 2014).

Es wird bei der Prüfung auf Diskriminanzvalidität daher zunächst ein Chi-Quadrat-Differenztest durchgeführt, bei dem das nicht-restringierte Modell, für das die Kovarianz zwischen den beiden Faktoren frei geschätzt wird, einem restringierten Modell gegenübergestellt wird, „wobei das restringierte Modell bei Vorliegen von Diskriminanzvalidität auf jeden Fall die schlechtere Güte aufweisen muss“ (Weiber & Mühlhaus, 2014, S. 164). Beim restringierten Modell wird folglich die Kovarianz zwischen den beiden Faktoren auf den Wert 1 fixiert. Ist der Chi-Quadrat-Differenztest für den Modellvergleich signifikant, spricht dies aus empirischer Sicht für Diskriminanzvalidität.

Die zweite Möglichkeit, Diskriminanzvalidität zu prüfen, ist die Verwendung des Fornell-Larcker-Kriteriums, eines Tests, bei dem das Verhältnis der DEV der Faktoren zur gemeinsam geteilten Varianz in Beziehung gesetzt wird (Farrell, 2010; Fornell & Larcker, 1981).

Dieses Kriterium ist dabei wesentlich strenger [als der Chi-Quadrat-Differenztest] und stellt die durchschnittlich durch einen Faktor erfasste Varianz (...) mit jeder quadrierten Korrelation (Φ^2_{ij}), die der betrachtete Faktor i mit einem anderen Faktor j aufweist, gegenüber. Da die quadrierte Korrelation zwischen zwei Faktoren als gemeinsame Varianz dieser Faktoren interpretiert werden kann, liegt nach Fornell/Larcker Diskriminanzvalidität dann vor, wenn diese gemeinsame Varianz kleiner ist als die DEV der jeweiligen Faktoren. (Weiber & Mülhhaus, 2014, S. 165)

Auch mit Blick auf die Prüfung der Diskriminanzvalidität ist an dieser Stelle der Skalenkonstruktion erneut auf das Verhältnis zwischen der Konzeption und dem Erkenntnisinteresse des *BDGS* und der empirischen Messung einzugehen. Für die Bewertung der Datenpassung zum vorgeschlagenen Modell haben sich die im vorigen Abschnitt beschriebenen Konventionen etabliert. Auch hier muss an der entsprechenden Stelle jedes Mal erneut abgewogen werden, zwischen strengem statistischen Test (der wohlgemerkt wie jede zur Konvention gewordene Daumenregel recht willkürlich gewählte Grenzen festlegt) und der konzeptionellen Unterscheidung in einzelne Dimensionen, denen innerhalb des *BDGS* eine bestimmte Bedeutung zukommt. Da die einzelnen Dimensionen des *BDGS* konzeptuell zusammenhängen und – insbesondere im Aggregat – eine Verbindung zwischen hoch ausgeprägten Wissensüberzeugungen und Nutzenerwartungen erwartet wird, werden nachfolgend die Prüfungen auf Diskriminanzvalidität entsprechend den Vorgaben aus der Literatur durchgeführt, jedoch nicht unhinterfragt auf Grundlage von Grenzwerten angenommen oder abgelehnt, sondern jeweils im Lichte der Skalenkonzeption reflektiert.

Die Prüfung der Diskriminanzvalidität hoch korrelierter und konzeptuell ähnlicher Dimensionen wird in Abschnitt 9.1.6 ergänzt um eine Diskussion und Prüfung der ‚dimensionalen Verschiedenheit‘ (DV) des spezifizierten Modells, bei der alle miteinander korrelierten latenten Konstrukte im Correlated-Factor-Modell (CFM) einem Second-Order-Factor-Modell (SOFM) gegenübergestellt werden (Gignac & Kretzschmar, 2017). Die Prüfung DV bezieht sich dabei auf das Gesamtmodell und die Plausibilität der Annahme separierbarer Dimensionen im Gesamtmodell.

9.1.4 Die Richtigkeit digitaler Daten – die Dimensionen Genauigkeit, Wahrhaftigkeit und Objektivität digitaler Daten sowie der hieraus resultierende Wissensgewinn

Ausgehend von der in Abschnitt 3.5.1 getroffenen Unterteilung der *Richtigkeit der digitalen Daten* in die zugehörigen Dimensionen *Genauigkeit, Wahrhaftigkeit und*

Objektivität sowie den *Wissensgewinn* wird die Datenauswertung durchgeführt. Aus Darstellungsgründen wird die Dimension *Wahrhaftigkeit und Objektivität* dabei nachfolgend nur als *Wahrhaftigkeit* bezeichnet. Die in Tabelle 2.C im Anhang im elektronischen Zusatzmaterial gelisteten 18 Items betrafen die Bewertung jener drei Subdimensionen der Richtigkeit im Entstehungs- und Verwertungskontext digitaler Daten.

Eine CFA, die mit den imputierten Daten für das Messmodell gerechnet wird und bei dem jeweils sechs Indikatoren in einem Single-Factor-Modell (SFM) auf einen zugrunde liegenden latenten Faktor laden, zeigt eine geringe Anpassungsgüte ($X^2(132) = 365.32$, $p < .001$; RMSEA = .083 [.073, .094]; TLI = .922) und wird abgelehnt. Mit einer Ausnahme¹¹ haben alle Items eine Faktorladung von mindestens $\lambda = .714$. Zwar zeigt sich mit Blick auf die interne Konsistenz der Faktoren, dass die jeweiligen Indikatoren der Dimensionen *Genauigkeit* ($\alpha = .908$; DEV = .636), *Wahrhaftigkeit* ($\alpha = .895$; DEV = .591) und *Wissensgewinn* ($\alpha = .914$; DEV = .649) hohe Reliabilitätswerte aufweisen. Doch auch hier haben etliche Items der distinkten Faktoren eine gemeinsam geteilte Varianz, die nicht durch den gemeinsamen latenten Faktor erklärt wird, so dass die Kovarianz der Fehlerterme freigegeben werden müsste, um frei zu variieren. Es zeigt sich mit Blick auf die beiden latenten Konstrukte *Genauigkeit* und *Wahrhaftigkeit* der digitalen Daten eine sehr hohe Korrelation ($r_{\text{Genauigkeit, Wahrhaftigkeit}} = .903$), die auch hier darauf hinweist, dass möglicherweise nicht von diskriminanter Validität zwischen den beiden Faktoren ausgegangen werden kann. Die Korrelationen zwischen dem Faktor *Genauigkeit* und dem Faktor *Wissensgewinn* ($r_{\text{Genauigkeit, Wissensgewinn}} = .874$) bzw. zwischen dem Faktor *Wahrhaftigkeit* und dem Faktor *Wissensgewinn* ($r_{\text{Wahrhaftigkeit, Wissensgewinn}} = .734$) sind ebenfalls hoch.

Diskriminanzvalidität der Richtigkeitsdimensionen

Es ist zu diskutieren, inwieweit mit Blick auf die drei unterschiedenen Faktoren im CFM noch von diskriminanter Validität ausgegangen werden kann, d. h., messen die Indikatoren tatsächlich auch empirisch die drei konzeptuell unterschiedlichen Dimensionen der *Richtigkeit* oder kann empirisch eben nicht unterschieden werden. Wie in Abschnitt 9.1.3 beschrieben, gibt es im Rahmen

---

¹¹ Das Item ‚Digitale Daten können als eine höhere Form von Intelligenz betrachtet werden‘ hat lediglich eine Faktorladung von $\lambda = .606$. Dies ist insofern erwähnenswert, als dass dieses Item unmittelbar auf den von boyd und Crawford benannten Kern des Mythos Big Data abstellt: “large data sets offer a higher form of intelligence and knowledge” (2012, S. 663). Unbenommen ist hier, dass die wortwörtliche Übersetzung aus dem Englischen ins Deutsche im Original anders verstanden werden kann und geläufiger ist als die zugegebenermaßen holprig klingende deutsche Version.

der CFA zwei Möglichkeiten der Prüfung auf Diskriminanzvalidität: 1) den Chi-Quadrat-Differenztest und 2) das Fornell-Larcker-Kriterium (Weiber & Mühlhaus, 2014). Wie erwähnt, ist das Fornell-Larcker-Kriterium der strengere Test und es wird schnell deutlich, dass die quadrierten Korrelationen der Faktoren *Genauigkeit* und *Wahrhaftigkeit* ($r^2_{\text{Genauigkeit, Wahrhaftigkeit}} = .815$) sowie *Genauigkeit* und *Wissensgewinn* ($r^2_{\text{Genauigkeit, Wissensgewinn}} = .764$), mit Blick auf die Angaben zur DEV weiter oben, die Annahme der Diskriminanzvalidität nicht bestätigen können, während er für die Faktoren *Wahrhaftigkeit* und *Wissensgewinn* angenommen wird ($r^2_{\text{Wahrhaftigkeit, Wissensgewinn}} = .539$).

Das bedeutet, dass zwar konzeptionell zwischen der *Genauigkeit* und der *Wahrhaftigkeit* der Daten unterschieden werden kann, dieser Unterschied jedoch von den Befragten in der Zustimmung zu den Frageitems nicht gemacht wird, eine getroffene Unterscheidung zwischen diesen Konzeptdimensionen also nicht beobachtbar ist. Jemand, der erwartet, dass Daten ein genaues Verständnis der Welt ermöglichen, bewertet diese Erkenntnis auf Datengrundlage auch als objektiv und wahrhaftig. So ist bspw. die mögliche Einschätzung, dass ein möglicher Bias in präzise dokumentierten Daten dazu führt, dass die Daten zwar genau das wiedergeben, was sie messen, aber dennoch nicht ‚objektiv‘ sind und diskriminieren, zumindest über alle Befragten hinweg auf dem Abstraktionsniveau der vorliegenden Indikatoren nicht zu beobachten.

Als Konsequenz wurden in einem ersten Schritt die Faktoren *Genauigkeit* und *Wahrhaftigkeit* zusammengelegt, so dass die zwölf Items, die diesen Faktoren zugrunde lagen, nun auf einen gemeinsamen Faktor luden und ein Modell mit zwei Faktoren ‚*Wahrhaftigkeit/Genauigkeit*‘ und ‚*Wissensgewinn*‘ geprüft wurde. Die CFA für dieses Modell, das mit den imputierten Daten gerechnet wurde, zeigt weiterhin eine geringe Anpassungsgüte ($X^2(135) = 640.598$, $p < .001$; RMSEA = .121 [.112, .131]; TLI = .835) und muss abgelehnt werden.

Das Modell wurde nun nach dem im vorigen Abschnitt dargestellten Vorgehen respezifiziert. Insbesondere wurden nur unzureichend auf die Faktoren ladende Items eliminiert und es wurde mit Blick auf die Zusammenlegung der beiden Dimensionen *Genauigkeit* und *Wahrhaftigkeit* darauf geachtet, dass letztendlich Items, die den Kern beider Dimensionen betrafen, mit in die finale Messung aufgenommen wurden. Hier wurden allerdings mit Rücksicht auf den möglichen Konflikt zwischen den Begrifflichkeiten Wahrheit und Wissen sowie die erforderliche Trennschärfe der Konstrukte *Objektivität* und *Wissensgewinn* die Items eliminiert, die den Wahrheitsbegriff beinhalten. Die Items für den Faktor ‚*Genauigkeit/Wahrhaftigkeit*‘ betreffen mithin die Genauigkeit der Abbildung realweltlicher Gegebenheiten, die auch mit Blick auf die Objektivität von einer Fehlerfreiheit und Unverzerrtheit der digitalen Daten ausgeht und dahingehend

eine präzise wahrhaftige Repräsentation der Welt in diesen abbildbar findet und erfahrbar macht.

Um in den nachfolgenden Beschreibungen nicht jedes Mal den umständlichen Doppelbegriff mitzuführen und mögliche Irritationen bezüglich der Zugehörigkeit zur Beschreibungsdimension *Veracity* zu vermeiden, wird die Dimension ‚*Genauigkeit/Wahrhaftigkeit*‘ lediglich als Dimension ‚*Genauigkeit*‘ bezeichnet. Die nachfolgend angegebenen Item-IDs werden von hier an durchgängig gebraucht, um das jeweilige Frageitem zu identifizieren. Das finale Modell besteht mithin aus den zwei zuvor genannten latenten Faktoren *Genauigkeit* und *Wissensgewinn*, auf die die folgenden sechs Indikatoren laden.

Genauigkeit

- Digitale Daten ermöglichen ein genaues Verständnis der Welt. (Item-ID *GE1*; $\lambda_{GE1} = .808$, Referenzindikator)
- Digitale Daten ermöglichen ein exaktes Verständnis der Welt. (ID *GE2*; $\lambda_{GE2} = .840$)
- ID *GE3*: Digitale Daten ermöglichen ein objektives Verständnis der Wirklichkeit. (ID *GE3*; $\lambda_{GE3} = .824$)

Wissensgewinn

- Digitale Daten verhelfen zu neuem Wissen. (ID *W1*; $\lambda_{W1} = .804$, Referenzindikator)
- Digitale Daten führen zu neuen Fakten. (ID *W2*; $\lambda_{W2} = .779$)
- Digitale Daten verhelfen zu neuen Erkenntnissen. (ID *W3*, $\lambda_{W3} = .886$)

Die CFA für dieses Modell, das mit den nicht-imputierten Daten gerechnet wurde, kommt zu einer plausiblen Lösung und zeigt eine zufriedenstellende Anpassungsgüte ($X^2(8) = 11.062$, $p = .198$; RMSEA = .039 [.000, .089]; TLI = .990) und wird folglich angenommen. Allerdings zeigt sich auch hier für *Genauigkeit* ($\alpha = .867$; DEV = .679) und *Wissensgewinn* ($\alpha = .853$; DEV = .679) eine hohe Interkorrelation beider Faktoren ($r = .829$, $r^2 = .687$), welche die Diskriminanzvalidität in Frage stellt. Da die DEV beider Faktoren mit jeweils genau .679 leicht niedriger liegt als die quadrierte Korrelation wird hier dennoch zugunsten der Zwei-Faktor-Lösung entschieden, wobei zu prüfen sein wird, inwieweit sich Diskriminanzvalidität in Folgeerhebungen zeigt oder nicht. Die Aufrechterhaltung der konzeptionellen und der empirischen Trennung in der Modellierung und der Erhebung der Items für die zwei Faktoren wird auch dahingehend fortgeführt,

als dass vorstellbar ist, dass je nach Studienschwerpunkt und Forschungsfragen in der Anschlussforschung eine Aufteilung sinnvoll sein könnte. Es muss hier jedoch auch bedacht werden, dass die Multi-Kollinearität beim *BDGS* ein generelles Problem darstellt, da sich zeigt, dass alle Dimensionen hoch miteinander korreliert sind. Diese Fragen betreffen nachfolgend auch die Trennung der nun folgenden Nutzendimension.

9.1.5 Die Nutzendimensionen des *BDGS*

Mit Blick auf den aus den Daten gezogenen Nutzen wird im *BDGS* und somit auch in den Befragungseitems zwischen einem persönlichen Nutzen für das Individuum (*„Individueller Nutzen“*) und dem Nutzen für die Gesamtgesellschaft (*„Gesellschaftlicher Nutzen“*) unterschieden (siehe Abschnitt 3.5.2). Mit Blick auf die Frageitems der letzteren Dimension blieb in der Einleitung und den Aussagen offen, wer oder was unter dem Begriff der Gesellschaft gemeint ist. Es wurde nicht zwischen einer vermeintlich vagen Allgemeinheit oder einer konkreten Gesellschaft (wie bspw. einer europäischen oder einer nationalen Gesellschaft) abgestuft. Das hat zur Folge, dass Befragte ganz verschiedene Dinge mit dem Gesellschaftsbegriff verbinden, mithin unterschiedliche Vorstellungen davon haben können, was Gesellschaft ausmacht, woraus und aus wem sie jeweils besteht (Kölbl & Althof, 2014). Es zeigt sich dann auch, dass mit Blick auf das Individuum und die menschliche Identitätswahrnehmung ganz unterschiedliche Verständnisse von Gesellschaft zwischen einem Zugehörigkeitsgefühl zu einer Weltgesellschaft und regionalen, gar lokalen Verbänden existieren (Herrmann et al., 2004). Dieses feinteilige Verständnis kann auch mit Blick auf die Messung der gesellschaftlichen Nutzenwahrnehmung nur unzureichend operationalisiert werden. Während also offenbleibt, was oder wen die Befragten nun zur situativ imaginierten Gesellschaft zählen, so sollte durch die Kontrastierung mit Items, die auf den *individuellen Nutzen* aus digitaler Datensammlung und -verwertung abstellen, deutlich werden, dass hier bezüglich eines Kollektivs gefragt wird, zu dem die Befragten in Beziehung stehen, als dessen vollständig zugehörige Mitglieder sich die Befragten jedoch nicht zwingend verstehen müssen. Die Modellierung der beiden Dimensionen und die Datenanalyse der Frageitems werden im folgenden Abschnitt nachvollzogen.

Der individuelle und gesellschaftliche Nutzen im Entstehungs- und Verwertungskontext digitaler Daten

Die ersten sieben der 15 in Tabelle 2.C gelisteten Items betrafen die Erwartung eines *individuellen Nutzens* aus dem Entstehungs- und Verwertungskontext digitaler Daten. Sprich: Glaubt bspw. eine bestimmte Befragte, dass sie persönlich einen Nutzen für sich selbst aus der Datensammlung und -verwertung zieht. Dabei wurde nicht zwischen dem Entstehungs- und Verwertungskontext unterschieden, sondern die Items fragten im Präsens formuliert, ob ein Nutzen wahrgenommen wird. Die letzten acht Items stellten dann auf den gesellschaftlichen Nutzen ab, der sich aus dem Entstehungs- und Verwertungskontext digitaler Daten ergibt.

Eine CFA, die mit den imputierten Daten für das Messmodell gerechnet wird und bei dem alle 15 Nutzen-Indikatoren im SFM auf einen latenten Faktor laden, zeigte eine unzureichende Anpassungsgüte ($X^2(90) = 299.515$, $p < .001$; RMSEA = .096 [.084, .108]; TLI = .924) und wird abgelehnt. Zwar haben alle Items eine Faktorladung von mindestens $\lambda = .709$. Doch auch hier haben etliche Items eine gemeinsame geteilte Varianz, die nicht durch den gemeinsamen latenten Faktor erklärt wird, so dass hier entweder die Kovarianz der Fehlerterme freigegeben werden müsste, um frei zu variieren, oder aber eine mehrdimensionale Lösung mit zwei oder mehr Faktoren gerechnet werden sollte.

Es wurde anschließend ein Modell geprüft, bei dem zwischen dem *individuellen und gesellschaftlichen Nutzen* unterschieden wurde, der aus digitalen Datenbeständen gezogen wird. Das Modell bestand aus den zwei latenten Faktoren *individueller Nutzen* (sieben reflektive Indikatoren) und *gesellschaftlicher Nutzen* (acht reflektive Indikatoren). Die jeweiligen Items luden allein auf den jeweiligen latenten Faktor. Das Modell sah keine Kreuzladungen vor und die latenten Faktoren konnten frei miteinander kovariieren. Die geschätzte Lösung der CFA ist zulässig. Das Modell, das mit den imputierten Daten gerechnet wurde, zeigt weiterhin eine insgesamt noch akzeptable Anpassungsgüte ($X^2(89) = 156.822$, $p < .001$; RMSEA = .055 [.040, .069]; TLI = .975), auch wenn es mit Blick auf den Chi-Quadrat-Test abgelehnt werden muss. Es zeigt sich darüber hinaus, dass zwischen den beiden latenten Konstrukten eine hohe Korrelation besteht ($r_{\text{individueller Nutzen, gesellschaftlicher Nutzen}} = .897$), was mit Blick auf die diskriminante Validität der beiden konzeptuell unterschiedlichen Dimensionen näher zu betrachten ist.

Zunächst wurde das Modell jedoch wiederum nach dem in Abschnitt 9.1.3 dargestellten Vorgehen respezifiziert, um Items zu eliminieren, die hoch miteinander kovariierten oder im Vergleich schlechter auf die jeweiligen latenten Faktoren luden. Das finale Modell besteht aus zwei latenten Faktoren, auf die jeweils die folgenden drei Indikatoren laden. Auch hier werden nachfolgend die angegebenen Item-IDs von nun an durchgängig vergeben und gebraucht, um das jeweilige Frageitem zu identifizieren.

Individueller Nutzen

- Digitale Daten erzeugen für mich persönlich einen Mehrwert. (Item-ID IN1; $\lambda_{IN1} = .838$, Referenzindikator)
- Digitale Daten führen für mich zu Vorteilen. (ID IN2; $\lambda_{IN2} = .817$)
- Digitale Daten haben für mich persönlich positive Auswirkungen. (ID IN3; $\lambda_{IN3} = .839$)

Gesellschaftlicher Nutzen

- Digitale Daten erzeugen einen Mehrwert für die Gesellschaft. (ID GNI; $\lambda_{GNI} = .823$, Referenzindikator)
- Digitale Daten führen zu gesellschaftlichen Vorteilen. (ID GN2; $\lambda_{GN2} = .809$)
- Digitale Daten sind ein gesellschaftlicher Gewinn. (ID GN3; $\lambda_{GN3} = .839$)

Die CFA für dieses Modell, gerechnet mit den nicht-imputierten Daten, kommt zu einer plausiblen Lösung und zeigt eine hohe Anpassungsgüte ($X^2(8) = 2.174$, $p = .975$; RMSEA = .000 [.000, .000]; TLI = 1.016), mit Faktorladungen von mindestens $\lambda = .801$.

Diskriminanzvalidität der Nutzendimensionen

Die jeweils drei Indikatoren der final modellierten Faktoren *individueller Nutzen* ($\alpha = .874$; DEV = .700) und *gesellschaftlicher Nutzen* ($\alpha = .863$; DEV = .676) zeigen eine hohe Reliabilität und konvergente Validität der Indikatoren, jedoch auch eine sehr hohe Korrelation zwischen den Faktoren ($r_{\text{individueller Nutzen, gesellschaftlicher Nutzen}} = .915$). Wie bereits im vorhergehenden Abschnitt angedeutet, ist zu diskutieren, inwieweit mit Blick auf die beiden Faktoren noch von diskriminanter Validität ausgegangen werden kann: D. h., messen die Indikatoren tatsächlich auch empirisch die zwei konzeptuell unterschiedlichen Dimensionen des Nutzens oder kann empirisch eben nicht unterschieden werden? Auch hier wurden, wie eingangs in Abschnitt 9.1.3 beschrieben, zur

Prüfung auf Diskriminanzvalidität 1) der Chi-Quadrat-Differenztest und 2) das Fornell-Larcker-Kriterium herangezogen (Weiber & Mühlhaus, 2014).

Es wurde daher zunächst ein Chi-Quadrat-Differenztest durchgeführt, bei dem das nicht-restringierte Modell einem restringierten Modell gegenübergestellt wird. Der Chi-Quadrat-Differenztest für den Modellvergleich ($\Delta X^2 = 6.56$, $df = 1$, $p < .05$) ist signifikant, was für Diskriminanzvalidität spricht.

Eine weitere Möglichkeit, Diskriminanzvalidität zu prüfen, ist die Verwendung des Fornell-Larcker-Kriteriums. Da die durchschnittlich extrahierten Varianzen beider Faktoren mit .700, bzw. .676 kleiner sind als deren quadrierte Korrelation ($r^2_{\text{individueller Nutzen, gesellschaftlicher Nutzen}} = .837$), wäre im vorliegenden Fall keine Diskriminanzvalidität gegeben.

Die sich widersprechenden Ergebnisse zeigen, dass es auch hier keine einheitliche Antwort auf die Frage nach der Diskriminanzvalidität gibt, da diese von den willkürlich gesetzten Cut-off-Kriterien abhängt. Auch hier ist mit Blick auf konzeptuelle Modellierung und die empirischen Ergebnisse abzuwägen, welcher Modellierungsvariante der Vorzug gegeben wird. Mit Blick auf die Nutzendimension des *BDGS* wird der weniger strengen Auslegung gefolgt und die Aufteilung in zwei Faktoren beibehalten, die mit persönlichem *individuellen* und dem allgemeinen *gesellschaftlichen Nutzen* zwei distinkte Dimensionen des Nutzengewinns auf Grundlage digitaler Daten unterscheidet. Das wird zum einen damit begründet, dass mit Blick auf die positive öffentliche Betrachtung der gesamtgesellschaftlichen Konsequenzen der Digitalisierung und der eigenen Erfahrung mit digitalen Werkzeugen eine hohe Korrelation des *individuell gesehenen Nutzens* mit dem *gesellschaftlichen Nutzen* zu erwarten ist. Haben Personen jedoch eine negative Bewertung der Folgen für die Gesellschaft, schlägt diese Einschätzung möglicherweise auch auf die Bewertung der eigenen Lage durch. Es werden an dieser Stelle noch keine Hypothesen aufgestellt oder Prognosen abgegeben. So gibt es zumindest keine Hinweise darauf, dass die Nutzenerwartungen für das Individuum systematisch vom Gesamtgesellschaftlichen entkoppelt sind. Auch wenn dies für bestimmte Personen und Digitalisierungskontexte durchaus möglich erscheint, müssten entsprechende Annahmen zunächst einmal argumentativ für den spezifischen Fall hergeleitet werden. Die Ausführungen zum Glaubenssystem in Kapitel 6 legen hingegen eher nahe, dass eine Aktivierung des *BDGS*, unter der Annahme eines nicht vollständig durchschaubaren Ursache-Wirkungs-Zusammenhangs digitaler Datenentstehung und -verwertung und die hieraus entstehenden Konsequenzen, eine im Aggregat relativ undifferenzierte Gesamtevaluation bezüglich der Nutzenerwartung zur Folge hat. Zudem soll ja gerade die feinteilige Unterscheidung in die diskreten Dimensionen des *BDGS* nachfolgend genutzt werden, um zu prüfen, inwieweit einzeln erhobene Glaubenssätze

als erklärende Variablen für die empirische Forschung dienen können, wobei es je nach Forschungszusammenhang und Befragungskontext sinnvoll ist, zwischen den Nutzendimensionen unterscheiden zu können. Es ist denkbar, dass es Lebensbereiche und mithin Forschungsfragen gibt, für die unterschiedliche Bewertungen und Auswirkungen der Nutzenerwartungen dokumentiert werden könnten. Aus diesen Gründen ist der konzeptuellen sowie der empirischen Trennung der Dimensionen in der Modellierung an dieser Stelle zunächst der Vorzug zu geben.

9.1.6 Datenauswertung für das finale Messmodell der BDGS-Skala

Die einzelnen Dimensionen wurden nun, wie in Abbildung 9.2 dargestellt, in einem CFM zusammengefasst spezifiziert.

Abbildung 9.2 Messmodell der vier Dimensionen des BDGS (CFM)

Um zu prüfen, ob das finale Messmodell auch zu den ursprünglichen nicht-imputierten Daten passt, wurde das finale Gesamt-Modell also noch einmal mit dem Ursprungsdatensatz berechnet. Hierbei kommt bei der Parameterschätzung die Full-Information-Maximum-Likelihood-Schätzung (FIML) zum Einsatz, für die in Amos die Funktion *‘Estimate Means and Intercepts’* ausgewählt werden muss, die Durchschnittswerte und Achsenabschnitte der Variablen im Modell berechnet. Dabei wurde für jede Dimension eines der drei jeweiligen Items als Referenzindikator festgelegt (siehe Tabelle 9.2), d. h., dass dessen Regressionsgewicht auf den Wert 1 restringiert wurde. Diese Referenzindikatoren werden auch in den weiteren Erhebungen beibehalten. Zudem wurden die Mittelwerte der latenten Konstrukte jeweils auf den Wert null restringiert. Um in den späteren Abschnitten im Rahmen der Hypothesenprüfung zu testen, ob sich bspw. die Werte der Achsenabschnitte signifikant vom Skalenmittelpunkt unterscheiden, wurde an den entsprechenden Stellen zusätzlich ein Bootstrapping in AMOS für *‘Bias-corrected confidence intervals’* nach Efron (1987) mit insgesamt 500 Stichproben durchgeführt. Dieses Bootstrapping-Verfahren wird auch in allen nachfolgenden Analysen entsprechend angewandt.

In Tabelle 9.2 finden sich nun die Itemformulierungen und Kennwerte der einzelnen Items der Dimensionsindikatoren sowie Kennwerte für die gesamte Dimension. Zudem sind dort die Variablennamen aus dem Originaldatensatz zu finden, so wie diese noch in der Tabelle 2.C im Anhang im elektronischen Zusatzmaterial vermerkt waren. Diese Variablennamen werden ab hier in eine standardisierte, oben bei den Items sowie im Anhang im elektronischen Zusatzmaterial erwähnte Indikatorbenennung geändert, so dass ab hier eine durchgängig konsistente Bezeichnung der einzelnen Indikatoren verwendet wird. Es ist somit an jeder Stelle in den nachfolgenden Ausführungen zu den Datenauswertungen der einzelnen Erhebungen deutlich, über welchen Indikator gesprochen wird.

Tabelle 9.2 Kennwerte der Dimensionen und Indikatoren – Erhebung 1.1: Finales Modell BDGS

Itemformulierung „Digitale Daten ...“	Item-ID alt	Item-ID neu	B (SE)	β (SE)	λ	Cronbachsches α (standardisiert)	Durchschnittlich extrahierte Varianz (DEV) ^b
Genauigkeit							
... ermöglichen ein genaues Verständnis der Welt.	AV01_38	GE1	4.118 (.096)	1 ^a	.812	.867 (n = 232)	.679
... ermöglichen ein exaktes Verständnis der Welt.	AV01_39	GE2	4.077 (.098)	1.061*** (.073)	.840		
... ermöglichen ein objektives Verständnis der Wirklichkeit.	AV01_48	GE3	3.955 (.093)	.976*** (.069)	.820		
Wissen							
... verhelfen zu neuem Wissen.	AV01_49	W1	4.931 (.089)	1 ^a	.798	.854 (n = 242)	.679
... führen zu neuen Fakten.	AV01_51	W2	4.620 (.092)	1.021*** (.077)	.783		
... verhelfen zu neuen Erkenntnissen.	AV01_54	W3	4.897 (.089)	1.111*** (.073)	.887		

(Fortsetzung)

Tabelle 9.2 (Fortsetzung)

Itemformulierung „Digitale Daten ...“	Item-ID alt	Item-ID neu	B (SE)	β (SE)	λ	Cronbachsches α (standardisiert)	Durchschnittlich extrahierte Varianz (DEV) ^b
Individueller Nutzen							
... erzeugen für mich persönlich einen Mehrwert.	AV01_55	IN1	3.920 (.099)	1 ^a	.834	.874 (n = 242)	.701
... führen für mich zu Vorteilen.	AV01_57	IN2	3.901 (.094)	.957*** (.063)	.833		
... haben für mich persönlich positive Auswirkungen.	AV01_60	IN3	3.910 (.090)	.929*** (.060)	.845		
Gesellschaftlicher Nutzen							
... erzeugen einen Mehrwert für die Gesellschaft.	AV01_62	GN1	4.112 (.086)	1 ^a	.814	.864 (n = 236)	.678
... führen zu gesellschaftlichen Vorteilen.	AV01_65	GN2	4.020 (.090)	1.029*** (.072)	.804		
... sind ein gesellschaftlicher Gewinn.	AV01_67	GN3	4.062 (.092)	1.114*** (.072)	.851		

Notizen. n = 255, ^a Referenzindikator, ^b Die DEV wurde auf Grundlage der standardisierten Regressionswerte (Faktorladungen) im Gesamtmodell berechnet. * p < .05, ** p < .01, *** p < .001

Die CFA für dieses Modell kommt zu einer plausiblen Lösung und zeigt eine hohe Anpassungsgüte ($X^2(48) = 39.138$, $p = .815$; RMSEA = .000 [.000, .026]; TLI = 1.007). Die Regressionskonstanten der einzelnen Indikatoren indizieren für die Dimensionen *Genauigkeit* sowie *individueller und gesellschaftlicher Nutzen* Werte nahe dem Skalenmittelpunkt. Dies ist so zu interpretieren, dass bei einer durchschnittlichen Ausprägung des latenten Faktors bei einem Befragten der Skalenmittelpunkt des entsprechenden Frageitems gewählt wird. Die Regressionskonstanten sind daher wie die Mittelwerte einer Stichprobe zu interpretieren, wobei es sich aufgrund der ML-Schätzung und der fehlenden Werte für manche Befragten nicht um einen beobachteten Mittelwert handelt, sondern eine durch FIML geschätzte Konstante der Regressionsgleichung. Für die Indikatoren der Dimension *Wissensgewinn* werden nun höhere Regressionskonstanten beobachtet, wenn der Mittelwert des latenten Faktors auf null restringiert wird. Im Durchschnitt liegen die Befragten einen halben bis fast einen ganzen Skalenpunkt von der Skalenmitte entfernt, was eine hohe Zustimmung zu den jeweiligen Frageitems signalisiert. In Tabelle 9.3 finden sich darüber hinaus die Angaben zu den Korrelationen der latenten Faktoren, die allesamt hoch miteinander korrelieren.

Tabelle 9.3 Korrelationsmatrix – Erhebung 1.1: Finales Modell BDGS

	Genauigkeit	Wissensgewinn	Individueller Nutzen
Genauigkeit	1		
Wissensgewinn	.830***	1	
Individueller Nutzen	.809***	.714***	1
Gesellschaftlicher Nutzen	.878***	.831***	.915***
Notizen. * $p < .05$, ** $p < .01$, *** $p < .001$			

Mit drei Indikatoren pro latentem Faktor ist zudem eine ausreichende Anzahl an Indikatoren gegeben, ohne dass die Erhebung des BDGS mit einer zwölf Items zählenden Itembatterie unnötig lang geraten ist.

Anmerkungen zur dimensionalen Struktur und Modellierung der BDGS-Dimensionen unter Berücksichtigung der dimensionalen Verschiedenheit der latenten Faktoren

Die Annahme eines SFM wurde zuvor an den entsprechenden Stellen in einem ersten Schritt geprüft und alle Modelle zeigen eine entsprechend schlechtere Anpassungsgüte, so dass die Annahme verworfen wurde. Beim finalen BDGS-Modell handelt es sich daher zunächst um ein CFM, bei dem davon ausgegangen

wird, dass es sich um eigenständige voneinander trennbare Dimensionen handelt, die hoch miteinander korrelieren, jedoch konzeptuell und datenanalytisch voneinander trennbar sind. Um zu prüfen, inwieweit diese Annahme haltbar ist, empfehlen sich zwei weitere Prüfschritte. Zunächst sollte das CFM einem Second-Order-Factor-Modell (SOFM) gegenübergestellt werden, bei dem ein Faktor zweiter Ordnung auf die vier Dimensionen lädt. In einem weiteren Schritt erlaubt die Spezifikation eines SOFM laut Gignac und Kretzschmar (2017) auch bei Annahme des CFM dann, ob die Annahme der DV haltbar ist. Diese erlaubt eine spezifische Interpretation der einzelnen Dimensionen innerhalb des *BDGS* und geht somit über die Prüfung der Diskriminanzvalidität zweier hoch korrelierter latenter Faktoren hinaus. Letztere fragt danach, ob man zwischen zwei Dimensionen unterscheiden kann. Die DV nimmt das Gesamtmodell aller latenten Faktoren in den Blick und untersucht die Plausibilität der Annahme separierbarer Dimensionen im Gesamtmodell. Die Analyse einzelner Dimensionen wird bspw. relevant, wenn sie als eigenständige Erklärungsvariablen in der Forschung eingesetzt werden oder wenn man zwischen Befragten mit Blick auf die einzelnen Dimensionen differenzieren möchte, ohne auf die Ausprägung des Gesamtmodells Bezug nehmen zu müssen.

Die Prüfung auf DV erfolgt also in einem ersten Schritt über den Vergleich mit einem SOFM, bei der die einzelnen latenten Faktoren im Gesamtmodell auf einen einzelnen Faktor zweiter Ordnung laden. Um die Stärke der einzelnen latenten Faktoren zu ermitteln, wird deren Varianz im SOFM auf Signifikanz getestet, selbst wenn dieses Modell eine schlechtere Anpassungsgüte zeigt als das CFM¹².

Mit Blick auf die vier Dimensionen des *BDGS* zeigt das SOFM, bei dem ein Faktor zweiter Ordnung auf die vier latenten Faktoren *Wahrhaftigkeit*, *Wissensgewinn*¹³, *individueller Nutzen* sowie *gesellschaftlicher Nutzen* lädt, eine gute Anpassungsgüte ($X^2(50) = 53.415$, $p = .344$; RMSEA = .016 [.000, .045]; TLI = .997). Im Vergleich der nicht-hierarchischen Modelle zeigt das SOFM jedoch

---

¹² Gignac und Kretzschmar (2017) schlagen weitergehend einen Test mittels des Koeffizienten Omega hierarchical subscale' (OmegaHS) vor, der das Verhältnis von Faktorvarianz zur Gesamtvarianz ermittelt und standardisiert, so dass Vergleiche zwischen latenten Faktoren möglich sind. Diese weitergehende Betrachtung der spezifischen Faktorvarianz wird nachfolgend nicht ermittelt, da lediglich die dimensionale Verschiedenheit geprüft werden soll.

¹³ Die unstandardisierte Faktorladung des Faktors zweiter Ordnung auf den Faktor *Wissensgewinn* wurde auf 1 restringiert, wodurch diese Dimension als Referenzindikator festgelegt wird. Dies wird für alle nachfolgenden Modelle zur Prüfung der DV beibehalten. Angaben zur Modellgüte des SOFM werden nachfolgend nicht mehr gemacht, da dem CFM konzeptuell und datenanalytisch an dieser Stelle der Vorzug gegeben wird. Das SOFM wird lediglich zur Analyse der DV herangezogen.

einen höheren AIC-Wert¹⁴ ($\Delta\text{AIC} = 10.277$), weshalb dem CFM nicht nur auf Grundlage der definitorisch abgeleiteten mehrdimensionalen Konzeption, sondern auch aus empirischer Sicht der Vorzug zu geben ist. An dieser Stelle richtet sich der Blick jedoch auf die Varianz der einzelnen Dimensionen des *BDGS*, die hier als latente Faktoren erster Ordnung vorliegen. Es zeigen sich signifikante Varianzen der Faktoren *Genauigkeit* ($\sigma^2_{\text{Genauigkeit}} = .268$, S.E. = .065, $p < .001$), *Wissensgewinn* ($\sigma^2_{\text{Wissensgewinn}} = .360$, S.E. = .068, $p < .001$) und *individueller Nutzen* ($\sigma^2_{\text{Ind. Nutzen}} = .311$, S.E. = .069, $p < .001$), jedoch nicht für den Faktor *gesellschaftlicher Nutzen* ($\sigma^2_{\text{Ges. Nutzen}} = .023$, S.E. = .041, $p = .577$). Für letztere Dimension kann die DV im vorliegenden Sample nicht angenommen werden. Das bedeutet, dass die Bewertung des gesellschaftlichen Nutzens sich nicht ausreichend von den drei anderen Dimensionen unterscheidet. Allerdings weisen Simulationen von Sinharay (2010) darauf hin, dass bei hoch ausgeprägten allgemeinen Faktoren (hier der Faktor zweiter Ordnung) bis zu zehn Indikatorvariablen notwendig sind, um die Signifikanz eines spezifischen Faktors erster Ordnung festzustellen. Mit Blick auf die gebotene Kürze der *BDGS*-Skala ist von dieser Erweiterung der Indikatoren jedoch im vorliegenden Fall aus forschungsökonomischen Gründen abzusehen. Die Prüfung der DV wird nachfolgend an der entsprechenden Stelle für die Erhebung 1.2 und Erhebung 1.3 wiederholt.

9.1.7 Der Zusammenhang der *BDGS*-Skala mit der Bewertungsdimension Volumen

Es ist nun im Rahmen der Validität der Messung des *BDGS* zu prüfen, inwieweit die vier Dimensionen der *BDGS*-Skala mit der Bewertungsdimension Volumen zusammenhängen. Dies wird auch als kriteriumsbezogene Validität bezeichnet (Bandalos, 2017). Die im Anhang 1 im elektronischen Zusatzmaterial dokumentierte Messung zur Bewertung des Ausmaßes zielte darauf ab, zu erheben, inwiefern die Befragten eine weitreichende Datensammlung und -auswertung begrüßen. Zum latenten Faktor gehören hierbei die folgenden vier Indikatoritems:

Bewertung Volumen ($\alpha = .905$; $\text{DEV} = .703$)

- Es ist gut, dass immer mehr digitale Daten ausgewertet werden. ($\lambda = .768$, Referenzindikator)

---

¹⁴ Zum AIC-Wert äußert Kline (2011): “The AIC and related indexes are generally used in SEM to select among competing nonhierarchical models estimated with the same data. Specifically, the model with the smallest AIC value is chosen as the one most likely to replicate” (Kline, 2011, S. 220).

- Es ist gut, dass weltweit Unmengen von digitalen Daten gespeichert werden. ($\lambda = .831$)
- Es ist gut, so viele digitale Daten wie möglich zu speichern. ($\lambda = .883$)
- Es ist gut, so viele digitale Daten wie möglich auszuwerten. ($\lambda = .867$)

Es steht zu erwarten, dass bei einem Glauben an die *Genauigkeit* und den *Wissensgewinn* der Daten sowie dem hieraus gezogenen *individuellen Nutzen* und *gesellschaftlichen Nutzen* eine positive Bewertung digitaler Datensammlung und -auswertung mit einhergeht. Dabei ist mithin zu prüfen, ob die vier Dimensionen des *BDGS* positiv mit einer positiven Bewertung einer umfassenden Sammlung digitaler Daten korrelieren.

Das Messmodell für diese Bewertungsdimension Volumen wurde nun zusammen mit dem *BDGS* im CFM modelliert. Die CFA kommt zu einer plausiblen Lösung, zeigt jedoch mit Blick auf den signifikanten Chi-Quadrat-Test eine unzureichende Anpassungsgüte ($X^2(94) = 130.533$, $p = .008$; RMSEA = .039 [.021, .054]; TLI = .982). Auch hier haben die hoch miteinander korrelierten Indikatoren noch gemeinsam geteilte Varianz, die nicht durch die Modellpfade abgebildet wird, sondern in der gemeinsamen Varianz der kovariierenden Fehlerterme begründet ist. Diese können im Modell eben nicht frei variieren, da sie auf null restringiert sind. An dieser Stelle wird, da das Modell mit der Bewertungsdimension Volumen vor allem Illustrationszwecken und der Prüfung der Validität dient, der konzeptuellen Modellspezifikation der Vorzug gegeben. Es zeigt sich mit Blick auf die Inter-Korrelationen der fünf Dimensionen in Tabelle 9.4, dass die positive Bewertung der Sammlung und Auswertung von digitalen Daten hoch mit der Zustimmung zu den vier Dimensionen des *BDGS* korreliert.

Tabelle 9.4 Korrelationsmatrix – Erhebung 1.1: Finales Modell *BDGS* mit Volumen

	Volumen	Genauigkeit	Wissensgewinn	Individueller Nutzen
Volumen	1			
Genauigkeit	.779***	1		
Wissensgewinn	.600***	.828***	1	
Individueller Nutzen	.702***	.809***	.714***	1
Gesellschaftlicher Nutzen	.795***	.879***	.832***	.914***
Notizen. * $p < .05$, ** $p < .01$, *** $p < .001$				

Je stärker zugestimmt wird, dass Daten ein genaues und objektives Verständnis der Welt ermöglichen, dass aus ihnen Wissen gewonnen und Nutzen gezogen wird, desto positiver wird eine umfangreiche Datenspeicherung und -auswertung bewertet. Dies spricht für die erwartete positive Valenz der Sichtweise auf die digitalen Daten im Rahmen des besprochenen Glaubenssystems und somit auch für die Validität der Messung des *BDGS*.

9.1.8 Diskussion der Ergebnisse der Skalenkonstruktion

Als erste Erkenntnis vorweg und es ist dabei nicht banal das festzustellen: Die befragten Personen haben sich in der überwiegenden Mehrheit zugetraut, die ihnen gestellten Fragen unabhängig der Dimensionszugehörigkeit der einzelnen Items zu beantworten, auch wenn die Möglichkeit bestand, die Ausweichkategorie „Kann ich nicht beantworten“ auszuwählen. Dies wurde mit Blick auf die in Abschnitt 9.1.3 besprochenen Resultate dort jedoch nur von höchstens 6,3 % der Befragten bei Item *GE3* („Digitale Daten ermöglichen ein objektives Verständnis der Wirklichkeit“) auch gewählt und lag unter diesem Wert für die anderen erfragten Items. Das ist zwar keine vernachlässigbar geringe Anzahl an Personen, die in der Befragungssituation ein generelles oder zumindest partielles Nicht-Wissen angegeben haben. Man könnte jedoch genauso gut erwarten, dass sich eine noch höhere Anzahl an Personen keine Beantwortung der Fragen zutraut. Basierend auf dem Digitalindex 2018/2019 von Initiative D21 e. V. (2019) lässt sich die deutsche Bevölkerung in ‚Digitale Vorreiter‘ (37 %), ‚Digital Mithaltende‘ (42 %) und ‚Digital Abseitsstehende‘ (21 %) einteilen. Die Abseitsstehenden zeichnen sich durch einen geringen Zugang, geringe Kompetenz, wenig vielfältiges Nutzungsverhalten und geringe Offenheit gegenüber digitalen Anwendungen aus. Hier ist eine hohe Quote von Non-Responses zu erwarten. Nun kann es sein, dass sich in den jeweiligen Stichproben weniger Personen dieser Gruppe sammeln und andere Subgruppen überrepräsentiert waren. Der Samplequalität und dem Zufall muss an dieser Stelle zumindest Rechnung getragen werden. Doch selbst wenn alle Befragten zu den Mitgliedern der als „Digitale Vorreiter“ definierten Gruppe entstammten, dann ist mit Blick auf die derzeitigen Zugangschancen und das Berührungspotential mit dem Phänomen Big Data (siehe Kapitel 3) in den Samples über alle Befragten hinweg ein Ausdruck des Selbstvertrauens in die Beantwortungsfähigkeit der gestellten Fragen zu erkennen, das nicht unerheblich ist.

Mit Blick auf die Verteilungswerte der Indikatoren der vier Dimensionen zeigt sich, dass eine variable Einschätzung bezüglich der *Genauigkeit* digitaler Daten sowie des hieraus gezogenen *individuellen* und *gesellschaftlichen Nutzens* besteht. Die Befragten haben hier im Mittel eine abwägende Wahrnehmung, die durchaus eine kritische Sicht auf die vermeintliche Genauigkeit und Objektivität der Daten und hieraus gezogene Nutzengewinne offenbart, da diesen Eigenschaften der Daten nicht automatisch zugestimmt wird. Geht es jedoch um den *Wissensgewinn*, so wird den jeweiligen Glaubenssätzen im Aggregat zugestimmt, die Überzeugungen von aus den digitalen Daten gewonnener Erkenntnis formulieren.

9.1.9 Skalenanpassungen auf Basis von Erkenntnissen der ersten Erhebung der Skalenkonstruktion

Die in den vorherigen Abschnitten angesprochene aus der ersten Erhebung gewonnene Erkenntnis ist, dass es nicht ganz unerheblich ist, eine „weiß nicht“-Option bei der Beantwortung der *BDGS*-Skala anzubieten – da ihre Frageitems zumindest in Teilen das Wissen und den Erkenntnisgewinn selbst betreffen. So wurde auch in Reviewerkomentaren zur Skala aus den nachfolgenden Erhebungen 2.1 und 2.2 der Wunsch geäußert, befragten Personen die Möglichkeit zu geben, Unsicherheit und Nicht-Wissen bezüglich des Erkenntnisgewinns gezielt anzugeben.

Das Anbieten einer solchen „weiß nicht“-Option könnte nach Krosnick et al. (2001) bei einigen Befragten jedoch auch zu ‚Satisficing‘ führen. Satisficing bedeutet, es wird aus unterschiedlichen Gründen (Vermeidung kognitiver Arbeit, Vermeidung des persönlichen Eingeständnisses von Unwissenheit) keine bedeutungsvolle Antwort gegeben, was als Datenqualitätsverlust gewertet werden kann. Die Messung kognitiver Arbeit in der Vorbereitung einer Antwort und der Überzeugungsstärke, mit der die Antwort gegeben wurde, ist jedoch mit der vorliegenden Messung nur bedingt möglich. Die (Wieder-)Aufnahme einer Ausweichoption wurde für die nachfolgenden Erhebungen aufgrund geringer Auswahlhäufigkeit verworfen, wobei ihr Vorhandensein bei zukünftigen Einsätzen der *BDGS*-Skala oder ähnlicher Messungen mit Blick auf Zielgruppe und Forschungsinteresse reflektiert werden sollte bspw. beim Einsatz der Skala bei älteren Befragten.

9.2 Validierung der Skalenkonstruktion in Erhebung 1.2 und Erhebung 1.3

Die zwölf Frageitems und die zugrunde liegende Modellstruktur aus Erhebung 1.1 wurden anschließend in zwei Erhebungen validiert, die im Folgenden auch als Erhebung 1.2 und Erhebung 1.3 bezeichnet werden. Ziel dieser Validierung ist zu prüfen, ob Item- und Skaleneigenschaften sowie die Modellstruktur stabil sind und sich die *BDGS*-Skala über unterschiedliche Untersuchungskontexte hinweg einsetzen lässt.

Zudem wurde in Erhebung 1.3 erneut die diskriminante Validität der einzelnen Dimensionen in den Fokus genommen, um zu erörtern, inwieweit die Präsentationsweise der *BDGS*-Skala einen Einfluss auf die Bearbeitung hat. Somit wurde dort neben einer zusätzlichen Validierung der Dimensionen und ihrer Items eine leicht veränderte Skalenpräsentation getestet, deren Ergebnisse mit Blick auf den zukünftigen Einsatz der Skala diskutiert werden.

Erhebung 1.2 – Masterseminar Digitalisierungsfolgen

Im Rahmen des MA-Seminars ‚Die dunkle Seite der Macht: Risiken der Digitalisierung‘ im Sommersemester 2017 und im Wintersemester 2017/2018 am Institut für Kommunikationswissenschaft in Münster bestand die Möglichkeit für eine Validierung der Skala mit einem repräsentativen Sample. Ziel des Seminars war die Untersuchung der Wahrnehmung von Privatheitsrisiken in der deutschen Online-Bevölkerung mit Blick auf die drei Digitalisierungskontexte E-Commerce, politische Online-Kommunikation und Self-Tracking. Das Sample wurde daher aufgeteilt, so dass je eine Teilgruppe zu einem der drei Anwendungsfälle befragt wurde. Weitere Informationen zu dieser zur Publikation angenommenen Studie finden sich bei Lünich et al. (2021) und bei Klinger et al. (2018).

Erhebung 1.3. – Open Data

In Kooperation mit Dr. Christian Wiencierz vom Institut für Kommunikationswissenschaft in Münster wurde im Jahr 2018 ein Forschungsprojekt zu Open Data durchgeführt, das von der Deutschen Gesellschaft für Online-Forschung e. V. (DGOF) finanziell gefördert wurde. Ziel des Projekts war es, die Frage zu beantworten, inwiefern explizite Hinweise auf eingehaltenen Datenschutz und die darüberhinausgehende Selbstverpflichtung zur Einhaltung ethischer Leitlinien im Umgang mit personensensiblen Daten durch Plattformbetreibende einen Einfluss auf das Vertrauen der Plattform-Nutzenden haben. Um zu testen, welchen Einfluss die Menge an bereitgestellten Informationen auf das entgegengebrachte Vertrauen hat, wurde eine Online-Befragung mit Experimentaldesign durchgeführt.

In diesem bekamen die Teilnehmer*innen eine eigens für das Projekt programmierte Webseite eines fiktiven kommerziellen Open-Data-Plattformanbieters gezeigt, die jeweils unterschiedlich viele Hinweise (a) zum Datenschutz und (b) zu den ethischen Leitlinien im Umgang mit personensensiblen Daten bereithielt. Anschließend wurden sie zu ihrer Wahrnehmung des Anbieters und seiner Vertrauenswürdigkeit befragt. Weitere Informationen finden sich bei Wiencierz und Lünich (2020).

Bedeutung der Erhebung 1.3 für die Skalenkonstruktion

Neben der erneuten Validierung der Modellstruktur wurde die dritte Erhebung zur Skalenkonstruktion dazu genutzt, den Einfluss der Präsentationsweise der *BDGS*-Skala auf die diskriminante Validität zu untersuchen. Die Ergebnisse der Erhebung 1.1 indizieren, dass es aufgrund der hohen Inter-Korrelation der latenten Faktoren im CFA-Modell mitunter fraglich ist, ob überhaupt von diskriminanter Validität ausgegangen werden kann. Ein Grund für die hohen Korrelationswerte könnte darin liegen, dass die Präsentationsweise der *BDGS*-Skala mit den zwölf Frageitems auf einer Seite zu sogenannten Reihenfolgeeffekten auf Seiten der Befragten führt (Scholl, 2018). Es wäre bspw. ein Konsistenzeffekt zu beobachten, wenn die Beantwortung der ersten präsentierten Frage einen Einfluss auf die Beantwortung der nachfolgenden Frageitems hat. Es käme womöglich zu einem Auseinanderfallen der vom Befragten erwarteten oder ‚geratenen‘ Dimensionierung und der konzeptuellen sowie mutmaßlich empirischen Dimensionierung. Aufgrund der Rotation der Fragebogenitems auf der Seite befand sich zwar immer eine andere der insgesamt zwölf Aussagen am Anfang der jeweiligen Skala, es könnten jedoch auch der Gesamteindruck der Skala und die von den Befragten erwartete Zusammengehörigkeit der Items zu einem allgemeinen Test einen kognitiven Befragungseffekt auslösen. Die Befragten würden dann erwarten, dass die Items – unabhängig von der konkreten Formulierung – alle in eine bestimmte Richtung abzielen, so dass man auch alle Items ähnlich beantworten müsse. Scholl (2018) merkt hierzu Folgendes an:

Zu vermeiden sind solche Kontexteffekte nicht, es sei denn, man verzichtet darauf, mehrere Fragen zum gleichen Thema zu stellen. Eine andere Möglichkeit besteht darin, die Fragen zu trennen und an unterschiedlichen Stellen im Fragebogen abzufragen. Allerdings gefährdet diese Maßnahme die Stringenz der Befragung. Außerdem ist ihre Wirkung zweifelhaft, weil das Gedächtnis des Befragten durch die Folgefrage(n) reaktiviert wird. Wenn bekannt ist, wie Kontexteffekte funktionieren, kann man sie je nach Forschungsziel nutzen. (S. 217)

Um erste Erkenntnisse über eben jene Kontexteffekte auf die Beantwortung der *BDGS*-Skala zu gewinnen und zu prüfen, inwieweit die Präsentation der *BDGS*-Skala einen Einfluss auf die diskriminante Validität hat, wurde die Skala im vorliegenden Fragebogen aufgeteilt: Die Frageitems der Dimensionen *Genauigkeit* und *Wissensgewinn* fanden sich auf einer gemeinsamen Seite und die der Dimensionen zum *individuellen* und *gesellschaftlichen Nutzen* nach ein paar dazwischengeschobenen Fragenblöcken ein paar Seiten später gemeinsam auf einer Seite. Auch wenn die beiden Nutzendimensionen hoch miteinander korrelieren, wurde davon abgesehen, gerade die thematisch ähnlichen Dimensionen auf unterschiedliche Seiten aufzuteilen. Grund hierfür war, Irritationen auf Seiten der Befragten zu vermeiden, die durch mögliche mangelnde Unterscheidungsfähigkeit von auf mehreren Seiten aufgeteilten Fragen zum epistemischen Charakter der Daten und dem hieraus gezogenen Nutzen hervorgerufen werden könnte. Eine direkte Kontrastierung erschien bei diesen Dimensionen geboten.

9.2.1 Feldphasen und Sample

Erhebung 1.2 Feldphase, Sample und Fragebogaufbau

Die Feldphase der Befragung in Erhebung 1.2 lief vom 5. bis zum 13. Dezember 2017. Die Rekrutierung der Befragten erfolgte erneut durch die *Respondi AG*, die hierzu auf ihr Online-Access-Panel zugriff. Die Teilnehmer*innen der Untersuchung, die durch eine Vergütung für ihre Teilnahme incentiviert wurden, wurden nach Geschlecht, Alter (ab 14 Jahre) und Bildungsniveau quotiert auf Grundlage der aktuellen Zusammensetzung der deutschen Bevölkerung mit Onlinezugang, womit das Sample repräsentativ für diese Population ist (Arbeitsgemeinschaft Online Forschung – agof, 2017). Es wurden 9266 Personen aus dem Panel eingeladen. Insgesamt wurde der Fragebogen 1921-mal aufgerufen. Mit der Beantwortung begannen 1807 Befragte, wovon wiederum 1320 Befragte den Fragebogen beendeten und die letzte Fragebogenseite erreichten. Die Antwortrate lag somit bei 14,2 %.

Insgesamt wurden also $n = 1320$ Personen befragt, für die vollständig ausgefüllte Fragebögen vorlagen. Im Anschluss an die Datenerhebung wurde der Datensatz bereinigt. Dabei wurden, wie in Abschnitt 9.1.2 dokumentiert, auch hier Befragte von der Auswertung ausgeschlossen, die auffällig schnell durch den Fragebogen geklickt haben. Die Zeit für den Screenout lag hier bei weniger als 7 Minuten und 34 Sekunden; die Zeit, die man nach internen Tests in

der Projektgruppe mindestens benötigte, um alle Fragen gewissenhaft zu bearbeiten. Das finale Sample enthält mithin $n = 1114$ Fälle und setzt sich wie folgt zusammen: Männer und Frauen waren in ungefähr gleichen Anteilen vertreten (50,7 % männlich, 49,3 % weiblich). Im Durchschnitt waren die Befragten 45 Jahre alt ($SD = 15,54$). Knapp 29,8 % der Untersuchungsteilnehmer*innen wiesen eine niedrige Bildung auf (Volks- oder Hauptschule), 33,7 % eine mittlere Bildung (mittlere Reife) sowie 36,5 % eine hohe Bildung (mindestens Hochschulreife oder Hochschulabschluss). Die Verteilung der drei Teilstichproben war annähernd gleich (E-Commerce: $n = 372$, 33,4 %; Politische Online-Partizipation: $n = 375$, 33,7 %; Self-Tracking: $n = 367$, 32,9 %).¹⁵

Die *BDGS*-Skala fand sich in der Mitte des Fragebogens auf der zwölften von 23 Seiten (siehe Fragebogen im Anhang 3.2 im elektronischen Zusatzmaterial). Die Frageitems der *BDGS*-Skala wurden rotiert, so dass die Reihenfolge der gezeigten Aussagen jeweils zufällig festgelegt wurde. Die drei untersuchten Kontexte wurden vor der Beantwortung der *BDGS*-Skala eingeführt und jeweils mit einem eigenen Einleitungstext erklärt. Da es in den drei Kontexten explizit um die Sammlung und Auswertung privater digitaler Daten ging, ist es wahrscheinlich, dass bedingt durch diesen geteilten Fragebogaufbau jeweils Befragungseffekte bei der Beantwortung der *BDGS*-Skala zu erwarten sind. Deshalb werden nachfolgend für die Kontexte E-Commerce (nachfolgend Erhebung 1.2a), politische Online-Kommunikation (Erhebung 1.2b) und Self-Tracking (Erhebung 1.2c) die jeweiligen Kennwerte für Indikatoren und Messmodelle gesondert berichtet, jedoch jeweils gemeinsam besprochen.

Der Aufbau und die Formulierungen der *BDGS*-Skala blieben identisch zu Erhebung 1.1, wobei nur noch 18 Items gezeigt wurden. Dazu zählen die zwölf Items der vier Dimensionen *Genauigkeit*, *Wissensgewinn*, *individueller Nutzen* und *gesellschaftlicher Nutzen* sowie jeweils drei Items zur Bewertung des Ausmaßes (Volume) und der Geschwindigkeit (Velocity) der Datenspeicherung und -auswertung. Letztere werden in der vorliegenden Arbeit jedoch aus oben genannten Gründen nicht weiter berücksichtigt. Zudem wurde die Ausweichoption „weiß nicht“ ab Erhebung 1.2 für alle *BDGS*-Items gestrichen, so dass nur noch eine Verortung auf der siebenstufigen Beantwortungsskala möglich war.

---

¹⁵ Die Befragten wurden mittels ‚Urnenziehung ohne Zurücklegen‘ einer der drei Befragungskonditionen zugelost.

Erhebung 1.3 – Feldphase, Sample und Fragebenaufbau

Die Feldphase der Befragung lief vom 5. bis zum 12. Juni 2018. Die Rekrutierung der Befragten erfolgte auch hier durch die *Respondi AG*, die hierzu auf ihr Online-Access-Panel zugriff. Die Teilnehmer*innen der Untersuchung, die durch eine Vergütung für ihre Teilnahme incentiviert wurden, wurden zufällig gezogen. Es wurden 2767 Einladungen an das Panel versendet. Insgesamt wurde der Fragebogen 1364-mal aufgerufen. Mit der Beantwortung begannen 1288 Befragte, wovon wiederum 836 Befragte den Fragebogen beendeten und die letzte Fragebogen-seite erreichten. Die Antwortrate lag somit bei 30,2 %. Hierbei ist anzumerken, dass kurz vor Ende der Befragung ein Attention-Check eingebaut wurde, der die Befragten aufforderte, bei einem Item in der Fragebatterie einen bestimmten Skalenpunkt anzukreuzen. 209 Personen kreuzten hier falsch und wurden ohne Vergütung zum Panelanbieter zurückgeleitet, da anzunehmen ist, dass sie nicht alle Fragen aufmerksam und gewissenhaft gelesen und bearbeitet haben.

Insgesamt wurden also $n = 836$ Personen befragt, für die vollständig ausgefüllte Fragebögen vorlagen. Im Anschluss an die Datenerhebung wurde der Datensatz bereinigt. Dabei wurden auch hier Befragte von der Auswertung ausgeschlossen, die auffällig schnell durch den Fragebogen geklickt haben. Die Zeit für den Screenout lag hier bei weniger als 2 Minuten und 30 Sekunden, also mindestens zwei Standardabweichungen weniger als die durchschnittliche Beantwortungszeit von 11 Minuten und 27 Sekunden. Dies traf jedoch nur auf einen Fall zu. Zudem wurden Befragte von der Datenanalyse ausgeschlossen, bei denen nicht sichergestellt werden konnte, dass der Fragebogen technisch einwandfrei funktioniert hat ($n = 19$). Schließlich wurden Befragte ausgeschlossen, die entgegen der ihnen zugewiesenen Experimentalkondition angaben, keine Informationen zum Datenschutz ($n = 9$) oder zu den ethischen Leitlinien im Umgang mit persönnensensiblen Daten ($n = 3$) gesehen zu haben. Das finale Sample enthält mithin $n = 784$ Fälle und setzt sich wie folgt zusammen: Männer und Frauen waren in etwa gleichen Anteilen vertreten (48,9 % männlich, 51,5 % weiblich). Im Durchschnitt waren die Befragten 47,7 Jahre alt ($SD = 14,52$). Knapp 13,3 % der Untersuchungsteilnehmer*innen wiesen eine niedrige Bildung auf (Volks- oder Hauptschule), 35,2 % eine mittlere Bildung (mittlere Reife) sowie 51,1 % eine hohe Bildung (mind. Hochschulreife oder Hochschulabschluss). Damit ist das Sample etwas höher gebildet als in Erhebung 1.2.

Die Frageitems der *BDGS*-Skala fanden sich mit *Genauigkeit* und *Wissensgewinn* auf Seite 3 und *individueller* und *gesellschaftlicher Nutzen* auf Seite 7 (siehe Fragebogen im Anhang 3.3 im elektronischen Zusatzmaterial). Auf den drei Seiten dazwischen fanden sich Fragen zu informationellen Privatheitsbedenken bei

der Internetnutzung, der Intensität der Online- und Smartphone-Nutzung sowie der Intensität der Nutzung von Medienberichterstattung über die Digitalisierung.

9.2.2 Datenauswertung und Ergebnisse

Die Datenauswertung der Erhebungen 1.2 und 1.3 folgt der Modellspezifikation des Gesamtmodells des *BDGS* aus Erhebung 1.1 (siehe Abbildung 9.2) und dem in Abschnitt 9.1.3 beschriebenen Vorgehen bei der Datenanalyse mit AMOS 23.

Alle Modelle wurden erneut mit der FIML-Schätzung geschätzt. Die Mittelwerte der latenten Faktoren wurden erneut auf den Wert null restringiert, so dass die Achsenabschnitte (Regressionskonstanten) aller beobachteten Indikatoren geschätzt werden konnten.

Die Datenauswertung für Erhebung 1.2 greift zudem die Logik der Kontextunterscheidung in der Untersuchungsanlage auf und berichtet nachfolgend die Ergebnisse für jeweils eine CFA für das Szenario E-Commerce, politische Online-Partizipation und Self-Tracking.

Die CFA für das jeweils identisch spezifizierte Modell kommt in allen drei Kontexten der Erhebung 1.2 zu einer plausiblen Lösung. Im Sinne der einfachen Vergleichbarkeit der drei Kontexte finden sich die jeweiligen Werte für die drei CFAs in den folgenden Tabellen nebeneinander abgetragen. Es zeigt sich mit Blick auf die in Tabelle 9.5 berichteten Fit-Maße, dass der strenge Chi-Quadrat-Test jeweils eine unzureichende Anpassungsgüte zeigt, die inkrementellen Fit-Maße des RMSEA und der TLI jedoch jeweils eine hohe Anpassungsgüte indizieren. Auch für diese Modelle würden freizusetzende Korrelationen der Fehlerterme den Modellfit jeweils verbessern, diese folgen jedoch mit Blick auf die Modifikationsindizes und gemeinsam geteilte Varianz keiner Systematik.

Die CFA für das spezifizierte Modell kommt auch für das Sample der Erhebung 1.3 zu einer plausiblen Lösung und zeigt mit Blick auf den strengen Chi-Quadrat-Test eine unzureichende Anpassungsgüte ($\chi^2(48) = 136.608$, $p < .001$; RMSEA = .049 [.039, .058]; TLI = .981). Der RMSEA und der TLI indizieren jedoch eine ausreichend hohe Anpassungsgüte. Auch für dieses Modell würden freizusetzende Korrelationen der Fehlerterme den Modellfit jeweils verbessern. Zudem ist hier ebenfalls zu bedenken, dass das vorliegende Modell bei ganzen 784 Beobachtungen zu einem signifikanten Chi-Quadrat-Test tendiert (siehe hierzu Abschnitt 9.1.3).

Tabelle 9.5 Fit-Maße des BDGS-Modell (CFM) – Erhebung 1.2

Erhebungskontext	$\chi^2(48)$	RMSEA (LO90; HI90)	TLI	<i>n</i>
Erhebung 1.2a – E-Commerce	103.835***	.056 (.041; .071)	.980	372
Erhebung 1.2b – Pol. Online-Kommunikation	92.037***	.050 (.034; .065)	.985	375
Erhebung 1.2c – Self-Tracking	102.060***	.055 (.040; .070)	.981	367

Notizen. * $p < .05$, ** $p < .01$, *** $p < .001$

In Tabelle 9.6 finden sich die jeweiligen Angaben zum Cronbachschen α und der DEV der jeweiligen Dimensionen in Erhebung 1.2 und 1.3. Die Werte indizieren für alle Modelle über alle vier Dimensionen hinweg eine hohe interne Konsistenz.

Darüber hinaus finden sich in der Tabelle 9.7 für Erhebung 1.2 und Tabelle 9.8 für Erhebung 1.3 jeweils die Regressionskonstanten, die Regressionswerte sowie die geschätzten Faktorladungen der einzelnen Indikatoren samt deren Dimensionszugehörigkeit innerhalb des *BDGS*. Es zeigt sich, wie auch in Erhebung 1.1, dass die Regressionskonstanten für die Indikatoren der Dimensionen *Genauigkeit*, *individueller Nutzen* und *gesellschaftlicher Nutzen* nahe dem Skalenmittelpunkt liegen. Für die Indikatoren der Dimension des Wissensgewinns hingegen liegen die Regressionskonstanten in der Erhebung 1.2 rund einen halben Skalenpunkt oberhalb der Skalenmitte, in Erhebung 1.3 sogar einen ganzen Skalenpunkt.

In Tabelle 9.9 finden sich die Inter-Korrelationen der vier Dimensionen des *BDGS* in den Erhebungen 1.2 und 1.3, die erneut hoch miteinander korrelieren.

Tabelle 9.6 Kennwerte Dimensionen BDGS-Modell – Erhebung 1.2 und 1.3

Dimension	Cronbachsches α (standardisiert)			Durchschnittlich extrahierte Varianz (DEV) ^{a)}				
	Erhebung 1.2a – E-Commerce	Erhebung 1.2b – Pol. Online-Kommunikation	Erhebung 1.2c – Self-Tracking	Erhebung 1.3	Erhebung 1.2a – E-Commerce	Erhebung 1.2b – Pol. Online-Kommunikation	Erhebung 1.2c – Self-Tracking	Erhebung 1.3
Genauigkeit	.883	.895	.893	.878	.719	.742	.739	.709
Wissensgewinn	.873	.883	.901	.877	.697	.718	.753	.714
Individueller Nutzen	.886	.917	.915	.872	.723	.790	.783	.682
Gesellschaftlicher Nutzen	.902	.911	.882	.865	.754	.776	.714	.697

Notizen. ^{a)} Die DEV wurde auf Grundlage der standardisierten Regressionswerte (Faktorladungen) im Gesamtmodell berechnet.

Notizen. ^a Die DEV wurde auf Grundlage der standardisierten Regressionswerte (Faktorladungen) im Gesamtmodell berechnet.

Tabelle 9.7 Kennwerte Dimensionen und Indikatoren – Erhebung 1.2: Modell BDGS

Item-ID	Itemformulierung „Digitale Daten ...“	Erhebung 1.2a – E-Commerce			Erhebung 1.2b – Pol. Online-Kommunikation			Erhebung 1.2c – Self-Tracking		
		B (SE)	β (SE)	λ	B (SE)	β (SE)	λ	B (SE)	β (SE)	λ
Genauigkeit										
GE1	... ermöglichen ein genaues Verständnis der Welt.	3.954 (.075)	1 ^a	.884	4.016 (.074)	1 ^a	.898	3.967 (.081)	1 ^a	.874
GE2	... ermöglichen ein exaktes Verständnis der Welt.	3.755 (.076)	.951*** (.041)	.837	3.899 (.079)	1.010*** (.044)	.856	3.771 (.084)	1.018*** (.045)	.866
GE3	... ermöglichen ein objektives Verständnis der Wirklichkeit.	3.901 (.071)	.877*** (.043)	.821	3.981 (.078)	.971*** (.046)	.828	3.959 (.077)	.910*** (.045)	.839
Wissen										
W1	... verhelfen zu neuem Wissen.	4.632 (.074)	1 ^a	.799	4.731 (.072)	1 ^a	.877	4.662 (.073)	1 ^a	.860
W2	... führen zu neuen Fakten.	4.546 (.072)	1.044*** (.058)	.853	4.464 (.074)	.966*** (.049)	.824	4.480 (.077)	1.056*** (.050)	.865
W3	... verhelfen zu neuen Erkenntnissen.	4.624 (.072)	1.048*** (.058)	.851	4.648 (.075)	1.007*** (.048)	.840	4.681 (.074)	1.030*** (.047)	.878
(Fortsetzung)										

(Fortsetzung)

Tabelle 9.7 (Fortsetzung)

Item-ID	Itemformulierung „Digitale Daten ...“	Erhebung 1.2a – E-Commerce		Erhebung 1.2b – Pol. Online-Kommunikation		Erhebung 1.2c – Self-Tracking	
		B (SE)	β (SE)	λ	B (SE)	β (SE)	λ
Individueller Nutzen							
IN1	... erzeugen für mich persönlich einen Mehrwert.	3.777 (.077)	1 ^a	.887	3.957 (.080)	1 ^a	.880
IN2	... führen für mich zu Vorteilen.	4.000 (.074)	.888*** (.043)	.823	3.949 (.077)	.947*** (.041)	.870
IN3	... haben für mich persönlich positive Auswirkungen.	3.981 (.072)	.884*** (.041)	.839	3.941 (.077)	1.007*** (.039)	.915
Gesellschaftlicher Nutzen							
GN1	... erzeugen einen Mehrwert für die Gesellschaft.	3.976 (.074)	1 ^a	.883	4.117 (.073)	1 ^a	.891
GN2	... führen zu gesellschaftlichen Vorteilen.	3.960 (.072)	.957*** (.041)	.871	3.971 (.075)	.987*** (.042)	.864
GN3	... sind ein gesellschaftlicher Gewinn.	3.944 (.074)	.970*** (.043)	.851	4.099 (.078)	1.060*** (.042)	.887

Notizen. ^a Referenzindikator, * $p < .05$, ** $p < .01$, *** $p < .001$

Notizen. ^a Referenzindikator, * $p < .05$, ** $p < .01$, *** $p < .001$

Tabelle 9.8 Kennwerte Dimensionen und Indikatoren – Erhebung 1.3: Modell BDGS

Item-ID	Itemformulierung „Digitale Daten ...“	<i>B</i> (<i>SE</i>)	β (<i>SE</i>)	λ
Genauigkeit				
GE1	... ermöglichen ein genaues Verständnis der Welt.	4.403 (.048)	1 ^a	.895
GE2	... ermöglichen ein exaktes Verständnis der Welt.	4.147 (.051)	.992*** (.033)	.832
GE3	... ermöglichen ein objektives Verständnis der Wirklichkeit.	4.205 (.048)	.887*** (.033)	.797
Wissen				
W1	... verhelfen zu neuem Wissen.	5.161 (.045)	1 ^a	.874
W2	... führen zu neuen Fakten.	4.797 (.045)	.865*** (.035)	.759
W3	... verhelfen zu neuen Erkenntnissen.	5.096 (.043)	.967*** (.030)	.896
Individueller Nutzen				
IN1	... erzeugen für mich persönlich einen Mehrwert.	3.978 (.045)	1 ^a	.833
IN2	... führen für mich zu Vorteilen.	4.065 (.042)	.924*** (.035)	.821
IN3	... haben für mich persönlich positive Auswirkungen.	3.995 (.041)	.898*** (.035)	.823
Gesellschaftlicher Nutzen				
GN1	... erzeugen einen Mehrwert für die Gesellschaft.	4.172 (.039)	1 ^a	.868
GN2	... führen zu gesellschaftlichen Vorteilen.	4.102 (.038)	.886*** (.034)	.782
GN3	... sind ein gesellschaftlicher Gewinn.	4.129 (.040)	1.000*** (.034)	.852
Notizen. ^a Referenzindikator; * $p < .05$, ** $p < .01$, *** $p < .001$				

Tabelle 9.9 Korrelationsmatrix – Erhebung 1.2 und 1.3: Finales Modell BDGS

	Genauigkeit	Wissensgewinn	Individueller Nutzen	Gesellschaftlicher Nutzen
Genauigkeit				
Erhebung 1.2a – E-Commerce	1	.817***	.857***	.895***
Erhebung 1.2b – Pol. Online-Kommunikation	1	.761***	.791***	.888***
Erhebung 1.2c – Self-Tracking	1	.819***	.784***	.895***
Erhebung 1.3	1	.727***	.512***	.528***
Wissensgewinn				
Erhebung 1.2a – E-Commerce		1	.790***	.807***
Erhebung 1.2b – Pol. Online-Kommunikation		1	.756***	.853***
Erhebung 1.2c – Self-Tracking		1	.735***	.853***
Erhebung 1.3		1	.458***	.478***
Individueller Nutzen				
Erhebung 1.2a – E-Commerce			1	.947***
Erhebung 1.2b – Pol. Online-Kommunikation			1	.886***
Erhebung 1.2c – Self-Tracking			1	.901***
Erhebung 1.3			1	.877***

Notizen. * $p < .05$, ** $p < .01$, *** $p < .001$

Prüfung auf Messinvarianz

Als Voraussetzung für die Vergleichbarkeit der Messung einzelner Erhebungen muss zunächst die Messinvarianz (synonym auch als Messäquivalenz bezeichnet) des Gesamtmodells überprüft werden, also, ob und inwieweit sich die Beziehungsstruktur des Messmodells für die vorliegenden Stichproben unterscheidet (van de Schoot et al., 2012; Weiber & Mülhhaus, 2014). Nur bei Messäquivalenz kann davon ausgegangen werden, dass die latenten Faktoren das gleiche

Konstrukt messen und sinnvoll miteinander verglichen werden können (Widaman & Reise, 1997). Dabei gibt es der Ausarbeitung von Weiber und Mühlhaus (2014) folgend unterschiedliche Grade an Messinvarianz: Konfigurale Messinvarianz bedeutet, dass das spezifizierte Modell über die verglichenen Gruppen hinweg eine gleich hohe Anpassungsgüte besitzt und indiziert einen schwachen Grad faktorieller Invarianz. Metrische Messinvarianz bedeutet, dass die Stärke des Zusammenhangs zwischen Indikator und latentem Konstrukt über die verglichenen Gruppen hinweg identisch ist und wird auch als mittlerer Grad faktorieller Invarianz bezeichnet. Ein starker Grad faktorieller Invarianz liegt weiterhin vor, wenn bei gleicher Ausprägung des Konstrukts auch die Regressionskonstanten aller Indikatoren gleich ausfallen. Nur bei starker faktorieller Messinvarianz können die Mittelwerte der latenten Faktoren über die Gruppen hinweg verglichen werden. Zudem kann noch auf strikte faktorielle Invarianz geprüft werden, für die von einer Invarianz der Messfehler ausgegangen wird.

Für die drei Erhebungen 1.1, 1.2 und 1.3 wurde eine entsprechende Prüfung durchgeführt. Die Ergebnisse in Tabelle 9.10 zeigen, dass mit Blick auf den Chi-Quadrat-Differenztest die Modelle mit metrischer und skalarer Invarianz eine signifikant schlechtere Anpassungsgüte zeigen. Allerdings gibt es Hinweise darauf, dass auch im Rahmen der Prüfung auf Messinvarianz die Stichprobengröße (hier $n = 2153$) sowie die Anzahl der verglichenen Gruppen sowie die Größe des geschätzten Modells dahingehend einen Einfluss auf den Test auf Messinvarianz haben, als dass der strenge Chi-Quadrat-Differenztest schnell dazu tendiert ein signifikantes Ergebnis zu erbringen (Putnick & Bornstein, 2016). Mithin werden auch hier häufig inkrementelle Fit-Maße zur Beurteilung der Messinvarianz herangezogen (F. F. Chen, 2007; Meade et al., 2008), die eine Einschätzung erlauben, in welchem Grad sich das Modell durch die Gleichheitsrestriktionen verschlechtert.

Tabelle 9.11 Mittelwerte der latenten Faktoren im BDGS-Modell – Erhebung 1.1, 1.2 und 1.3

Dimension	<i>M (SE)</i>				
	Erhebung 1.1	Erhebung 1.2a E-Commerce	Erhebung 1.2b Pol. Online-Kommunikation	Erhebung 1.2c Self-Tracking	Erhebung 1.3
Genauigkeit	0 ^a	.080 ^b (.126)	.170 ^c (.127)	.110 ^d (.130)d	.487 ^{a,b,c,d} (.114)
Wissensgewinn	0 ^a	-.097 ^b (.111)	-.084 ^c (.111)	-.093 ^d (.113)d	.332 ^{a,b,c,d} (.098)
Individueller Nutzen	0 ^a	.133 ^b (.121)	.174 ^c (.124)	.055 ^d (.127)d	.239 ^{a,d} (.106)
Gesellschaftlicher Nutzen	0 ^a	.110 ^b (.119)	.216 ^c (.120)	.047 ^d (.120)d	.285 ^{a,b,d} (.103)

Notizen. ^a Werte, die einen gleichen Buchstaben teilen, unterscheiden sich anhand des Critical-Ratio-Tests signifikant mit $p < .05$.

So wird mit Blick auf Tabelle 9.10 ersichtlich, dass das Modell mit der skalaren Invarianz bezüglich der inkrementellen Fit-Maße des TLI und des RMSEA eine höhere bzw. mindestens gleich hohe Anpassungsgüte zeigt, verglichen mit den Modellen mit konfiguraler Invarianz bzw. metrischer Invarianz. Das bedeutet, dass das Modell mit den Gleichheitsrestriktionen eine signifikant, jedoch nicht substantiell schlechtere Anpassungsgüte zeigt und mithin von starker Messinvarianz ausgegangen wird.

Durch die Festsetzung der Gleichheitsrestriktionen lassen sich bei starker Messinvarianz die Mittelwerte der latenten Konstrukte über die Gruppen hinweg vergleichen, wobei die Stichprobe der Erhebung 1.1 als Referenzgruppe festgelegt wird, deren Faktor-Mittelwerte auf 0 restringiert sind (siehe Tabelle 9.11).

Es zeigt sich mit Blick auf die Mittelwerte, dass für alle Dimensionen des *BDGS* kein Unterschied zwischen den Erhebungen 1.1 und 1.2 besteht, die Mittelwerte der Dimensionen in Erhebung 1.3 jedoch signifikant höher ausfallen als in Erhebung 1.1 und in Teilen auch als in der Erhebung 1.2.

Prüfung der dimensionalen Verschiedenheit

Wie im Rahmen der Erhebung 1.1 diskutiert, stellt sich die Frage der DV. Daher wurde zur erneuten Prüfung der DV auch in Erhebung 1.2 sowie Erhebung 1.3 ein SOFM spezifiziert, das Aussagen über die spezifische Varianz der latenten Faktoren erlaubt.

Für alle drei Kontexte in Erhebung 1.2 zeigen sich signifikante Varianzen der Faktoren

- *Genauigkeit* ($\sigma^2_{\text{Genauigkeit\_1.2a}} = .263$, SE = .048, $p < .001$; $\sigma^2_{\text{Genauigkeit\_1.2b}} = .348$, SE = .051, $p < .001$; $\sigma^2_{\text{Genauigkeit\_1.2c}} = .345$, S.E. = .062, $p < .001$),
- *Wissensgewinn* ($\sigma^2_{\text{Wissensgewinn\_1.2a}} = .377$, SE = .055, $p < .001$; $\sigma^2_{\text{Wissensgewinn\_1.2b}} = .398$, SE = .055, $p < .001$; $\sigma^2_{\text{Wissensgewinn\_1.2c}} = .385$, S.E. = .053, $p < .001$),
- *Individueller Nutzen* ($\sigma^2_{\text{Ind. Nutzen\_1.2a}} = .153$, SE = .041, $p < .001$; $\sigma^2_{\text{Ind. Nutzen\_1.2b}} = .391$, SE = .054, $p < .001$; $\sigma^2_{\text{Ind. Nutzen\_1.2c}} = .454$, S.E. = .063, $p < .001$).

Die Varianz des Faktors *gesellschaftlicher Nutzen* ist jedoch in allen drei Kontexten nicht signifikant ($\sigma^2_{\text{Ges. Nutzen}_{1.2a}} = .055$, $\text{SE} = .032$, $p = .088$; $\sigma^2_{\text{Ges. Nutzen}_{1.2b}} = .006$, $\text{SE} = .032$, $p = .854$; $\sigma^2_{\text{Ges. Nutzen}_{1.2c}} = -.003$, $\text{S.E.} = .036$, $p = .932$).¹⁶

Auch in Erhebung 1.2 indizieren die Ergebnisse daher, dass sich mit Blick auf die DV die Dimensionen *Genauigkeit*, *Wissensgewinn* und *individueller Nutzen* voneinander unterscheiden lassen, die Bewertung des *gesellschaftlichen Nutzens* sich jedoch nicht ausreichend von den drei anderen Dimensionen innerhalb des *BDGS* unterscheidet (Gignac & Kretzschmar, 2017).

Daher wurde zur erneuten Prüfung der DV in Erhebung 1.3 ebenfalls ein SOFM spezifiziert, das Aussagen über die spezifische Varianz der latenten Faktoren erlaubt. Es zeigen sich hier signifikante Varianzen der Faktoren *Genauigkeit* ($\sigma^2_{\text{Genauigkeit}} = .898$, $\text{S.E.} = .068$, $p < .001$), *Wissensgewinn* ($\sigma^2_{\text{Wissensgewinn}} = .852$, $\text{S.E.} = .062$, $p < .001$), *individueller Nutzen* ($\sigma^2_{\text{Ind. Nutzen}} = .176$, $\text{S.E.} = .033$, $p < .001$) sowie, anders als bei den Erhebungen 1.1 und 1.2 zuvor, auch des *gesellschaftlichen Nutzens* ($\sigma^2_{\text{Ges. Nutzen}} = .115$, $\text{S.E.} = .026$, $p < .001$). Die Varianz der Nutzendimensionen fällt jedoch im Vergleich mit den anderen beiden Dimensionen trotz Signifikanz deutlich geringer aus und begründet sich auch in der größeren Stichprobe.

An dieser Stelle der Skalenkonstruktion erfolgt daher folgendes Fazit bezüglich der dimensional Struktur des Modells: Einem CFM mit Korrelationen zwischen den vier latenten Faktoren wird der Vorzug vor einem Single-Factor-Modell oder einem Second-Order-Factor-Modell gegeben. Die einzelnen Dimensionen lassen sich als eigenständige latente Faktoren identifizieren und voneinander unterscheiden, auch wenn es leichte Einschränkungen mit Blick auf die Stärke der dimensional Verschiedenheit gibt. Bei ausgeprägtem *BDGS* variieren die Befragten stärker mit Blick auf ihre Einschätzungen der Wahrhaftigkeitsdimensionen als mit Blick auf den wahrgenommenen Nutzen.

Die Prüfung von Reihenfolgeeffekten auf die Diskriminanzvalidität

Explizites Ziel der Erhebung 1.3 war zudem die Prüfung von Reihenfolgeeffekten auf die Diskriminanzvalidität der latenten Faktoren. Die Inter-Korrelationen zwischen den latenten Faktoren in den Erhebungen 1.1 und 1.2 erreichten in Teilen so hohe Werte, dass gemäß dem Fornell-Larcker-Kriterium keine Diskriminanzvalidität festgestellt werden konnte. Durch eine Aufteilung der Skala

¹⁶ Die Varianz des latenten Faktors in Kontext 1.2c ist negativ, so dass es aufgrund dieses Heywood Cases (siehe Abschnitt 9.1.3) zu einer unzulässigen Lösung des Schätzalgorithmus im SOFM kommt.

auf zwei Befragungsseiten sollte der Einfluss der Skalenpräsentation auf einer Seite geprüft werden, wobei zwischen den beiden Seiten der nun aufgeteilten Skala weitere Itembatterien positioniert wurden, um einen Ablenkungseffekt zu erzielen.

Zunächst richtet sich der Blick auf die Dimension *Genauigkeit* und deren quadrierte Korrelation mit der Dimension des *individuellen Nutzens* ($r^2_{\text{Genauigkeit, individueller Nutzen}} = .262$) und der Dimension des *gesellschaftlichen Nutzens* ($r^2_{\text{Genauigkeit, gesellschaftlicher Nutzen}} = .279$). Die in Tabelle 9.6 berichtete DEV der Faktoren zeigt nun deutlich, dass von diskriminanter Validität ausgegangen werden kann, da die jeweilige DEV größer als die gemeinsam geteilte Varianz der Faktoren ist. Schaut man auf die Dimension *Wissensgewinn* und *individueller Nutzen* ($r^2_{\text{Wissensgewinn, individueller Nutzen}} = .210$) und *gesellschaftlicher Nutzen* ($r^2_{\text{Wissensgewinn, gesellschaftlicher Nutzen}} = .228$), fallen die Werte für die quadrierten Korrelationen ebenfalls deutlich niedriger aus als jeweils beide DEV-Werte. Vergleicht man gemeinsam geteilte Varianz der latenten Faktoren, die sich auf unterschiedlichen Fragebogenseiten befanden, mit denen aus Erhebung 1.1 (siehe Tabelle 9.3) und Erhebung 1.2 (siehe Tabelle 9.9), die auf einer Seite angezeigt wurden, so lässt sich rund die Hälfte bis zwei Drittel der Varianz durch den Reihenfolgeeffekt erklären.

9.2.3 Diskussion der Ergebnisse

Die Anpassungsgüte des *BDGS*-Modells wird auch in den beiden Erhebungen 1.2 und 1.3 der Studie 1 als ausreichend bis zufriedenstellend bewertet. Zwar verfehlt die Modellierung des *BDGS* in beiden Erhebungen den strengen Chi-Quadrat-Test. Dennoch wurde auf Grundlage der vorliegenden Konzeption die Freisetzung der restringierten Parameter, also die Zulässigkeit korrelierter Fehlerterme, vermieden, da die inkrementellen Fit-Maße eine gute Passung indizieren.

Erhebung 1.2

Die Mittelwerte der Dimensionen des *BDGS* zeigen im Vergleich der Erhebung 1.1 mit der Erhebung 1.2 über alle drei Kontexte einen ähnlichen Gesamteindruck. Diese Ähnlichkeit der Resultate führt zu der Einschätzung, dass die Messung des *BDGS* mit dem vorgeschlagenen Skalen-Instrument zum einen reliabel möglich ist und die in Erhebung 1.1 gewählten Indikatoren insgesamt geeignet

sind, die konzeptuelle Dimensionierung abzubilden. Zum anderen indizieren die Ergebnisse, dass das *BDGS* als ein stabiles Wahrnehmungsmuster erscheint. Die Ausprägung der Dimensionen zeigt, dass mit Blick auf die *Genauigkeit* der Daten sowie auf den hieraus gezogenen *individuellen* und *gesellschaftlichen Nutzen* von den Befragten durchaus abgewogen und im Durchschnitt keine eindeutig negative oder positive Bewertung abgegeben wird. Die Einschätzung der Datensammlung und -auswertung ist mithin durchaus ambivalent, während die Befragten durchweg einen *Wissensgewinn* aus den Daten erwarten.

Erhebung 1.3

Die Ähnlichkeit der Resultate von Erhebung 1.3 führt auch hier zunächst zu der Einschätzung, dass die Messung des *BDGS* mit dem vorgeschlagenen Skalen-Instrument reliabel möglich ist und die gewählten Indikatoren geeignet sind, die konzeptuelle Dimensionierung abzubilden. Es zeigt sich mit Blick auf die erste Forschungsfrage (siehe Kapitel 7), die danach fragt, inwieweit sich ein Glaube an digitale Daten in seinen mehrdimensionalen Bezugspunkten von Überzeugungen empirisch messen lässt, dass die vorgeschlagene Operationalisierung und Datenanalyse geeignet ist, das theoretisch hergeleitete *BDGS* abzubilden.

Bezüglich der zweiten aufgeworfenen Forschungsfrage, die nach der Ausprägung der Überzeugungen und ihrer Beziehungen untereinander fragt, zeigt sich, dass die Mittelwerte der einzelnen Dimensionen in Erhebung 1.3 hierbei etwas größer ausfallen als in den beiden Erhebungen zuvor. Dieses Ergebnis indiziert, dass trotz der generellen Stabilität des *BDGS* je nach Stichprobe unterschiedlich hohe Ausprägungen des *BDGS* beobachtet werden können. Es zeigt sich auch hier erneut, dass der erwartete *Wissensgewinn* positiv ausgeprägt ist, während die Ausprägungen der anderen Dimensionen um den Skalenmittelpunkt liegen.

Im Aggregat erfahren die Aussagen, dass digitale Daten zu einem *Wissensgewinn* führen, durchweg eine hohe Zustimmung. Dies ist als erstes Indiz dafür zu werten, dass der Mythos eines weitreichenden Erkenntnisgewinns durch die Sammlung und Auswertung großer digitaler Datenbestände einen Niederschlag in der Wahrnehmung der Befragten gefunden hat. Diese positive Konnotation von Wissen vor dem Hintergrund der Wissensgesellschaft ist dabei eng verbunden mit Erwartungen an einen *individuellen* und *gesellschaftlichen Nutzen*.

Insgesamt zeigen sich hohe Inter-Korrelationen zwischen den einzelnen Dimensionen des *BDGS*, was indiziert, dass ihre jeweilige Ausprägung untereinander in einer engen Beziehung steht. Dabei gibt Erhebung 1.3 entscheidende Hinweise zur Diskriminanzvalidität der latenten Faktoren des Modells. Die Präsentation der Items führt zu einem kognitiven Befragungseffekt. Mit Blick auf die Ergebnisse sind Befragte eher dazu geneigt, zwischen den einzelnen Dimensionen

des *BDGS* zu unterscheiden, wenn diese auf unterschiedlichen Fragebogenseiten zu finden sind.

Es wird allerdings aus diversen Gründen nur in den wenigsten Befragungsstudien möglich sein, eine Skala, wie sie mit der Messung des *BDGS* vorliegt, so in den Fragebogenverlauf einzubetten, dass ihr Einsatz gänzlich unproblematisch ist. Dies wird vermutlich nur dort der Fall sein, wo eine ganz spezielle Dimension von Forschungsinteresse ist. Sind alle Dimensionen gleich bedeutend, so stellt sich das Problem, dass (a) die *BDGS*-Skala aus forschungsökonomischen Gründen nicht auf mehrere Seiten aufgeteilt werden kann, (b) berücksichtigt werden muss, dass auch eine Aufteilung Reihenfolgeeffekte nach sich ziehen kann und (c) eine Aufteilung über mehrere Seiten immer dann problematisch ist, wenn die gesamte Befragung auf ein Phänomen oder Artefakt im Digitalisierungskontext abstellt und der restliche Fragebogen weitere Skalen mit Bezug zur Digitalisierung beinhaltet. Diese Problematik sollte bei jedem Einsatz der *BDGS*-Skala erneut sorgfältig geprüft werden.

Mit dem Wissen um die Reihenfolgeeffekte des Fragebogaufbaus bedeutet das Folgendes: In Studien, in denen die Wechselwirkungen zwischen den einzelnen Dimensionen im Vordergrund stehen und einer Aufteilung aus forschungsökonomischen Gründen nichts entgegensteht, können und sollten die Dimensionen auf unterschiedlichen Fragebogenseiten präsentiert werden, wobei sich auch hier die Frage stellt, in welcher Reihenfolge die Dimensionen abgefragt werden. So ließen sich womöglich konvergente und diskriminante Validität besser sicherstellen, als wenn alle Items rotiert auf einer Seite zu finden sind.¹⁷ Letzteres ist jedoch aus Gründen der Fragebogenlänge eine zu präferierende Lösung, wenn wie in den nachfolgenden Erhebungen die *BDGS*-Skala in Fragebögen mitläuft, die noch andere Forschungsziele verfolgen. Deshalb und aus forschungsökonomischen sowie Konsistenzgründen wurde daher in den nachfolgenden Erhebungen auf eine Aufteilung der *BDGS*-Skala verzichtet und diese jeweils auf einer Seite abgebildet, wobei die Reihenfolge der Items zufällig rotiert wurde.

---

¹⁷ Da dies tatsächlich nur für die beiden Dimensionen der Richtigkeit und der Nutzendimensionen geprüft wurde und nicht zwischen den jeweils zu den beiden Oberdimensionen gehörenden Subdimensionen, liegen noch keine umfassenden empirischen Belege vor und es bleibt zunächst bei einer Vermutung, dass dies für alle Dimensionen gelten könnte. Dies beinhaltet zudem noch keine Aussagen oder Erkenntnisse über mögliche Reihenfolgeeffekte bedingt durch die Abfolge der einzelnen Dimensionen.

9.3 Reflexion über die Skalenerstellung im Anschluss an Skalenkonstruktion und Validierung

Um eine Vergleichbarkeit der Ergebnisse aus den einzelnen Erhebungen der Skalenerstellung herzustellen, wurde die Skala wie dokumentiert in allen Erhebungen in nahezu gleicher Weise eingesetzt und wurden dabei die Einleitung der Fragebatterie, die Formulierungen der einzelnen Items sowie die Skalierung und die Benennung aller Skalenpunkte durchweg konstant gehalten. Es liegt in der Natur der Skalenerstellung, dass durch den Einsatz der Skala Erkenntnisse gewonnen wurden, die nach wiederholtem Einsatz für eine Anpassung sprächen. Aufgrund des Anspruchs an Vergleichbarkeit und weiterer Gründe, wie etwa der vermeintlichen Geringfügigkeit der Bedeutung vieler einzelner Probleme, wurden Änderungen jedoch nicht umgesetzt. Allerdings sollen Erkenntnisse, die u. a. aus der Rückmeldung der Befragten und Reviewer*innen gewonnen wurden, sowie die gesammelten Erfahrungen des federführenden Autors (die über die Diskussion der Diskriminanzvalidität im vorherigen Abschnitt hinausgehen) in diesem Abschnitt dokumentiert und reflektiert werden. Zum einen zur Dokumentation, um ein besseres Verständnis für die Verwendung der Skala herzustellen und etwaige Rückmeldungen der Befragten in zukünftigen Studien zu antizipieren. Zum anderen als Empfehlung, um bei zukünftigen Skalenentwicklungen frühzeitig bedacht zu werden und mögliche Befragungsprobleme zu vermeiden.

Man könnte dann aufgrund dieser Erfahrungen und weitergehendem Interesse mit Blick auf die Methodenforschung der Befragung bei der zukünftigen Verwendung mit Hilfe eines Split-Sample-Designs etwaige Änderungen testen und mögliche Auswirkungen auf die Beantwortung der Skala prüfen. Mögliche Probleme und entsprechende Änderungen lassen sich dabei ganz grob unterscheiden nach (a) der Problematik der Formulierungen und (b) der Präsentation der Fragebogenitems.

9.3.1 Problematik der Formulierungen der Fragebogenitems

Das Gros der Rückmeldungen betraf problematische Formulierungen innerhalb des Einleitungstextes sowie der Itemformulierungen selbst, die einen Einfluss auf das Verständnis der Frageitems haben könnten und dadurch auch auf die Ergebnisse der Testbatterie. Probleme konnten hier mit Blick auf (a) Synonyme und subjektiv identische Begriffe sowie (b) Einflüsse aktueller Ereignisse und Modewörter unterschieden werden.

Synonyme und subjektiv identische Begriffe

Manche Formulierungen waren laut einigen wenigen Rückmeldungen für manche Befragte subjektiv nicht unterscheidbar, was durch die hierdurch ausgelöste Irritation zu einer Verzögerung in der Bearbeitung führen könnte. So erfolgte hin und wieder der Hinweis (v. a. in der nachfolgenden *Erhebung 2.4 KI & Hochschule*), dass die Worte *exakt* und *genau*, die in den Frageitems mit den IDs *GE1* und *GE2* der Subdimension *Genauigkeit* vorkamen, identisch sind oder synonym verwendet werden. Dies schließt sich an übliche Rückmeldungen aus Befragungen an, dass bei den Befragten das Gefühl bestand, eigentlich immer das Gleiche gefragt zu werden. Im Sinne der Drei-Item-Messung, die für Indikatoren von latenten Konstrukten empfohlen wird (J. C. Anderson & Gerbing, 1984, 1988; Weiber & Mühlhaus, 2014) und in den vorliegenden Erhebungen eingesetzt wurde, besteht durchgängig die Schwierigkeit, Items so zu formulieren, dass sie zwar sprachlich zu unterscheiden sind, gleichzeitig jedoch das Wesen und den Kern des zugrunde liegenden Konstrukts so genau und repräsentativ wie möglich reflektieren sollten. Die Items sollen gut, also hoch, auf das jeweilige latente Konstrukt laden und gleichzeitig nicht zu ähnlich verfasst sein, was mitunter zur Monotonie des Fragebogens führt (Petersen, 2014; Weiber & Mühlhaus, 2014). Dies muss bei der Formulierung fortlaufend abgewogen werden und führte im vorliegenden Fall letztendlich dennoch zur Wahl synonyme Begrifflichkeiten. Das geringe Ausmaß entsprechender Rückmeldungen der Befragten in den einzelnen Erhebungen veranlassen hier dann auch nicht zur Beunruhigung bezüglich der Reliabilität und Validität der entsprechenden Items. Jedoch hat die sprachliche Nähe synonyme Formulierung einen Einfluss auf die Modellgüte der Messmodelle in der konfirmatorischen Faktorenanalyse, was mit Verweis auf die Modifikationsindizes nachfolgend an den entsprechenden Stellen vermerkt wird. Je nach Rotation der Items innerhalb der Skala kommt es zudem zu Unterschieden in der Stärke geteilter Varianz einzelner Frageitems einer Dimension. Mit Blick auf die Fallzahlen der Stichprobengrößen sollten sich diese zufällig zustande kommenden geteilten Varianzen aufgrund der Rotation zwar nur selten ergeben, auszuschließen sind sie jedoch nicht.

Mögliche Einflüsse aktueller Ereignisse und Diskurse

In der nachfolgend besprochenen *Erhebung 2.1* merkten zwei Befragte nach der Teilnahme an, dass die Formulierung „Digitale Daten führen zu neuen Fakten“ (Item-ID *W2*) vor dem Hintergrund der Debatte um sogenannte Fake News und

alternative Fakten als problematisch wahrgenommen wurde.¹⁸ Der Begriff ‚Alternative Fakten‘ fand im Januar 2017 öffentliche Aufmerksamkeit, als Kellyanne Conway, Beraterin des 45. US-Präsidenten Donald Trump, mit diesem Ausspruch widersprüchliche Angaben eines Pressesprechers des Präsidenten rechtfertigte. Das Wort erfuhr im Jahresverlauf wiederholt große Beachtung und wurde in Deutschland und Österreich zum Unwort des Jahres 2017 gewählt (Der Standard, 2018; Frankfurter Allgemeine Zeitung, 2018). Es ist nicht auszuschließen, dass einzelne Befragte sich während der Beantwortung des Items an jene Debatte erinnern. Allerdings ergibt die Datenauswertung der Erhebung 1.1 bis 1.3 keine Hinweise darauf, dass dies zu einer schlechteren Ladung des entsprechenden Items auf das latente Konstrukt geführt hat. Wie sich jedoch nachfolgend zeigen wird, weist das entsprechende Item ab *Erhebung 2.1* tatsächlich eine entsprechend schlechtere Ladung auf, worauf an der entsprechenden Stelle hingewiesen wird.

9.3.2 Problematik der Präsentation der Fragebogenitems

Die Aufteilung der *BDGS*-Skala auf zwei Fragebogenseiten in Erhebung 1.3 verfolgte das Ziel, Erkenntnisse über den Einfluss von Reihenfolgeeffekten auf die diskriminante Validität zu gewinnen, und zeigt, dass die Präsentation der Fragebogenitems einen Einfluss auf die Beantwortung der Skala hat.

Zusammen mit der zuvor diskutierten begrifflichen Ähnlichkeit der Fragebogenitems könnte überlegt werden, die Frageitems der Skala über verschiedene Fragebogenseiten aufzuteilen oder die Anzahl der Fragebogenitems zu reduzieren. Beide Entscheidungen können jedoch nur in Abhängigkeit von der jeweiligen Forschungsfrage und dem Ziel des Einsatzes der *BDGS*-Skala oder einzelner benötigter Dimensionen der *BDGS*-Skala getroffen werden. Muss der Fragebogen möglichst kurz ausfallen, kann zudem eine Reduktion der *BDGS*-Skala auf die zentralen Frageitems einer Dimension (bspw. die Referenzindikatoren) vorgenommen werden, die auf einer Seite präsentiert werden.

---

¹⁸ Weitere Informationen zur wissenschaftlichen Untersuchung von Fake News finden sich bei D. M. J. Lazer et al. (2018). Zur Bedeutung von Fake News in der politischen Kommunikation siehe Schulz et al. (2017).

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.