5 Methodische Konzeption und Durchführung der empirischen Studie

Um ein umfängliches Wissen über den Forschungsgegenstand zu generieren, das sowohl in einer fallorientierten Sichtweise die alltäglichen Handlungserfahrungen der Beteiligten abbildet als auch aussagekräftige Aussagen erlaubt, wurde ein Mixed-Methods-Ansatz verwendet. Diese Vorgehensweise wird in Abschnitt 5.1 erläutert und begründet. Zur Beantwortung der Forschungsfragen wurde ein exploratives sequenzielles Design (Verallgemeinerungsmodell) gewählt, das mit einer qualitativen Erhebungsphase begann. Diese wird in Abschnitt 5.2 vorgestellt. Anschließend greift Abschnitt 5.3 die quantitative Erhebungsphase auf. Mit Abschnitt 5.4 wird dargestellt, wie die qualitativen und quantitativen Ergebnisse zusammengeführt wurden. Im letzten Abschnitt 5.5 wird das eigene Vorgehen mit Bezug auf Qualitätskriterien von Mixed-Methods-Studien zusammenfassend reflektiert.

5.1 Der Mixed-Methods-Ansatz als methodologische Basis der Studie

Konkrete und präzise Fragestellungen sind für eine gut funktionierende Forschung wesentlich. Mit ihnen wird das Forschungsinteresse notwendigerweise eingegrenzt und auf die interessierenden Aspekte fokussiert. Zudem können sich mit ihnen bestimmte Theoriebezüge und methodische Forschungsperspektiven a priori ergeben. Anders als für quantitative Fragestellungen ist es für qualitative Studien zulässig, die Ausgangsfragen im Verlauf des Forschungsprozesses weiter zu spezifizieren, was der offen-interpretativen Forschungslogik entspricht (Flick, 2017; Przyborski & Wohlrab-Sahr, 2014). Wie in der Einleitung der Arbeit begründet, lauten die übergreifenden Fragen dieser Arbeit:

1. Wie werden die gesetzlichen Vorgaben der WMVO und die Beteiligungsrechte der Werkstatträte in den WfbM auf praktischer Ebene ausgestaltet?
2. Was charakterisiert eine gelingende Werkstattratsarbeit und womit steht diese im Zusammenhang?

Die erste Frage zielt darauf, eine Bestandsaufnahme zu erlangen, sprich, den aktuellen Zustand der Werkstattratsarbeit auf praktischer Ebene zu beschreiben. Die zweite Frage fragt nach den Bedingungen einer gelingenden Beteiligung. Vor dem Hintergrund der Erkenntnisse zur betrieblichen Beteiligung und der Rechtssoziologie (siehe Kapitel 4) können die WMVO, DMVO und Caritas-WMO nicht als Status quo der Werkstattratsarbeit angenommen werden.

Zusatzmaterial online

Zusätzliche Informationen sind in der Online-Version dieses Kapitels (https://doi.org/10.1007/978-3-658-35383-4_5) enthalten.

© Der/die Autor(en) 2022

V. Schachler, *Partizipation durch Werkstatträte*, Beiträge zur Teilhabeforschung, https://doi.org/10.1007/978-3-658-35383-4_5

Es besteht wenig Detailwissen über die Arbeit von Werkstattträtinnen in WfbM. Offen bleibt, was eine gelingende Werkstattträtinnenarbeit charakterisiert.

Damit ist ein quantitativer Methodeneinsatz zunächst ungeeignet. Ein entwickeltes Erhebungsinstrument kann die tatsächliche Situation nur schlecht erfassen. Abschlüsse bei der Inhaltsvalidität wären zu erwarten. Jedoch kann auch ein qualitativer Erhebungszugang die Forschungsfragen nur teilweise beantworten. Eine zufriedenstellende Aussagenreichweite für das bundesweite Werkstattdiensystem wäre damit nicht möglich. Die skizzierten Fragen sind nicht eindeutig mit einem quantitativen oder qualitativen Forschungszugang verbunden, die als etablierte Methodenstränge der Sozialforschung vorliegen.⁵⁴ Neben diesen Forschungslinien bestehen Ansätze, die den qualitativen und quantitativen Methodenkanon kombinieren, was unter methodologischen Aspekten als Mixed Methods und teilweise als „third methodological movement“ (Tashakkori & Teddlie, 2003, S. 5) diskutiert wird. Die Stärke dieser Mixed-Methods-Ansätze liegt darin, dass diese ihre methodischen Vorgehensweisen nicht mit dogmatischen Forschungshaltungen verbinden, sondern in den zu erkundenden Forschungsgegenständen begründet sehen, womit eine passendere Analyse möglich wird (Creswell, 2003, S. 11). Die Anwendung eines solchen Ansatzes erlaubt es, die Fragen umfassend zu bearbeiten und zugleich so zu beantworten, dass die explorativen und konfirmatorischen Züge erhalten bleiben.

Mixed Methods als Verbindung komplementärer Forschungsstrategien

Creswell (2003), der als zentrale Figur innerhalb der Mixed-Methods-Forschung gilt, beschreibt diese als pragmatisch orientierte Vorgehensweise, in der sowohl numerische (quantitative) als auch textbasierte (qualitative) Daten erhoben und zur Erkenntnisgenerierung angewandt werden (S. 18 ff.). Neben dieser methodisch orientierten Definition bestehen zahlreiche mehr, die bspw. normative Grundhaltungen, eine Verbindung der Forschungsansätze zu mehreren Zeitpunkten oder bestimmte Durchführungskriterien betonen (Kuckartz, 2014, S. 31; Tashakkori & Teddlie, 2003, S. 9 ff.). Gemeinsamer Kerngedanke bleibt die kombinierte

⁵⁴ Die Charakteristika von qualitativer und quantitativer Forschung sind an anderen Stellen vielfältig diskutiert und dargestellt worden. Verweisen lässt sich z. B. auf Przyborski und Wohlrab-Sahr (2014) und Schnell, Hill und Esser (2013). Grob gesprochen wird qualitative Forschung mit der Beantwortung von explorativen Fragen assoziiert, wohingegen die Stärke von quantitativer Forschung in der Beantwortung von konfirmatorischen (bestätigenden) Fragen liegt (Tashakkori & Teddlie, 2003, S. 15). Es gibt jedoch auch Positionen, die diese dichotome Typisierung nicht als gegeben ansehen, wie dies bspw. Kuckartz (2014) formuliert: „Die Mehrzahl der Forschungsfragen besitzt ... sowohl qualitative wie quantitative Aspekte“ (S. 71).

Umsetzung von qualitativen und quantitativen Forschungsbestandteilen innerhalb eines Forschungsprojekts. Die Verwendung eines passenden, spezifischen Forschungsdesigns wird angeraten.

Ein Forschungsdesign kann als Planungshilfe oder Konzept verstanden werden, das offenlegt, mit welchen Vorgehensweisen eine Forschungsfrage beantwortet wird und welche Entscheidungen dazu vollzogen werden (Ragin, 1994, S. 26). Für Mixed Methods sind Forschungsdesigns besonders wichtig. Einerseits richtet sich die Wahl des Designs an dem Forschungsgegenstand aus und ist nicht methodenbezogen, wie dies für randomisierte kontrollierte Studien als Goldstandard der Fall ist (Kuckartz, 2014, S. 157). Andererseits sind weitere Entscheidungen zu treffen. Als mögliche Designtypen liegen verschiedene Klassifikationen vor. Um diese optisch abzubilden, wird teilweise auf ein Notationssystem zurückgegriffen, das die entstehenden Zusammenhänge und Betonungen verdeutlicht, z. B. Groß- und Kleinschreibungen als Kürzel für den Zeitpunkt oder die Bedeutung des Einsatzes (QUAN = quantitativer Studienteil, QUAL = qualitativer Studienteil) oder die Symbole + / → (Tashakkori & Teddlie, 2003, S. 27).

Creswell (2003, S. 16 f.) differenziert zwischen drei Ansätzen integrierter Forschung:

1. parallel verlaufende Ansätze, bei denen zeitgleiche Methodeneinsätze einander ergänzen, z. B. in der Datenerhebung, -auswertung oder -analyse;
2. sequenzielle Ansätze, in denen qualitative auf quantitative oder quantitative auf qualitative Methoden folgen;
3. transformative Ansätze, die entweder einen sequenziellen oder parallelen Methodeneinsatz aufweisen, im Gegensatz zu den ersten beiden Vorgehensweisen jedoch durch eine übergreifende Theorie gerahmt werden, was z. B. die Zielsetzung einer emanzipativ-transformativ ausgerichteten Forschung sein kann (Kuckartz, 2014, S. 50).

Für alle kombinierten Forschungsstrategien sind vier wesentliche Entscheidungen zu treffen: In einer methodenübergreifenden Perspektive ist festzulegen, ob mit einem bestimmten Theoriebezug gearbeitet und ob den qualitativen oder quantitativen Forschungsbestandteilen eine prioritäre oder eine gleichwertige Behandlung zugesprochen wird. Bezogen auf die praktische Umsetzung ist die zeitliche Abfolge der Erhebungen festzulegen sowie das Stadium, in dem die Forschungsergebnisse integriert werden (Creswell, 2003, S. 211 ff.). Die Integration der Ansätze kann zu verschiedenen Zeitpunkten erfolgen: zu Beginn bereits mit der Datenerhebung, während des Analyseprozesses, für die Interpretationen oder als integrierende

Schlussfolgerung (Kuckartz, 2014, S. 65; Mayring, 2001, S. 2 f.). Mit diesen Entscheidungsrichtlinien entstehen nun verschiedenste Möglichkeiten, integrierte Forschung umzusetzen, was in einer Vielzahl von möglichen Designtypen zum Ausdruck kommt. Als besonders markant gelten die folgenden vier Typen: concurrent design, explanatory sequential design, exploratory sequential design, embedded sequential / concurrent design (Creswell & Zhang, 2009, S. 613 ff.).

Basisdesign und Ziele der Mixed-Methods-Studie

Für die eigene Mixed-Methods-Studie wurde zunächst eine explorative qualitative Methode angewandt, um ein grundlegendes Wissen über den Forschungsgegenstand zu generieren, das auf dem Alltagserleben und den Handlungs- und Interaktionserfahrungen der Beteiligten beruht. Die gewonnenen Inhalte waren eine Grundlage für die quantitative Instrumentenerstellung, mit der die qualitativen Resultate weiterverfolgt und im Hinblick auf ihre Aussagereichweite überprüft werden konnten (sequenzielle Abfolge). Dies bedeutet, dass die qualitativen Ergebnisse in die Datengenerierung der quantitativen Methode eingingen, was in der neueren Mixed-Methods-Literatur als sequenzbasierte Integration beschrieben wird und eine Schnittstelle der Forschungsphasen ist (Kuckartz, 2018, S. 168). Integriert wurden die qualitativen und quantitativen Forschungsphasen weiter resultatsbasiert in Form einer gemeinsamen abschließenden Ergebnisanalyse und -interpretation. Hierbei wurden im eigenen Ansatz die qualitativen und quantitativen Forschungsergebnisse als gleichwertig betrachtet. Beide waren für ein mehrperspektivisches, tiefergehendes und verallgemeinerbares Verständnis des Forschungsgegenstandes notwendig. Eine explizite Theorie wurde nicht angewandt, da ein bestimmter theoretischer Bezug in den Forschungsfragen nicht angelegt war, die im Theorieteil der Arbeit skizzierten Inhalte waren die rahmenden Bezugspunkte.

Mit dieser Entscheidung kann der eigene Mixed-Methods-Ansatz nach den Vorschlägen von Creswell als sequenzielles exploratives Design (exploratory sequential design) charakterisiert werden, was in der deutschsprachigen Methodenliteratur als Verallgemeinerungsmodell bezeichnet wird (Burzan, 2016, S. 35 f.; Mayring, 2001, S. 5). Orientiert an den Entscheidungskriterien von Creswell (2003, S. 211) und den möglichen Integrationsstrategien nach Kuckartz (2018) ergibt sich für den Forschungsansatz das in Tabelle 8 dargestellte Bild.

Tabelle 8: Behandlung der Forschungsphasen im eigenen Ansatz

Implementierung	Sequenziell, qualitativer Beginn
Behandlung der Ansätze	Gleichwertig
Integration	Während der quantitativen Instrumentenkonstruktion und als resultatsbasierte Ergebnisanalyse/-interpretation
Theoretische Perspektive	Integriert

Ein Vorteil des gewählten Verallgemeinerungsdesigns ist, dass ein Forschungsgegenstand umfänglich erkundet und beschrieben werden kann. Gerade für ein neu zu erkundendes Forschungsthema ist dieses geeignet. Ein Nachteil liegt in einem benötigten großen Zeitaufwand, um beide Erhebungsteile abzuschließen und der qualitativen Studie nicht nur einen „Sprungbretteffekt“ (Kuckartz, 2018, S. 178) zuzuschreiben. Gerade in der Ergebnisintegration können Schwierigkeiten liegen (Creswell, 2003, S. 215 f.). Zudem ist generell für Mixed Methods ein höherer Ressourceneinsatz notwendig (Kuckartz, 2014, S. 54).

Abbildung 3: Mixed-Methods-Strategie zur Beantwortung der Forschungsfragen – exploratives sequenzielles Design (Verallgemeinerungsmodell)

Nichtsdestotrotz bot sich dieses Design für den eigenen Forschungsgegenstand als angemessen an: Ein breiter offener Ansatz war die Basis zur Wissensgenerierung. Darauf aufbauend wurde der Forschungsblick enger gefasst und die Erkenntnisse wurden bundesweit untersucht.

Damit wurde ein quantitatives Erhebungsinstrument generiert, das aus den Alltagserfahrungen der Beteiligten entstanden ist. Abschließend wurden die erhobenen Forschungsergebnisse integriert, um so Werkstatttratsarbeit umfassend empirisch darzustellen. Für diese angewandte Strategie ist das in Abbildung 3 dargestellte Pyramidenbild passend. Die qualitative Erhebungsphase war die Grundlage und fokussierte das Erkenntnisinteresse der zweiten Erhebungsphase auf bestimmte Untersuchungsbereiche. Zuletzt wurden die Ergebnisse auf dem Unterbau der qualitativen und quantitativen Forschungsphasen in ihrem Erkenntniswert integrierend zugespielt.

Ein übergeordnetes Ziel dieser Arbeit liegt darin, fundierte Erkenntnisse über die Ausgestaltung der Beteiligung in WfbM zu erarbeiten, mit denen die Aktivitäten der Werkstattträte in den Partizipationsdiskurs eingeordnet werden können. Empirisch angestrebt ist, sowohl die Sichtweisen und das Erfahrungswissen der beteiligten Akteure abzubilden als auch Aussagen über den Zustand der Werkstatttratsarbeit in den WfbM in Deutschland zu treffen. Mit der Anwendung des skizzierten Mixed-Methods-Ansatzes untergliedert sich das übergreifende Ziel in die folgenden empirischen Absichten:

1. Erhebung von für Werkstatttratsarbeit relevanten Bedingungen aus Sicht der Beteiligten (QUAL)
2. Konstruktion standardisierter Befragungsinstrumente auf Basis der qualitativen Forschungsergebnisse (QUAL => QUAN)
3. Bestimmung der Merkmale und Reichweiten von Beteiligungsbemühungen in WfbM (QUAN)
4. Ermittlung von Faktoren einer gelingenden Werkstatttratsarbeit und Überprüfung im Hinblick auf korrelierende Kontextfaktoren (QUAL + QUAN)

Eine fallweise Vermischung der qualitativen und quantitativen Erhebungsdaten innerhalb der zwei Erhebungsphasen war nicht angedacht; die qualitativen und quantitativen Forschungsdaten wurden singulär gewonnen.⁵⁵ Ein Methodentransfer wurde lediglich im Rahmen der

⁵⁵ Der Ablauf der gezielten Vermischung der Forschungsansätze, bspw. in dem qualitative in quantitative Daten transformiert werden, wird im deutschsprachigen Raum z. B. als „Transferdesign“ (Kuckartz, 2014, S. 87) oder „Daten-Triangulation“ (Mayring, 2001, S. 6) bezeichnet. Mit der qualitativen Inhaltsanalyse und QDA-Software ist solch ein Datentransfer möglich. Teilweise wird dieser auch unbewusst angewandt, in dem z. B. in qualitativer Forschung Angaben zu Häufigkeiten erfolgen oder quantitative Daten qualitativ transformiert werden (Kuckartz, 2014, S. 88).

üblichen Vorgehensweisen praktiziert (z. B. wurden Kategorienlabels bei metrischen Variablen vergeben). Die Erhebungen waren als in sich geschlossene Phasen angelegt, wonach sich der Erhebungsablauf als linear verlaufender Prozess ausgestaltete (siehe Abbildung 4).

Abbildung 4: Ablauf des Erhebungsprozesses als lineares Modell

5.2 Qualitative Erhebungsphase

Die postulierte Offenheit des Forschungsprozesses qualitativer Forschung heißt nicht, dass dieser ohne Planung erfolgt. Die eigene qualitative Erhebungsphase wurde als Momentaufnahme konstruiert. In diesem Unterkapitel werden das konzipierte Design (Abschnitt 5.2.1), die gewählte Erhebungsmethode der Gruppendiskussion (Abschnitt 5.2.2), das gewählte und realisierte Sampling (Abschnitt 5.2.3), die Erhebungsdurchführung (Abschnitt 5.2.4) sowie die Wahl der Auswertungsmethode (Abschnitt 5.2.5) und deren Anwendung (Abschnitt 5.2.6) schrittweise vorgestellt. Abschließend wird auf Qualitätskriterien qualitativer Forschung eingegangen (Abschnitt 5.2.7).

5.2.1 Forschungsdesign

Unabhängig von dem gewählten Methodenzugang sind Untersuchungskonzepte hilfreich, um einen Forschungsprozess zu strukturieren. Zudem dienen sie dazu, Forschung literatur- und theoriebegründet durchzuführen und mit den Ergebnissen anschlussfähig an den Fachdiskurs zu bleiben (Creswell, 2003, S. 3). Die Besonderheiten qualitativer Forschungsdesigns werden etwa von Przyborski und Wohlrab-Sahr (2019) herausgearbeitet. Gängige Varianten finden sich z. B. bei Creswell (1998, S. 79 ff.) oder Mayring (2010a, S. 225 f.) klassifiziert.

Für die eigene qualitative Erhebungsphase bot es sich an, diese als Momentaufnahme zu konstruieren, sprich als punktuelle Beschreibung und Analyse des Zustandes zu dem gegebenen Erhebungszeitpunkt (Flick, 2017, S. 255). Die weiteren Designkomponenten sind, orientiert an Flick (2017, S. 252), in Tabelle 9 dargestellt. Die der Arbeit zugrundeliegenden Ziele wurden den jeweiligen Forschungsstärken entsprechend zwischen den Erhebungsphasen aufgeteilt. Für die qualitative Studie wurden der deskriptiv-explorative Aspekt und die offenen Fragen übernommen. Die angewandte Vorgehensweise ist im Spektrum qualitativer Forschung als eher straff und kontrolliert zu bezeichnen, was durch die Samplingstrategie und die angewandten Methoden zum Ausdruck kommt. Sie empfahl sich, da die Forschungsergebnisse als Ausgangspunkt der quantitativen Erhebungsphase in dem Mixed-Methods-Design punktuell abgeschlossen sein mussten.

Tabelle 9: Forschungsdesign der qualitativen Erhebungsphase

Ziele	<ul style="list-style-type: none"> • Erhebung relevanter Bedingungen zur Arbeit von Werkstatträten aus Sicht der Beteiligten (Deskription) • Ermittlung von Vorstellungen und Faktoren einer gelingenden Werkstatttratsarbeit (Exploration) • Hypothesengenerierung zur Entwicklung standardisierter Erhebungsinstrumente (Exploration)
Forschungsfragen	<ul style="list-style-type: none"> • Wie werden die Verordnungsvorgaben und die Beteiligungsrechte der Werkstatträte in den WfbM ausgestaltet? • Was charakterisiert eine gelingende Werkstatttratsarbeit und womit steht diese im Zusammenhang?
Methodische Umsetzung	<ul style="list-style-type: none"> • Erhebung: Gruppendiskussionen • Auswertung: Qualitative Inhaltsanalyse
Sampling	<ul style="list-style-type: none"> • Inhaltliche Kriterien als Grundlage und Themenbreite als Ziel • Intensitäts-Sampling (Patton, 2002, S. 234) anhand von drei untersuchungsgegenständlichen Kriterien (eingenommene Perspektive, Aktivitätsebene, gültige Mitwirkungsverordnung)
Maß der Offenheit	<ul style="list-style-type: none"> • Gewählte Sampling- und Auswertungsstrategie entsprechen einer eher strafferen Vorgehensweise
Durchführung	<ul style="list-style-type: none"> • April bis Juni 2018 • Simultane Durchführung mit einem Forschungsauftrag zur Umsetzung der reformierten WMVO (Schachler 2018a, 2018b)

Die Samplingstrategie begründete sich anhand von inhaltlichen Kriterien, die nach der gewählten Erhebungsmethode vorgestellt werden.

5.2.2 Auswahl der Erhebungsmethode

Zur methodischen Umsetzung der qualitativen Erhebung wurde die ermittelnde Gruppendiskussion gewählt. Diese lässt sich als Methode charakterisieren, mit der Informationen über den fokussierten Inhalt sowie damit verbundene Einstellungen und Argumentationsmuster von teilnehmenden Personen gewonnen werden. In der Ausgestaltung handelt es sich um ein vorstrukturiertes, moderiertes Diskussionsverfahren, bei welchem Themenfelder mit einer überschaubaren Anzahl von Personen diskutiert werden. Dabei erlaubt es der diskursive Prozess, dass sich die Teilnehmenden gegenseitig befragen und Argumente wechselseitig ausgehandelt

werden (Lamnek, 1998, S. 29 f., 2005, S. 413). Teilweise wird diese Methode mit dem Begriff der Fokusgruppendiskussion (engl. focus groups, Morgan, 1997) gleichgesetzt, wobei diese Bezeichnung eher mit der angloamerikanischen Marktforschung (Bohnsack, 2010, S. 105) oder mittlerweile im deutschsprachigen Raum mit der drittmittelfinanzierten Sozialforschung (Schulz, 2012, S. 8) verbunden wird. Vorteile von Gruppendiskussionen liegen z. B. darin, dass mit ihnen auf einfache Weise eine große Themenbreite und vielfältige Meinungen erhoben werden können (Lamnek, 1998, S. 74 ff.). Für die Zielgruppe von Menschen mit geistigen Beeinträchtigungen werden Gruppendiskussionen als angemessene Methode empfohlen, da es mit der alltagsnahen Gesprächssituation eher gelingen kann, persönliche Sichtweisen zu gewinnen, als dies bspw. mit Einzelinterviews möglich ist (Wagner-Willi, 2002).

Als zwei methodische Vertreter der Gruppendiskussion lassen sich Bohnsack (2010) und Lamnek (1998) aufführen. Während Lamnek eine pragmatische und ressourcenschonende Methode präsentiert, hebt Bohnsack die methodologische Fundierung, die Analyse gruppenspezifischer Prozesse und kollektive Orientierungsmuster hervor. Dementsprechend kritisiert Bohnsack (2010, S. 105) Lamneks Entwurf als verkürzt und inadäquat. Auch Przyborski und Wohlrab-Sahr (2014, S. 89) beschreiben Gruppendiskussionen ohne interaktive Analyse als methodisch problematisch, nicht valide und wenig reliabel. Zur Ideen- und Hypothesengenerierung vor quantitativen Erhebungen sehen sie diese weniger kritisch. Da der Zweck der eigenen Gruppendiskussionen mit diesen Zielen verbunden war, es nicht um gruppenspezifische Orientierungsmuster ging und die qualitative Erhebung als Teilkomponente des Forschungsvorhabens realisierbar bleiben musste, wurde bewusst der heuristische Ansatz von Lamnek gewählt.

In seinem Methodenentwurf folgt Lamnek (1998, S. 27) der Definition von Morgan, wonach die Diskussionsthematik vorgegeben und die Daten interaktiv durch die Gruppenmitglieder erzeugt werden. In seinen weiteren Beschreibungen nehmen vor allem die Planungsschritte, die Umsetzungs-, Auswertungs- und Analysemöglichkeiten sowie die damit verbundenen Entscheidungen für eher offene oder stärker kontrollierte Vorgehensweisen breiten Raum ein.

5.2.3 Samplingstrategie und realisierte Gruppenbesetzungen

Auswahlentscheidungen sind zentrale Momente im Forschungsablauf. Die an diesen Stellen vorgenommenen Präferenzen bestimmen maßgeblich, was im Verstehensprozess jeweils relevant und konstituiert wird (Flick, 2016, S. 154 ff.). Die potenzielle Vielfalt an Auswahlmöglichkeiten spiegelt sich in einer Vielzahl von klassifizierten Vorgehensweisen zur Fallauswahl wider (Gerring & Cojocaru, 2015, S. 2).⁵⁶

Als typische Strategie des Auswahlprozesses der qualitativen Forschung gilt das theoretische Sampling nach Glaser und Strauss (2005). In seiner genuinen Form geht dieses nicht von einem vorab festgelegten Stichprobenplan und einer bestimmten Stichprobengröße aus, sondern die Fallauswahl wird während des Erhebungs- und Auswertungsprozesses anhand des vermuteten Mehrwerts entschieden und so lange „durch die im Entstehen begriffene ... Theorie kontrolliert“ (ebd., S. 53), bis diese thematisch gesättigt ist. Für die eigene Stichprobenbildung war dieses Vorgehen nicht praktikabel, da die Fallauswahl im Rahmen von Mixed Methods anwendbar und für das zweiteilige Vorhaben realisierbar bleiben musste. Jedoch wurde die Grundidee übernommen, nach theoriegeleiteten Kriterien (Glaser & Strauss, 2005, S. 53 ff.) vorzugehen. Hierzu wurde die Strategie des Intensitäts-Samplings angewandt. Bei dieser Vorgehensweise orientiert sich die Fallauswahl an informationsreichen Fällen, die das zu untersuchende Phänomen vertreten, jedoch nicht an besonders außergewöhnlichen oder etwa abweichenden Fällen. Diese heuristische Strategie setzt Vorwissen über das Untersuchungsphänomen voraus und bezieht Fälle anhand von Kriterien mit ein, die sich in der Vorarbeit als bestimmende Elemente des Untersuchungsgegenstands ergaben (Patton, 2002, S. 234).

⁵⁶ Zur Systematisierung vorhandener Techniken existieren verschiedene Vorschläge. Mit einer übergreifenden Typologie charakterisieren Gerring und Cojocaru (2015) fünf Haupttypen und 18 untergeordnete Subtypen, die zur Fallauswahl in empirischen Studien herangezogen werden: Erstens – *Repräsentativ*: Die ausgewählten Fälle bilden eine Bevölkerungsgruppe nach relevanten Aspekten ab. Zweitens – *Anomal*: Die ausgewählten Fälle weichen von einer gängigen Erscheinung / einem gängigen Themenverständnis ab. Drittens – *Most-different (method of agreement)*: Die ausgewählten Fälle haben eine Gemeinsamkeit, unterscheiden sich jedoch hinsichtlich betonter Faktoren. Viertens – *Entscheidend (crucial)*: Die ausgewählten Fälle bilden Einzelfallstudien, bei denen ein Faktor zu verschiedenen Zeitpunkten erforscht wird. Fünftens – *Most-similar (method of difference)*: Die ausgewählten Fälle ermöglichen den Vergleich zwischen einer Experimental- und einer Kontrollgruppe (ebd., S. 5 ff.).

Orientiert daran wurde für die Untersuchung ein Sampling anhand der folgenden drei Kriterien gewählt:

1. der Perspektive, mit der die Diskussionsteilnehmer*innen mit dem Thema verbunden waren (als Einrichtungsleitung, Werkstatttratsmitglied oder Vertrauens-/Unterstützungsperson);
2. der Aktivitätsebene der Interessenvertretung (werkstattbezogen oder überregional);
3. der drei Verordnungsgrundlagen der Werkstatttratsarbeit, sprich nach der Trägerschaft der Einrichtung.

So konnten multiperspektivische Sichtweisen und ein umfassendes Situationsbild zur Werkstatttratsarbeit eingeholt werden. Für das Fallverstehen wurde davon ausgegangen, dass die Teilnehmenden sich nicht nur als Person präsentieren, sondern als Repräsentant*innen unterschiedlicher Funktionen und Einrichtungsformen auftreten (Flick, 2016, S. 168). Przyborski und Wohlrab-Sahr (2014) beschreiben dies als „Präsentationsfassaden“ (S. 6).

Die Beeinträchtigungsart war kein spezielles Auswahlkriterium der Teilnehmenden. Stattdessen wurden diese lediglich in ihrer Funktion als Werkstatttratsmitglied (und -leitung oder Vertrauensperson) angesprochen, und es wurde auf eine barrierearme Gestaltung der Erhebungssituationen und -instrumente geachtet. So ist es mittlerweile für Menschen mit kognitiven Beeinträchtigungen nicht mehr gängig, auf Stellvertreteraussagen aus dem Umfeld zurückzugreifen (Buchner, 2008, S. 518). Gefolgt wird der Devise, dass diese am besten selbst Auskunft über ihre Lebenssituationen geben können (z. B. Niediek, 2016; zur Reflexion forschungsethischer Spannungsfelder Schuppener, Heusner & Müller, 2020).⁵⁷

Realisierte Samplingstrategie

Die gewählte Samplingstrategie wurde in vier Gruppendiskussionen transformiert. Diese wurden mit vier bis fünf Personen überschaubar gehalten. In der Besetzung waren Werkstatttratsmitglieder und Werkstattleitungen/Geschäftsführungen bewusst getrennt, da in den vertretenen Funktionen „Werkstattbeschäftigte*r“ und „Leitung“ ein Statusgefälle besteht und angenommen wurde, dass sich teilnehmende Personen in getrennten und kleinen Runden freier

⁵⁷ Zur Erhebungsdurchführung werden unterstützende Techniken empfohlen. Für qualitative Interviews wird etwa ein offenes Konzept angeraten, das die empathischen Fähigkeiten bei der Gesprächsführung betont (Buchner, 2008, S. 518 ff.) oder die Anpassung der Interviewsprache an vorher ermittelte Kommunikationsmuster (Hagen, 2002, S. 296 ff.).

äußern können. Die Anzahl der Diskussionen wurde zu Beginn der Erhebungen zunächst offengelassen. Da mit vier Gruppen bereits vielfältiges und umfangreiches Material gewonnen und die Samplingstrategie saturiert werden konnte, wurde es bei dieser Anzahl belassen.

Realisierte Gruppenbesetzungen

Folgende Gruppendiskussionen wurden realisiert:

- Gruppe 1: Einrichtungsververtretungen und Vertrauens-/Unterstützungspersonen mit einrichtungsbezogener (lokaler) und überregionaler Tätigkeit;
- Gruppe 2: Werkstatttratsmitglieder und Vertrauens-/Unterstützungspersonen mit einrichtungsbezogener und überregionaler Tätigkeit;
- Gruppe 3: Werkstatttratsmitglieder mit überregionaler Tätigkeit;
- Gruppe 4: Werkstatttratsmitglieder mit einrichtungsbezogener Tätigkeit.

Über offene Aufrufe der Interessenverbände und gezielte Anfragen wurden die Teilnehmenden akquiriert. Die ersten zwei Gruppen waren künstlich zusammengestellt, die anderen zwei waren natürliche, durch Arbeitszusammenhänge bestehende Gruppen. Tabelle 10 stellt dar, wie sich die Diskussionsgruppen jeweils zusammensetzten, um die Personen anonym zu halten, wurde auf genauere personenbezogene Angaben verzichtet.

Unter den 18 Teilnehmenden sind die Sichtweisen von Menschen mit Beeinträchtigungen mit 11 Werkstatttratsmitgliedern explizit vertreten. Alle beteiligten Assistenzen/Vertrauenspersonen verfügen über Erfahrungen in der überregionalen Interessenvertretung, teilweise obliegen ihnen dabei geschäftsführende Funktionen. Ihre Aufgaben nehmen sie in Teilzeit mit jeweils über 20 Stunden wahr. Die teilnehmenden Einrichtungsvertreter sind als Geschäftsführung oder Werkstattleiter tätig, alle verfügen über langjährige Erfahrung in WfbM. Auch die Werkstatttratsmitglieder. Ihre Zugehörigkeit zu Werkstätten reicht von 8 bis zu 40 Jahren. Alle Diskussionsteilnehmenden sind über 30 Jahre alt. Auffällig ist, dass acht der elf Werkstatttratsmitglieder über Arbeitserfahrungen auf dem allgemeinen Arbeitsmarkt verfügen. Für die meisten ist es die dritte Amtszeit, nur ein Werkstatttrat ist erst kürzlich in den Werkstatttrat gewählt worden. Die Teilnehmenden stammen aus zehn verschiedenen Bundesländern. Drei der vertretenen Einrichtungen gehören der Diakonie und eine der Caritas an, die restlichen Teilnehmenden arbeiten nach der WMVO. Mit den Personen sind vier kleinere (< 350 Werkstattbeschäftigte), fünf mittlere (350 bis < 1.000 Werkstattbeschäftigte) und fünf große WfbM (1.000 und mehr Werkstattbeschäftigte) vertreten. In fünf der Einrichtungen gibt es separate Zweigwerkstattsräte, teilweise auch einen zusätzlichen Gesamtwerkstatttrat.

Tabelle 10: Profil der realisierten Gruppendiskussionen

	Gruppe 1 (n = 5)	Gruppe 2 (n = 5)	Gruppe 3 (n = 4)	Gruppe 4 (n = 4)
Teilnehmende Personen	Zwei Assistenz-/Vertrauenspersonen. Drei Leitungen aus unterschiedlichen Werkstätten, ein Werkstattdirektor ist in der überregionalen Interessenvertretung aktiv.	Drei Mitglieder aus verschiedenen Werkstattdirektoren, darunter sind zwei in der überregionalen Interessenvertretung aktiv. Zwei Assistenz-/Vertrauenspersonen.	Vier Personen, die aus unterschiedlichen Werkstattdirektoren stammen und in der überregionalen Interessenvertretung aktiv sind.	Vier Mitglieder eines Werkstattdirektors.
Arbeits Erfahrung	Vier bis über 30 Jahre im Bereich WfbM.	Fünf bis elf Jahre als Mitglieder des Werkstattdirektors bzw. Assistenz.	Langjährige Erfahrungen als Mitglieder des Werkstattdirektors.	Ein Werkstattdirektorsmitglied übt das Amt seit 2001 aus, ein anderes Mitglied erst seit sieben Monaten.
Größe der WfbM	~ 170 bis 1.500 Werkstattbeschäftigte	~ 300 bis 1.800 Werkstattbeschäftigte	~ 120 bis > 3.000 Werkstattbeschäftigte	~ 400 Werkstattbeschäftigte
Verordnung	WMVO, DWMV	WMVO, Caritas-WMO	WMVO, DWMV	WMVO

Die Teilnehmenden der Gruppen 1–3 sind in der Werkstattdirektorsarbeit sehr engagiert, was bspw. durch die vielfältigen Aktivitäten der Werkstattdirektorsmitglieder, die umfangreichen Berichte der Leitungen oder die Stellenumfänge der Vertrauenspersonen/Assistenzen zum Ausdruck kommt. Diese Personen verfügen über ein breites Erfahrungswissen in ihrem Tätigkeitsbereich. Sie können nicht als typische Vertreter*innen im Rahmen ihrer Funktionen verstanden werden, da ein entsprechendes Engagement nicht charakteristisch für WfbM anzunehmen ist. Gruppe 4 lässt sich demgegenüber in Anlehnung an Breit und Kotthoff (1990, S. 149 ff.) als Gremium mit einem schwächeren Mitwirkungsprofil charakterisieren. Dieser Werkstattdirektor nimmt im Einrichtungsgefüge der Werkstatt bisher geringe soziale Bedeutung ein und weist eine eingeschränkte Selbstständigkeit auf.

5.2.4 Durchführung und Ablauf der Gruppendiskussionen

In Anlehnung an vorhandene Empfehlungen zur Durchführung von Gruppendiskussionen wurde ein thematischer Leitfaden erstellt, um die Erhebungssituationen grob zu strukturieren und thematisch mit erzählgenerierenden Fragen zu fokussieren (Lamnek, 1998, S. 87 ff.). Wie

Tabelle 11 zeigt, zielten die Stimuli zunächst auf Erzählungen und Beschreibungen und erst im letzten Teil auf Begründungen und Meinungen.

Tabelle 11: Erstellter Leitfaden für die Gruppendiskussionen (Ausschnitt)

Gesprächs- bereich	Erkenntnisinteresse	Beispielfrage
Vorstellungs- runde	„Warming-up“	Wer sind Sie und in welcher Form haben Sie mit Werkstatt- räten zu tun? Oder: Wer sind Sie und wie lange sind Sie schon im Werkstattatrat?
Einstieg	Erfahrungsbericht	Sie arbeiten also alle in einer Werkstatt oder mit Werkstätten. Wie ist das denn so mit den Werkstattträten? Können Sie mir erzählen, was Sie in der Arbeit mit Werkstattträten/als Werk- statttrat so erleben?
Themen zur Umsetzung und Ausge- staltung der WMVO / Werkstatt- ratsarbeit	Erfahrungsbericht	Was funktioniert in der Arbeit der Werkstattträte gut?
	<u>Bewertung</u>	
	Gelingensbedingun- gen	Bitte vervollständigen Sie diese Sätze: Für eine gelingende Ar- beit des Werkstattrats braucht es ...
Abschluss	Personenbezogene Faktoren	Was muss ein*e Werkstattleiter*in können, um gut mit dem Werkstatttrat zusammenzuarbeiten?
	Bewertung	Wenn Sie drei Wünsche für die Arbeit von Werkstattträten hät- ten, was würden Sie sich wünschen?

Um die Gruppendiskussionen für Personen mit kognitiven Beeinträchtigungen zu ermöglichen, wurde der Leitfaden durch ein Büro für Leichte Sprache übersetzt und überprüft. Im Kern betraf die Übersetzung kleine Formulierungen und zwei optionale Statements, die situa-
tiv als Diskussionshilfen dienten. In den Erhebungen wurden die Sprachvarianten situations-
adäquat eingesetzt.

Ablauf der Gruppendiskussionen

Die Gruppendiskussionen wurden im Zeitraum April bis Juni 2018 durch die Autorin als Mo-
deratorin durchgeführt. Die Durchführung war an den Prinzipien von Przyborski und Wohl-
rab-Sahr (2014, S. 96 ff.) orientiert, wonach sich die Rolle der Moderation auf immanente
Nachfragen, an die Gruppe gerichtete Erzähl- und Diskussionsstimuli, ein aktives Zuhören
und diskussionsrahmende Aspekte (z. B. Einbezug aller Teilnehmenden) beschränkte. Der ge-
sprächsunterstützende Leitfragebogen wurde insgesamt spärlich verwendet. Vor Beginn der
Diskussionen fand ein gesprächsaufwärmendes „Joining“ (ebd., S. 80 f.) statt. Die Gespräche
wurden per Audioaufnahme dokumentiert, um sie anschließend dezidiert auswerten zu

können. Gruppe 2 wurde zusätzlich mit Videokameras aufgenommen, um die Sprecherzuordnung zu erleichtern. An Gruppe 2 nahm zusätzlich eine personenbezogene Assistenzperson teil, die sich inhaltlich nicht an der Gesprächsrunde beteiligte. Eine Person aus Gruppe 1 musste die Diskussionsrunde terminbedingt vorzeitig verlassen. Die Diskussionen dauerten zwischen 102 (Gruppe 1) und 75 Minuten (Gruppe 3).

Nach dem Ende der Diskussionen wurden die Teilnehmenden gebeten, einen Kurzfragebogen auszufüllen, der personenbezogene Angaben (z. B. Amtstätigkeit in Jahren) und einrichtungsbezogene Angaben (Größe der Werkstatt etc.) enthielt. Gleichfalls wurde durch die Moderatorin ein Postskriptum zu prägnanten Themen und Besonderheiten im Diskussionsverlauf angefertigt.

Um die erhobenen Gruppendiskussionen textbasiert auswerten zu können, wurden diese vollständig transkribiert und anonymisiert. Die Transkriptionen wurden durch einen professionellen Schreib- und Büroservice nach den einfachen Transkriptionsregeln von Dressing und Pehl (2018, S. 21 ff.) mit dem Programm f4-transkript durchgeführt. Anschließend erfolgte ein Korrekturabgleich des Transkripts mit der Audio-Aufnahme durch die Autorin dieser Arbeit. Die Zuordnung der sprechenden Personen war i. d. R. möglich. Ausnahmen sind kurze Bestätigungssätze bzw. kleinere Sprecherüberlappungen. *Anhang 1* enthält eine Übersicht über die verwendeten *Transkriptionsregeln und die Sprecherkennzeichnung* nach Gruppen. Alle Personennamen, Bundesländer, Gegenden, Städte, Einrichtungsnamen und sonstigen Eigennamen wurden anonymisiert. Öffentlich bekannte Gegebenheiten wurden hingegen nicht weiter verfremdet (z. B. Lebenshilfe Marburg als Anbieter von Schulungen, SPD-Werkstatträte-Konferenz in Berlin).

5.2.5 Auswahl einer gegenstandsangemessenen Auswertungsmethode

Zur Auswertung von Gruppendiskussionen werden verschiedene Methoden und Analyseformen angewandt und verschiedene Verfahrenswege vorgeschlagen. Lamnek (1998, S. 162 ff.) differenziert zwischen einem inhaltlich-thematischen und einem gruppendynamischen Erkenntnisinteresse an das Datenmaterial, welches jeweils in deskriptiven, reduktiven oder explikativ-intensiven Formen bearbeitbar ist. In der Methodologie lässt sich Lamnek als ein Vertreter charakterisieren, der qualitative Forschung in Abgrenzung zu quantitativen Forschungszugängen und weniger als autarke Verfahren begründet (Przyborski & Wohlrab-Sahr, 2014, S. 11). Für eine eigenständige qualitative Forschungslogik wird demgegenüber postuliert, dass diese dann beginnt, wenn die Ebene einer ersten Analyseperspektive überschritten wird, d. h.,

wenn sich nach einer erfolgten Rekonstruktion der Äußerungen und der vorhandenen Alltagstheorien eine weitergehende Analyseperspektive anschließt, welche dem Handeln bzw. Wissen zugrundeliegende Prozess- und Sinnstrukturen rekonstruiert (die sogenannte Analyse zweiten Grades; ebd., S. 27). Zur Erarbeitung einer Analyse zweiten Grades wird bspw. die objektive Hermeneutik oder Grounded Theory empfohlen. Demgegenüber gilt die qualitative Inhaltsanalyse zur Offenlegung weitergehender Interpretationen als ungeeignet, da sie eher klassifiziert, „als dass sie Sinnstrukturen rekonstruiert, sie ist nicht in der Lage bzw. nicht darauf angelegt, implizite Bedeutungen ... zu erfassen“ (ebd., S. 183). Gleichwohl ist sie ein pragmatisches qualitatives Verfahren, das sich für das eigene Vorgehen angesichts der regelgeleiteten, systematischen Anwendung und des Mixed-Methods-Designs empfahl.

Die Auswertungsmethode der qualitativen Inhaltsanalyse ist wesentlich mit dem Namen Mayring verbunden. Dieser publizierte Anfang der 1980er-Jahre unter dieser Bezeichnung seine Auswertungsmethode, die verschiedene, stark regelgeleitete Techniken zur Analyse von textbasiertem Material beinhaltet. Mayring begründet seinen Ansatz mit den Grenzen quantitativer Forschungsmethoden und dem Bedarf von offenen Erhebungszugängen. Sein besonderes Anliegen ist es, in der qualitativen Auswertung methodisch kontrolliert und systematisch vorzugehen, um so Verdachtsmomenten einer beliebigen und vagen Vorgehensweise entgegenzuwirken (Mayring, 2010b, S. 9 ff.).⁵⁸ Die genaue (vorherige) Festlegung des Selektionskriteriums (als die theoriegeleitete Frage danach, was im Material genau erforscht wird) und des Abstraktionsniveaus (dieses spannt den Rahmen, mit welchem Blickwinkel gearbeitet wird, der weiter oder enger gefasst werden kann) in seinem Prozessmodell der induktiven Kategorienbildung sowie die strikte Trennung zwischen einer deduktiven und induktiven Vorgehensweise (ebd., S. 83 ff.) muten in der Anwendung mitunter wenig offen an. Im Gesamtblick beschreibt Mayring (ebd., S. 8) seine Auswertungsmethodik als eine Art „Zwischenposition“, die zwar ein (qualitativer) Interpretationsvorgang bleibe, aber auch quantitative Auswertungsschritte erlaube und als eine Verbindung von quantitativen und qualitativen Ansätzen verstanden werden könne. So eignet sie sich bspw. mit dem Ablaufmodell der deduktiven Kategorienbildung zur Bearbeitung von größeren Mengen schriftlichen Materials.

⁵⁸ Disziplinär verortet sich Mayring in einem stark quantitativ geprägten Umfeld. So lassen sich seine verwendeten Techniken auch mit einer quantitativen Prägung assoziieren. Bspw. gibt es ein skalierendes Verfahren, Codes bekommen Kürzel zugewiesen und der Berechnung der Intercoderreliabilität wird hohe Relevanz zugeschrieben (Mayring, 2010b, S. 114 ff.).

Neben Mayrings Entwurf gibt es weitere Varianten der qualitativen Inhaltsanalyse, die zur Komprimierung von qualitativem Erhebungsmaterial angewandt werden (für einen Überblick siehe z. B. Schreier, 2014). Unterschiede liegen bspw. in eher materialorientierten (induktiven), in mehr theoriegeleiteten (deduktiven) oder gemischt deduktiv-induktiven Vorgehensweisen. Auch Kuckartz begründet einen eigenständigen Entwurf, in dem er drei mögliche Verfahrensarten zur inhaltsanalytischen Auswertung skizziert (inhaltlich-strukturierende, evaluative oder typenbildende qualitative Inhaltsanalyse). Im Unterschied zu Mayring sind seine Strategien stärker zirkulär hermeneutisch geprägt und erlauben eine Kategorienbildung auf Grundlage des Materials in mehreren Auswertungsschleifen als induktiv-deduktives Vorgehen.

Vor diesem Hintergrund begründet sich die Auswahl einer gegenstandsangemessenen Auswertungsmethode wie folgt:

- Für das Vorgehen steht die themenzentrierte Analyse der Gesprächsinhalte im Fokus.
- Das Erkenntnisinteresse beschränkt sich auf das Nachzeichnen und Verstehen des Forschungsgegenstandes nach den Äußerungen der Beteiligten. Eine weitergehende Explikation der Sinn-, Handlungs- oder Prozessstrukturen im Sinne einer Analyse zweiten Grades wird nicht angestrebt.
- Stattdessen geht es um die Klassifikation von Inhalten im Themenbereich, wofür sich im Auswertungskanon der qualitativen Verfahren die qualitative Inhaltsanalyse anbietet.
- Unter den verschiedenen Varianten qualitativer Inhaltsanalyse wurde das inhaltlich strukturierende Verfahren nach Kuckartz (2016) gewählt, da so eine anwendungsbezogen-systematische Vorgehensweise möglich ist, die zugleich verhältnismäßig offen und materialorientiert bleibt (S. 51).

5.2.6 Die Anwendung der inhaltlich strukturierenden qualitativen Inhaltsanalyse

In der Auswertung von qualitativen Erhebungen sind Kategorien und Codes wesentliche Begriffe, wobei diese sehr unterschiedlich angewandt und interpretiert werden. Mangels klarer Abgrenzungen schlägt Kuckartz (2016, S. 37) eine synonyme Verwendung vor, wobei diese erst mittels dezidierter Beschreibung zu inhaltsanalytischen Kategorien werden. Dieser Beschreibungs- oder Definitionsvorgang bildet den für die Methode der qualitativen Inhaltsanalyse zentralen Prozess der Kategorienbildung. Diese ist auf verschiedene Weisen möglich. Die Erarbeitung mittels einer vorher festgelegten, klar definierten, abgrenzbaren und plausiblen Systematik (deduktive Kategorienbildung) und die materialorientierte Konstruktion

(induktive Kategorienbildung) werden als gegensätzliche Verfahrenswege unterschieden, wobei in der Forschungsanwendung häufig deduktiv-induktive Mischformen zu entdecken sind (ebd., S. 63 ff.). Der gewählte inhaltsanalytische Auswertungsvorgang enthält spezielle Begriffe, die bei Kuckartz (2016, S. 30 ff.) ausführlich dargestellt sind (z. B. Auswahlinheit, Analyseeinheiten, Codiereinheit etc.). In Anlehnung an seine Methodik (ebd., S. 83) wurden zunächst die Ziele der Kategorienbildung sowie die dazugehörigen Basiselemente bestimmt. Die Zielsetzung des eigenen Kategoriensystems lässt sich als deskriptiv sowie hypothesenbildend-explorativ charakterisieren. Dieses sollte eine Beschreibung ermöglichen, wie sich Werkstatttratsarbeit nach Ansicht der Beteiligten ausgestaltet. Als weiteres Ergebnis waren Hypothesen über eine gut funktionierende Werkstatttratsarbeit angestrebt. Durch diese Zielsetzungen und das Vorwissen zu den Arbeitsgrundlagen der Werkstattträte wurde die Kategorieenbildung mitgeprägt.

Kategorien können unterschiedliche Formen annehmen (bspw. evaluative, analytische oder formale, ebd., S. 34 f.). Im eigenen Vorgehen wurden primär thematische Kategorien gebildet, die im Material auf bestimmte Themen hinweisen. Vereinzelt waren diese um faktenorientierte Kategorien ergänzt, die sich tatsachenorientiert bspw. durch die Funktion als Werkstatttratsmitglied ergeben. Das Abstraktionsniveau (ebd., S. 83) war relativ allgemein gehalten, d. h., die Kategorien waren weniger abstrakt gefasst. Erstellte Fallzusammenfassungen und der Leitfaden dienten als Ausgangspunkt für mögliche Kategorien, die um weitere Aspekte ergänzt wurden, die aus der sequenziellen Materialsichtung stammten. Hieraus ergab sich verschiedene Begriffe, die in genau definierte Hauptkategorien überführt wurden. Das genauere Vorgehen ist ausführlich bei Kuckartz (2016, S. 97 ff.) dargestellt. An einer Teilmenge des Materials (ca. 50 %) wurde die Codierung geschärft. Anschließend wurden für Kategorien mit vielen Einheiten induktive Subkategorien gebildet, die als strukturgebendes Mittel das Material ordnen. Um die üppigen Inhalte von zwei Subkategorien zu sortieren, wurden diese noch einmal mittels einer weiteren Untergliederungsebene differenziert. Auch die Subkategorien wurden anhand der transkribierten Texte erprobt und komprimiert.⁵⁹

⁵⁹ Die Bemühungen um eine intersubjektive Nachvollziehbarkeit des Systems sind in Abschnitt 5.2.7 dargestellt. Zur inhaltsanalytischen Auswertung wurde mit dem Programm f4-analyse gearbeitet.

Tabelle 12: Beispiel aus dem erstellten Kategoriensystem zur Werkstatttratsarbeit: Hauptkategorie „Arbeitsformen, -inhalte und -erfolge“

Inhaltsbeschreibung	Bezieht sich auf konkrete Arbeitsformen, die mit der Amtsausführung als Werkstatt*rätin verbunden sind, Inhalte, die bearbeitet werden oder wurden, sowie genannte Erfolge.	
Subkategorien	Arbeitsformen	Konkrete Arbeitsinhalte
Inhaltsbeschreibung	Bezieht sich auf übergreifende Arbeitsformen des Werkstatttrats.	Beschreibt konkrete Arbeitsinhalte, die Werkstatttratsmitglieder bearbeiten oder durchgeführt haben.
Ankerbeispiele	„Was ich einfach oft erlebe, ist so dieses selbst/ dieses Organisieren von Werkstatttratsarbeit, dass das vielen sehr, sehr schwerfällt, das wirklich selbstständig zu machen, Versammlungen einzuberufen und an die Dinge zu denken, die da notwendig sind, Informationen weiterzugeben, zu verschriftlichen, das ist für viele einfach sehr schwer bei uns. (...)“ (Gruppe_1, Absatz 18)	„Wir haben zum Beispiel auch erreicht, letztes Jahr, Mitte letztes Jahr konnte man bei uns noch keine halben Tage Urlaub nehmen. Wir haben uns dafür eingesetzt, wir haben/ wir sind sozusagen dem Werkstattleiter ein bisschen auf den Nerv gegangen damit, weil wir gesagt haben ‚Wir wollen auch, so wie die normalen Mitarbeiter, wollen wir auch unsere halben Tage.‘“ (Gruppe_4, Absatz 37)
Abgrenzung zu anderen Hauptkategorien	<ul style="list-style-type: none"> • Beteiligte Stellen und Personen – Werkstatttrat: bezieht sich auf die Stärke des Gremiums, dessen Zusammensetzung sowie Werkstatttratsmitglieder als Einzelpersonen. • Erfahrungsbericht Werkstatttratsarbeit: Bezieht sich auf die Motivation zur Amtsausführung sowie die Arbeitsbewertung durch Werkstatttratsmitglieder oder allgemeine Erfahrungen/Eindrücke ohne Bezug zu einem bestimmten Arbeitsinhalt. 	

Kuckartz (2016, S. 100 ff.) beschreibt diesen Auswertungsvorgang als zirkulär verlaufenden Prozess, an dessen Ende die vollständige Materialcodierung steht. In diesem deduktiv-induktiven Prozess wurden letztlich 10 Haupt- und 25 Subkategorien (inklusive „Sonstiges“ und 6 Kategorien auf der dritten Differenzierungsebene) gebildet, die in den Diskussionen besonders prägnant und zur Beantwortung der forschungsleitenden Fragen geeignet schienen. Mit genauen inhaltlichen Beschreibungen, Anwendungshinweisen sowie Ankerbeispielen aus den Texten wurden diese als Kategoriensystem festgelegt. Tabelle 12 zeigt ein Beispiel aus dem erstellten Kategoriensystem. Mit diesem ausdifferenzierten Codierleitfaden wurde anschließend das komplette Material der Analyseeinheiten bearbeitet, wobei die Gruppen (nicht die teilnehmenden Personen) als Fälle definiert waren.

An die abgeschlossene vollständige Materialcodierung schloss sich die Ergebnisauswertung an. Gemäß der inhaltlich strukturierenden qualitativen Inhaltsanalyse stand die materialkomprimierte Themendarstellung und -analyse im Vordergrund. Die inhaltliche Ergebnispräsentation der Gruppendiskussionen ist in Kapitel 6 enthalten.

5.2.7 Qualitätskriterien und Reflexion der qualitativen Erhebungsphase

In der qualitativen Methodendiskussion gibt es keinen Konsens darüber, ob und wenn ja, welche Kriterien sich zur Beurteilung qualitativ hochwertiger Forschung eignen. Vielmehr liegen unterschiedliche Positionen und Bewertungskriterien vor, was sich mit dem unstandardisierten Anliegen des Forschungsstranges (Flick, 2017), der dortigen „Heterogenität methodologischer Positionen“ (Helsper, Kelle & Koller, 2016, S. 739) und der Kontextgebundenheit vorliegender Kriterien (Flick, 2019) begründen lässt. In der eigenen Arbeit wurde versucht, den Kernkriterien gerecht zu werden, die Steinke (2017, S. 323 f.) für qualitative Forschung skizziert. Das heißt, es wurde u. a. darauf geachtet, dass der Forschungsablauf personenunabhängig plausibel und nachvollziehbar und eine Ergebnisbeurteilung im Kontext der Studienbedingungen möglich ist. Dies wird als „intersubjektive Nachvollziehbarkeit“ (ebd., S. 324) beschrieben. Um diese zu erreichen, wurde die Vorgehensweise sorgfältig dokumentiert und die getroffenen Entscheidungen genau begründet (Kapitel 5.2.1 bis 5.2.6).

Gemäß dem Kriterium der „Indikation des Forschungsprozesses“ (Steinke, 2017, S. 326 ff.) lässt sich reflektieren, inwieweit die gewählte Erhebungsmethode der Gruppendiskussion und die angewandte Auswertungsmethode der qualitativen Inhaltsanalyse angemessen waren. Diese erwiesen sich als erkenntnis- und ergebnisreich. Die Ziele der qualitativen Erhebungsphase, relevante Bedingungen zur Arbeit von Werkstatträten aus Sicht der Beteiligten zu erheben, Vorstellungen und Faktoren einer gelingenden Werkstattratsarbeit zu ermitteln und Hypothesen über den Untersuchungsgegenstand zu entwickeln, konnten damit gut erreicht werden. Die methodischen Auswahlentscheidungen waren für das eigene Vorgehen zielführend. Nichtsdestotrotz sind mit ihnen Begrenzungen verbunden. Kritische Aspekte zur Methode der ermittelten Gruppendiskussion in der Auslegung Lamneks wurden bereits in Abschnitt 5.2.2 angesprochen. Eine gruppendynamische Analyse wurde, wie in dem genannten Abschnitt begründet, nicht vorgenommen, da ein solches Vorgehen im Hinblick auf die Ziele der Arbeit als weniger zentral und eher redundant eingeschätzt wurde. Stattdessen wurden die Ergebnisse mit weiteren Erhebungen verknüpft, was den Defiziten der Methode begegnen

kann (zu den Vor- und Nachteilen von Gruppendiskussionen siehe Lamnek, 1998, S. S. 65 ff.).

Die begründete Auswahl der qualitativen Inhaltsanalyse und des inhaltlich strukturierenden Verfahrens nach Kuckartz (2016) als gegenstandsangemessene Auswertungsmethode wurde in Abschnitt 5.2.5 expliziert. Im Hinblick auf die Gütekriterien, die für die qualitative Inhaltsanalyse benannt werden (siehe ebd., S. 201 ff.), besteht in der vorliegenden Arbeit eine Schwachstelle darin, dass eine Kategorienzuzuweisung durch verschiedene Personen nicht möglich war. Aufgrund fehlender Ressourcen konnte die Intercoder-Übereinstimmung nicht geprüft werden. Ersatzweise wurde vor dem Festzurren des Kategoriensystems und der endgültigen Materialcodierung der sogenannte „Intracoder-Vorgang“ praktiziert und zur Schärfung der Kategorien verwendet. Das heißt, im Abstand von ein paar Wochen wurden die codierten Textsegmente erneut betrachtet und überprüft (ebd., S. 207).

Die qualitative Inhaltsanalyse ist nicht darauf ausgerichtet, eine Analyse zweiten Grades zu ermöglichen und weitergehende Sinnstrukturen offenzulegen (Przyborski & Wohlrab-Sahr, 2014, S. 18 ff.). Im Gesamtkontext der Mixed-Methods-Studie kann die gewählte Vorgehensweise als angemessen gelten; für eine weitergehende Erforschung impliziter Bedeutungen wäre die Anwendung offenerer Methoden hingegen passender gewesen. So bleibt fraglich, ob die gewählte Auswertungsmethode dazu in der Lage ist, das Qualitätskriterium der „empirischen Verankerung“ (Steinke, 2017, S. 328), d. h. der datengesicherten Theorieprüfung, auszufüllen. Von Steinke wird sie unter den dazu geeigneten Methoden nicht gelistet. Demgegenüber konnte dem Aspekt, Forschung hinsichtlich der erreichten Relevanz zu beurteilen (ebd., S. 330), mit dem eigenen Vorgehen gut begegnet werden. Dass die untersuchte Forschungsthematik relevant ist, wurde auch durch diverse Nachfragen nach den Erhebungsergebnissen von Werkstatttratsmitgliedern, von Vertrauenspersonen, Werkstattleitungen und weiteren Praxisakteuren sowie durch gezielte Publikationsanfragen zur Thematik deutlich.

Bezogen auf das Merkmal der „Limitation“ (ebd., 329) diente die qualitative Erhebung wiederum bewusst als punktuelle Analyse und erste Annäherung an den Forschungsgegenstand. Hiermit konnte eine alltagsnahe Grundlage für die quantitative Erhebungsphase geschaffen werden; ein Generalisierungsanspruch wurde für die qualitativen Forschungsergebnisse nicht erhoben.

5.3 Quantitative Erhebungsphase

Dieses Unterkapitel widmet sich der Durchführung der quantitativen Erhebungsphase. Einleitend wird ein Überblick über das gewählte Design gegeben (Abschnitt 5.3.1). Anschließend werden die verschiedenen Komponenten einer quantitativen Erhebungsdurchführung von der Operationalisierung der Untersuchungsdimensionen (Abschnitt 5.3.2) über die Instrumentenerstellung (Abschnitt 5.3.3) bis zur Beschreibung der Pretests und der praktischen Durchführung (Abschnitt 5.3.4) schrittweise vorgestellt. Deziidiert betrachtet wird, inwieweit sich der intendierte Zugang der Befragung des Werkstatttrats als Gremium realisieren ließ (Abschnitt 5.3.5) und wie die Erhebungsrückläufe der drei Teilstudien ausgefallen sind (Abschnitt 5.3.6). Die Rückläufe werden im Hinblick auf zentrale Strukturmerkmale der teilnehmenden Werkstätten diskutiert (Abschnitt 5.3.7). Ausführlich wird auf die Aufbereitung der Daten eingegangen und offengelegt, welche Entscheidungen im Umgang mit den Daten getroffen wurden und wie die erstellten Indizes konstruiert sind (Abschnitt 5.3.8). Zuletzt wird das eigene Vorgehen im Hinblick auf die zentralen Gütekriterien quantitativer Forschung reflektiert (Abschnitt 5.3.9).

5.3.1 Forschungsdesign

Die Funktion der quantitativen Erhebungsphase liegt darin, die Forschungsergebnisse der qualitativen Erhebungsphase auf breiter Basis weiterzuverfolgen und zu überprüfen. Angestrebt war eine generalisierte Deskription der derzeitigen Arbeit von Werkstattträten und die Prüfung, ob bestimmte Faktoren im Zusammenhang mit unterschiedlichen Ausprägungen der Werkstatttratsarbeit stehen. Tabelle 13 fasst das erstellte Forschungsdesign zusammen.

Zur quantitativen Bearbeitung wurden die Ausgangsfragen der Mixed-Methods-Studie auf Basis der qualitativen Ergebnisse weiter spezifiziert. Unter der Dimension „Strukturmerkmale der Werkstatttratsarbeit“ (siehe Abschnitt 5.3.2) bezog sich die erste Frage „Wie wird Werkstatttratsarbeit organisiert und ausgestaltet?“ auf spezifische Aspekte der Gremienarbeit. Mit den Fragen 2 und 3 wurde auf drei auffällige Argumentationsfiguren der qualitativen Ergebnisse Bezug genommen, die des (in)aktiven, (un)autonomen und (wenig) durchsetzungsstarken Werkstatttrats. Die vierte Frage zielte darauf, abschließend zu prüfen, inwieweit geäußerte Zusammenhänge zwischen einer aktiven bzw. autonomen Werkstatttratsarbeit und

bestimmten Kontextfaktoren statistisch nachweisbar sind.⁶⁰ Die quantitative Erhebungsphase war im einmalig stattfindenden Querschnittsdesign (Häder, 2010, S. 115 ff.) als Ex-post-facto-Anordnung angelegt, wonach die Untersuchungsgruppe erst im Nachgang der Studie – ohne entsprechende Kontrollgruppe – im Hinblick auf vorhandene Unterschiede eingeteilt wurde (Schnell et al., 2013, S. 222 ff.).

Wie in der qualitativen Erhebungsphase standen als Zielpopulation Werkstatträte im Vordergrund, jedoch wurden auch die Perspektiven von Vertrauenspersonen und Leitungen erhoben. Insofern wurden diese Zielgruppen in drei separaten Teilstudien berücksichtigt.

Zur Befragung von Betriebsräten ist es gängig, Anschreiben an die Gremien zu richten und ein Mitglied des Betriebsrats oder speziell die Vorsitzenden stellvertretend zu befragen (siehe etwa Minssen & Riese, 2007; Müller-Jentsch & Seitz, 1998, S. 362). Mögliche Diskrepanzen in den Einschätzungen der Auskunftsperson zu den anderen Gremiumsmitgliedern werden i. d. R. nicht reflektiert. Um nicht nur Einzelmeinungen der vorsitzenden Personen zu erfassen, sondern die Einschätzungen als Gesamtgremium einzuholen, wurde der eigene Erhebungszugang auf Werkstatträte auf Kollektivebene ausgerichtet. Dies entsprach dem expliziten Wunsch von Werkstatratsmitgliedern im Vorfeld der quantitativen Erhebungsphase.⁶¹

⁶⁰ In Anlehnung an die ICF wurden unter Kontextfaktoren Komponenten der „materiellen, sozialen und einstellungsbezogenen Welt“ (WHO, 2005, S. 14) und personbezogene Faktoren verstanden, die Aktivitäten beeinflussen (ebd., S. 21f.).

⁶¹ Dieser Wunsch wurde vehement auf der Mitgliederversammlung von Werkstatträte Deutschland Anfang 2019 geäußert, bei der die Autorin ihr Forschungsvorhaben präsentierte. Die Funktionalität dieses Zugangs wurde im Rahmen unterschiedlicher Pretests (siehe Abschnitt 5.3.4) ausführlich erprobt und bewertet.

Tabelle 13: Forschungsdesign der quantitativen Erhebungsphase

Ziele	<ul style="list-style-type: none"> • Beschreibung der Umsetzung der formalisierten Beteiligung in WfbM (generalisierte Deskription). • Untersuchung von Merkmalen einer aktiven, autonomen und durchsetzungsstarken Werkstatttratsarbeit (generalisierte Deskription). • Ermittlung von begünstigenden Faktoren (konfirmatorische Analysen). • Berücksichtigung unterschiedlicher Perspektiven.
Forschungsfragen	<ul style="list-style-type: none"> • Wie wird Werkstatttratsarbeit organisiert und ausgestaltet? • Wie aktiv sind die Gremien? • Wie autonom und durchsetzungsstark sind diese? • Welche Zusammenhänge bestehen zwischen Kontextfaktoren und dem ermittelten Aktivitätsniveau bzw. der autonomen Arbeitsweise eines Werkstatttrats?
Erhebungsdesign	<ul style="list-style-type: none"> • Querschnittstudien, Ex-post-facto-Anordnung
Grundgesamtheit und Zielpopulationen	<ul style="list-style-type: none"> • Grundgesamtheit: Hauptwerkstätten der anerkannten WfbM in Deutschland • Befragung von drei Zielpopulationen: <ol style="list-style-type: none"> 1. Werkstattträte (Gremien) → <i>Teilstudie 1</i> 2. Vertrauenspersonen → <i>Teilstudie 2</i> 3. Werkstattleitungen/Geschäftsführungen → <i>Teilstudie 3</i>
Stichprobenbildung	<ul style="list-style-type: none"> • Vollerhebung auf Basis des Werkstättenverzeichnisses (BA, 2019)
Erhebungs-instrumente	<ul style="list-style-type: none"> • Standardisierte schriftliche Befragungen • Teilstudie 1: schriftliche Befragung in einfacher Sprache • Teilstudie 2 und 3: personalisierte Online-Umfragen
Durchführung ⁶²	<ul style="list-style-type: none"> • September – Oktober 2019 • Postalische Einladung
Auswertungsmethoden	<ul style="list-style-type: none"> • Deskriptiv- und Inferenzstatistik, bi- und multivariate Analysen • Indexkonstruktion, Typenbildung • Prüfung von Zusammenhangs- und Unterschiedshypothesen unter Zusammenführung der Teilstudien 1 und 2

⁶² Teilstudie 3 wurde in Kooperation mit Prof. Dr. Gudrun Wansing (Humboldt-Universität zu Berlin) und Jun-Prof. Dr. Mario Schreiner (Otto-von-Guericke-Universität Magdeburg) im Rahmen eines Projekts zur Untersuchung der reformierten Werkstätten-Mitwirkungsverordnung durchgeführt.

Alle drei Teilstudien wurden mit standardisierten schriftlichen Befragungen realisiert. Dieses „Standardinstrument empirischer Sozialforschung“ (Schnell et al., 2013, S. 314) zur Datengenerierung war in dem gegebenen Rahmen lediglich als schriftliches Verfahren möglich. Für die Befragung der Werkstatträte wurde ein Fragebogen in einfacher Sprache in Papierform erstellt. Die Befragungen der Vertrauenspersonen und Werkstattleitungen bzw. Geschäftsführungen (Teilstudien 2 und 3) waren als personalisierte Online-Befragungen konzeptualisiert.⁶³ Die dazu eingesetzten Zugriffscores zur Befragung waren zugleich ein Pseudonym, mit dem die drei Teilstudien pro Werkstatt zusammengeführt werden konnten, wenn für eine Einrichtung Angaben aus den verschiedenen Teilstudien vorlagen (siehe Abbildung 5).

Abbildung 5: Pseudonyme Zusammenführung der drei Teilstudien

Im „Verzeichnis anerkannter Werkstätten für behinderte Menschen“ mit Stand vom 29.05.2019 der BA sind 733 amtlich anerkannte WfbM gelistet. Diese Werkstätten setzen sich zum Großteil aus verschiedenen Betriebsstätten zusammen, wobei jeweils eine Hauptadresse

⁶³ Für Online-Studien werden anonyme Umfragen, an denen jeder teilnehmen kann, und personalisierte Umfragen unterschieden (Gräf, 2010, S. 16). Bei personalisierten Umfragen erhält jede angeschriebene Befragungsperson einen eindeutigen Zugang. Nur unter Eingabe des Codes kann auf die Befragung zugegriffen und an dieser teilgenommen werden, wobei auch Unterbrechungen und spätere Fortsetzungen möglich sind. Nur mit einem personalisierten Zugang kann der Erhebungsrücklauf kontrolliert werden.

(i. d. R. die Hauptwerkstatt) im Verzeichnis angegeben ist. Wie viele Werkstatttratsgremien in diesen Werkstätten bestehen, ist nicht bekannt. Eine Listung aller Werkstatttratsgremien an den Haupt- und Zweigwerkstätten in Deutschland schied somit aus. Um Werkstatttratsarbeit zu untersuchen und gleichzeitig die Aussagereichweite der Befragung kontrollieren zu können, richteten sich die Teilstudien deswegen ausschließlich an die Hauptwerkstätten aller anerkannten WfbM in Deutschland. Ziel war es, damit eine detaillierte Datenbasis zu erhalten. Angesichts des heterogenen Forschungsgegenstands bot sich der Verzicht auf eine Stichprobenziehung an (Schnell et al., 2013, S. 258). Werkstatttratsarbeit kann verschiedenste Ausprägungen erfahren. Allein verordnungsrechtlich sind drei unterschiedliche Ausprägungen möglich. Als Vollerhebung wurden alle Hauptwerkstätten angeschrieben und in die dargestellte dreiteilige Sample-Beziehung einbezogen.

5.3.2 Operationalisierung der Untersuchungsdimensionen

Zur Operationalisierung der im vorangegangenen Abschnitt dargestellten Fragen wurde mit vier zentralen Dimensionen gearbeitet.

Dimension Strukturmerkmale der Werkstatttratsarbeit

Diese Generaldimension zielte darauf, die Gremien hinsichtlich ihrer organisationalen Einbindung, Ausstattung und Abläufe zu beschreiben sowie Arbeitsinhalte, personelle Verhältnisse der Gremien und die (Arbeits-)Beziehungen zwischen den Beteiligten zu charakterisieren.

Dimension Aktivitäten des Werkstatttrats

Unter dieser Dimension wurden in der Erhebung zum einen Tätigkeiten des Gremiums gefasst, wie sie in den Verordnungsgrundlagen angelegt sind, sowie deren Quantität und Ausgestaltung unter Einbindung der weiteren Werkstattakteure. Zum anderen war hierunter die Wahrnehmung der gestaltenden Funktion der Mitbestimmung abgebildet – in Form von Komponenten, die in der Mitbestimmungsforschung als Ausdruck einer wahrnehmbaren oder effektiven Arbeitsweise gelten.

Dimension autonome Arbeitsweise

Individuelle Autonomie wird in der Behindertenbewegung mit der Anerkennung als autonom handelndes Subjekt und von Handlungsfähigkeit sowie der Forderung nach Selbstbestimmung verbunden (Waldschmidt, 2003). Neben dem Aspekt, dass Autonomie eine (gesellschaftlich)

anerkenkende Komponente beinhaltet, gehen in den Autonomiegedanken Überlegungen ein, die einen vom Subjekt ausgehenden aktiven Impetus betonen, was sich weitergehend in der Polarität von „Befreiung“ (ebd., 2012, S. 32) aus fremdbestimmten Zwängen und (neoliberaler) „Verpflichtung“ (ebd., S. 33) zur Selbstbestimmung kontrastiert.

In der vorliegenden Erhebung bezieht sich die Dimension „autonome Arbeitsweise“ auf den Werkstatttratt als Gremium und nicht auf die individuelle Autonomie einzelner Personen. Übernommen wurde die Polarität des Autonomiegedankens insofern, als in die Dimension „autonome Arbeitsweise“ zwei Komponenten eingingen: erstens strukturelle Bedingungen, die so gegeben sein müssen, dass Abhängigkeiten vermieden und die Gremien eigenständig arbeiten können; zweitens Handlungsweisen, die ein autonomes Agieren der Gremien aufzeigen, in dem Sinne, dass Inhalte und Abläufe der Gremienarbeit selbst bestimmt und eigene Meinungen geäußert oder durch auftretende Konflikte indiziert werden können, was in Form der intermediären Gremienposition (Müller-Jentsch, 1999, S. 9) tendenziell angelegt ist.

Dimension Durchsetzungsstärke

Ein Wesenskern, den Partizipationsbetrachtungen teilen, ist die Zielrichtung als einflussnehmendes Handeln. In Anlehnung an Kißler et al. (2011, S. 40) lässt sich dieses im betrieblichen Kontext als Interessendurchsetzung verstehen, die mit einer Verschiebung von Machtverhältnissen zugunsten der Arbeitnehmerseite verbunden ist. Im Sinne des „Wechsel[s] von einer ausgelieferten zu einer selbstbewussten Position“ (Baer, 2017, S. 223) wurden mit der Dimension „Durchsetzungsstärke“ Handlungen erfasst, die ein Anspruchswissen über die durchsetzbaren Rechte des Gremiums Werkstatttratt indizieren oder formale Verfahrensschritte zur Rechtsmobilisierung gemäß den Verordnungsgrundlagen abbilden. Ergänzt wurden diese Indikatoren durch subjektive Einschätzungen zu den vorhandenen Einflussmöglichkeiten des Werkstatttratts.

Subdimensionen

Die Operationalisierung beinhaltete zudem die folgenden zwölf Subdimensionen: Trägerstrukturen; Organisation der Gremien; Ausstattung der Gremienarbeit; Rahmenbedingungen der Arbeit und Arbeitszeitregelungen der Vertrauenspersonen; Arbeitsinhalte; Zusammensetzung des Gremiums; Weiterbildung; Kooperation und Vernetzung (intern; extern); Verhältnisse zwischen den Beteiligten; Interessenvertretung; Informationsqualitäten; Vermittlungsstelle.

Diese waren in der Erhebung ein Arbeitsmittel, um die Untersuchung zu strukturieren und einen Rahmen für die deskriptive Ergebnisdarstellung vorzugeben. Anzumerken ist, dass diese (Sub-)Dimensionen nicht immer trennscharf verlaufen. Als mehrdimensionale Komponenten des Konstrukts „Werkstattratsarbeit“ lassen sie sich theoretisch verschiedenen Bereichen zuordnen.

Operationalisierung der Dimensionen

Die Grundlage für die erstellten und verwendeten Fragen waren die qualitativen Erhebungsergebnisse, die mit Bezug zu vorhandenen Erkenntnissen und Annahmen aus der betrieblichen Mitbestimmungsforschung, der WMVO, der Rechtssoziologie und mit vorliegenden Praxisberichten operationalisiert wurden. Die vorherigen Ausführungen zur Partizipation bildeten hingegen keine gezielte Grundlage für die Operationalisierung, da sich diese – im Sinne des angewandten Mixed-Methods-Ansatzes – konsequent an den Forschungsfragen und den qualitativen Untersuchungsergebnissen orientierte, und in diesen der Partizipationsbegriff keine explizite Erwähnung fand. Die Auswahl zur Messung der latenten Sachverhalte erfolgte anhand der „kausal-analytische[n] Lösung“ (Schnell et al., 2013, S. 122), d. h. auf Basis behelfstheoretischer Annahmen über beobachtbare Indikatoren (manifeste Variablen), die als Ausprägungen latenter Variablen zu verstehen sind. Die Eingrenzung des „Itemuniversums“ (Guttman, 1950 nach Schnell et al., 2013, S. 123) wurde hypothesengeleitet, Annahmen überprüfend und im Hinblick auf methodische Aspekte zur Fragebogenkonstruktion vorgenommen,⁶⁴ im Fall der Teilstudie 1 zudem unter Berücksichtigung der Befragungsgruppe, sodass die Fragen in einfacher Sprache beantwortbar blieben.

Tabelle 14 zeigt zwei Beispiele aus dem vorgenommenen Operationalisierungsschema. Eine vollständige Auflistung zur Ableitung und Begründung der Indikatoren sowie ggf. deren Entlehnung ist im *Anhang 2 Operationalisierung der Fragestellungen* im elektronischen Zusatzmaterial enthalten. Geringfügig sind die Fragen aus vorhandenen Erhebungsinstrumenten entnommen oder Praxismaterialien entlehnt, um auf bewährte Formulierungen zurückzugreifen und Ergebnisse indirekt vergleichen zu können.

⁶⁴ Beispielsweise war die Frage „Wann war bei Ihnen die letzte Wahl zum Werkstattrat?“ (Teilstudie Werkstatträte) ein Anker und Referenzwert für verschiedene Fragen. So ist der letzte Wahlzeitpunkt laut WMVO zwar regelmäßig vorgegeben, jedoch kann es zu Abweichungen kommen, die in der Auswertung überprüft wurden. Zudem diente diese Frage auch dazu, auf die anschließenden Fragen zum Wahlüberhang und der -beteiligung hinzuweisen.

Tabelle 14: Beispiel aus der Operationalisierung

Dimensionen	Indikatoren	Begründung der Auswahl	Annahmen und Hypothesen	Sonstiges
Trägerstrukturen	Bewertung der Produktionsorientierung in WfBM	Gruppendiskussionen: Haltungen als zentrale Erklärung für Unterschiede; Mitbestimmungsforschung: „in produktionsorientierten WfB's hat der Werkstattrat den schwersten Stand“ (Breit & Kotthoff, 1990, S. 131).	Hypothese: Wenn die Produktionsorientierung hoch ist, dann hat der Werkstattrat ein niedrigeres Aktivitätsniveau / eine niedrigere autonome Arbeitsweise.	
Verhältnis zwischen den Beteiligten: Teamarbeit	Absprache zur Aufgabenteilung	Gruppendiskussionen: Teamarbeit als wichtiger Faktor; Mitbestimmungsforschung: Betriebsratsinterne Arbeitsorganisation hat hohe Bedeutung für die Aufgabenbewältigung (Minssen & Riese, 2007, S. 19).	Annahme: Nicht in allen Gremien vorhanden. Hypothese: Wenn eine Absprache zur Aufgabenteilung vorhanden ist, dann hat der Werkstattrat ein höheres Aktivitätsniveau / eine höhere autonome Arbeitsweise.	Frage orientiert an Bieneck, Engelmeyer & Hofmann, 2003, Nr. 12 Material 2.

5.3.3 Instrumentenerstellung und -beschreibung

Die Erhebungsinstrumente wurden anhand vorliegender Richtlinien für Fragebögen erstellt (z. B. Bortz & Döring, 2006, S. 255; Schnell et al., 2013, S. 352 ff.) und mit dem Fragebewertungssystem von Faulbaum, Prüfer und Rexroth (2009, S. 114 ff.) überprüft.⁶⁵ Beachtet wurde zudem die „Übersicht über spezifische Probleme und Gegenmaßnahmen in Interviews“ zur Befragung von Menschen mit sogenannten geistigen Behinderungen von Schäfers (2008, S. 179).

Die Skalengestaltung folgte den empirischen Empfehlungen von Menold und Bogner (2015), die den methodologischen Forschungsstand im Themenbereich bündeln. Demgemäß wurde u. a. eine Mittelkategorie und i. d. R. eine Skala mit fünf Abstufungen in horizontaler Ausrichtung verwendet. Die Skalenabstufungen wurden (voll-)verbalisiert, was mit einer größeren Alltagsnähe assoziiert wird und eine leichtere Beantwortung ermöglichen kann

⁶⁵ Hierbei wurde das Bewertungssystem als Hilfsmittel interpretiert, um die Fragequalität zu verbessern. Eine sehr akribische Handhabung der Checkliste wäre dem Untersuchungsgegenstand sowie einer einfachen Instrumentenhandhabung nicht gerecht geworden.

(Faulbaum et al., 2009, S. 68). Auf die eher abstrakte numerische Bezifferung (Menold & Bogner, 2015, S. 3) wurde verzichtet. Um die Differenzierungsfähigkeit der Befragungsgruppe mit kognitiven Beeinträchtigungen nicht zu überlasten (siehe Schäfers, 2008, S. 179), wurden überwiegend fünfstufige Ratingskalen verwendet. Hierzu wurde auf die Abstufungen *nie – selten – gelegentlich – oft – immer* (zur Abfrage von Häufigkeiten) und *gar nicht – kaum – mittelmäßig – ziemlich – außerordentlich* (zur Abfrage von Intensitäten) zurückgegriffen, die als weitgehend äquidistant gelten (Bortz & Döring, 2006, S. 177). Da sich zur standardisierten Befragung in Einrichtungen der Behindertenhilfe jedoch die mittlere Kategorie *manchmal* bewährt hat (Schäfers, Schachler, Schneekloth, Wacker & Zeiler, 2016, S. 51), wurde diese anstatt *gelegentlich* verwendet und die Abstufung *außerordentlich* in *sehr* abgeändert, was im Sprachbezug der Formulierung für die Zielgruppe einfacher erschien.

Bei den Items des Likert-Typs wurde der empirischen Empfehlung gefolgt, auf Zustimmungsfragen zu verzichten und stattdessen direkte Einschätzungsskalen zu verwenden (Menold & Bogner, 2015, S. 8). Es wurden positiv und negativ gepolte Items verwendet, was eine Technik ist, um inhaltsunabhängigen Antworttendenzen entgegenzuwirken (siehe etwa Schnell et al., 2013, S. 177 f.; Schäfers 2008, S. 199).

Die Antwortoptionen wurden überwiegend im geschlossenen oder gemischten Format gehalten. Ergänzend offene Fragen ermöglichten ungebundene Akzentsetzungen.

Inhaltlich tangierten die Fragebögen Sachverhalte der Gremienarbeit, deren Gegebenheiten nicht immer als bekannt vorausgesetzt werden konnten. Die Antwortkategorie *weiß nicht* bot deswegen für weniger alltägliche Fragen eine Ausweichmöglichkeit (Schnell et al., 2013, S. 328; zum Einsatz von nicht-inhaltlichen Kategorien siehe Menold & Bogner, 2015, S. 6 ff.). Diese Auswahlmöglichkeit konnte zudem Auskunft darüber geben, inwieweit die formalen Gremienstrukturen unter den Befragten (nicht) bekannt sind.

Erhebungsinstrument Teilstudie I (Werkstatträte)

Sprachlich war im Hinblick auf die Mitglieder der Werkstatträte intendiert, dass sich die Fragen und Antwortoptionen auf einem leicht verständlichen Niveau ohne komplexe Satzkonstruktionen bewegen. Auf den Einsatz von bildlichen Antworthilfen und ergänzende Bilder, wie sie zur schriftlichen Kommunikation in Leichter Sprache angewandt und empfohlen werden (Netzwerk Leichte Sprache, 2014, S. 53), wurde bewusst verzichtet. So sind diese in Erhebungsinstrumenten per se nicht immer unterstützend, z. B. wenn der Frageinhalt und die Ikonizität einer optischen Antworthilfe kontextspezifisch divergieren und durch einen

vorliegenden Interpretationsspielraum (Schäfers, 2008, S. 157) mögliche Messfehler evozieren. Linguistische Erörterungen verweisen unter dem sogenannten „Split-Attention-Effekt“ darauf, dass ein Leseverständnis durch gleichzeitige Text- und Bildinformationen sinken kann (Bredel & Maaß, 2016, S. 282 f.). Statt Bilder zu verwenden, wurde der Bogen sprachlich möglichst einfach formuliert und einem ausführlichen Pretesting unterzogen. Dies berücksichtigte den Spagat der Befragung, zum einen Menschen mit kognitiven Beeinträchtigungen zu erreichen, zum anderen den Personenkreis mit psychischen oder körperlichen Beeinträchtigungen adäquat anzusprechen, die sich durch eine einfache Sprachwahl mitunter despektierlich behandelt fühlen.

Im Ergebnis umfasste das Erhebungsinstrument 44 Fragen, die sich auf einem zwölfseitigen Fragebogen verteilten. Inhaltlich waren in dem Instrument Fragen zu den Trägerstrukturen, den Werkstattratswahlen, den Arbeitsressourcen, den Aktivitäten des Gremiums, der internen Zusammenarbeit und Haltungen, den Unterrichtsrechten sowie zum Auftreten von Konflikten enthalten. Auf dem Titelblatt des Erhebungsinstrument waren das Logo und ein Unterstützungshinweise von Werkstattträte Deutschland e. V. integriert. Mit einem Erhebungsanschreiben wurden die Befragten gemäß den Bestimmungen der EU-Datenschutz-Grundverordnung (EU-DSGVO) über die Zwecke der Erhebungsdurchführung, die Verwendung der Daten (Artikel 13 EU-DSGVO) etc. informiert. Das Einverständnis der Befragten zur Verwendung der Angaben wurde eingeholt und das nachträgliche Widerspruchsrecht (Artikel 21 EU-DSGVO) aufgezeigt. Das Anschreiben der Teilstudie 1 war in einfacher Sprache formuliert. Da das Erhebungsinstrument für sehbeeinträchtigte Personen oder bei eingeschränkter Lesekompetenz nicht barrierefrei war, wurde eine Telefonnummer angegeben, unter der auch mündlich teilgenommen werden konnte. Dem Anschreiben war ein Anweisungsbogen mit ergänzenden Informationen zum Ausfüllen beigelegt.

Erhebungsinstrumente Teilstudien 2 und 3 (Vertrauenspersonen und Werkstattleitungen)

Die Fragebögen umfassten mit allen unterschiedlichen Filterführungen 45 (Teilstudie 2) bzw. 52 Fragen (Teilstudie 3). Die tatsächliche Anzahl der Fragen variierte je nach Verlauf des Interviews und fiel geringer aus. In der Teilstudie 3 waren 21 Fragen Inhalt des kooperativ beteiligten Projekts zur Umsetzung der reformierten Werkstätten-Mitwirkungsverordnung (siehe Fußnote 62). Diese wurden für die vorliegende Arbeit nicht verwendet. Thematisch betrafen diese die Finanzierung der Werkstattratsarbeit und die Frauenbeauftragte nach § 39a WMVO.

Für die eigene Teilstudie wurden an die Leitungen Fragen zu den Einrichtungsstrukturen, zur Umsetzung und Ausgestaltung der Mitwirkungs(ver)ordnungen und zu eingerichteten Vermittlungsstellen gestellt. Zudem wurden Einschätzungen über die Gremien erbeten. Ergänzend dazu lag der inhaltliche Schwerpunkt der Befragung 2 auf den Rahmenbedingungen der Arbeit als Vertrauensperson, der Arbeitsweise und der Zusammensetzung des Werkstattrats und dessen Ausstattung bzw. Einbindung in die Werkstatt. Auch hier wurden Einschätzungen erbeten.

Durchgeführt wurden die Online-Befragungen mit dem Befragungsserver www.soscisurvey.de, mit dem sich die Fragebogengestaltung mittels vorgegebener Funktionen und einfacher PHP-Skriptsprache realisieren ließ.

5.3.4 Pretests und Erhebungsdurchführung

Die Vorgaben zu einer angemessenen Fallzahl der probeweisen Bearbeitung von Fragebögen im Vorfeld einer Befragung variieren (etwa Schnell et al., 2013, S. 342). Sie sind abhängig von den gegebenen Erhebungsumständen und eine zentrale Qualitäts- und eine Ressourcenfrage. Um die neu konstruierten Erhebungsinstrumente zu verbessern, den Theoriegehalt der gewählten Indikatoren zu überprüfen (Häder, 2010, S. 377) und den intendierten Zugang der Gruppenbefragung zu erproben, wurden verschiedene Pretestverfahren angewandt. Zeitlich wurde dazu ein viermonatiger Zeitraum verwendet, da sich durch die Gremienstrukturen und die Zielgruppe der Werkstattkräfte lange Kommunikationswege ergaben.

Vertiefter Pretest / kognitives Verfahren in der Teilstudie 1 (Werkstattkräfte)

Als vertiefter Pretest wurden mit drei Werkstattratsgremien kognitive Verfahren durchgeführt (Werkstatt I: vier Personen mit psychischen Beeinträchtigungen und deren Vertrauensperson; Werkstatt II: drei Mitglieder des Werkstattrats mit kognitiven Beeinträchtigungen und deren Vertrauensperson; Werkstatt III: vier Personen mit verschiedenen Beeinträchtigungen und zwei Vertrauenspersonen). Hierbei handelte es sich um mehrstündige Termine, bei denen das Erhebungsinstrument den Mitgliedern des Werkstattrats persönlich vor Ort in den Werkstätten übergeben und um eine Sichtung und Bearbeitung der Erhebungsunterlagen gebeten wurde. Durch die Besonderheit der Befragung als Gremium war ein diskursives Aushandeln der Fragen und der passenden Antwortoption notwendig, sodass in den Argumentationen bereits deutlich wurde, was die Befragten mit den Fragen assoziierten und womit sie ihre

Antwort begründeten. Vergleichbar mit einem kognitiven Pretest (Häder, 2010, S. 393 ff.) verdeutlichte dies Abweichungen von den Intentionen der Fragen, fehlende Präzisierungen, vorhandene Verständnisprobleme und Schwierigkeiten bei der Antwortfindung. Zusätzlich wurden durch die Forscherin im Sinne der „Probing- oder Nachfragetechnik“ (ebd., S. 394) Ergänzungsfragen gestellt, um das Frage- und Antwortverständnis weiter zu ergründen. Die Teilnehmenden äußerten vorhandene Schwierigkeiten auch direkt und schlugen mögliche alternative Formulierungen im Sinne eines „Paraphrasings“ (ebd.) vor. Bei der Befragungsorts als Gruppe zeigten sich keine offensichtlichen Schwierigkeiten: Vorhandene Meinungsdiskrepanzen oder differierende Wissensbestände wurden begründet oder kollegial – mitunter nach dem Abstimmungsprinzip – ausgehandelt. Zudem wurde ein Weiterbildungseffekt durch die Fragen des Fragebogens positiv hervorgehoben.

Expertenbewertung und Standard-Pretest in der Teilstudie 1 (Werkstatträte)

In einem frühen Stadium wurde das Erhebungsinstrument in einer Telefonkonferenz mit einem Teil des Vorstands und der Assistenz von Werkstatträte Deutschland e. V. diskutiert und Verbesserungshinweise gegeben. Unter den verschiedenen Pretestformen wird diese Strategie als Expertenbewertung (Häder, 2010, S. 397) aufgeführt. Eine zweite inhaltliche Expertise wurde in einem späteren Stadium in Form eines persönlichen Termins mit einem ehemaligen Vorstandsmitglied von Werkstatträte Deutschland e. V. eingeholt. Hierbei wurden das finale Wording sowie der Ein- bzw. Ausschluss einzelner Fragen auf Grundlage der vermuteten Alltagsrelevanz bewertet.

Als dritte Möglichkeit zur Form der Instrumentenoptimierung wurde ein Standard-Pretest durchgeführt, bei dem die Erhebungsunterlagen an Werkstatträte verschickt und ein probeweises Ausfüllen und Kommentierung von Unklarheiten erbeten wurde. Bei dieser klassischen Pretestform gingen neun Rückmeldungen ein. Die Teilnehmenden wurden in einem Vortrag der Autorin, durch persönliche Kontakte sowie durch spontane Anfragen in Werkstätten akquiriert. Es wurden unterschiedliche Bundesländer (Bayern, Berlin, Niedersachsen, NRW, Rheinland-Pfalz, Schleswig-Holstein und Thüringen), Personenkreise, Einrichtungsträger, -größen und Verordnungsgrundlagen einbezogen.

Im Vergleich der Bögen, die von Einzelpersonen ausgefüllt worden waren, mit denjenigen, die im Gremium beantwortet waren, wiesen die singular beantworteten Bögen deutlich geringere Antwortquoten auf. Beim Standard-Pretest wurde zusätzlich telefonisch nacherfasst, wie die Befragung im Gremium und die Antwortfindung verlief. Hierbei wurde eine kritische Meinung geäußert, die den hohen Zeitaufwand bei der gemeinsamen Beantwortung betraf. Im

Hinblick auf die anderen Pretestergebnisse wurde dieser Befragungszugang dennoch angewandt, da die Adressierung an den Werkstatttrat als Gremium dem Charakter des interessierenden Forschungsgegenstands entsprach.

Pretest in den Teilstudien 2 und 3 (Vertrauenspersonen und Werkstattleitungen)

Auch in den Teilstudien der Vertrauenspersonen und Werkstattleitungen wurden die skizzierten Pretestverfahren angewandt. Beide Instrumente wurden als Expertenbewertung mit thematisch bewanderten Personen diskutiert und durch diese kommentiert. Vertiefte Pretests bzw. kognitive Verfahren wurden mit einer Vertrauensperson und zwei Werkstattleitungen durchgeführt.

Die 16 Befragungen des Standard-Pretests der Werkstattleitungen konnten in 12 verschiedenen Bundesländern realisiert werden. Bis auf Thüringen, Niedersachsen und Sachsen beteiligten sich Werkstattleitungen bzw. Geschäftsführungen aus allen Bundesländern daran (in Hamburg wurde kein Pretest durchgeführt, da hier nur zwei anerkannte WfbM bestehen).

Bei den 14 Pretest-Interviews der Vertrauenspersonen wurde eine große Unterstützungsbereitschaft deutlich. So wurde erfreut geäußert, sich endlich einmal mit jemandem über die Arbeit austauschen zu können, es erfolgten umgehende Bearbeitungen der Fragebögen und die Vertrauenspersonen akquirierten von sich aus weitere Kolleg*innen für das Pretesting.

Maßnahmen für eine hohe Erhebungsteilnahme

Um eine hohe Erhebungsbeteiligung zu generieren, wurden verschiedene Maßnahmen angewandt. Im Vorfeld wurde das Forschungsvorhaben in Gremien- und Veranstaltungszusammenhängen bekannt gegeben (Vorstellung auf der Mitgliederversammlung von Werkstattträte Deutschland am 05.02.2019; Vortrag auf der Werkstätten:Messe 2019), ein wissenschaftlicher Beitrag mit Verweis auf die Erhebungen wurde in der Mitgliederzeitschrift der BAG WfbM publiziert (Schachler, 2019), Erhebungsinformationen über die Verbände gestreut und zur Beteiligung aufgerufen (Newsletter Anthropoi Bundesverband; Newsletter der Bundesvereinigung Lebenshilfe; Homepage BAG WfbM; Homepage Werkstattträte Deutschland e. V.). Zudem wurden die LAGen der Werkstätten und der Werkstattträte in den 16 Bundesländern vorab informatorisch angeschrieben, was diese wiederum teilweise an ihre Mitglieder weitertrugen.

Erhebungsdurchführung

Der Zeitpunkt der Durchführung war auf September 2019 terminiert. Zu diesem Zeitpunkt befanden sich die Werkstattträte in der Mitte ihrer Amtszeit seit den letzten regulären Wahlen.

Alle im Werkstättenverzeichnis 2019 der BA aufgeführten Hauptwerkstätten erhielten postalisch die folgenden Erhebungsunterlagen:

- Anschreiben zur Online-Teilnahme für Leitungen mit individuellen Zugangsdaten;
- verschlossener Umschlag mit einem Anschreiben an den Werkstattrat, einem Anweisungsbogen zum Ausfüllen, einem Fragebogen sowie einem Rücksendeumschlag;
- separat verschlossener Umschlag mit einem Anschreiben zur Online-Teilnahme für Vertrauenspersonen mit individuellen Zugangsdaten.

Adressiert waren die Erhebungsunterlagen an die Geschäftsführung oder Werkstattleitung (bei mehreren Betriebsstätten: an die Geschäftsführung/Leitung der Hauptwerkstatt). Diese wurden gebeten, die verschlossenen Umschläge an den Werkstattrat und die Vertrauensperson weiterzugeben (bei mehreren Werkstattratsgremien an den Werkstattrat der Hauptwerkstatt und an dessen Vertrauensperson). Dieser gebündelte Versand wurde zum einen aus Kostengründen gewählt. Zum anderen berichteten Werkstattratmitglieder vorab, dass sie über den direkten Postweg nicht zu erreichen sind und die Post verschwinde.

In den Erhebungsunterlagen an den Werkstattrat und die Vertrauenspersonen waren genaue Anweisungen gegeben, an wen sich die Befragung richtete, wenn mehrere Werkstattratsgremien oder Vertrauenspersonen bestehen. Durch die Codierung der Bögen war sichergestellt, dass die Teilnahme für jede anerkannte Werkstatt jeweils nur einmal gezählt wurde.

Mit verschiedenen Nachfassaktionen wurde an die Erhebung erinnert und die Erhebungsfristen bis Ende Oktober 2019 verlängert. Vereinzelt gingen Fragebögen von Werkstatträten noch im November (11 Bögen) und Dezember 2019 (1 Bogen) ein.

In der Teilstudie 1 wurden – sofern eine Absenderadresse angegeben war – in geringem Umfang Nacherfassungen durchgeführt, wenn ganze Seiten im Fragebogen übersprungen waren, was auf ein versehentliches Überblättern der Seite hindeutete.

5.3.5 Die Gremien als Auskunft gebende Befragungsgruppe

Teilstudie 1 war auf die Befragung von Werkstattratmitgliedern auf Kollektivebene ausgerichtet.⁶⁶ Tabelle 15 zeigt, wie sich dieser Zugang realisieren ließ. Am häufigsten wurden die

⁶⁶ In den begleitenden Informationen zum Ausfüllen war dies wie folgt beschrieben: „Wer soll die Fragen beantworten? Am besten der Werkstattrat als Gremium. Das heißt als Gruppe. Zum Beispiel auf einer Werkstattrats-Sitzung. Wenn das nicht geht, soll der Vorsitzende oder ein Mitglied des Werkstattrats das übernehmen.“

Erhebungsfragen durch den Werkstatttrat als Gremium beantwortet, mit 18% folgt an zweiter Stelle die (stellvertretende) vorsitzende Person des Werkstatttrats. In geringem Umfang stammen die Angaben von nicht intendierten Personen:

- in drei Fällen von Gesamtwerkstattträten, die zusätzlich zu Zweigwerkstattträten bestehen. Diese Fragebögen wurden als ungültige Fälle behandelt und von der Auswertung ausgeschlossen.
- Zwei Befragungen wurden von Vertrauenspersonen beantwortet. Deren Angaben zur Arbeit des Werkstatttrats wurden verwendet, da sich diese auf die Grundgesamtheit der Erhebung (Werkstattträge an den Hauptstandorten) bezogen. Ausgeschlossen wurden die Angaben bei den subjektiven Bewertungsfragen zur Zusammenarbeit mit Vertrauenspersonen.

Tabelle 15: Die Fragen beantwortende Personen der Teilstudie 1

	<i>n</i>	In Prozent
Der ganze Werkstatttrat als Gremium	206	60 %
Nur der*die (stellvertretende) Vorsitzende des Werkstatttrats	60	18 %
Ein Teil des Werkstatttrats als Gremium	25	7 %
Der*die Vorsitzende und der*die stell. Vorsitzende oder die Vorsitzenden oder ein/zwei Mitglieder aus dem Werkstatttrat	26	8 %
Nur ein Mitglied aus dem Werkstatttrat	12	4 %
Der ganze Werkstatttrat als Gremium und die Frauenbeauftragte	6	2 %
Sonstige Personen (darunter zwei Vertrauenspersonen)	6	2 %
Keine Angabe	3	1 %
Gesamt	344	100 %
Der Gesamtwerkstatttrat	3	–

Anmerkungen. Datenquelle Teilstudie 1; Frageformulierung „Wer hat die Fragen dieser Umfrage beantwortet?“.

Gemäß dem beabsichtigten Zugang ist ein Großteil der Fragebögen von mehreren Werkstatttratsmitgliedern beantwortet worden. Um mögliche Unstimmigkeiten bei der Beantwortung zu berücksichtigen, war in dem Anweisungsbogen zur Erhebung ein Hinweis enthalten,

wie mit Meinungsdivergenzen umzugehen sei.⁶⁷ Von der Möglichkeit, gegensätzliche Auffassungen zu vermerken, machten lediglich drei Gremien Gebrauch.⁶⁸ Die Anmerkungen verweisen darauf, dass der gremienspezifische Zugang gemeistert werden konnte und die Erhebungsfragen als schwierig erlebt wurden. Dementsprechend haben 76 % der befragten Gremien Hilfestellung durch eine Unterstützungsperson beim Beantworten der Umfrage bekommen.⁶⁹ In den erfragten Kategorien bezieht sich die Unterstützung zu je zwei Dritteln der Fälle auf zusätzliche Informationen (67 %) und auf eine Hilfestellung beim Antworten (66 %). Ein Werkstattrat (0,4 %) gab an, dass die Unterstützungsperson die Antworten bestimmt hat. In 13 % der Fälle wird auf eine sonstige Unterstützung verwiesen (etwa vorgelesen, $n=8$; geschrieben, $n=8$; in Leichter Sprache erklärt / frei formuliert, $n=5$).

Insgesamt wurden die Bögen sorgfältig ausgefüllt. Bei den fehlenden Werten (*weiß nicht*, keine Angabe) fällt auf, dass diese bei der Beantwortung als Gremium seltener sind als bei der Ausfüllung des Bogens durch Einzelpersonen. Angesichts der durchaus anspruchsvollen Frageinhalte, die mit einem Wissensstand als Gruppe einfacher beantwortet werden können, kann der Zugang als gelungen bezeichnet werden. Zur Vermeidung von Missings hat sich der gremienbezogene Ansatz zudem besonders bewährt.

5.3.6 Erhebungsrückläufe und Ausschöpfungsquoten

Für drei WfbM wurde von den angeschriebenen Leitungen/Geschäftsführungen zurückgemeldet, dass dort kein Werkstattrat vorhanden war.⁷⁰ In diesen Fällen handelte es sich um stichprobenneutrale Ausfälle, die von der Brutto-Ausgangsstichprobe subtrahiert werden konnten,

⁶⁷ „Haben Sie verschiedene Meinungen über die Antwort auf eine Frage? Dann diskutieren Sie und stimmen Sie ab, welche Antwort passt. Wenn Sie sich nicht einig werden, können Sie am Ende angeben, welche Meinungen es gab.“

⁶⁸ Einmal ist generell geäußert „Die Fragen sind nicht immer leicht verständlich“ (WR_Case 291). Zweimal sind zu einzelnen Fragen spezifische Anmerkungen gegeben. Zur Frage „Wie ernst nehmen die Gruppenleiter den Werkstattrat?“ ist angemerkt: „Die Aussagen der WRat-Mitglieder wichen voneinander ab. Wir haben einen Mittelwert ermittelt“ (WR_Case 330) und bei den Fragen zu den Informationen: „hier herrscht Uneinigkeit von oft bis selten“ (WR_Case 280).

⁶⁹ Datenquelle Teilstudie 1; Frageformulierung „Hat eine Unterstützungsperson beim Beantworten geholfen?“; gültige Angaben von 339 Gremien; Folgefrage: „Die Unterstützungsperson hat ...“; Mehrfachnennungen möglich; gültige Angaben von 247 Gremien.

⁷⁰ In einem Fall hatte sich der Werkstattrat gerade aufgelöst, in zwei Fällen werden die Aufgaben durch andere Gremien abgedeckt. So lagen nach Angabe des Arbeitgebers in einer WfbM Arbeitsverhältnisse mit Gültigkeit des BetrVG vor.

womit die bereinigte Stichprobe aus 730 WfbM bestand. In Anlehnung an Porst (1985) wurden zur Berechnung der Ausschöpfungsquoten lediglich die vollständig ausgefüllten und auswertbaren Fragebögen verwendet.⁷¹ Tabelle 16 stellt die Anzahl der verwertbaren Fragebögen der Teilstudien dar. Wie im vorherigen Abschnitt dargestellt, waren in drei Fällen der Teilstudie 1 die Fragebögen von übergreifenden Gesamtwerkstatträgern ausgefüllt worden, die von der Auswertung ausgeschlossen wurden. Die Ausschöpfungsquote von Teilstudie 2 ließ sich näherungsweise über eine Angabe aus Teilstudie 1 rekonstruieren.⁷² So benannten in dieser Teilstudie fünf Werkstatttratsgremien (1.5 %), dass sie ohne Vertrauenspersonen arbeiten (siehe Abschnitt 7.1.1.11), womit sich die bereinigte Stichprobe für Teilstudie 2 schätzen lässt.

Tabelle 16: Rückläufe der Teilstudien

Teilstudie	Versickte Fragebögen	Bereinigte Stichprobe	Anzahl verwertbare Fragebögen	Ausschöpfungsquote
1 – Werkstattträge	733	730	344	47 %
2 – Vertrauenspersonen	733	719*	230	32 %*
3 – Werkstattleitungen	733	730	140	19 %

Anmerkung. * Schätzwert auf Basis der Angabe „es gibt keine Vertrauensperson“ der Teilstudie 1.

Mit Ausschöpfungsquoten von 19 % bis 47 % (bzw. Ausfallraten von 81 % bis 53 %) fallen die Rückläufe sehr unterschiedlich aus. Hohe Ausfallraten sind in Befragungen der Regelfall (Schnell et al., 2013, S. 300 f.). Insbesondere für Online-Befragungen sind gute Rücklaufzahlen schwer zu erzielen (Maurer & Jandura, 2009).⁷³ Für die eigene Erhebung wurde über

⁷¹ Nach Porst (1985, S. 91) lässt sich die Ausschöpfungsquote wie folgt berechnen:

$$\text{Ausschöpfungsquote} = \frac{\text{Bereinigte Stichprobe} - \text{Ausfälle} - \text{nicht auswertbare Interviews}}{\text{Bereinigte Stichprobe}} \times 100$$

⁷² Nach § 39 Abs. 3 WMVO besteht ein Anspruch des Werkstatttrats auf eine unterstützende Vertrauensperson, jedoch keine Verpflichtung, diese zu beanspruchen. Auch Personalmangel oder andere Gründe können dazu führen, dass nicht für jeden Werkstatttrat Vertrauenspersonen tätig sind.

⁷³ Gegenüber persönlichen und postalischen Interviews sind hier weitere Hürden zu nehmen, die zu Lasten der Beteiligungsbereitschaft gehen (Bandilla, Kacmirek, Blohm, & Neubarth, 2009, S. 134; Maurer & Jandura, 2009, S. 67). Nach einer Metaanalyse methodenkomparativer Studien, die zwischen 1998 und 2006 englischsprachig publiziert wurden, erzielten die 39 betrachteten Umfragen im Mittel Rückläufe von 45 % in der schriftlich-postalischen Befragung ($SD = 0.22$) im Vergleich zu 34 % ($SD = 20$) bei der Online-Version (Shih & Fan, 2008, S. 257 ff.). Für Online-Befragungen, die auf echten Zufallsstichproben beruhen, werden Werte von

Schwierigkeiten bei der Zustellung der postalischen Unterlagen innerhalb der Werkstätten berichtet: Vermeintlich waren die Erhebungsunterlagen bei Leitungen und Werkstattträten nicht angekommen. Teilweise waren diese nach gestarteten Suchaktionen innerhalb der Einrichtungen noch auffindbar oder bereits durch andere Personen ausgefüllt worden.⁷⁴ So gingen erhebungsbereite Personen mutmaßlich verloren. Ein systematischer Ausschluss ist jedoch nicht erkennbar. Im Vergleich zu älteren Studien, die in Werkstätten durchgeführt wurden (Breit & Kotthoff, 1990; Detmar et al., 2008), fällt der Rücklauf der Teilstudie 3 gering aus. Allerdings lagen sehr unterschiedliche Studienbedingungen vor: etwa ein zwischenzeitlich sensiblerer Umgang mit Befragungsdaten gemäß der EU-DSGVO und im Herbst 2019 parallel laufende Befragungen im Themenbereich der Werkstätten, womit eine gewisse Befragungsermüdung einhergehen kann. Explizit begründet wurde das Nonresponse-Verhalten auch mit einer hohen zusätzlichen Arbeitsbelastung durch die Änderungen des Bundesteilhabegesetzes.

Durch die pseudonyme Codierung der Befragungen konnte die Erhebungsbeteiligung vollständig kontrolliert und die verschiedenen Teilstudien pro anerkannter Werkstatt zusammengeführt werden. Tabelle 17 zeigt, wie viele Fragebögen jeweils in den verschiedenen Kombinationsmöglichkeiten der Teilstudien vorhanden sind. Insgesamt liegen für 420 der anerkannten WfbM in mindestens einer der drei Teilstudien Informationen vor (58 % der Grundgesamtheit). Die größten Überschneidungen sind für die Teilstudien 1 und 2 (Werkstattträte und Vertrauenspersonen) gegeben. Hier lassen sich für 182 Werkstätten die Angaben zusammenführen. Demgegenüber sind es lediglich 81 WfbM, in denen sich die Leitung und die Vertrauensperson beteiligten. In der dreifachen und kompletten Kombination der Erhebungsrückläufe der Teilstudien 1 bis 3 liegen schließlich nur für 63 WfbM Angaben vor. Auch singuläre Beteiligungen in lediglich einer der drei Teilstudien sind zu konstatieren.

weniger als 5 % für schlecht verlaufende Access-Panels oder Rückläufe von 50–80 % bei kleiner Grundgesamtheit mit hoher Internetanbindung benannt (Maurer & Jandura, 2009, S. 66).

⁷⁴ Aus Datenschutzgründen war nicht dokumentiert, welche Werkstatt welches Pseudonym erhalten hatte, sodass die Unterlagen nicht erneut zur Verfügung gestellt werden konnten. Um den Erhebungszugang (je eine Teilnahme pro Teilstudie für jede anerkannte Werkstatt) nicht zu verfälschen, wurden Ersatzunterlagen nicht ungeprüft herausgegeben. Lediglich in zwei Fällen wurden Werkstattleitungen Ersatzcodes übermittelt, da die Erhebungsunterlagen verschollen blieben. Für eine Werkstatt wurden die kompletten Erhebungsunterlagen noch einmal zur Verfügung gestellt. Rückläufe aufgrund einer fehlerhaften Adresse gab es nicht, nur in einem Fall konnten die Unterlagen erst nach einer Korrektur der Adresse des Werkstättenverzeichnisses zugestellt werden.

Tabelle 17: Kombinierte Erhebungsbeteiligungen in den Teilstudien (verwertbare Bögen)

Teilstudie/n	Beteiligte WfbM
1, 2 oder 3 – Werkstattträte, Vertrauenspersonen oder Werkstattleitungen	420
1 und 2 – Werkstattträte und Vertrauenspersonen	182
1 und 3 – Werkstattträte und Werkstattleitungen	93
2 und 3 – Vertrauenspersonen und Werkstattleitungen	81
1, 2 und 3 – Werkstattträte, Vertrauenspersonen und Werkstattleitungen	63

5.3.7 Strukturvariablen der teilnehmenden Werkstätten (Teilstudien 1 und 3)

Da bei Ausschöpfungsquoten ein „mehr“ nicht gleichbedeutend mit „besser“ ist“ (Koch, 1998) und die Höhe eines erzielten Rücklaufs noch keinen Rückschluss auf die Studienqualität erlaubt (Diekmann, 2012, S. 422), wird im Folgenden diskutiert, inwieweit die erzielten Erhebungsrückläufe der Teilstudie 1 und 3 der Grundgesamtheit im Hinblick auf gegebene Strukturvariablen entsprechen. Teilstudie 2 lässt sich in diese Diskussion nicht mit einbeziehen. In dieser Teilstudie wurden keine Strukturvariablen erfasst. Eine abschließende Reflexion und Einordnung der Datengüte des Erhebungsrücklaufs ist in Abschnitt 5.3.9 enthalten.

Ausschöpfungsquoten in den Bundesländern

Angesicht der unterschiedlichen (Finanzierungs-)Strukturen in den Bundesländern sind diese ein oft erwähntes Unterscheidungsmerkmal in der Werkstatttratsarbeit. Tabelle 18 zeigt die erzielten Erhebungsrückläufe und Ausschöpfungsquoten nach Bundesländern. In Teilstudie 1 liegen für die Bundesländer Bayern, Sachsen, Saarland, Berlin und Thüringen die höchsten Beteiligungsquoten vor. Für Brandenburg, Schleswig-Holstein und Mecklenburg-Vorpommern fallen die Rückläufe mit 31 bis 39 % am niedrigsten aus. Ein anderes Bild zeigt sich hingegen in Teilstudie 3. Hier ist die Ausschöpfungsquote in Schleswig-Holstein – neben Bayern – überproportional, wohingegen für Thüringen mit 7 % ein sehr schlechter Rücklauf zu konstatieren ist. Festhalten lässt sich, dass die Beteiligungsquoten innerhalb der Teilstudien nach Länderebene variieren. Bayern weist in beiden Teilstudien die höchsten Beteiligungsquoten auf.

Tabelle 18: Erhebungsrücklauf und Ausschöpfungsquoten nach Bundesländern

	Anzahl	Anzahl Fragebögen		Ausschöpfungsquote	
	WfbM*	Teilstudie 1	Teilstudie 3	Teilstudie 1	Teilstudie 3
Baden-Württemberg	103**	43	21	42 %	20 %
Bayern	118	72	34	61 %	29 %
Berlin	17	9	3	53 %	18 %
Brandenburg	26	8	3	31 %	12 %
Hessen	49 ¹	20	7	41 %	14 %
Mecklenburg-Vorp.	23	9	5	39 %	22 %
Niedersachsen / Bremen / Hamburg***	83	39	16	47 %	19 %
Nordrhein-Westfalen	103**	42	15	41 %	15 %
Rheinland-Pfalz	36	15	8	42 %	22 %
Saarland	11	6	2	55 %	18 %
Sachsen	60	34	9	57 %	15 %
Sachsen-Anhalt	33	16	4	48 %	12 %
Schleswig-Holstein	37	14	10	38 %	27 %
Thüringen	31	16	2	52 %	7 %
Ohne Angabe		1	1		
Gesamt	730	344	140	47 %	19 %

Anmerkungen. Frageformulierungen: „In welchem Bundesland ist Ihre Werkstatt?“ (Teilstudie 1); „In welchem Bundesland befindet sich Ihre Werkstatt?“ (Wenn sich die WfbM über mehrere Bundesländer erstreckt, bitte den Standort der Hauptwerkstatt angeben.) (Teilstudie 3).

* Angaben nach dem Werkstättenverzeichnis (BA, 2019).

** Ohne stichprobenneutralen Ausfall. *** Zur Wahrung der Anonymität gemeinsam erfasst.

Um Unter- und Übererfassungen nach Länderebene einfacher zu erkennen, visualisiert Diagramm 2 die Anteile der WfbM nach Bundesländern in der Grundgesamtheit (bereinigte Stichprobe) sowie in den erzielten Erhebungsrückläufen der Teilstudien 1 und 3. Mit dieser Visualisierung wird zunächst die Spannweite der WfbM-Anteile in den Ländern deutlich, die von 2 % im Saarland bis zu 16 % in Bayern reicht. Im trilateralen Vergleich zeigt sich, dass die Rückläufe in den Bundesländern ihrem Anteil in der Grundgesamtheit in etwa entsprechen.

Anmerkung. Datenquellen Teilstudie 1 und 2; Verzeichnis der anerkannten WfbM 2019 (BA, 2019).

Diagramm 2: WfbM nach Bundesländern in der Grundgesamtheit sowie in den erzielten Erhebungsrückläufen der Teilstudien 1 und 2

Lediglich Bayern ist in beiden Teilstudien deutlich übererfasst. Die Rückläufe liegen rund 5 bzw. 8 Prozentpunkte über dem tatsächlichen Anteil der bayerischen WfbM.⁷⁵ Leicht untererfasst sind in der Teilstudie 3 NRW und Thüringen, die in den Daten der Leitungen rund 3 Prozentpunkte unterrepräsentiert sind.

Trägerschaft der teilnehmenden Werkstätten

Die Trägerschaft der Einrichtungen ist in den Teilstudien 1 (Werkstattträte) und 3 (Werkstattleitungen) mit sieben Bezeichnungen als Mehrfachangabe erfragt, in der Teilstudie der Werkstattträte zusätzlich mit einer offenen Option, mit der sich eine nachträgliche Zuordnung der Trägerschaft ermöglichte.⁷⁶ In der Teilstudie 3 benannten 96 % der Befragten einen Werkstattträger. Für 4 % der Werkstätten wurden zwei Träger aufgeführt. In Teilstudie 1 liegt dieser Anteil mit 5 % geringfügig höher. Für vier WfbM (1 %) wurden drei Träger angegeben. Bleibt die Angabe *Anderer* (siehe Tabelle 19) unberücksichtigt, haben sich in beiden Teilstudien am häufigsten Lebenshilfe-Werkstätten beteiligt, gefolgt von Werkstätten der Diakonie und Caritas. Die AWO- und DRK-Werkstätten sind hingegen nur geringfügig vorhanden.

Tabelle 19: Werkstattträger in den Teilstudien 1 und 3 (Mehrfachnennungen möglich)

	Teilstudie 1 (n= 343)	Teilstudie 3 (n= 140)
Lebenshilfe	37 %	28 %
Diakonie (evangelischer Träger*)	24 %	21 %
Caritas (katholischer Träger*)	17 %	18 %
Anthroposophischer Träger	6 %	4 %
AWO	2 %	2 %
DRK	2 %	1 %
Anderer	19 %	30 %

Anmerkung. * Nachträglich konstruierter Zusatz.

⁷⁵ Diese Verzerrung kann sich damit erklären, dass die Landesarbeitsgemeinschaft der Werkstätten für behinderte Menschen (LAG WfbM) Bayern ihren Mitgliedern vorab empfahl, sich an der Erhebung zu beteiligen und in diesem Bundesland ein höherer Bekanntheitsgrad der Autorin vorliegt (Schachler, 2014).

⁷⁶ Frageformulierungen: „Wer ist der Träger von Ihrer Werkstatt?“ (Teilstudie 1); „In welcher Trägerschaft befindet sich die WfbM?“ (Teilstudie 3). Mit der Kategorisierung der offenen Antworten wurden die Zusätze *katholischer Träger* (in der Kategorie Caritas) und *evangelischer Träger* (in der Kategorie Diakonie) ergänzt.

Im paarweisen Vergleich der Teilstudien 1 und 3 sind Werkstattträte von Lebenshilfe-Werkstätten deutlich häufiger vertreten als deren Leitungen. Dieses Bild zeigt sich gleichfalls für anthroposophische Einrichtungen, bei denen in Teilstudie 1 eine höhere Beteiligung als in Teilstudie 3 gegeben ist.

Zum Abgleich dieser Angaben lassen sich lediglich Referenzwerte der Dachverbände verwenden. Beispielsweise benannte die BAG WfbM (2019) im Jahr 2019 die Trägerschaft ihrer Mitgliedswerkstätten wie folgt: Paritätischer 43 %, Diakonie 26 %, Caritas 17 %, AWO und DRK je 4 %. Im gleichen Jahr lag der Anteil der Lebenshilfe-Werkstätten bei ca. 30 %⁷⁷ und WfbM in anthroposophischer Trägerschaft gab es zu ca. 7 %.⁷⁸ Diese Referenzen gleichen teilweise den Erhebungsrückläufen (etwa bei der Trägerschaft der Caritas-Werkstätten), andere Angaben weichen ab. So sind in der Teilstudie 3 Werkstätten der Diakonie seltener vertreten, als diese unter den Mitgliedern der BAG WfbM und in der Teilstudie 1 vorhanden sind, jedoch könnten unter der Angabe *Andere* noch weitere diakonische Einrichtungen enthalten sein. Mit 37 % sind Lebenshilfe-Werkstätten in der Studie der Werkstattträte besonders häufig vorhanden, was über der Vergleichsangabe des Bundesverbandes liegt. Insgesamt bleiben die indirekten Vergleiche jedoch spekulativ, da die Erfassungsweisen der Dachverbände nicht genau bekannt sind. Zudem kann die eigene Klassifizierung mit ihrer einfachen Sprachwahl für Werkstattleitungen als etwas unpräzise gelten, worauf der hohe Anteil anderer Träger hindeutet.

Anzahl der Betriebsstätten

In Teilstudie 3 wurde die „Anzahl der Betriebsstätten der Werkstatt zum Stichtag 01.01.2019 (Hauptwerkstatt, Zweigwerkstatt, Außenstellen)“ als struktureller Faktor der Werkstätten erfasst. Im Mittel wurden rund 4 Betriebsstätten benannt, wobei das Spektrum von 1 Betriebsstätte als Modalwert bis zu 32 Betriebsstätten reicht. Direkte Vergleichsangaben im Hinblick auf die Anzahl der Betriebsstätten liegen nicht vor. Lediglich die Angaben aus der Studie von Detmar et al. (2008, S. 27) können als Näherungswerte herangezogen werden. Aus dieser Befragung liegen Angaben aus den Jahren 2007 und 2008 zu 482 von 701 WfbM vor, so etwa

⁷⁷ A. Meinert (persönl. Mitteilung, 14.01.2020): Zum Zeitpunkt des Erhebungsversands gab es 221 Hauptwerkstätten in Trägerschaft von Lebenshilfen.

⁷⁸ D. Steinel (persönl. Mitteilung, 02.08.2019): Unter ihren Mitgliedseinrichtungen befinden sich über 80 Angebote zur Teilhabe am Arbeitsleben, darunter etwa 50 WfbM (bei der Angabe enthalten sind auch Einrichtungen, die mit anderen Trägern kooperieren). Bei 733 anerkannten WfbM mit Stand 2019 entspricht dies ~ 7 %.

zur „Anzahl der Betriebsstätten/Zweigwerkstätten, einschließlich Hauptwerkstatt“ (ebd., Anhang 2, S. 1). Tabelle 20 zeigt diese Ergebnisse gemeinsam mit den Angaben aus dem erzielten Rücklauf der Teilstudie 3.

Tabelle 20: Anzahl der Betriebsstätten in der Teilstudie 3 und bei Detmar, Gehrman, König, Momper, Pida und Radatz (2008)

	Teilstudie 3* (n = 137)	Studie Detmar et al. (2008)** (n = 468)
Nur Hauptbetriebsstätte	24 %	24 %
Hauptbetriebsstätte + 1 weitere Betriebsstätte	14 %	22 %
Hauptbetriebsstätte + 2–3 weitere Betriebsstätten	29 %	30 %
Hauptbetriebsstätte + 4 oder mehr Betriebsstätten	33 %	24 %
Gesamt	100 %	100 %

Anmerkungen.

* Frageformulierung: „Anzahl der Betriebsstätten der Werkstatt zum Stichtag 01.01.2019 (Hauptwerkstatt, Zweigwerkstatt, Außenstellen)“.

** Frageformulierung: „Anzahl der Betriebsstätten der Werkstatt, also Hauptwerkstatt sowie ggf. weitere Betriebsstätten/Zweigwerkstätten. Ausgelagerte Gruppen- oder Einzelarbeitsplätze (Außenarbeitsplätze) sind hier nicht gemeint und daher nicht mitzuzählen.“ (Detmar et al., 2008, Anhang 3/11).

Die Klassifikation in die vier Größenklassen ist der Studie von 2008 entlehnt. Während sich die Anteile der WfbM mit einem singulären Hauptstandort und derjenigen mit einem Hauptstandort und zwei bis drei weiteren Betriebsstätten in den Studien gleichen, weichen die Anteile in der Gruppe der Hauptbetriebsstätte + 1 weitere Betriebsstätte und der größten Werkstätten (Hauptbetriebsstätte + 4 oder mehr Betriebsstätten) voneinander ab. Genauer bleibt dabei wiederum hypothetisch; so haben seit 2007 verschiedene WfbM fusioniert und die Erfassungsweisen beinhalten unterschiedliche Spekulationsspielräume.

Anzahl der Beschäftigten

Die Anzahl der Beschäftigten in einer Werkstatt ist als Indikator der Einrichtungsgröße ein oft erwähntes Unterscheidungsmerkmal. In der Studie der Werkstattleitungen wurde diese als „Anzahl belegter Plätze im Arbeitsbereich zum Stichtag 01.01.2019“ mit dem Zusatz „Gemeint ist die Anzahl der Plätze in allen Betriebsstätten der anerkannten WfbM (Arbeitsbereich inkl. Fördergruppen mit Sozialversicherung)“ erhoben, was dem Zuständigkeitsbereich des Werkstatttrats laut WMVO entspricht. Im Ergebnis wurden Größenordnungen von 23 bis 2.570

Werkstattplätze benannt. Der Mittelwert liegt bei rund 417 Plätzen im Arbeitsbereich ($SD = 394$ Plätze).

Tabelle 21 unterscheidet die Werkstätten nach sechs unterschiedlichen Größenklassen. Im Rücklauf der Teilstudie 3 befinden sich am häufigsten kleine bis mittelgroße Werkstätten. Über 40 % lassen sich hier mit 120–239 und 240–359 Plätzen verorten. Anfragen lässt sich, warum Werkstätten in der Größenordnung von < 120 Plätzen vorliegen (darunter 13 WfbM mit < 90 Plätzen). So ist nach der WVO die Werkstattgröße i. d. R. mit mindestens 120 Plätzen (§ 7 Abs. 1 WVO) vorgesehen, wobei dies auch über einen Werkstattverbund erfüllt werden kann (§ 7 Abs. 2 WVO).

Tabelle 21: Anzahl der Beschäftigten in der Teilstudie 3 und belegte Werkstattplätze bei Detmar, Gehrman, König, Momper, Pida und Radatz (2008)

	Teilstudie 3 (<i>n</i> = 139)	Studie Detmar et al. (2008) (<i>n</i> = 482)
Unter 120	14 %	14 %
120–239	23 %	22 %
240–359	20 %	17 %
360–479	14 %	17 %
480–599	10 %	10 %
600 + mehr	19 %	20 %
Gesamt	100 %	100 %

Für einen Abgleich der Zahlen kann wiederum nur die bereits erwähnte Studie von Detmar et al. (2008, S. 28) dienen. Zwar erlauben sich – angesichts unterschiedlicher Erfassungsweisen⁷⁹ erneut nur vorsichtige Vergleiche, jedoch ähneln sich die Größenklassen in allen Kategorien erstaunlich (siehe Tabelle 21), und auch hier sind die kleinsten WfbM in einer ähnlichen Größenordnung zu finden.

⁷⁹ Detmar et al. (2008, S. 28) erhoben die Anzahl der Beschäftigten im Eingangsverfahren, Berufsbildungs- und Arbeitsbereich. Die eigene Studie erfasste demgegenüber, wie erwähnt, die Anzahl belegter Plätze im Arbeitsbereich (Werkstattstatus mit Sozialversicherung). Die Plätze im Eingangsverfahren/Berufsbildungsbereich entsprechen einem Anteil von rund 10 % der Gesamtplätze in WfbM (Arbeitsbereich inkl. Fördergruppen mit Sozialversicherung, Eingangsverfahren, Berufsbildungsbereich; BAG WfbM 2020a).

5.3.8 Aufbereitung und Auswertung der Daten

Zur Auswertung der Daten wurde mit der Statistik-Software SPSS gearbeitet. Die Auswertungen wurden in der Befehlssprache von SPSS per SPSS Syntax dokumentiert. In Abschnitt 5.3.8.1 werden die vorgenommenen Auswertungsentscheidungen beschrieben. Zur Ergebnisanalyse von Teilstudie 1 (Befragung der Werkstatträte) wurden Indizes konstruiert, die einen Teil der Dimensionen vertiefend untersuchen. In Abschnitt 5.3.8.2 wird zunächst allgemein beschrieben, was unter einer Indexbildung zu verstehen ist und wie diese in dieser Arbeit vorgenommen wurde. In den Abschnitten 5.3.8.3 bis 5.3.8.5 sind anschließend die drei verwendeten Indizes und die angewandten Voraussetzungsprüfungen dokumentiert.

5.3.8.1 Auswertungsentscheidungen

Die nicht-inhaltlichen Ausweichkategorien (*weiß nicht; es gibt keinen Betriebsrat / Personalrat / keine MAV* etc.) wurden in der Auswertung i. d. R. wie fehlende Werte behandelt.⁸⁰ Sofern diese inhaltlich relevant sind, ist dies im Ergebnisteil kommentiert.

Die offenen Antwortoptionen wurden zur Auswertung iterativ in induktiv-deduktiv gebildete Kategorien überführt. Wortgetreue Wiedergaben veranschaulichen im Ergebnisteil das Abstraktionsniveau und geben Einblicke in die Anmerkungen. Angesichts der Datenfülle konnte eine Auswertung der weiteren offenen Nennungen im Rahmen dieser Arbeit nicht mehr realisiert werden. Dies betrifft z. B. die Subdimension „Arbeitsinhalte“ (erhoben u. a. als abgeschlossene Betriebsvereinbarungen und eingebrachte Verbesserungsvorschläge). Auch auf die Darstellung von weiteren Detailspekten (etwa Befugnisse der Gesamtwerkstatträte bei mehreren anerkannten WfBM des Trägers, Besetzung der Vermittlungsstellen) wurde im Rahmen dieser Arbeit verzichtet, um die primären Fragestellungen zu fokussieren.

Wie im Abschnitt zur Instrumentenerstellung (5.3.3) begründet, wurden in der Erhebung bewusst Ratingskalen mit fünf Abstufungen verwendet. Diese wurden in der Auswertung unter Vorbehalt intervallskaliert behandelt. In der empirischen Sozialforschung ist dies für ordinale Daten anlässlich fehlender Auswertungsalternativen gängig. Um systematischen Messfehlern zu begegnen, wurden Skalenwerte verwendet, die als ausbalanciert gelten, sodass eine Gleichabständigkeit (Äquidistanz) unterstellt werden kann (Bortz & Döring, 2006, S. 177 ff.).

⁸⁰ Beachtenswert sind in der Teilstudie 2 die hohen Antwortquoten. Obwohl die Ausweichoption *weiß nicht* gegeben war, finden sich häufig keinerlei ausweichende Antworten oder fehlende Werte, was für ein sorgfältiges Ausfüllen oder eine passende Frageformulierung spricht (Schulz & Renn, 2009, S. 38).

Gleichfalls wurde mit fünf Abstufungen die Mindestanzahl erreicht, die als weitere notwendige Voraussetzung für ein unterstelltes Intervallniveau gilt (etwa Baur, 2008, S. 287; Faulbaum et al., 2009, S. 26).

Inferenzstatistik in der quantitativen Ergebnisanalyse

In der quantitativen Ergebnisanalyse werden in Abschnitt 5.3.8 Indizes eingeführt, um verschiedene Zusammenhangs- und Unterschiedshypothesen statistisch zu testen (zu den Verfahren siehe etwa Häder, 2010, S. 424 ff.). Zur Anwendung von Inferenzstatistik bei Vollerhebungen gibt es verschiedene Ansichten.⁸¹ In der eigenen Studie ist der Erhebungszugang als Vollerhebung angelegt, jedoch konnte diese nicht vollständig realisiert werden, es liegt eine teilrealisierte Vollerhebung vor. Die erzielten Rückläufe in den drei Teilstudien stellen lediglich Teilmengen der interessierenden Grundgesamtheit dar. Unter Berücksichtigung möglicher Rücklaufverzerrungen (siehe Abschnitt 5.3.6) wurde Inferenzstatistik folglich angewandt, wenn es um Rückschlüsse auf Werkstatträte an den Hauptwerkstätten aller anerkannten WfbM geht. Die Anwendung wurde auf die Teilstudien 1 und 2 sowie auf Werkstatträte mit regelhaftem Wahlturnus beschränkt, um den Aussagegehalt des Ex-post-facto-Designs nicht zu überladen (Schnell et al., 2013, S. 228).

Untersucht wurden Hypothesen, die in den Ergebnissen der qualitativen Erhebungsphase besonders prägnant erschienen (siehe Abschnitt 6.2.2) und sich unter Bezugnahme auf den Theorieteil der Arbeit begründen sowie operationalisieren ließen (siehe Anhang 2 *Operationalisierung der Fragestellungen* im elektronischen Zusatzmaterial). Die angewandten Verfahren richteten sich nach der Beschaffenheit der Variablen, der Daten sowie der Art der formulierten Hypothesen aus. Zur Anwendung kamen parametrische und nichtparametrische Verfahren. Die Voraussetzungsprüfungen (Normalverteilung, Varianzhomogenität etc.) wurden in digitalen Anhängen dokumentiert. Zur statistisch korrekten Ausführung wurde das Angebot der Lernplattform Statistik-Akademie (<https://www.statistik-akademie.de>; Keller, 2016) genutzt und die Analysen mit Beratung durch die Diplom-Mathematikerin Keller durchgeführt.

⁸¹ Einem eher handwerklichen Pragmatismus, der für Sozialforscher*innen konstatiert wird, steht eine statistische Orientierung gegenüber, die Methoden der statistischen Inferenz für Vollerhebungen etwa nur unter methodisch reflektierten und inhaltlich relevanten Voraussetzungen als geeignet begreift (Behnke, 2005, O5).

5.3.8.2 Indexbildung (Teilstudie 1)

Eine Indexbildung lässt sich als Auswertungsverfahren bezeichnen, dessen Konstruktion mit der Instrumentenentwicklung angelegt wird. Inwieweit Indizes konzeptionelle Messinstrumente sind, ist strittig (Latcheva & Davidov, 2019, S. 893). Die Indexerstellung ist zunächst „eine rein theoretische Arbeit“ (Schnell et al., 2013, S. 158). Objektive Beurteilungskriterien zur Vollständigkeit eines Index existieren nicht. Nach der dimensionalen Festlegung erfolgt die Indexbildung, bei der die Antwortkategorien der Fragen bzw. Items einen Summenscore erzeugen. Eine angemessene Gewichtung gilt als Qualitätsmerkmal (Bortz & Döring, 2016, S. 278). Möglich ist diese theoriegeleitet, durch ein Expertenrating oder durch die Ergebnisse von Faktorenanalysen, bei denen die Korrelationen zwischen dem Gesamtindex und den einzelnen Items die Wertung bestimmen. Empfohlen wird eine gleichwertige Behandlung, da gewichtete Indizes willkürlichen Komponenten unterliegen (Schnell et al., 2013, S. 163). Demgemäß erfolgt in der vorliegenden Arbeit die Indexbildung additiv. Jede Antwort ging gleichwertig summiert in die Zieldimension ein.

Um fehlende Werte ausgleichen zu können und dabei möglichst geringfügig in die Daten einzugreifen (siehe etwa Engel & Schmidt, 2019), wurde mit Mittelwertindizes gearbeitet, bei denen sich der fallbezogene Mittelwert aus der Summe der vorhandenen Werte, dividiert durch die Anzahl der vorhandenen Indikatoren, generiert.⁸² Das heißt, lagen nur vier und nicht fünf vorgesehene Indikatoren vor, wurde die erzielte Summe durch vier geteilt, bei fünf Indikatoren hingegen durch fünf. Die so gebildeten Mittelwertindizes drücken die Informationsqualitäten an den Werkstattrat, das vorliegende Aktivitätsniveau eines Gremiums und dessen autonome Arbeitsweise numerisch aus. Im Folgenden werden die konstruierten Indizes genauer dargestellt.

5.3.8.3 Index „Informationsqualitäten“

Zur zusammenfassenden Darstellung der praktizierten Informationsübermittlung an den Werkstattrat wurden die in Abschnitt 7.1.3.3 behandelten Items zu einem Index „Informationsqualitäten“ zusammengefasst. Zunächst wurde geprüft, ob alle Indikatoren zur

⁸² Im Umgang mit itembezogener Nonresponse sind verschiedene Strategien möglich. So können fehlende Werte etwa ausgleichend gewichtet, imputiert oder ignoriert werden (siehe hierzu Engel & Schmidt, 2019; Spieß, 2010). Die jeweilige Handhabung hängt nicht zuletzt von dem angenommenen Fehlertyp ab (etwa MCAR – Missing Completely At Random etc.).

Zusammenfassung geeignet sind.⁸³ Die angewandte Voraussetzungsprüfung ist in Tabelle 22 dargestellt. Die Item-Response bewegt sich mit 97 % bis 99 % auf einem hohen Niveau. Um in den Indizes nur aussagekräftige Items zu verwenden, lässt sich die Itemschwierigkeit betrachten.⁸⁴ Diese bewegt sich zwischen 0.3 und 0.78 und entspricht mittleren Bestätigungsraten. Damit weisen alle Items eine ausreichende Streuung auf.

Zur Reliabilitätsbeurteilung einer Messung wird oft Cronbachs α (Cronbachs Alpha Koeffizient) als Ausdruck der internen Konsistenz einer Skala angewandt. Die Maßzahl eignet sich für eindimensionale, äquivalente Items (Danner, 2015, S. 6). Da Cronbachs α als aggregierter Wert jedoch von der Anzahl der verwendeten Items abhängt, wurde er in der vorliegenden Analyse nicht als ausschlaggebendes Kriterium interpretiert (Kopp & Lois, 2014, S. 98 f.). Stattdessen wurde betrachtet, wie sich der Schätzer verhält, wenn ein Item weggelassen wird. Weiter wurde mit der Trennschärfe (Item-Skala-Korrelation) gearbeitet. Als (korrigierte) Korrelation eines Items mit dem Gesamtergebnis einer Skala zeigt diese an, wie stark oder schwach ein Item mit dem zu messenden Konstrukt zusammenhängt und dieses durch das einzelne Item bereits repräsentiert wird. In der Reliabilitätsanalyse erfüllten alle Items den kritischen Grenzwert der Trennschärfe von > 0.3 .⁸⁵ Für drei Items liegt die Trennschärfe in hohen Bereichen. Dementsprechend würde Cronbachs α nicht im Gesamtwert von 0.76 steigen, wenn ein Item entfällt.

⁸³ Bei den vorliegenden Items handelt es sich um reflektive Indikatoren, die ein übergeordnete Konstrukt mit feinen Unterschieden widerspiegeln. Zur Unterscheidung von reflektiven und formativen Indikatoren siehe Latcheva und Davidov (2019, S. 898).

⁸⁴ Die Itemschwierigkeit berechnet sich wie folgt: $\frac{\text{Summe erreichter Punkte eines Items}}{\text{Theoretisch mögliche Summe des Items}}$ (Bortz & Döring, 2006, S. 219). Sie kann sich im Bereich von 0 (völlige Ablehnung, tritt nicht auf o. Ä.) bis 1 (völlige Zustimmung; Lösung o. Ä.) bewegen und drückt den aggregierten Zustimmungswert zu einer Frage aus. Nach Bortz und Döring (2006) gelten Werte von 0.2 und 0.8 als praktikabel (S. 219).

⁸⁵ Für die Trennschärfe – liegen im Gegensatz zu Cronbachs α – klarere Konventionen vor. Der Wertebereich 0.3–0.5 gilt als mittelmäßiges, > 0.5 als hohes Ergebnis. Zu beachten ist, dass die Trennschärfe auch dadurch beeinflusst wird, wie schwer oder leicht ein Item bestätigt werden kann. Ausschließlich hohe Trennschärfen zu verwenden – die bei mittelschweren Items zu finden sind – ist deswegen nicht empfehlenswert (Bortz & Döring, 2006, S. 220; Kopp & Lois, 2014, S. 99).

Tabelle 22: Index Informationsqualitäten: Itemschwierigkeit und Reliabilitätsanalysen

Item: Der Werkstattrat bekommt ...	<i>n</i>	It.-R.	Its.	Tr.	Cr. α -It.
• die Infos so, dass er sie verstehen kann	337	98 %	0.78	0.61	0.68
• alle wichtigen Infos	337	98 %	0.71	0.64	0.67
• die Infos zum Arbeitsergebnis der Werkstatt verständlich übermittelt	334	97 %	0.61	0.43	0.77
• die Infos zwischen Tür und Angel	342	99 %	0.30	0.42*	0.75
Item: Der Werkstattrat erfährt Infos erst im Nachhinein	338	98 %	0.38	0.60*	0.69

Anmerkungen. Datenquelle Teilstudie 1; $n = 344$; Cronbachs $\alpha = 0.76$; listenweiser Fallausschluss.

It.-R. = Itemresponse in Prozent.

Its. = Itemschwierigkeit, Berechnung mit 0 (*nie*) als niedrigster und 4 (*immer*) als höchster Wert.

Tr. = Trennschärfe (Item- Skala-Korrelation).

Cr. α -It = Cronbachs α , wenn Item weggelassen.

* In der Berechnung umkodiert.

Umgang mit fehlenden Werten und Indexbildung

Insgesamt sprach aus der Voraussetzungsprüfung nichts gegen den Einbezug aller Items. Für 321 Gremien (93 %) lagen alle Items und in zwei Fällen keine Angaben zu den Informationsqualitäten vor. Bei 20 Gremien fehlte ein Variablenwert und in einem Fall fehlten zwei Werte. Wie in Abschnitt 5.3.8.2 beschrieben, lassen sich ein oder zwei fehlende Werte in Form eines Mittelwertindex ausgleichen, indem sich der fallbezogene Wert aus den vorhandenen Items generiert. Gleichzeitig blieb der Eingriff in die Daten damit möglichst gering. Vor der Berechnung wurden zwei invertierte Items transformiert, sodass die Abstufungen alle in eine identische Richtung wiesen.

5.3.8.4 Index „Aktivitätsniveau“

In diesem Auswertungsvorgang wurde ein mehrdimensionaler Index gebildet, der aus dem Zusammenspiel der formativen Indikatoren (Latcheva & Davidov, 2019, S. 898) entsteht und das latente Konstrukt „Aktivitätsniveau des Werkstatttrats“ numerisch ausdrückt. Die Indikatoren umfassen formale Aktivitäten und Organisationsabläufe (Sitzungsturnus, Werkstattversammlungen⁸⁶), Arbeitsinhalte (eingebrachte Vorschläge), Vernetzungsstrukturen (externe Kontakte) und Tagungs- bzw. Schulungsteilnahmen. Nicht berücksichtigt wurden Aspekte, die sich in Form von manifesten Variablen weniger konkret erfassen lassen (bspw. individuelle Sinnsetzungen). Die Indikatoren entstammten den qualitativen Erhebungsergebnissen und den rechtlichen Grundlagen der Werkstatttratsarbeit und finden sich in der betrieblichen Mitbestimmungsforschung wieder (siehe *Anhang 2 Operationalisierung der Fragestellungen* im elektronischen Zusatzmaterial). In der Pretestphase (Abschnitt 5.3.4) wurden die Indikatoren inhaltlich diskutiert und weniger valide Variablen von der Indexkonstruktion ausgeschlossen.

Um gleiche Messbedingungen zu schaffen, gingen in den Index nur Werkstatttratsgremien ein, die zuletzt im regelhaften Wahlturnus Oktober/November 2017 (plus/minus drei Monate) gewählt wurden. Gremien mit abweichendem Wahlzeitpunkt oder unpräziser Angabe wurden von der Auswertung ausgeschlossen.

Die Antwortformate lagen entweder bereits mit Skalenwerten von 0 bis 4 vor oder wurden in diese Abstufungen mit äquidistanten Skalenwerten transformiert, sodass die Variablen quasi-metrisch behandelt werden konnten (siehe Tabelle 23). Ausreißer und Extremwerte wurden nach inhaltlicher Plausibilität dem nächstgelegenen Skalenwert zugeordnet (zur Dokumentation siehe Tabelle 23), was auch im Umgang mit Item-Nonresponse Anwendung findet (Engel & Schmidt, 2019, S. 398).

⁸⁶ Für die Anzahl der durchgeführten Werkstattversammlungen wurde berücksichtigt, dass gemäß § 9 Abs. 2 WMVO Abteilungsversammlungen zulässig sind. Das heißt, es wurde eine neue Variable gebildet und bei null durchgeführten Werkstattversammlungen die Anzahl der erhobenen Abteilungsversammlungen ersatzweise substituiert. Dies betraf elf Fälle.

Tabelle 23: Indexbildung „Aktivitätsniveau des Werkstattrats“

Dimension	Indikator	Skalenbildung	n	M	Rundungen / Zwischenwerte
Organisation der Gremien	Sitzungsumus	< als 1-mal im Monat = 0 Punkte	303	1.92	> als jede Woche = 4 Punkte (n = 2) 3-mal im Monat = 3.5 Punkte (n = 5)
		1-mal im Monat = 1 Punkt			
		Alle drei Wochen = 2 Punkte			
		Alle zwei Wochen = 3 Punkte			
Arbeitsinhalte	Anzahl eingebrachter Verbesserungsvorschläge	Jede Woche = 4 Punkte	273	2.41	> als 4 Vorschläge = 4 Punkte (n = 14)
		0 Vorschläge = 0 Punkte			
		1 Vorschlag = 1 Punkt			
		2 Vorschläge = 2 Punkte			
		3 Vorschläge = 3 Punkte			
Kooperation und Vernetzung: intern (neu gebildete Variable)	Anzahl Werkstattversammlungen (inklusive Abteilungsversammlungen bei 0 Werkstattversammlungen in 11 Fällen)	4 Vorschläge = 4 Punkte	301	2.10	> als 4 Versammlungen = 4 Punkte (n = 27)
		0 Versammlungen = 0 Punkte			
		1 Versammlung = 1 Punkt			
		2 Versammlungen = 2 Punkte			
		3 Versammlungen = 3 Punkte			
Weiterbildung	Anzahl besuchter Tagungen	4 Versammlungen = 4 Punkte	295	2.05	> als 8 Tagungen = 4 Punkte (n = 32)
		0 Tagungen = 0 Punkte			
		1–2 Tagungen = 1 Punkt			
		3–4 Tagungen = 2 Punkte			
		5–6 Tagungen = 3 Punkte			
		7–8 Tagungen = 4 Punkte			
	Externe Schulungsdauer	0 Schulungstage = 0 Punkte	300	2.26	1.5 Schulungstage = 1.5 Punkte (n = 1) > als 4 Tage = 4 Punkte (n = 29)
		1 Schulungstag = 1 Punkt			
		2 Schulungstage = 2 Punkte			
		3 Schulungstage = 3 Punkte			
Kooperation und Vernetzung: extern	Anzahl Kontakte zu anderen Gremien	4 Schulungstage = 4 Punkte	305	1.51	Keine
		Pro Kontakt 1 Punkt			

Anmerkungen. Datenquelle Teilstudie I; berücksichtigt wurden 306 Gremien mit regelhaften Wahlen 2017.

In Tabelle 24 sind der Anteil der fehlenden Werte und die Itemschwierigkeit der sechs Indikatoren angegeben. Abgesehen von dem Indikator „Anzahl der Verbesserungsvorschläge“ liegen sehr hohe Antwortraten vor. Die Antwortschwierigkeit der Indikatoren variiert von 0.38 bis 0.60. Dies sind Bestätigungsraten, die unterschiedlich leicht bzw. schwer zu erzielen und für die Analyse aufschlussreich sind (zur weiteren Bedeutung siehe Index „Informationsqualitäten“ im vorangegangenen Abschnitt).

Tabelle 24: Index Aktivitätsniveau: Fehlende Werte und Itemschwierigkeit

Indikator	<i>n</i>	Anzahl fehlende Werte	Itemschwierigkeit*
Sitzungsturnus	303	3	0.48
Anzahl Verbesserungsvorschläge	273	33	0.60
Anzahl Versammlungen	301	5	0.52
Anzahl besuchter Tagungen	295	11	0.51
Exterbe Schulungsdauer in Tagen	300	6	0.57
Anzahl Kontakte	305	1	0.38

Anmerkungen. Datenquelle Teilstudie 1, Gremien mit regelhaften Wahlen 2017.

* Berechnung nach Bortz und Döring (2006, S. 219).

Angesicht der mehrdimensionalen Struktur des Index war Cronbachs α kein geeigneter Schätzer (Danner, 2015, S. 6) für die Indexreliabilität. Gleichfalls schied eine explorative Faktorenanalyse zur Konstruktvalidierung (Rammstedt, 2014, S. 18 ff.) aus, da formative Indikatoren verschiedene Inhalte messen und nicht zwingend korrelieren müssen (Latcheva & Davidov, 2019, S. 899). Jedoch kann zur Legitimität der Konstruktbildung eine Prüfvariable zeigen, ob alle Indikatoren in eine identische Richtung weisen (ebd., S. 900). Hierzu wurde die Variable „Eigene Büronutzung in Jahren“ verwendet. Wie Tabelle 25 zeigt, korrelieren alle Indikatoren positiv mit der Prüfvariablen. Das heißt, ein höherer Wertebereich geht bei diesen Indikatoren mit einer längeren Dauer der Büronutzung des Werkstatttrats einher. Auch untereinander korrelieren alle Indikatoren positiv, wobei geringe (signifikante) Zusammenhänge bestehen.

Tabelle 25: Index „Aktivitätsniveau“: Korrelationen der Indikatoren und Prüfvariable

Indikator ¹	Sitzungs- turnus	Verb.vor- schläge	Versamm- lungen	Tagun- gen	Externe Schu- lungsd.	Büronutzung in Jahren ²
Sitzungsturnus						.20** (<i>n</i> = 283)
Verbesserungsvor- schläge	.04					.03 (<i>n</i> = 255)
Versammlungen	.09	.09				.05 (<i>n</i> = 280)
Tagungen	.16**	.17**	.07			.23** (<i>n</i> = 276)
Externe Schulungs- dauer	.09	.13**	.04	.25**		.11* (<i>n</i> = 280)
Kontakte	.20**	.11*	.06	.28**	.16**	.25** (<i>n</i> = 284)

Anmerkungen. Datenquelle Teilstudie 1; Gremien mit regelhaften Wahlen 2017; Korrelationen nach Kendall, paarweiser Fallausschluss.

¹ Wertebereich der Variablen jeweils von 0 bis 4. ² Wertebereich der Variablen von 0 bis 32.

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

Umgang mit fehlenden Werten und Indexbildung

In toto ergaben sich aus den statistischen Betrachtungen keine Hinweise, die gegen den Einbezug aller Indikatoren in den Index sprachen. Für 255 (83 %) der 306 Gremien mit regelhaftem Wahlturnus lagen alle sechs Indikatoren vor. Bei 14 % (44 Gremien) fehlte eine Messung, in sechs Fällen zwei und in einem Fall drei Variablenwerte. Die meisten Missings bestanden für die Anzahl der Verbesserungsvorschläge seit der letzten Werkstattratswahl. Angesichts der hohen Bedeutung für eine aktive Gremienarbeit wurde der Indikator dennoch verwendet und zum Ausgleich der fehlenden Werte erneut ein Mittelwertindex gebildet. Dieses Verfahren war zulässig, da die Gesamtmittelwerte der Indikatoren im Wertebereich von 1.92 bis 2.41 (siehe Tabelle 23) variablenbezogen nicht zu stark variieren und fehlende Werte sich so ega-lisieren konnten.

5.3.8.5 Index „autonome Arbeitsweise“

Wie in Abschnitt 5.3.2 dargestellt, wurden unter dem mehrdimensionalen Konstrukt „autonome Arbeitsweise“ Komponenten vorliegender Arbeitsbedingungen, die ein autonomes Handeln ermöglichen, sowie Aspekte einer eigenständigen Arbeitsweise zusammengefasst. Strukturelle Bedingungen für ein autonomes Agieren werden im Index mit Indikatoren zur Beschaffenheit der Informationen und dem Zugang zu Arbeitsmitteln (Büronutzung) angezeigt. Eine eingenommene unabhängige Position des Gremiums wird hingegen über die eigene Meinungsvertretung und die nicht direktive Rolle der Vertrauensperson sowie über die Anzahl der Widersprüche in Angelegenheiten der Mitwirkung und Mitbestimmung indiziert (zur Herkunft und detaillierten Begründung der Indikatoren siehe *Anhang 2 Operationalisierung der Fragestellungen* im elektronischen Zusatzmaterial). Diese formativen Indikatoren wurden in Fünfer-Skalen des Wertebereichs von 0 bis 4 transformiert. Die genaue Transformation ist mit der Behandlung von Ausreißern und Zwischenwerten in Tabelle 27 angegeben.

Die Item-Nonresponse der Indikatoren (siehe Tabelle 26) bewegt sich mit 2 bis 18 fehlenden Werten in einem praktikablen und überwiegend geringen Bereich. Bei der Itemschwierigkeit zeigt sich hingegen, dass der Indikator „Anzahl der Widersprüche“ mit 0.14 nicht brauchbar ist, da die Frage nur von sehr wenigen Gremien „gelöst“ und der zu erzielende minimale Grenzwert von 0.2 (Bortz & Döring, 2006, S. 219) nicht erreicht werden kann.

Tabelle 26: Index „autonome Arbeitsweise“: Fehlende Werte und Itemschwierigkeit

Indikator	<i>n</i>	Fehlende Werte	Itemschwierigkeit*
Dauer Büronutzung	288	18	0.36
Informationsqualitäten	304	2	0.70
Direktives Verhalten Vertrauensperson	296	10	0.70
Andere Meinung haben	302	4	0.43
Andere Meinung sagen	300	6	0.80
Anzahl Widersprüche	288	18	0.14

Anmerkungen. Datenquelle Teilstudie 1, berücksichtigt wurden die Angaben von 306 Gremien mit regelhaften Wahlen 2017.

* Berechnung nach Bortz und Döring (2006, S. 219).

Tabelle 27: Indexbildung „autonome Arbeitsweise“

Dimension	Indikator	Skalenbildung	n	M	Rundungen / Zwischenwerte
Verfügung über Arbeitsmittel	Dauer der Nutzung eines eigenen Büros	Kein Büro = 0 Punkte	288	1.44	Seit Beginn / schon immer = 4 Punkte > als 8 Jahre = 4 Punkte (n = 50)
		Seit 1–2 Jahren (2019/2018) = 1 Punkt			
		Seit 3–4 Jahren = 2 Punkte			
		Seit 5–6 Jahren = 3 Punkte			
Informations- und Beteiligungsrechte	Informationsqualitäten*	Seit 7–8 Jahren = 4 Punkte	304	2.80	
		M von 1 bis 1.8 = 0 Punkte			
		M von 1.9 bis 2.6 = 1 Punkt			
		M von 2.7 bis 3.4 = 2 Punkte			
Verhältnis zwischen den Beteiligten	Direktives Verhalten der Vertrauensperson	M von 3.5 bis 4.2 = 3 Punkte	296	2.81	
		M von 4.3 bis 5 = 4 Punkte			
		Immer = 0 Punkte			
		Oft = 1 Punkt			
Interessenvertretung	Anzahl Widersprüche in Angelegenheiten der Mitwirkung und Mitbestimmung*	Manchmal = 2 Punkte	302	1.71	Oft bis manchmal = 2.5 Punkte Manchmal bis selten = 1.5 Punkte
		Selten = 1 Punkt			
		Nie = 0 Punkte			
		Immer = 4 Punkte			
Interessenvertretung	Anzahl Widersprüche in Angelegenheiten der Mitwirkung und Mitbestimmung*	Oft = 3 Punkte	300	3.20	Oft bis manchmal = 2.5 Punkte
		Immer = 4 Punkte			
		Kein Widerspruch = 0 Punkte			
		1 Widerspruch = 1 Punkt			
Interessenvertretung	Anzahl Widersprüche in Angelegenheiten der Mitwirkung und Mitbestimmung*	2 Widersprüche = 2 Punkte	288	0.55	x-mal = 4 Punkte (n = 1) > als 4 Widersprüche = 4 Punkte (n = 4)
		3 Widersprüche = 3 Punkte			
		4 Widersprüche = 4 Punkte			
		Kein Widerspruch = 0 Punkte			

Anmerkungen. Datenquelle Teilstudie 1; berücksichtigt wurden die Angaben von 306 Gremien mit regelhaften Wahlen 2017.

* Nach der Voraussetzungsprüfung von der Indexkonstruktion ausgeschlossen.

Als weitere Eignungsbetrachtung sind in Tabelle 28 die Korrelationen zur Variablen „Anzahl der besuchten Tagungen“ aufgeführt. Diese Prüfvariable kann zeigen, ob alle Indikatoren in eine identische Richtung weisen.

Tabelle 28: Index „autonome Arbeitsweise“: Korrelationen der Indikatoren¹ und Prüfvariable

Indikator ¹	Informations- qualitäten	Büro- nutzung	Direktives Verhalten	Meinung haben	Anzahl Tagun- gen ²
Informationsqualitäten					-.07 (<i>n</i> = 293)
Dauer Büronutzung	-.01				.19** (<i>n</i> = 279)
Direktives Verhalten Vertrauensperson	-.01	.16**			.01 (<i>n</i> = 286)
Andere Meinung haben	-.24**	.10*	.11*		.16** (<i>n</i> = 289)
Andere Meinung sagen	-.14**	.05	.18**	-.001	.07 (<i>n</i> = 280)

Anmerkungen. Datenquelle Teilstudie 1; berücksichtigt wurden die Angaben von 306 Gremien mit regelhaften Wahlen 2017; Korrelationen nach Kendall, paarweiser Fallausschluss.

¹ Wertebereich der Variablen jeweils von 0 bis 4.

² Wertebereich der Variablen von 0 bis 25.

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

So zeigt sich, dass die Variable „Informationsqualitäten“ für den Index gleichfalls ungeeignet ist, da diese – anders als die anderen Indikatoren – mit den Tagungsbesuchen negativ korreliert. Auch die Beziehung zu den anderen Indikatoren signalisiert negative Zusammenhänge, d. h., hohe Informationsqualitäten treten tendenziell mit geringeren Werten der anderen Variablen auf. Dies wird bei den Meinungsindikatoren mit signifikanten Zusammenhängen offensichtlich und scheint insofern inhaltlich stimmig, da ein autonomeres Handeln dazu führen kann, dass Informationsqualitäten kritischer bewertet oder Informationen weniger bereitwillig gegeben werden. Zwischen den Indikatoren „andere Meinung haben“ und „andere Meinung sagen“ besteht eine weitere negative Korrelation, die nahezu bei null liegt. Dies bedeutet, dass

kein Effekt bzw. linearer Zusammenhang zwischen den Variablen feststellbar ist.⁸⁷ Die anderen Werte zeigen hingegen positive Zusammenhänge und mitunter kleinere Effekte der Stärke $> .1$ an.

Umgang mit fehlenden Werten und Indexbildung

Die statistische Voraussetzungsprüfung zeigte, dass zwei Variablen nicht zur Indexkonstruktion geeignet sind. Folglich wurden lediglich die vier formativen Indikatoren „Dauer der Büronutzung“, „direktives Verhalten der Vertrauensperson“, „eine andere Meinung als die Leitung haben“ und „sich trauen, eine andere Meinung zu sagen“ zur Bildung des Index „autonome Arbeitsweise“ herangezogen. Für 273 Gremien (89 %) lagen dazu alle Variablen vor, für weitere 29 (10 %) drei Werte. Durch die Bildung eines Mittelwertindex (additiver Index dividiert durch die vorhandene Anzahl der Indikatoren) konnten diese 302 Werkstatträte einbezogen werden. Lediglich die verbleibenden vier Fälle, in denen mehr als ein Indikator fehlte, wurden von der Analyse ausgeschlossen.

5.3.9 Gütekriterien und Reflexion der quantitativen Erhebungsphase

Für eine hohe Güte quantitativer Forschungsergebnisse müssen Messungen gültig sein und zuverlässig funktionieren. Das heißt, es gilt bei Messvorgängen sowohl ohne Abstriche das zu erfassen, was es zu erfassen gilt, was als Validität bezeichnet wird, als auch bei wiederholten Erhebungen identische (reliable) Ergebnisse zu generieren (Schnell et al., 2013, S. 140). Neben der Validität und Reliabilität ist die Objektivität das dritte klassische Gütekriterium quantitativer Forschung (Häder, 2010, S. 108). Diese Kriterien können sich auf das Forschungsdesign, auf Messinstrumente oder statistische Auswertungen beziehen und werden in weitere Unterformen unterteilt, wobei disziplinär unterschiedliche Schwerpunktsetzungen feststellbar sind (Krebs & Menold, 2019, S. 489).

Objektivität

Unter dem Gütekriterium der Objektivität wird die Unabhängigkeit der Ergebnisse von den beteiligten Personen angesprochen. Sowohl die Durchführung einer Erhebung als auch die

⁸⁷ Im Folgenden wird der Einteilung nach Cohen (1988) für r (Pearsons Produkt-Moment-Korrelationskoeffizient) gefolgt: $.1$ bis $.3$: entspricht einem kleinen Effekt; $.3$ bis $.5$ entspricht einem mittleren Effekt; $.5$ und höher entspricht einem starken Effekt.

Datenauswertung und die Interpretation müssen möglichst objektiv und intersubjektiv nachvollziehbar bleiben (Häder, 2010, S. 108). Während quantitative Erhebungen sich gut standardisiert durchführen und transparent auswerten lassen, wird eine *Interpretationsobjektivität* für die empirische Sozialforschung als grundsätzlich nicht möglich erachtet (Krebs & Menold, 2019, S. 490). In der Instrumentenkonstruktion wurde das Fragebewertungssystem von Faulbaum et al. (2009) als Hilfestellung für ein objektives Wording angewandt. Ein Schwachpunkt der *Durchführungsobjektivität* kann in der Befragung des Werkstatttrats (Teilstudie 1) als Gremium liegen. Dieser Befragungszugang beinhaltet bei Einschätzungsfragen im Detail methodologische Aspekte, die den Rahmen der vorliegenden Arbeit übersteigen. Gleichfalls kann für Teilstudie 3 bemängelt werden, dass diese an die Geschäftsführung oder Werkstatteleitung gerichtet war und nicht erhoben wurde, welche Personengruppen die Befragung ausgefüllt haben. Auf diesen Aspekt wurde jedoch bewusst verzichtet, da damit keine Hypothesen oder Annahmen verbunden waren und sich die Leitungsstrukturen in WfbM sehr unterschiedlich ausgestalten.

Für eine hohe *Auswertungsobjektivität* wurden die Indexkonstruktionen und die statistischen Auswertungsschritte sorgfältig dokumentiert (siehe Abschnitt 5.3.8). Für eine objektive Vergleichbarkeit der Angaben gehen in die Indizes „Aktivitätsniveau“ und „autonome Arbeitsweise“ nur Gremien ein, die zuletzt im regelhaften Wahlturnus Oktober/November 2017 (plus/minus drei Monate) gewählt wurden. Dies schafft vergleichbare Messbedingungen: Untersucht wird ein ca. zweijähriger Zeitraum und die Gremien befinden sich zum Untersuchungszeitraum in der Mitte ihrer vierjährigen Amtszeit. Zu beachten ist, dass sich die Ergebnisse damit nur auf die Gremien mit regelhaftem Amtsturnus beziehen.

Zur Einstufung der praktischen Bedeutsamkeit der statistischen Signifikanz wurde in der Ergebnisanalyse mit Effektgrößen gearbeitet. Dies gilt als ein Kriterium qualitativ hochwertiger Sozialforschung (Krebs & Menold, 2019, S. 502).

Validität

Validität ist im Detail ein vielfältiges Gütekriterium, mit der die Gültigkeit einer Messung thematisiert wird. So lässt sich bspw. die *interne* und *externe Validität* von Forschungsergebnissen/-designs betrachten (Bortz & Döring, 2016, S. 94 f.) oder die Validität von Messungen als *Inhalts-*, *Kriteriums-* und *Konstruktvalidität* weiter spezifizieren (zur Unterscheidung siehe etwa Diekmann, 2012, S. 258). Nur Einzelaspekte der Validität lassen sich statistischen Betrachtungen unterziehen. Die Kriteriumsvalidität kann etwa anhand eines Außenkriteriums

überprüft werden, das in der Praxis nur selten vorhanden ist (Krebs & Menold, 2019, S. 497), so auch bei den vorliegenden eigenen Ergebnissen. Zur Konstruktvalidierung eines Instruments kann u. a. eine Faktorenanalyse angewandt werden (Rammstedt, 2014, S. 18 ff.). Dies setzt wiederum bestimmte Eigenschaften der Daten sowie eine Vielzahl äquivalenter Items voraus, die gleiche Inhalte und ein übergreifendes Konstrukt messen und sich hierbei nur in feinen Nuancen unterscheiden (Latcheva & Davidov, 2019, S. 898). Dies war für die eigenen Messungen nicht gegeben. So wäre eine Vielzahl äquivalenter Items zu Lasten des Befragungsumfanges gegangen. Damit wäre die Dauer der Befragung den Befragten nicht mehr zumutbar gewesen oder die Erhebung hätte sich thematisch auf Einzelaspekte fokussieren müssen. Um mit den Befragungen dennoch möglichst genaue Messungen zu erzielen, wurden Maßnahmen angewandt, wie sie zur Validitätssicherung beschrieben werden (Krebs & Menold, 2019, S. 496). So wurden vorab die Untersuchungsdimensionen als Hilfstheorie genau beschrieben, die verwendeten Indikatoren benannt, deren Auswahl präzise begründet und vermutete Zusammenhänge transparent dargestellt (siehe *Anhang 2 Operationalisierung der Fragestellungen* im elektronischen Zusatzmaterial). Zudem gelten Indizes, wie sie in der Ergebnisdarstellung angewandt wurden, für die Abbildung von „multidimensionalen Sachverhalten“ (Häder, 2010, S. 92) als geeignet. So kann ein komplexes Konstrukt in Form partieller Einzelmessungen valider erfasst werden als durch einen Einzelindikator und führt durch die ausgleichende Wirkung zufälliger (jedoch nicht systematisch verzerrender) Messfehler zu einer höheren Reliabilität der Messung (Latcheva & Davidov, 2019, S. 896 f.). Damit in die Auswertung nur Indikatoren eingehen, die in der Lage sind, die Untersuchungsdimensionen stellvertretend abzubilden, wurden die theoretisch entwickelten Indizes vor der Auswertung einer genauen Prüfung unterzogen.

Reliabilität

Die Reliabilität drückt aus, inwieweit Ergebnisse unter gleichen Bedingungen reproduzierbar sind, sprich das gleiche Messresultat erzielen. Methodisch kann diese mit wiederholten Messungen, mit parallel eingesetzten oder gesplitteten Erhebungsinstrumenten und den Korrelationen zwischen den Ergebnissen bestimmt werden (Häder, 2010, S. 109). Angesichts des erheblichen Aufwandes konnten diese Verfahren in der vorliegenden Arbeit nicht angewandt werden. Lediglich bei dem Index „Informationsqualitäten“, der aus reflektiven Indikatoren besteht, konnten Reliabilitätsanalysen durchgeführt werden, die Weiterentwicklungen der gesplitteten Testhalbierung sind (Diekmann, 2012, S. 250). Gearbeitet wurde dazu mit

Cronbachs α als Ausdruck der internen Konsistenz einer Skala und der Trennschärfe (Item-Skala-Korrelation).

(Alltagssprachliche) Repräsentativität und externe Validität

Unter dem Stichwort repräsentative Studie werden u. a. Zufalls- oder Quotenstichproben bezeichnet oder Erhebungsrückläufe, die im Hinblick auf bestimmte Merkmale einer definierten Grundgesamtheit entsprechen (Diekmann, 2012, S. 430). Das heißt, diese merkmalsbezogen repräsentieren, wobei sich die Frage stellt, welche Merkmale berechtigterweise herangezogen werden sollen. Wissenschaftlich betrachtet bleibt somit unklar, was mit dem Ausweis von Repräsentativität genau bezeichnet wird. Der Begriff entstammt eher einem alltäglich-laienhaften als einem statistischen Sprachgebrauch (von der Lippe, 2011, S. 3). Auch Schnell et al. (2013, S. 297 f.) plädieren dafür, in Studien der empirischen Sozialforschung nicht einen vermeintlichen Repräsentativitätsnachweis durchzuführen, sondern Stichproben hinsichtlich der methodischen Umsetzung genau zu beschreiben und anhand des Ablaufprozesses zu beurteilen.

Methodologisch gesprochen wird die Frage danach, ob Forschungsergebnisse über eine Stichprobe hinaus für eine Grundgesamtheit verallgemeinert werden können, unter dem Gütekriterium der externen Validität thematisiert (etwa Krebs & Menold, 2019, S. 500). Die externe Validität eines Forschungsdesigns wird durch Rücklaufverzerrungen und unnatürliche Untersuchungssituationen negativ beeinflusst (zur Diskussion siehe etwa Bortz & Döring, 2016, S. 104 ff.). Demgemäß gilt es, die Durchführung und den erzielten Rücklauf der quantitativen Erhebungsphase noch einmal genau zu reflektieren.

Reflexion der Durchführung und des Erhebungsrücklaufs

Die Grundgesamtheit der quantitativen Erhebung waren alle Hauptstandorte der anerkannten WfbM in Deutschland. Als Vollerhebung wurden alle Standorte angeschrieben. Die Auswahlwahrscheinlichkeit lässt sich somit berechnen und war für alle Hauptstandorte gleichermaßen gegeben. An den Hauptstandorten wurden Werkstattleitungen, Werkstattträte und Vertrauenspersonen zur Teilnahme an den drei separaten Befragungen eingeladen. Durch die Zugriffsbeschränkungen und die pseudonymisierten Fragebögen konnte der Rücklauf durchgängig kontrolliert und Mehrfachteilnahmen pro anerkannter Werkstatt ausgeschlossen werden.

Als ungünstig kann für die Teilstudien 2 und 3 der Wechsel des Befragungsmediums von der postalischen Einladung hin zur Online-Befragung gelten. So wurde die händische Eingabe der Befragungslinks (www.soscisurvey.de/Werkstattleitungen bzw. www.soscisurvey.de/

Werkstattrat-VP) als schwierig beurteilt. Internetzugänge können als Grundlage der Teilnahme (Bandilla et al., 2009, S. 130) in den beruflichen Kontexten der Befragungsgruppen jedoch als gegeben angenommen werden. Ein offensichtlicher systematischer Fehler ist in der Durchführung nicht zu erkennen. Im Ergebnis sind sehr gute (Teilstudie 1) und gängige (Teilstudie 2 und 3) Ausschöpfungsquoten erreicht. Die Ausschöpfungsquote für Teilstudie 2 beruht auf einem Näherungswert, der berücksichtigt, dass nicht alle Werkstatträte mit Vertrauenspersonen arbeiten.

In den realisierten Stichproben der Teilstudien 1 und 3 zeigt sich hinsichtlich der Bundesländer mehrheitlich eine gute Übereinstimmung mit der Grundgesamtheit. Lediglich für Bayern liegt eine Übererfassung und in der Teilstudie 3 eine Untererfassung der Werkstätten in NRW und Thüringen vor. Auf eine korrigierende Gewichtung dieser Rücklaufmerkmale wird dennoch verzichtet, da solche Korrekturverfahren letztlich immer strittig bleiben (Schnell et al., 2013, S. 305 ff.). Stattdessen gilt es, potenzielle Verzerrungen für Bayern (und NRW bzw. Thüringen in Teilstudie 3) bei Anteils- und Mittelwerten zu betrachten (Diekmann, 2012, S. 425). Merkmalsbezogene Ländervergleiche können dann darüber Auskunft geben, ob länderspezifische Besonderheiten feststellbar sind.

Abgesehen von den Bundesländern bestehen keine weiteren belastbaren Angaben zum Abgleich der Rückläufe mit der Grundgesamtheit. Es sind lediglich indirekte Vergleiche für die Teilstudien 1 und 3 mit Angaben möglich, die teilweise veraltet sind oder mit divergierenden Erfassungsweisen operieren. Die indirekten Vergleiche zeigen, dass zumindest keine sehr großen Abweichungen bestehen. Die erzielten Rückläufe spiegeln die typischen WfbM-Konstellationen wider. Offen bleibt, ob in Teilstudie 1 eine höhere Beteiligung von Lebenshilfe Werkstätten vorliegt, als diese in der Grundgesamtheit tatsächlich gegeben sind. Merkmalsbezogene Vergleiche können auch hier zeigen, ob durch eine eventuelle Übererfassung Verzerrungen entstehen. Bei den betrachteten Merkmalen gilt es zu beachten, dass ein „Non-response-Bias ... eine merkmalspezifische Größe“ (Koch, 1998, S. 67) bleibt. Das heißt, während bei einer Variablen keine Verzerrung zwischen den antwortenden und den nicht-antwortenden Einrichtungen oder Personen feststellbar ist, kann eine solche bei anderen Merkmalen vorliegen oder unentdeckt bleiben (ebd.).

Angeichts der skizzierten Erhebungsdurchführung und der diskutierten Merkmale kann angenommen werden, dass die Erhebungsrückläufe aussagekräftige Ergebnisse über Werkstattratsarbeit an den Hauptwerkstätten der anerkannten WfbM in Deutschland liefern und zumindest Teilstudie 1 eine ausreichende externe Validität aufweist, sodass sich die Ergebnisse

für die Grundgesamtheit verallgemeinern lassen. Zu beachten ist, dass sich mit der späteren Präzision der Indizes auf die Gremien mit regelhaftem Wahlturnus (siehe Abschnitt 5.3.8) die Grundgesamtheit verändert und sich die statistischen Analysen auf diese Werkstattträte beziehen.

Im Rückblick wäre es für die Teilstudie 3 günstiger gewesen, a priori eine verlässliche Fallzahlplanung festzulegen und diese mit persönlichen Nachfassaktionen zu erzielen. Dies konnte a posteriori nicht realisiert werden, da aus Anonymitätsgründen nicht dokumentiert war, welcher Zugangscode an welche Werkstatt ging. Angesichts der Fallzahl wird bei den Daten der Teilstudie 3 nicht mit statistischen Hypothesenprüfungen gearbeitet. Damit entfallen die geplanten Zusammenhangsanalysen zwischen der Anzahl der Werkstattbeschäftigten und dem Aktivitätsniveau bzw. der autonomen Arbeitsweise eines Werkstatttrats und zur Ausstattung eines Werkstatttrats. Ersatzweise kann jedoch auf die Anzahl der Mitglieder in einem Werkstatttratsgremium zurückgegriffen werden, da diese Angabe aus Teilstudie 1 vorliegt.

5.4 Verknüpfung der qualitativen und quantitativen Erhebungsergebnisse

Wie in den vorangegangenen Abschnitten dargestellt, wurden die Daten der qualitativen und quantitativen Erhebungsphasen separat erhoben, ausgewertet und analysiert. Darauf aufbauend galt es, die Erhebungsergebnisse miteinander in Beziehung zu setzen, was in der Mixed-Methods-Forschung als Integration bezeichnet wird. Studien, die eine Ergebnisintegration unterlassen, sind nach Kuckartz (2018, S. 168) keine echten Mixed-Methods-Studien. Die Form der Integration wird im Kern durch drei Faktoren bestimmt: „die Begründung (‘justification’) für die Wahl eines Mixed-Methods-Ansatzes, das Mixed-Methods-Sampling und das Mixed-Methods-Design“ (ebd., S. 164).

In der vorliegenden Arbeit wurde ein Mixed-Methods-Ansatz gewählt, um die Forschungsfragen adäquat zu beantworten, den Forschungsgegenstand umfassend zu beleuchten und um passende quantitative Instrumente zu konstruieren. Diese Gründe werden als Expansion und Entwicklung bezeichnet (ebd., S. 161). Mit ihnen und dem gewählten sequenziell angelegten Design war die erste Integration bereits vorgegeben: Die qualitativen Ergebnisse wurden dazu verwendet, die Fragen der quantitativen Erhebung und die Untersuchungsdimensionen zu spezifizieren, zu operationalisieren und Hypothesen aufzustellen.

Bezogen auf das realisierte Sampling waren die Stichproben in den Forschungsphasen nicht identisch. Wenngleich diese den gleichen Personenkreisen angehörten, waren die Erhebungen nicht so angelegt, dass eine fallbezogene Zusammenführung der qualitativen und

quantitativen Daten möglich war. Eine datenbasierte Integrationsstrategie (ebd., S. 172) war somit ausgeschlossen.

Um dennoch die Ergebnisse in Verbindung zu bringen, wurde eine resultatsbasierte Integration verfolgt, bei der die qualitativen und quantitativen Daten gemeinsam analysiert und interpretiert werden (ebd., S. 169 f.) Dargestellt ist dies in Kapitel 8. Den Bezugsrahmen bildeten die übergeordneten Forschungsfragen der Arbeit sowie die vorgenommene tabellarische Operationalisierung der Fragestellungen für die quantitative Erhebung (siehe *Anhang 2 Operationalisierung der Fragestellungen* im elektronischen Zusatzmaterial). In dieser Tabelle sind die qualitativen Ergebnisse und die dazu gebildeten Indikatoren und Variablen aufgelistet und miteinander in Verbindung gebracht, so wie dies als Vorstufe zur Ergebnisintegration von Verallgemeinerungsstudien empfohlen wird (Kuckartz, 2014, S. 116). Eine zunächst begonnene resultatsbasierte Integration mittels MAXQDA und thematischem Codieren (ebd., S. 128) wurde angesichts des erheblichen Aufwandes und der fehlenden Struktur wieder verworfen. Stattdessen wurde die Gegenüberstellung der Ergebnisse als Strategie gewählt (ebd., S. 170).

5.5 Reflexion der realisierten Mixed-Methods-Studie

Der Ausgangspunkt von Mixed-Methods-Studien liegt darin, Methodendesigns konsequent an Forschungsfragen auszurichten. Methodische Nachteile von quantitativer und qualitativer Forschung können berücksichtigt und deren Stärken so verbunden werden, dass eine Forschungsfrage valide und umfassend beantwortet werden kann. Um zu erörtern, inwieweit dies letztlich gelungen ist, sind zunächst die qualitativen und quantitativen Forschungsbestandteile im Hinblick auf die tradierten Gütekriterien dieser Forschungsstränge zu diskutieren (Bortz & Döring, 2016, S. 114). Dies wurde in den Abschnitten 5.2.7 und 5.3.9 vollzogen. Zudem gilt es, die Durchführung der Mixed-Methods-Studie und deren Ergebnisse insgesamt in den Blick zu nehmen. Hierzu lässt sich auf Empfehlungen und Richtlinien zurückgreifen, die für Mixed-Methods-Forschung vorliegen. Die Ausarbeitung von eigenständigen Gütekriterien für diesen Forschungsstrang ist noch in der Entwicklung und wird als Zukunftsaufgabe begriffen (Baur, Kelle & Kuckartz, 2018, S. 16; Kuckartz, 2014, S. 161; zur Diskussion multipler Validitätsformen von Mixed Methods siehe z. B. Johnson, 2014, S. 41 ff.).

Tabelle 29 zeigt die Fragestellungen und Ziele der Mixed-Methods-Studie und der qualitativen und quantitativen Erhebungsphasen und fasst bestehende Hürden bzw. die betrachteten Qualitäts- und Gütekriterien von qualitativer und quantitativer Forschung zusammen. Die

Tabelle verdeutlicht die aufwendigen Arbeiten, die mit Mixed Methods verbunden sind. So galt es, mehrere Erhebungsdesigns auszuarbeiten, zu begründen und festzuzurren, mehrere Erhebungsphasen zu durchlaufen, mehrere Erhebungsstränge auszuwerten etc. In der Richtlinie „Best practices for mixed methods research in the health sciences“ (Creswell, Klassen, Plano Clark & Smith, 2011) werden die aufwendigen Arbeiten als Ressourcenproblem geschildert (ebd., S. 8). Ein intensiver Zeiteinsatz und die für Mixed Methods benötigten Fertigkeiten in qualitativer, quantitativer und Mixed-Methods-Forschung werden benötigt. In dem eigenen Vorhaben konnten die benötigten Zeitressourcen zur Fertigungsaneignung und zur Forschungsdurchführung erst mit der Nutzung einer Promotionsförderung durch die Hochschule Fulda geschaffen werden.

Eine weitere Herausforderung lag darin, mit der erzielten Datenfülle umzugehen. Dies ist gleichfalls ein typisches Problem von Mixed-Methods-Studien. Von Creswell et al. (2011) wird dies unter dem Stichwort „page and word limitations“ (S. 9) thematisiert. Gelöst wurde dieses in der eigenen Studie dahingehend, dass – wie in Abschnitt 5.3.8.1 dargestellt – vereinzelte Aspekte (primär offene Fragen der quantitativen Studie) im Rahmen der vorliegenden Arbeit nicht verwendet wurden.

Besonders anspruchsvoll ist in Mixed-Methods-Forschung die Integration der qualitativen und quantitativen Studienbestandteile. Bei dem gewählten qualitativ-quantitativen Vertiefungsdesign können bspw. bei der Überführung der Ergebnisse in die darauffolgende Untersuchungsphase Validitätsprobleme entstehen. Die Überführung ist inhaltlich sachgerecht und methodisch adäquat auszufüllen (Creswell & Zhang, 2009, S. 615). Um mit dieser Herausforderung umzugehen, wurde die Transformation der qualitativen Ergebnisse in manifeste Untersuchungsvariablen jeweils explizit begründet und die Auswahl mithilfe von Literaturhinweisen und bewährten Indikatoren weiter untermauert. Integrationsstrategien sind eine Übungssache und „ein wichtiger Gradmesser für den Reifegrad des Mixed-Methods-Ansatzes“ (Kuckartz, 2018, S. 180). Gleichzeitig sind diese ein vergleichsweise junges Thema (ebd., S. 158 f.). Es liegen erst wenige Publikationen und methodische Anwendungshinweise dazu vor. So war in der eigenen Arbeit die Findung einer geeigneten Strategie für die resultatsbasierte Integration zunächst schwierig. Wie bereits angesprochen, wurde die anfänglich getestete Vorgehensweise mittels MAXQDA und thematischem Codieren (ebd., 2014, S. 128) wieder verworfen, da dies in einem feingliedrigen, aufwendigen Prozess gemündet hätte, der in einem ungünstigen Verhältnis zu dem erwarteten Ergebnis gestanden hätte. Stattdessen

wurde eine komprimierte Integration verfolgt, um mit der Fokussierung auf die Forschungsfragen zentrale Ergebnisse zu analysieren und eine kompakte Interpretation zu ermöglichen.

Tabelle 29: Fragestellungen, Ziele und Hürden bzw. Qualitäts-/Gütekriterien der Mixed-Methods-Studie und der Erhebungsphasen

Fragestellungen	Ziele	Hürden, Qualitäts-/Gütekriterien
Mixed-Methods-Studie – exploratives sequenzielles Design (Verallgemeinerungsmodell)		
<ul style="list-style-type: none">• Wie werden die gesetzlichen Vorgaben der WMVO und die Beteiligungsrechte der Werkstatträte in den WfbM auf praktischer Ebene ausgestaltet?• Was charakterisiert eine gelingende Werkstattarbeit und womit steht diese im Zusammenhang?	<ul style="list-style-type: none">• Empirische Beantwortung der Forschungsfragen durch deren Spezifizierung in einer qualitativen und einer quantitativen Erhebung.• Konstruktion standardisierter Befragungsinstrumente auf Basis qualitativer Ergebnisse.• Gemeinsame Ergebnisanalyse/-interpretation.	<ul style="list-style-type: none">• Hoher Ressourceneinsatz.• Integration der qualitativen und quantitativen Studien.• Erzielter Erkenntnisgewinn als Maßstab (Kuckartz, 2014, S. 164).
Qualitative Erhebungsphase – Momentaufnahme		
<ul style="list-style-type: none">• Wie werden die Verordnungsvorgaben und die Beteiligungsrechte der Werkstatträte in den WfbM ausgestaltet?• Was charakterisiert eine gelingende Werkstattarbeit und womit steht diese im Zusammenhang?	<ul style="list-style-type: none">• Erhebung relevanter Bedingungen zur Arbeit von Werkstatträten aus Sicht der Beteiligten.• Ermittlung von Vorstellungen und Faktoren einer gelingenden Werkstattarbeit.• Hypothesengenerierung zur Entwicklung standardisierter Erhebungsinstrumente.	<ul style="list-style-type: none">• Nachvollziehbares Vorgehen.• Angemessenheit.• Relevanz der Thematik.• Bewusste Begrenzung (Steinke, 2017, S. 323 f.).
Quantitative Erhebungsphase – Querschnittstudien, Ex-post-facto-Anordnung		
<ul style="list-style-type: none">• Wie wird Werkstattarbeit organisiert und ausgestaltet?• Wie aktiv sind die Gremien?• Wie autonom und durchsetzungstark sind diese?• Welche Zusammenhänge bestehen zwischen Kontextfaktoren und dem Aktivitätsniveau bzw. der autonomen Arbeitsweise eines Werkstatttrats?	<ul style="list-style-type: none">• Beschreibung der Umsetzung der formalisierten Beteiligung in WfbM.• Untersuchung von Merkmalen einer aktiven, autonomen und durchsetzungstarken Werkstattarbeit.• Ermittlung von begünstigenden Faktoren.• Berücksichtigung unterschiedlicher Perspektiven.	<ul style="list-style-type: none">• Reliabilität, Objektivität, Validität (insbesondere externe Validität des Forschungsdesigns).• Reflexion der Erhebungsdurchführung und des erzielten Erhebungsrücklaufs.

Festgehalten werden kann, dass sich in Mixed-Methods-Forschung die zu treffenden Auswahlentscheidungen potenzieren. Generell sind Auswahlprozesse in Forschungsabläufen keine leichtfertigen Angelegenheiten. Notwendig sind diese dennoch bspw. zur Fall-, zur Materialauswahl und bei der Ergebnisdarstellung (Flick, 2016, S. 154 ff.). Um den vielen Auswahlentscheidungen reflexiv zu begegnen, wäre die Arbeit im Team vorteilhaft gewesen. Da dies nicht möglich war, wurde ersatzweise auf das Literaturstudium, den Besuch von Methodenseminaren und fachliche Austauschgespräche zurückgegriffen, um reflexiv und methodisch angemessen zu handeln. So waren die Auswahlentscheidungen sehr langwierige Prozesse.

Letztlich muss sich „die Sinnhaftigkeit eines Mixed-Methods-Design[s] am Erkenntnisgewinn für die Forschungsfrage erweisen“ (Kuckartz, 2014, S. 164). Im Hinblick darauf war der Ansatz die richtige Wahl (siehe Abschnitte 4.1 bis 5.3). Erst das Zusammenspiel der qualitativen und quantitativen Forschungsbestandteile erlaubte es, die Forschungsfragen adäquat zu behandeln und einen verstehens- und wissensvertiefenden Mehrwert zu schaffen.