Methodik der Skalenkonstruktion für das Big-Data-Glaubenssystem und erste empirische Evidenzen

8

An dieser Stelle beginnt der empirische Teil der vorliegenden Arbeit, der zum einen in diesem Kapitel 8 einen Vorschlag unterbreitet, wie man das Big-Data-Glaubenssystem (*BDGS*) einer empirischen Messung zuführen kann und wie die einzelnen konzeptuellen Dimensionen des *BDGS* im Rahmen dieser zu entwickelnden Messung operationalisiert werden können. Zum anderen sollen Untersuchungsergebnisse referiert werden, die erste Erkenntnisse über die Ausprägung des *BDGS* in der deutschen Bevölkerung (siehe Kapitel 9) und den Einfluss des *BDGS* auf kognitive, affektive und konative Komponenten der Einstellung in unterschiedlichen Lebensbereichen berichten, in denen KI-Anwendungen zum Einsatz kommen (siehe Kapitel 10).

Daher wird in den folgenden Abschnitten zunächst die Durchführung der Skalenkonstruktion dokumentiert und eine Messung vorgeschlagen, deren Einsatz in einer ersten Studie mit drei Erhebungen entwickelt und getestet wurde. Das empirische Vorgehen soll nachfolgend detailliert und intersubjektiv nachvollziehbar beschrieben werden, um den Aufbau und die Verwendung der Skala als zentrales Werkzeug der Arbeit zu erläutern und eine mögliche Datenanalyse für die produzierten Beobachtungen vorzuschlagen. Die Ausführungen orientieren sich dabei an den in der Literatur zur Testkonstruktion empfohlenen Richtlinien aus der Literatur zur Methodenforschung der standardisierten Befragung, allen voran an denen von (Bühner, 2011), sowie den Empfehlungen zur Skalenentwicklung von Carpenter (2017) und Bandalos (2017).

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_8.

Generell sollen die Beschreibungsdimensionen der Definition des Phänomens Big Data und ihr Wesensgehalt, der sich in einer Erwartungshaltung bezüglich der Potentiale von Big Data und digitalen Daten ausdrückt, in messbare Indikatoren überführt werden, deren jeweilige Messung eine Auskunft über das Vorhandensein oder eben die Abwesenheit der jeweiligen Überzeugungen gibt. Dabei sind die Bürger*innen Merkmalsträger*innen und deren Erwartungshaltung mit Bezug auf die digitalen Daten das zu untersuchende Merkmal (Brosius et al., 2008). Das *BDGS* und seine zugehörigen Dimensionen lassen sich – wie bei sozialwissenschaftlichen Konstrukten üblich – nicht direkt und unmittelbar beobachten; ihre Messung muss daher empirisch operationalisiert werden (Bandalos, 2017; Weiber & Mülhhaus, 2014).

8.1 Die Operationalisierung und Messung von Glauben und Glaubenssätzen mit Hilfe der standardisierten Befragung

Die vorliegende Arbeit zielt darauf ab, Personen nach der Stärke der Ausprägung des individuellen Glaubens an Charakteristika und Konsequenzen digitaler Daten und ihrer Verwertung zu unterscheiden. Wie zuvor ausgeführt, gibt es innerhalb des *BDGS* mitunter Überzeugungen, die dazu führen, eine ganz unterschiedliche Bewertung dieser digitalen Daten in ihrem Entstehungs- und Verwertungskontext vorzunehmen. Doch wie gelingt diesbezüglich eine Unterscheidung, kann man die Existenz und Stärke des Glaubens den jeweiligen Personen ja eben nicht von den Lippen ablesen. Das *BDGS*, das sich auf Glauben an digitale Daten bezieht, kann nicht aus eben jenen bestehenden großen digitalen Datenbeständen herausgelesen werden, sondern muss anhand geeigneter Indikatoren erhoben werden (Hubbard, 2010). Entsprechende Daten müssen aktiv produziert werden. Dabei ist das Glaubenssystem immateriell oder in anderen Worten: nicht greifbar. Es ist jedoch über die Reproduktion von Aussagen bezüglich eigener Glaubensüberzeugungen und die Zustimmung oder Ablehnung zu diesen zu operationalisieren und zu messen. Eine Person, die einen ausgeprägten Glauben an einzelne Dimensionen des Wesens und der Konsequenzen von Big Data hat, sollte in der Befragungssituation entsprechend Aussagen, die den digitalen Daten nun bspw. eine Eigenschaft wie detailgetreue Realitätsabbildung oder positive Konsequenzen für das Selbst zuschreiben, entsprechend zustimmen. Die (Nicht-)Zustimmung zu den Glaubenssätzen, also manifeste Äußerungen der Merkmalsträger, dienen als Indikatoren für die jeweilige konkrete Ausprägung des *BDGS*, die dokumentiert werden und Rückschlüsse zulassen. Genauer werden den Bürger*innen Aussagen zu digitalen

Datenbeständen vorgelegt und sie sollen als Befragte selbst dazu Auskunft geben, inwieweit sie den Aussagen zustimmen oder auch nicht.

Der Ansatz folgt dabei sozialwissenschaftlicher Forschung, die ebenfalls darum bemüht ist, die Messung von Glaubenssätzen mittels standardisierter Itembatterien für Befragungsstudien zu ermöglichen. Entsprechende Befragungsinstrumente wurden unter dem Stichwort der Messung von Glaubensüberzeugungen u. a. für die Erwartungshaltung mit dem zukünftigen Umgang mit positiven Emotionen vorgeschlagen (F. Bryant, 2011) und werden auch bei der Analyse politischer Überzeugungen eingesetzt (Baldassarri & Goldberg, 2014; Boutyline & Vaisey, 2017; Brandt et al., 2019).

Zuvor wurde dabei der Begriff des Messens unter den Bedingungen von Daten und seine Bedeutung für diese Arbeit bereits ausführlich aus analytischer Perspektive reflektiert. Für das handwerkliche Verständnis des Messens bei der empirischen Erhebung des *BDGS* im Rahmen von Sozialforschung wird sich nachfolgend hingegen an folgender Definition von Friedrichs (1990) orientiert. Hier bedeutet

Messen die systematische Zuordnung einer Menge von Zahlen oder Symbolen zu den Ausprägungen einer Variable, mithin auch zu den Objekten. (...) Die Zuordnung (oder genauer: Abbildung) sollte so erfolgen, dass die Relationen unter den Zahlenwerten den Relationen unter den Objekten entsprechen. (S. 97, Hervorh. im Orig.)

Bei der standardisierten Befragung kommen also Fragen zum Einsatz, die in ihrem Einsatz in einer Skala oder Fragenbatterie auch als Items bezeichnet werden. Diese Fragen werden mithin im Folgenden als (Frage-)Items oder Indikatoren bezeichnet und die entsprechenden Begriffe synonym verwendet.¹

8.2 Die Operationalisierung von Messung von Glauben als reflektives latentes Konstrukt

Dabei folgt diese Messung von Glauben der Logik reflektiver Messmodelle, d. h., ein nicht unmittelbar beobachtbares *latentes* Konstrukt wird über die

¹ Es wird an der jeweiligen Stelle im Text deutlich gemacht, ob von der Gesamt- oder Sub-Skala, also der gesamten Itembatterie, oder einem Teilausschnitt dieser Batterie gesprochen wird. Ein Teilausschnitt umfasst dabei meist immer eine konkrete Dimension des *BDGS*. Skala kann sich jedoch auch auf das Messniveau und die den Befragten vorgegebenen standardisierten Antwortoptionen der Fragen einer Itembatterie beziehen (siehe Abschnitt 8.4.6).

Erhebung von Messindikatoren (sogenannten *manifesten* oder auch beobachtbaren Variablen) operationalisiert. Die Ausprägung des latenten Konstrukts auf Seiten des untersuchten Subjekts beeinflusst die jeweilige Ausprägung eines Indikators. Dabei werden alle Befragten mit den gleichen vorformulierten Glaubensüberzeugungen des *BDGS* konfrontiert und es verorten sich alle Befragten auf der gleichen abgestuften Antwortskala (Bandalos, 2017). Die Ausprägung der Erwartungshaltung und ihrer zugehörigen Dimensionen wird auch als ein empirisches Relativ verstanden, das nun mittels Messung möglichst strukturgleich in ein numerisches Relativ überführt wird (Brosius et al., 2008; Rössler, 2017).² D. h., der jeweiligen Ausprägung des empirischen Relativs wird systematisch ein Zahlenwert zugewiesen und erlaubt für registrierte Äußerungen im Aggregat eine vergleichende statistische Auswertung mittels Strukturgleichungsmodellierung und konfirmatorischer Faktorenanalyse (Kline, 2011; Weiber & Mülhhaus, 2014). Anhand eines Beispiels mit Bezug auf das *BDGS* verdeutlicht: Je ausgeprägter also die Glaubensüberzeugung bei einer Person, dass die digitalen Daten einen Nutzen für die Gesellschaft haben, desto eher stimmt diese Person auch in der Befragungssituation einer Aussage zu, die diesen Nutzen für die Gesellschaft postuliert. Die befragte Person signalisiert die Zustimmung durch ein vorgegebenes Antwortschema, was im vorliegenden Fall digital registriert wird. Der Zustimmung wird ein hoher Zahlenwert zugewiesen. Wird hingegen in der Befragungssituation Nicht-Zustimmung signalisiert und ein Nutzen verneint, wird ein niedrigerer Zahlenwert festgehalten. Die Wirkungsrichtung läuft also vom zu messenden, nicht unmittelbar zu beobachtenden Konstrukt hin zur tatsächlich geäußerten Zustimmung, die im Rahmen des standardisierten Instruments registriert wird.

Der Glaube, dass digitale Daten bestimmte epistemische Charakteristika besitzen und im Stande sind, einen Nutzengewinn herbeizuführen, soll also mittels Befragung und somit Selbstauskunft der Befragten erhoben werden. Zwar könnte auch in Erwägung gezogen werden, die Ausprägung des *BDGS* durch Beobachtung des Verhaltens und der Handlungen von Personen zu erheben oder durch eine Dokumentation und anschließende Inhaltsanalyse ihrer aufgezeichneten Äußerungen. Die Befragung, und hier insbesondere eine standardisierte Form der Befragung auf Grundlage eines psychometrischen Tests, scheint jedoch besonders geeignet, einen ersten Aufschluss über die Valenz und Stärke der individuellen Erwartungshaltung zur epistemischen Qualität und dem hieraus gezogenen

² Zum Unterschied zwischen isomorpher und homomorpher Abbildung des empirischen Relativs in ein numerisches Relativ siehe Brosius et al. (2008, S. 49): Im vorliegenden Fall ist die Abbildung homomorph, was bedeutet, dass die Zuordnung „nicht mehr eindeutig umkehrbar ist“.

Nutzen digitaler Daten zu geben und diese über Befragte und unterschiedliche Stichproben hinweg vergleichbar zu machen.

Allein aus ökonomischen Gründen ist die zunächst zwar aufwändige, dann jedoch beliebig oft anzuwendende Erarbeitung eines standardisierten Befragungsinstruments sinnvoll, bei dem einige wenige Fragen mit Hilfe sich wiederholender vorgegebener Antwortmöglichkeiten in kurzer Zeit aus der Ferne ausgespielt werden können. Diese Erhebungs- und Messkonzeption ermöglicht daher besonders gut das Festhalten einer standardisiert quantifizierbaren Messung reproduzierter Überzeugungen des *BDGS*.

Bevor jedoch die Testkonstruktion im Detail erläutert wird, sei bereits ein Hinweis gegeben, der auch unter Reflexion und Methodenkritik aufgegriffen wird: Sozialwissenschaftliche Messungen sind nicht perfekt und stellen lediglich Annäherungen an das abzubildende, nicht unmittelbar zu beobachtende Konstrukt dar, da es nicht *die eine richtige* Messung gibt (Bandalos, 2017). Testkonstruktionen sind mitunter fehlerbehaftet und setzen ein Bewusstsein und eine stetige Reflexion über diese Fehlbarkeit voraus, die sich aus unterschiedlichen Fehlerquellen ergeben kann (Weisberg, 2009). Auch wenn die nachfolgende Auswertung die Möglichkeit des Messfehlers zulässt (Kline, 2011; Weiber & Mühlhaus, 2014): In den nachfolgenden Ausführungen müssen alle Entscheidungen auf dem Weg zur Messung des *BDGS* ausführlich begründet und ihre Pro- und Contra-Argumente abgewogen werden.³

Das Testziel der ‚Big-Data-Glaubenssystem‘-Skala

Bevor auf die konkreten Entscheidungen bei der Konstruktion der Skala und der Formulierung ihrer Items eingegangen werden kann, muss zunächst ein allgemeines Testziel festgelegt werden, das die einzelnen Entscheidungen leitet, die bei der Skalenentwicklung getroffen werden. Das Testziel besteht in der vorliegenden Arbeit darin, die Perzeptionen der Bürger*innen von Big Data und der mit Big Data verbundenen Datafizierung und Quantifizierung messbar zu machen. Es wird hier auf eine Erwartungshaltung abgestellt, die als *BDGS* im vorigen Teil der Arbeit konzeptuell beschrieben wurde. Die Messung durch standardisierte Befragung stellt hierbei also insbesondere auf Glaubensüberzeugungen bezüglich der in Abschnitt 3.5.1 bzw. Abschnitt 3.5.2 besprochenen Beschreibungsdimensionen *Richtigkeit* (Veracity) und *Nutzen* (Value) großer digitaler Datenbestände ab, die auch als Big Data bezeichnet werden.

³ Vermeintliche Fehler sind dem Verfasser anzulasten, wobei die Dokumentation der empirischen Datenerhebung und der spezifizierten Modelle es Dritten mit Hilfe der Daten im Anhang im elektronischen Zusatzmaterial erlauben, das empirische Vorgehen bei Datenerhebung und -auswertung zu prüfen.

In der Anfangsphase der vorliegenden Forschungsarbeit wurde zunächst auch erwogen, noch zusätzlich zwischen den drei Momenten der *Entstehung*, *Speicherung* und *Auswertung* von digitalen Daten unterscheiden zu wollen. Zudem wurde eine Bewertung des Ausmaßes, der Geschwindigkeit und der Vielfalt digitaler Datenentstehung und -verwertung erfragt. Diese feinteilige konzeptuelle Unterscheidung ist empirisch jedoch kaum oder gar nicht abzubilden, da etwaige Feinheiten in den entsprechenden (Item-)Formulierungen nur unzureichend wahrgenommen werden und sich im Aggregat nicht abbilden lassen. Siehe hierzu die entsprechenden Ergänzungen in Abschnitt 1.1 im Anhang im elektronischen Zusatzmaterial. Auf die jeweiligen Unterscheidungen wird nachfolgend nur jeweils an Stellen eingegangen, an denen ihr diskutierter Bezug zur Messung wichtig wird und zum Verständnis der empirischen Untersuchung beiträgt. Dies ist der Notwendigkeit der transparenten Dokumentation durchgeführter Forschung geschuldet.

8.3 Befragungsisems als manifest beobachtbare Indikatoren der Messung der latenten Konstrukte des BDGS

Bei der Festlegung der Art der Indikatoren soll nach Bühner (2011), „die Frage beantwortet werden, ob der Test anhand objektiver oder subjektiver Indikatoren ein Konstrukt erfassen soll“ (S. 85). Da das zu erhebende Charakteristikum – anders als bei einem Wissenstest – keine eindeutige Falsch- oder Richtig-Klassifizierung erlaubt, sondern über Ratings und somit sogenannte Q-Daten (*Questionnaire data*) abgebildet wird, handelt es sich um subjektive Indikatoren, die mittels Selbstauskunft der Befragten deren Überzeugungen messen sollen (Cattell, 1973). Diese Überzeugungen, als zu messendes mehrdimensionales Konstrukt, meint die individuelle Erwartungshaltung bezüglich großer digitaler Datenbestände, also das zuvor beschriebene Big-Data-Glaubenssystem mit seinen Dimensionen *Richtigkeit* und *Nutzen*. Es gibt mithin keine richtigen oder falschen Aussagen, sondern es wird die jeweilige individuelle Zustimmung der Befragten zu allgemein gehaltenen Aussagen über die zuvor besprochenen Beschreibungsdimensionen von Big Data erfragt. Die Aussagen fungieren dann wie beschrieben als Indikatoren des zu messenden Konstrukts: Die konkrete Einschätzung der Aussagen durch die Befragten bildet die Ausprägung des jeweiligen Indikators ab. Wird einer allgemein gehaltenen Aussage aus einer Dimension umfassend zugestimmt, signalisiert dies das Vorhandensein einer Erwartungshaltung mit Bezug zu den digitalen Daten. Wird dieser Aussage jedoch nicht

zugestimmt, indiziert dies eine nicht vorhandene Erwartungshaltung auf Seiten der befragten Person. Die Messindikatoren, über die das latente Konstrukt operationalisiert wird, reflektieren dessen Ausprägung. Sie werden daher nachfolgend in ein reflektives Messmodell überführt, das eine Prüfung ihrer dimensional Struktur erlaubt.

8.4 Zur Operationalisierung der latenten Konstrukte des Big-Data-Glaubenssystems – Voraussetzungen der Konzeption von geeigneten Indikatoren

Nachdem nun das Ziel und die Logik der Konstruktmessung erörtert und die Art der geeigneten Indikatoren festgelegt wurden, folgen nun allgemeine Überlegungen zur Ausgestaltung der Indikatoren in der Befragung. Bei der Formulierung der Indikatoren muss zunächst bedacht werden, welche Zielgruppe später einmal mit dem Befragungs-Instrument konfrontiert wird, um mögliche problematische Einflüsse auf die Messung durch bestimmte Eigenheiten der Befragten (bspw. Einflüsse der Soziodemografie, des Sprachverständnisses sowie des Erlebens- und Verhaltensspektrums) zu vermeiden (Bühner, 2011). Weiterhin müssen weitere Entscheidungen getroffen werden, die die Messung betreffen, wie bspw. die Skalierung der Items und die Benennung der Skalierungspunkte sowie die Einleitung, Darstellung und Aufteilung der Skala im Fragebogen (Scholl, 2018). Die folgenden Abschnitte formulieren die einzelnen Ansprüche, die an Skala und Items gestellt werden, und dokumentieren, wie diese bei der Testkonstruktion des *BDGS* umgesetzt wurden.

8.4.1 Festlegung der Zielgruppe des Tests

Bereits zu Beginn der Entwicklung eines Tests sollte nach Bühner (2011) festgelegt werden, „welche Zielgruppe den zu entwickelnden Test später einmal bearbeitet sowie deren Merkmale zu beschreiben, die für die Testbearbeitung relevant sein könnten“ (S. 87). Denn: „Für die Testkonstruktion werden schon durch die Definition der Zielgruppe die Weichen für das Format des Tests, die Itemschwierigkeit und die Itemformulierung gelegt“ (S. 91). Grundsätzlich sollte ein Test für die Prävalenz des *BDGS* für alle deutschsprachigen Bürger*innen der Bundesrepublik Deutschland geeignet sein, so dass eine Verwendung in Repräsentativbefragungen grundsätzlich möglich ist. Es ist also weitgehend unerheblich, wie ausgeprägt die Kenntnisse der Befragten über die Digitalisierung

und allgemeine Erfahrung mit Internetnutzung sind oder ob konkretes technisches Wissen über mögliche Aspekte ebenjener digitalen Daten oder Big Data vorliegt, die über basale Vorstellungen und lose Assoziationen hinausgehen.

Gerade eine tiefergehende Auseinandersetzung mit digitaler Technologie und Daten wird für das Gros der deutschen Bevölkerung eben nicht angenommen. Diese kommt in Beruf und Alltag nicht unmittelbar mit großen digitalen Datenmengen und der (Daten-)Analyse von Big Data in Berührung. Selbst Expert*innen wie bspw. hauptberuflich Forschende oder etwa privatwirtschaftlich angestellte Data Scientists bekommen immer nur einen überschaubaren Teil der Daten zu Gesicht und haben unmöglich einen vollständigen Überblick über Bestand und Einsatz der Daten in Wirtschaft und Verwaltung sowie die aus den Daten gewonnen Erkenntnisse und die Auswirkungen, die sich aus diesen Daten ergeben. Das hat zur Konsequenz, dass womöglich nicht alle Bürger*innen überhaupt etwas mit diesem Begriff *Big Data* anfangen können. So stellen auch Orwat und Schankin (2018) fest: “The many definitions, understandings, and general ambiguity of the term ‘big data’ do not allow for direct work with this term in a survey” (S. 1). Daher wird an den entscheidenden Stellen in der Befragung wie etwa Einleitungstexten und Itemformulierungen von ‚großen digitalen Daten(beständen)‘ gesprochen.

8.4.2 Berücksichtigung möglicher Herausforderungen bei der Testkonzeption durch sozio-demografische Merkmale der Befragten

Die allgemein gehaltene Zielgruppe der deutschsprachigen Bundesbürger*innen bedeutet mithin auch, dass der Test für alle Alters- und Bildungsstufen anwendbar sein sollte. Das hat vor allem Konsequenzen für die sogenannte *psychometrische Itemschwierigkeit*, da bspw. die Nutzung digitaler Medientechnologien durch das Alter beeinflusst wird (Seifert, 2016). Junge und hochgebildete Personen haben wahrscheinlich einen anderen Zugang zu einem Thema, das einen gewissen technischen Anspruch hat, und können aufgrund eines Wissensvorsprungs, der aus größerer Nutzungserfahrung entspringt, bestimmten Items nicht nur eher, sondern möglicherweise auch anders zustimmen als technisch unerfahrene Personen. Es besteht mithin die Möglichkeit, dass für spezielle Untergruppen der Zielgruppe Probleme bei der Skalenkonstruktion und der letztendlichen Befragung zum *BDGS* zu erwarten sind, die über die üblichen Probleme und Herausforderungen der Befragungsforschung hinausgehen (Scholl, 2018).

Allerdings sollte eine Vertrautheit mit den allgemeinen technischen Begrifflichkeiten im Bereich Internet und Computer vorauszusetzen sein. Es wird erwartet, dass die Begriffe ‚Digitalisierung‘ bzw. ‚digital‘ und auch ‚Daten‘ bekannt sind und ein Alltags- bzw. Laienverständnis vorliegt, ohne dass die jeweiligen Befragten dazu befähigt sein müssen, eine präzise Definition wiederzugeben. Da es ja um die allgemeinen Perzeptionen von Big Data und eine diesbezügliche individuelle Erwartungshaltung geht und nicht bspw. um einen Wissenstest, sollte der Einfluss von Alter und Bildung zumindest auf die allgemeine Beantwortbarkeit der gestellten Fragen minimal sein. Alle Befragten sollten ein Verständnis von Datenerzeugung und -speicherung durch digitale Medien aus persönlicher Erfahrung sowie medienvermittelten Erzählungen mitbringen. Dass digitale Medientechnik Daten erzeugt und speichert, ist seit Jahren und auch aktuell Gegenstand öffentlicher Diskussionen (Knorre et al., 2020; Zeller et al., 2010). Alle potentiellen Befragten sollten durch die regelmäßigen Hinweise zu Datenerhebung und Datenschutz bei der Nutzung digitaler sowie nicht-digitaler Angebote auch damit vertraut sein, dass sie in gewissem – wenn auch nicht vollumfänglich nachvollziehbarem – Umfang laufend ihre Daten preisgeben und sich Unternehmen und Dienstleister vorbehalten, diese Daten für eigene Zwecke zu nutzen, ohne dass hiermit gleich eine kommerzielle Nutzung gemeint ist. So wird man beim Arztbesuch bedingt durch das Inkrafttreten der DSGVO im Mai 2018 neuerdings um die Einwilligung in Datenerhebung und -speicherung gebeten (Wensing, 2018).

Lediglich mit Bezug auf die technische Auswertbarkeit der Daten kann es durchaus unterschiedliche Vorstellungen geben. Allerdings sollte auch hier weitgehend klar sein, dass etwas mit den erzeugten und gespeicherten Daten geschieht, selbst wenn sich das explizite Wissen um die konkreten Auswertungsmöglichkeiten zwischen einzelnen Gruppen unterscheiden mag.

Herausforderungen der BDGS-Messung bei älteren Befragten

Es muss jedoch diskutiert werden, inwieweit ältere Befragte – eine Gruppe, in der sich viele Nicht-Nutzer digitaler Medientechnologien befinden (Seifert & Schelling, 2018) – sich möglicherweise mit der Beantwortung schwertun und es zu Verständnisproblemen in der Befragungssituation kommt.

Die Mehrheit der älteren Befragten sollte unter normalen Umständen in der Lage sein, die Befragungssitems sinnhaft zu beantworten. Um dies sicherzustellen, sollten die Frageitems mit einer Einleitung versehen werden, die ein Grundverständnis für das herstellt, was im Fragebogen unter digitalen Daten verstanden wird. Der Einleitungstext zum Test wird in Abschnitt 8.4.7 dokumentiert. Es sind hier des Weiteren keine oder vernachlässigbare Probleme mit Blick auf ältere

Befragte zu erwarten, da „besondere Problemgruppen unter den Älteren üblicherweise in Surveys gar nicht oder nur selten befragt werden: die ganz Alten (Personen über 75 oder 80 Jahre), die institutionalisierten Alten und die (meist geistig) Labilen (Demente)“ (Kühn & Porst, 1999, S. 27). Mit Blick auf den nachfolgend dokumentierten Einsatz der *BDGS*-Skala in diversen Erhebungen wird deutlich werden, dass jene Personen der ganz Alten in vernachlässigbarer Anzahl teilgenommen haben. Dies liegt schon darin begründet, dass die durchgeführten Befragungen online stattfanden, was nachfolgend diskutiert wird. Dennoch kann diese Befragtengruppe nicht gänzlich vernachlässigt werden, da sie doch gleichsam zur Grundgesamtheit zählt und einen beachtlichen und zunehmenden Anteil an der Wahlbevölkerung umfasst (Falter & Gehring, 1998). Etwaige Problemlagen müssen mithin mit Blick auf den zukünftigen Einsatz einer *BDGS*-Skala je nach Forschungsinteresse berücksichtigt werden. Zusammengefasst bedeutet dies zunächst, dass es keine Gruppe unter den potentiellen Befragten geben sollte, die aufgrund soziodemografischer Merkmale durch das Instrument systematisch benachteiligt wird und keine valide Beantwortung des Tests erbringen könnte.

8.4.3 Berücksichtigung möglicher Herausforderungen bei der Testkonzeption durch das Erlebens- und Verhaltensspektrum der Befragten

Es gibt ebenso eine ganze Reihe von Personen mit einer spezifischen Perspektive auf ihr eigenes Erlebens- und Verhaltensspektrum im Rahmen digitaler Daten. Bspw. diejenigen, die aus der eigenen intensiven Beschäftigung mit digitalen Daten eine spezielle Wahrnehmung entwickelt haben, die auf die Beantwortung der zu entwickelnden Skala durchschlägt. Gemeint sind hier allen voran Technikaffine aus dem Spektrum der *Self-Tracker* und der *Quantified-Self*-Bewegung, für die ebenfalls die Itemschwierigkeit bedacht werden muss (Nafus & Sherman, 2014; Neff & Nafus, 2016). Diese Personen entstammen möglicherweise einem anderen Erlebens- und Verhaltensspektrum und es könnte für sie mithin einfacher sein, die Testfragen zu beantworten. Gleiches könnte auch für Expert*innen auf dem Gebiet der Digitalisierung zutreffen, die aufgrund ihres beruflichen Hintergrunds und der Nähe zu Themen der Digitalisierung ein entsprechend verändertes Erlebens- und Verhaltensspektrum aufweisen (van der Aalst, 2014).

Allerdings sollte bei der Skalenkonstruktion so allgemein über die Qualität von großen Datenmengen für einen Erkenntnis- und Nutzengewinn befragt werden, dass es analog zu Alter und Bildung keinen fundamentalen Unterschied zwischen

dieser Gruppe und anderen Gruppen gibt, die einer reliablen und vor allem validen Beantwortung des Tests in diesen speziellen Gruppen entgegensteht. Es liegt in der Natur des Tests, dass eben solche Personen, die sich ausgiebig beruflich oder privat mit der Digitalisierung oder mit digitalen Daten beschäftigen, eine andere Sicht auf die Befragungsisems haben. Jedoch ist dies noch kein Hinweis darauf, dass das *BDGS* bei diesen Personen, wenn auch möglicherweise anders ausgeprägt, mit der vorgeschlagenen Messung überhaupt nicht zu erfassen sein sollte, etwa, weil diese Personen etwas gänzlich anderes unter digitalen Daten verstehen, als es die vorliegende Skalenkonstruktion untersucht.⁴

Herausforderungen der BDGS-Messung bei Offlinern

Es ist hingegen durchaus vorstellbar, dass Personen, ganz gleich welchen Alters, die sich der Digitalisierung aktiv oder zufällig entziehen, bspw. sogenannte *Offliner* (Cachelin, 2015), mit den auf digitale Daten abzielenden Formulierungen nichts anfangen können, weil ihnen möglicherweise die Begrifflichkeiten einfach unbekannt sind. Auch hier ist allerdings davon auszugehen, dass der Großteil der möglichen Befragten aus dieser Gruppe zumindest über allgemeine (Offline-) Medienrezeption durchaus mit den Begrifflichkeiten bezüglich der Digitalisierung vertraut ist.

Diese Annahme stützt sich auf aktuelle Umfragen zur Internetnutzung und Wahrnehmung der Digitalisierung. Während je nach Quelle im Jahr 2017 zwischen 81 % (Statista, 2018b) und 89,8 % (Koch & Frees, 2017) der deutschen Bevölkerung zu den Internetnutzer*innen zählen, zeigen europaweite Umfragen zur öffentlichen Meinung zur Digitalisierung, dass sich ein Großteil der Bürger*innen ein Urteil über die Auswirkungen der Digitalisierung auf die Gesellschaft zutraut (Europäische Kommission, 2010). Gefragt nach ihrer Einschätzung zum Einfluss digitaler Technologien auf die Wirtschaft geben lediglich 2 % der befragten Personen in Europa offen zu, nicht genug über digitale Technologien zu wissen, um eine Einschätzung geben zu können, und weitere 8 % wählen die begründungslose „weiß nicht“-Option. Könnten sich Personen, die keine oder kaum persönliche Berührungspunkte mit der Digitalisierung haben,

⁴ In den nachfolgenden Abschnitten wird an den entsprechenden Stellen eine Prüfung auf Messinvarianz durchgeführt, die prüft, ob die Messung von unterschiedlichen Gruppen gleich verstanden und somit mess-äquivalent ist (Weiber & Mülhhaus, 2014).

kein Urteil bilden, müssten eben jene Ausweichkategorien wie „keine Angabe“, „weiß nicht“ und insbesondere das qualifizierte Nicht-Wissen zu digitalen Technologien stärker besetzt sein.⁵

So zeigen auch aktuelle Umfragen zu KI, einer speziellen datenbasierten Digitalisierungsanwendung (siehe Abschnitt 3.6), der in jüngster Zeit ein großes Interesse zukommt, dass sich auch hier ein Großteil der befragten Personen in Deutschland Aussagen zutraut: Laut einer Studie von Bitkom aus dem Herbst 2018 konnten nur 12 % der Befragten überhaupt nichts mit dem Begriff der KI anfangen (Bitkom, 2018b). Dies sagt an dieser Stelle noch nichts über die tatsächliche Validität der Messungen in den genannten Befragungen aus, sondern dokumentiert lediglich, dass für die *BDGS*-Skala nicht damit zu rechnen ist, dass in Befragungen mit einer heterogenen Zusammensetzung hohe Ausfallquoten (sprich: höher als ca. 10 %) zu erwarten sind. Die Möglichkeit von Befragungsartefakten wie Non-Opinions, also das Produzieren von Meinungsartefakten, obwohl keine Meinung oder kein Wissen über das Meinungsobjekt vorliegt, bleibt hiervon zunächst einmal unberührt (Bachleitner & Aschauer, 2008; Brosius et al., 2008). Es wird daher davon ausgegangen, dass der Pool der Personen, die mit den Items zu digitalen Daten in Kontakt kommen und hiermit nichts anfangen können, zwar nicht vernachlässigbar, aber doch so gering ist, dass das Unterfangen einer allgemeinen Skala zum Glaubenssystem, die in der empirischen Sozialforschung eingesetzt werden kann, zunächst einmal nicht als aussichtslos erscheint. Allerdings wurde aus Vorsicht in der ersten Erhebung noch die Möglichkeit einer „weiß nicht“-Option gegeben. Auf das Ausmaß der Wahl dieser Option wird in Abschnitt 9.1.3 eingegangen und dokumentiert, inwiefern die Unmöglichkeit einer Antwort bei einem heterogenen Befragtensample zu erwarten ist.

⁵ Eine kurze Anmerkung zum zu erwartenden Ausmaß der Non-response-Quote auf einzelne Befragungssitems: Ferber (1966) berichtet aus einer postalischen Paper-and-Pencil-Studie, in der lediglich 37,5 % der Befragten vollständige Antworten gaben, also im Umkehrschluss 62,5 % der Befragten mindestens eine Frage nicht beantworteten. Mit Blick auf einzelne Items berichtet die Methodenforschung Non-response-Quoten zwischen 2,9 und 8 % (Craig & Mccann, 1978). Anzumerken ist hierbei, dass in Befragungsstudien zu persönlich sensiblen Themen wie dem individuellen Gesundheitszustand durchaus hohe Item-Nonresponse-Quoten von bis zu 50 % auftreten können (Elliott et al., 2005). Auf dieser Grundlage erscheinen Item-Nonresponse-Quoten von ca. 10 % durchaus üblich und lassen demnach nicht erwarten, verzerrte oder nicht aussagekräftige Erkenntnisse bezüglich Digitalisierungs-Prädispositionen in der Bevölkerung zu produzieren.

8.4.4 Der Einfluss der Sprachbeherrschung der Zielgruppe auf die Testkonstruktion

Die Einschränkung auf die deutschsprachige Bevölkerung in Deutschland exkludiert naturgemäß nicht-deutschsprachige Personen. Nur so kann allerdings sichergestellt werden, dass ein breites umfassendes Verständnis der Fragen zu einem techniknahen Phänomen besteht. Mithin sollten der Test und seine Fragen so konstruiert sein, dass Befragte jedes Sprachniveaus, ab der Ebene der kompetenten Sprachverwendung (mind. Level C1 des Gemeinsamen Europäischen Referenzrahmens für Sprachen), den Fragebogen beantworten können. Vor allem im Hinblick auf das vorausgesetzte Wissen zu technischen Begriffen sollte also sichergestellt sein, dass die verwendete Sprache der Fragen des Tests der gesamten Testzielgruppe zugänglich ist. Dies wird ebenfalls u. a. durch den Einführungstext sichergestellt.

Die Beschränkung auf Deutsch als einziger Sprache, in der der Test zur Messung des *BDGS* in der vorliegenden Arbeit konzipiert wurde, schließt eine nachfolgende Übersetzung des Tests nicht aus. Allerdings müssen hier sprachliche und kulturelle Unterschiede in den zu untersuchenden Populationen berücksichtigt werden, um sicherzustellen, dass Äquivalenz und Messinvarianz bei der Erhebung des Konstrukts gegeben sind (Hatlevik et al., 2017; Steinmetz et al., 2009). Es bestehen dann auch durchaus strenge Ansprüche an die Entwicklung und Validierung einer Messung, die für komparative Forschung herangezogen werden kann und die eine einfache Übersetzung und den ungeprüften Einsatz der final verwendeten *BDGS*-Skala unzulässig erscheinen lässt (Harkness, 1998; Harkness et al., 2010).

8.4.5 Item- und Testfairness

Schwierigkeitsunterschiede einzelner Frageitems zwischen einzelnen Befragten-Gruppen können auch auf eine mangelnde Item- und Testfairness eines Tests hinweisen, die bestimmte Gruppen systematisch benachteiligt, bspw. durch schwer erkennbare Schriftgrößen für ältere Befragungsteilnehmer*innen (Bühner, 2011; Kühn & Porst, 1999). Auch hier sollte mangelnde Testfairness dadurch verhindert werden, dass mit einem standardisierten und somit für alle Befragten-gruppen gleich zugänglichen Fragebogendesign sowie allgemein verständlich formuliert über die Glaubensüberzeugungen bezüglich großer digitaler Datenmengen befragt wird und Frageinhalt und Frageformulierung es allen Befragten ermöglichen, ihre persönliche Einschätzung abzugeben. Eine Bevorzugung oder

Benachteiligung durch die Vertrautheit mit ähnlichen Tests oder den Einfluss kultureller Zugehörigkeit sollte durch die Verwendung eines standardisierten Tests, der Q-Daten (d. h. Selbst- oder Fremdratings) mit allgemeinverständlicher Formulierung erhebt und sich auf die Zielgruppe der deutschsprachigen Bevölkerung beschränkt, vermieden sein.

Mithin wurde bei der folgenden Itemerstellung berücksichtigt, dass Format, Formulierung und Schwierigkeit des zu konzipierenden Tests weder einzelne Befragte noch ganze Teilgruppen der Zielgruppe systematisch benachteiligen. Da die Befragungsstudien in dieser Arbeit durchweg mit Hilfe der Fragebogen-Software *SoSci Survey*⁶ auf digitalen Computermedien durchgeführt wurden, waren so gut wie alle Fragebögen und die Fragebogenseite mit der *BDGS*-Skala barrierefrei zugänglich. So konnten auch Personen mit eingeschränktem Sehvermögen problemlos teilnehmen (bspw. durch Verwendung eines *ScreenReaders*).

8.4.6 Die Wahl der Antwortdimensionen der *BDGS*-Skala

Neben Einflüssen auf die Beantwortung der Frageitems einer Skala, die durch spezielle Eigenschaften der Zielgruppe bedingt sein können, müssen auch allgemeine von der Zielgruppe unabhängige Entscheidungen bezüglich der Itemformulierung und insbesondere auch der Skalierung bei der Skalenkonstruktion berücksichtigt werden. Die gewählte Skala wird im Sinne der vorgeschlagenen Ratingskala von Likert (1932) erhoben, so dass mehrere Aussagen (im Folgenden auch Items genannt) als Indikatoren einer Dimension dienen. Die Antwortdimensionen der Skalen sollen den Grad der Zustimmung zu den Aussagen über das Wesen digitaler Daten und ihrer Speicherung und Auswertung abbilden, weshalb der „Grad der Zustimmung“ (Faulbaum et al., 2009, S. 23) in der vorliegenden Arbeit durch eine Zustimmungsskala abgefragt wird. Die Aussagen, im Folgenden auch Glaubenssätze genannt, sind keine Fragen im eigentlichen Sinne, sondern als Aussagesätze formuliert. Als Antwort auf die vermeintlichen Fragen ist dann der explizite Zustimmungsgrad der Aussage zu werten, die von den Befragten erbeten wird. Die Zustimmung (ab hier immer auch gedacht als mögliche Nicht-Zustimmung) kann dann auf einer mehrstufigen Skala je nach Ausprägungsintensität abgestuft werden. Diese Stufen werden als Skalenpunkte bezeichnet. Es ist mit Blick auf die Skalenpunkte u. a. zu entscheiden, ob (a) die Skalierung gerade oder ungerade ausfallen soll, (b) auf wie vielen Skalenpunkten

⁶ Die browserbasierte Befragungssoftware steht für nicht-kommerzielle Studien im wissenschaftlichen Kontext kostenlos zur Verfügung (D. J. Leiner, 2019a).

die Befragten ihre Antworten abstufen können und (c) ob und wie die einzelnen Skalenpunkte benannt werden.

Die Entscheidung für eine gerade oder ungerade Skalierung

Bei einer ungeraden Anzahl an Skalenpunkten gibt es eine Mischkategorie, bei der man einen Übergang zwischen Zustimmung und Nicht-Zustimmung hat. Dies erleichtert vor allem unsicheren Befragten die Beantwortung und lässt Ambivalenz zu, führt allerdings auch häufig dazu, dass der Skalenmittelpunkt als eben jene Ausweichkategorie gesehen wird, die er eigentlich gar nicht darstellen soll. Zudem gibt es bei Befragten eine gut dokumentierte ‚Tendenz zur Mitte‘ (Menold & Bogner, 2014). Es würden dann Personen, die eigentlich überhaupt keine Zustimmung zur Aussage geben können, mit jenen zusammengeworfen, die zwar eine Antwort geben können, jedoch zwischen Zustimmung und Ablehnung unentschieden sind und gute Argumente für beide Positionen haben. Die Itemformulierungen werden unter Verwendung von Ausweichmöglichkeiten, seien es Skalenmittelpunkte bei ungeraden Skalenpunkten oder eine „weiß nicht“-Option, möglicherweise nicht in der beabsichtigten Bearbeitungstiefe verarbeitet, da sich Befragte eben auf ihre Unentschiedenheit zurückziehen. Dieses Phänomen ist auch als *Satisficing* bekannt, eine Strategie der Befragten, um sich kognitive Arbeit bei der Beantwortung von Fragen zu ersparen (Krosnick, 1991; Krosnick et al., 1996; Krosnick et al., 2001). Bei gerader Anzahl an Skalenpunkten sind die Befragten hingegen gezwungen, sich zu entscheiden. Diese Festlegung könnte zum einen dazu führen, dass Befragte verunsichert werden können, sie insgesamt länger nachdenken und sich somit die Bearbeitungszeit verlängert. Dies in Kombination mit möglicher Reaktanz könnte schneller zu einer Ermüdung führen und die Abbruchquote erhöhen.

Unter Berücksichtigung des angesprochenen Für und Wider bei der Wahl zwischen einer geraden oder ungeraden Skalierung wurde sich für eine ungerade Skalierung entschieden, da vermeintliche Probleme wie Reaktanz und Verzögerung bei der Bearbeitung schwerer wiegen als die Tendenz zur Mitte sowie Satisficing und auch eher dazu führen könnten, auf die weitere Beantwortung in einer Befragungsstudie auszustrahlen, in der die *BDGS* eingesetzt wird. Eine finale Klärung der Frage nach der präferierten Art der Skalierung ist hier entsprechender Methodenforschung vorbehalten.

Die Entscheidung über die Anzahl der Skalenpunkte, Skalenpolung und -orientierung

Ein Blick in die Methodenliteratur zur Befragungsforschung offenbart etliche unterschiedliche Empfehlungen zur Anzahl und visuellen Ausrichtung der Skalenpunkte (Faulbaum et al., 2009, S. 65–66; Menold & Bogner, 2014; Scholl, 2018, S. 167 ff.). Eine Skalierung mit sieben Skalenpunkten scheint im vorliegenden Falle sinnvoll, da dies mit Verweis auf die Erkenntnisse der Methodenforschung zur Befragung eine ausreichend hohe Reliabilität gewährleistet (Alwin & Krosnick, 1991; Groves et al., 2011). Die einzelnen Skalenpunkte werden im vorliegenden Fall von links nach rechts von Nicht-Zustimmung zu Zustimmung abgestuft, so dass bei der Zuweisung der numerischen Werte für die Datenanalyse eine 1 für Nicht-Zustimmung der Aussage steht und mit Zwischenschritten in ganzen Zahlen hin zu 7 verläuft, was für vollkommene Zustimmung steht. Die Skala ist somit unipolar und die Skalenorientierung verläuft von links nach rechts. Während es zu diesen Entscheidungen laut Menold und Bogner (2014) keine klaren Empfehlungen gibt, wurden die dokumentierten Entscheidungen hier entsprechend getroffen, um einen einheitlichen und weitgehend standardisierten Skalenaufbau zu gewährleisten.

Die Entscheidung über die Benennung der Skalenpunkte durch adverbale Modifikationen

Zudem wird die Zustimmung durch adverbale Modifikationen abgestuft, wobei alle Skalenpunkte und nicht nur die Skalenendpunkte beschriftet werden (Faulbaum et al., 2009). Es wurde zwar zunächst beabsichtigt, nur die Skalenendpunkte zu beschriften, nach einem Hinweis aus einem persönlichen Interview des Pre-Tests (siehe Abschnitt 8.5.3) wurde die vollständige Beschriftung der Skala beschlossen. In der Literatur der Methodenforschung finden sich auch hierzu Empfehlungen: Laut Krosnick (1999) können hierdurch die Reliabilität und Validität der Beantwortung erhöht werden. So kommen auch Weijters et al. (2010) zu dem Schluss, dass für die Skalenentwicklung vollständig beschriftete Skalenpunkte verwendet werden sollten. Dies kann zudem zu einer höheren Befragtenzufriedenheit führen (Dickinson & Zellinger, 1980), Ambiguität reduzieren und die Test-Retest-Reliabilität verbessern (Weng, 2004).

Mithin wird für den Skalenpunkt 1 die Antwort „stimme überhaupt nicht zu“ und für den Skalenpunkt 7 die Antwort „stimme voll und ganz zu“ vergeben, so dass die Valenz der Skalierung mit aufsteigender Zustimmung von links nach rechts verläuft. Der Skalenmittelpunkt, also Punkt 4 wurde mit „Teils/teils“ beschriftet, um Ambivalenz zu signalisieren (Menold & Bogner, 2014). Die zur Nicht-Zustimmung tendierenden Zwischenpunkte auf der Skala werden für Punkt

2 mit „stimme größtenteils nicht zu“ und Punkt 3 mit „stimme eher nicht zu“ beschriftet. Die zustimmenden Zwischenpunkte auf der Skala werden für Punkt 5 mit „stimme eher zu“ und Punkt 6 mit „stimme größtenteils zu“ beschriftet. Damit der Skalenmittelpunkt von den Befragten nicht als Fluchtkategorie verwendet wird, besteht in der ersten Erhebung für die Befragten die Möglichkeit, durch die Antwortoption „Kann ich nicht beantworten“ anzuzeigen, dass das Item eben nicht beantwortet werden kann. Dies könnte dann ein Hinweis auf Formulierungsprobleme und die Beantwortungsschwierigkeit sein und wurde bei der Auswahl der finalen Items berücksichtigt.

Die Zuweisung der numerischen Werte erfolgte lediglich für die Auswertung. Den Befragten selbst wurden keine numerischen Werte für die jeweiligen Skalenpunkte angezeigt, da es bei ihnen hierdurch zu unbeabsichtigten Interpretationsprozessen bezüglich der Wertigkeiten der einzelnen Skalenpunkte kommen kann, was vermieden werden sollte (Krosnick, 1999; Schwarz et al., 1991).

8.4.7 Einleitungstext

Wie bereits zuvor erwähnt, soll an dieser Stelle mit Blick auf mögliche Verständnisschwierigkeiten des Befragungsgegenstandes *Big Data* auf den Einleitungstext des konzipierten Tests eingegangen werden. Die Fragenbatterie der *BDGS*-Skala wurde in allen Erhebungen mit diesem Einleitungstext eingeführt, um ein Grundverständnis bei allen Befragten über das herzustellen, was in der Einleitung und den Frageitems als ‚Digitale Daten‘ beschrieben wurde. Mit Hilfe interner Diskussionen am Arbeitsbereich von Prof. Dr. Marcinkowski samt Kollegen (siehe nachfolgend Abschnitt 8.5.2) und persönlicher Interviews mit der Methode des *lauten Denkens*, die genutzt wurden, um die Verständlichkeit der Frageitems für die erste Erhebung zu prüfen (siehe nachfolgend Abschnitt 8.5.3), wurden allgemeinverständliche Formulierungen und Beispiele gesammelt. Hier sollte ein allgemeines Verständnis über den Entstehungs- und Verwendungszusammenhang digitaler Daten im Sinne der vorliegenden Untersuchung vermittelt werden. Dabei sollte der Einleitungstext so einfach und kurz wie möglich sein und sollten die Beispiele allgemein gehalten werden, um die Befragten zwar zu informieren und ein einheitliches Verständnis des Gegenstandes herzustellen, zu dem befragt wird, eine Beeinflussung jedoch so weit möglich ausschließen.

Der formulierte Einleitungstext lautete wie folgt:

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. Diese können gespeichert, miteinander kombiniert und ausgewertet werden. Die hieraus entstehenden großen digitalen Datenbestände werden häufig auch als Big Data bezeichnet.

Die Einleitung orientiert sich dabei in Länge, Aufmachung und Inhalt (insbesondere der Verwendung und Erklärungen von Fachwörtern) an vergleichbaren Skalen wie bspw. der Kurzska *Technikbereitschaft* von Neyer et al. (2016). Der Einleitungstext fand sich auf der jeweiligen Fragebogenseite oberhalb der Skala. Die Positionierung sowie Aufmachung kann anhand der Dokumentation der Fragebögen im Anhang 3 im elektronischen Zusatzmaterial nachvollzogen werden.

8.5 Vorgehen bei der Itemgenerierung

Im Folgenden werden die einzelnen Schritte der Skalenkonstruktion beschrieben, insbesondere der Erstellung des Itempools, der in einer ersten Befragung auf die zugrunde liegende und aus der Literatur abgeleitete Dimensionalität des *BDGS* geprüft werden soll. Auf Grundlage der Definitionen der einzelnen Beschreibungsdimensionen von Big Data aus der wissenschaftlichen Literatur wurden Wörter extrahiert, die *Volumen*, *Vielfalt*, *Geschwindigkeit*, *Richtigkeit* und den *Nutzen* von Big Data sowie die jeweiligen Subdimensionen beschreiben, wobei eine direkte Nennung der Dimensionsbezeichnungen vermieden werden sollte. In einem weiteren Schritt wurden daher Synonyme und Umschreibungen zu diesen Wörtern gesucht. Eine Übersicht über den entstandenen Wortkorpus findet sich in Tabelle 2.A im Anhang im elektronischen Zusatzmaterial.

Dieser Wortkorpus diente als Vorlage für die Formulierung der Items. Hierbei wurden nicht alle Begriffe auch tatsächlich in Itemformulierungen überführt. Gründe hierfür waren bspw., dass einfachere Begriffe bevorzugt wurden oder Begriffe mehrdeutig und vermeintlich missverständlich sein konnten. So wurde ‚korrekt‘ als nicht synonym für *Genauigkeit* aufgenommen, da das Wort häufig wertend verwendet wird („Er ist ein korrekter Typ“).

8.5.1 Die notwendige Anzahl der Frageitems

Es stellt sich bereits bei der ersten Formulierung die Frage, wie viele Items, die später als Indikatoren für die Operationalisierung der jeweiligen Dimension eingesetzt werden, denn nun tatsächlich gebraucht werden und wie viele Items man nun für den ersten Schritt der Skalenkonstruktion vorsorglich formulieren müsse. Insbesondere im Rahmen des datenanalytischen Auswertungsverfahrens der konfirmatorischen Faktorenanalyse (‘Confirmatory Factor Analysis’ – kurz: CFA) gilt zu bedenken, dass für die Identifikation der geschätzten Lösung eines reflektiven Messmodells eine ausreichende Anzahl von Indikatorvariablen in die Modellspezifikation eingehen muss:

A common question about CFA concerns a minimum number of indicators per factor. In general, the absolute minimum for CFA models with two or more factors is two indicators per factor, which is required for identification. However, CFA models – and SR [Structural Regression] models, too – with factors that have only two indicators are more prone to problems in the analysis, especially in small samples. Also, it may be difficult to estimate measurement error correlation for factors with only two indicators, which can result in a specification error. Kenny’s (1979) rule of thumb about the number of indicators is apropos: ‘Two *might* be fine, three is better, four is best, and anything more is gravy’ (p. 143; emphasis in original). (Kline, 2011, S. 114–115)

Auch Weiber und Mühlhaus (2014) greifen die Diskussion um die Anzahl auszuwählender Indikatoren auf und empfehlen „bei Mehr-Konstrukt-Modellen reflektive Konstrukte mindestens mit jeweils zwei Indikatoren zu messen“ (S. 113). Diese Minimalanforderung scheint jedoch mit Blick auf erstmalige Testkonstruktion und den erwogenen Einsatz einer *BDGS*-Skala in diversen Forschungsprojekten mit unterschiedlichen Stichprobenzusammensetzungen und -größen als zu gering. Insbesondere bei kleineren Stichproben (bspw. in Experimentalstudien) empfehlen Marsh et al. (1998) eine größere Anzahl an Indikatoren, wobei ihrer Ansicht nach mehr Indikatoren generell besser sind als weniger Indikatoren. Gleichzeitig muss bedacht werden, dass eine große Anzahl von formal identischen Likert-Skalen von den Befragten als monoton wahrgenommen wird, die zu konstruierende Skala also nicht zu lang werden darf (Petersen, 2014; Ring, 1976). Nach Abwägung der Argumente wurde entschieden, mindestens drei Indikatoren zur Messung der jeweiligen Dimension des *BDGS* anzustreben.

Es muss an dieser Stelle jedoch nicht nur die final gewünschte Anzahl an Indikatoren pro Faktor bedacht werden, sondern auch, wie viele Items zunächst für einen ersten Aufschlag zur Testkonzeption vorbereitet werden, die dann womöglich aufgrund mangelnder Messgüte wieder eliminiert werden. Sowohl hier am

Beginn der Testkonstruktion als auch mit Blick auf den letztendlichen Einsatz des Tests zur *BDGS* ist daher zwischen einer notwendigen Anzahl von Indikatoren für die jeweilige zu messende Dimension und der gebotenen Sparsamkeit bei der Fragebogenlänge abzuwägen. Hierbei muss neben Rücksicht auf die Konzentrationsfähigkeit auch der ‚Respondent Fatigue‘, also der generellen Unlust, an Befragungsstudien teilzunehmen (S. R. Porter et al., 2004), durch die Verwendung möglichst kurzer Skalen entgegengewirkt werden. Es erschien daher sinnvoll, mindestens drei Items für die jeweiligen Subdimensionen zu formulieren, um sowohl zumindest die Unterscheidungsfähigkeit der Subdimensionen *Entstehung*, *Speicherung* und *Auswertung* zu prüfen (siehe Seite 93) als auch die Möglichkeit zuzulassen, ggf. unpassende Items zu eliminieren, weshalb immer mindestens vier, vereinzelt jedoch auch acht Items formuliert wurden.

8.5.2 Diskussion der Items mit Fachkollegen

Der so entstandene Itempool wurde zusammen mit zwei Wissenschaftlern mit langjähriger Erfahrung mit empirischer Sozialforschung und insbesondere Befragungsstudien in einem gemeinsamen Workshop am 23. August 2017 diskutiert.⁷ Die dort gemachten Anmerkungen zum besprochenen Itempool wurden aufgegriffen und der Itempool wurde erneut überarbeitet. Folgende Aspekte und Anmerkungen waren die Grundlage der Überarbeitung:

- Formulierungen wurden aus der Perspektive eines normalen Nutzers auf Verständlichkeit geprüft und ggf. entsprechend vereinfacht.
- Es wurde diskutiert, inwieweit jedes einer Dimension zugewiesene Item als repräsentativ für die Dimension gewertet werden könnte.
- Items, deren Formulierung einander zu sehr ähnelten, wurden gestrichen.

Im Anschluss an diese Überarbeitung der Frageformulierungen fanden sich 82 Frageitems im Itempool für die *BDGS*-Skala.

⁷ An dieser Stelle möchte ich Herrn Christopher Starke und Pero Došenović für ihre hilfreichen Anmerkungen danken.

8.5.3 Befragung von typischen Nutzer*innen digitaler Medien mittels der Methode des lauten Denkens

In einem weiteren Schritt wurden, wie in der Literatur vorgeschlagen, persönliche Interviews mit einfachen Nutzer*innen digitaler Medientechnik mit dem Ansatz des *lauten Denkens* durchgeführt, um die 82 Items auf Verständlichkeit zu prüfen und mögliche Probleme bei der Beantwortung der Items zu identifizieren (Prüfer & Rexroth, 2000, 2005). Beim *lauten Denken* „werden Personen aufgefordert, ihre Gedanken laut auszusprechen, während sie sich einer Aufgabe oder Tätigkeit widmen“ (Bilandzic, 2017, S. 406).

Insgesamt wurden im Zeitraum zwischen dem 30. August und dem 6. September 2017 sechs Interviews geführt. Bei der Rekrutierung der Gesprächsteilnehmer lag der Fokus insbesondere auf Internetnutzer*innen, die auf eine gewisse Online-Erfahrung zurückblicken konnten und ein Grundverständnis der Begriffe ‚Digitalisierung‘ und ‚digitale Daten‘ mitbrachten, da ihnen an dieser Stelle noch kein standardisierter Einleitungstext vorgelegt wurde. In Tabelle 2.B im Anhang im elektronischen Zusatzmaterial finden sich Informationen zu den persönlichen Angaben der Personen und der Interviewlänge. Insgesamt wurde in jedem Interview aus Zeitgründen lediglich ein Drittel der Items vorgelegt, so dass jeweils zwei Interviewte annähernd die gleichen Aussagen evaluiert haben.⁸ Zu diesem Zweck wurden die Items vor dem ersten Interview randomisiert, um zu gewährleisten, dass nicht nur Items aus einer bestimmten Dimension besprochen, sondern von allen Befragten möglichst viele Aussagen aus jeder der vorgeschlagenen Dimensionen eingeschätzt wurden.

Die Gespräche wurden mit Hilfe eines Tonbandgeräts aufgezeichnet. Zudem wurden handschriftliche Notizen zu den jeweiligen Items gemacht, die die Anmerkungen der Interviewten dokumentierten. Darüber hinaus wurde langes Zögern bei der Beantwortung einzelner Items notiert, um festzuhalten, welche Items möglicherweise problematisch sein könnten. Die ursprünglich formulierten Items wurden auf Grundlage der Interviews überarbeitet und angepasst. Folgende Anmerkungen waren die Grundlage der Überarbeitung:

- Missverständliche Items wurden verständlicher formuliert.
- Items, deren Formulierung als identisch wahrgenommen wurde, wurden gestrichen oder angepasst.

⁸ Aufgrund direkter Überarbeitung von missverständlichen und fehlerhaften Items, die im Anschluss an jedes Interview stattfanden, wichen auch die jeweils gleichen vorgestellten Itemlisten in Teilen voneinander ab.

Insgesamt ergab sich mit Blick auf die vorgelegten Itemlisten, dass die Items insgesamt als verständlich und erfolgreich zu beantworten wahrgenommen wurden. Es zeigten sich bei den verwendeten Ausdrücken keinerlei Schwierigkeiten für die Beantwortung, so dass nach Streichung als identisch wahrgenommener Items 69 Frageitems vorlagen. Auch eine Verortung auf der geplanten 7-Punkt-Likert-Skala konnte von allen Interviewten für die Items vorgenommen werden.

8.5.4 Zusammensetzung der Frageitems für die erste Erhebung

Alle 69 Frageitems sowie deren Zuordnung zu den einzelnen Dimensionen finden sich in Tabelle 2.C im Anhang im elektronischen Zusatzmaterial. Für die Dimensionen der Bewertung des *Volumens*, der *Geschwindigkeit* und der *Vielfalt* fanden sich letztendlich jeweils zwölf Frageitems. Bei der Formulierung der Items wurde zunächst der Idee gefolgt, man könne mit Blick auf den Entstehungs- und Verwertungskontext der Daten bei diesen Dimensionen noch zwischen Entstehung, Speicherung und Auswertung der Daten differenzieren. Die Formulierung der Frageitems greift daher diese Aufteilung in den genannten Dimensionen auf, so dass sich jeweils zwei bzw. drei hierauf abzielende Items finden. Da diese Frageitems (wie auf S. 94 besprochen) nicht zentral für die nachfolgende Auswertung sind, finden sich weitere Informationen hierzu im Anhang Abschnitt 1.1 im elektronischen Zusatzmaterial.

Für die Dimension *Richtigkeit* der Daten fanden sich 18 Frageitems, von denen jeweils sechs Frageitems die spezifischen Charakteristika *der Genauigkeit*, *des Wissensgewinns* sowie *der Wahrhaftigkeit* und *der Objektivität* betrafen. Für den *Nutzen* der digitalen Daten fanden sich insgesamt 15 Items, wobei sieben Items den *individuellen Nutzen* und acht Items den *gesellschaftlichen Nutzen* adressierten.

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.