5. Vom theoretischen Modell zum Untersuchungsdesign

Im vorangegangen Kapitel wurde die theoretische Argumentation über die administrative Performanz sowie für den Wirkungszusammenhang zwischen administrativer Performanz und möglichen Einflussfaktoren entwickelt. In diesem Kapitel geht es nun darum, diese theoretischen Ausführungen für die empirische Untersuchung aufzubereiten. Entsprechend werden im Detail nachstehende Punkte spezifiziert und etwaige Probleme in der Umsetzung diskutiert: (1) Design der empirischen Untersuchung, (2) Begründung der Fallauswahl, (3) Datengrundlage, (4) Operationalisierung des theoretischen Modells und (5) Methoden der Datenerhebung.

5.1 Untersuchungsdesign

Der unbefriedigende Forschungsstand zu den Determinanten administrativer Performanz wird wesentlich durch Methoden- und Fallauswahl verursacht. Bisherige empirische Studien können wie folgt klassifiziert werden: (a) *Einzelfallbetrachtungen* oder *qualitative Vergleiche* häufig mit einem inhaltlichen *Reformfokus*. (b) Länder- oder aufgabenübergreifende *Querschnittsvergleiche* mit weitgehend *unkontrollierten Drittvariablen*. Beide Zugänge können wertvolle Beiträge zur Organisations- und Verwaltungsforschung liefern. Wie zu zeigen sein wird, sind sie jedoch zur *Erklärung* administrativer Performanz eher ungeeignet. Aufgabe und Voraussetzung einer erfolgreichen Analyse ist folglich, die Ursachen dieser Defizite zu erkennen und ein sie vermeidendes Forschungsdesign zu entwickeln.

5.1.1. Kritik der performanzorientierten Organisationsforschung

Qualitative Vergleichsdesigns sind der am weitesten verbreitete methodische Ansatz der verwaltungswissenschaftlichen Organisationsforschung. Noch stärker als in der Politikwissenschaft (vgl. Naßmacher 1991; Peters 1998; Landmann 2000) haben sie sich hier als unentbehrlich für die Theorieentwicklung gezeigt. Allerdings ist eine wesentliche methodische Erkenntnis aus den vorangegangenen umfangreichen *vergleichenden* Fallstudien über Verwaltungsregime und ihre Performanz (vgl. Bauer et al. 2007a; Kuhlmann et al. 2011), dass ein extern valider Test der aus Fallstudien und Literatur abgeleiteten Hypothesen, ein dedukti-

https://doi.org/10.5771/9783845247137-130, am 09.10.2024, 02:22:58
Open Access – (cc) BY-NC-ND – https://www.nomos-elibrary.de/agb

ves, quantitativ angelegtes Forschungsverfahren erfordert. Fallstudien ermöglichen zwar detaillierte Fallbeschreibungen und -vergleiche; die Generalisierbarkeit der Befunde bleibt jedoch eng begrenzt, da lediglich für das gewählte, zwangsläufig kleinzahlige Sample valide Aussagen getroffen werden können.

Mit der Entscheidung für ein quantitativ orientiertes Forschungsverfahren verlässt diese Untersuchung den *mainstream* der qualitativ orientierten klassischen Organisationsforschung, ohne die Bedeutung dieser Ansätze zu negieren. Im Gegenteil: Erst durch diese Vorarbeiten und die dabei gewonnenen Detailkenntnisse über Politikinhalte, Arbeitsweisen und Machtverhältnisse der vollziehenden Verwaltung konnte eine informierte Hypothesenbildung, eine präzise Spezifikation des Untersuchungsmodells und eine valide Interpretation der Befunde erfolgen. Die zentrale Herausforderung eines quantitativ-hypothesentestenden Designs liegt darin, jene Schwierigkeiten quantitativer Untersuchungen zu überwinden, welche als ursächlich für die weite Verbreitung fallorientierter qualitativer Ansätze anzusehen sind. Insbesondere muss es darum gehen, die Problematik einer regelmäßig großen Zahl an zu berücksichtigenden Variablen bei gleichzeitig einer sehr geringen Zahl an tatsächlich vergleichbaren Organisationen zu lösen (vgl. Lijphart 1971: 685-690; Wollmann 2004b; Wollmann/Kuhlmann 2011). Hierzu sind methodisch kreative Lösungen gefragt.

5.1.1.1 Kritik des fallzentrierten Ansatzes

Insbesondere die anwendungsorientierte Verwaltungsforschung vergleicht Vollzugsarrangements oder Behördentypen miteinander. Konkret werden in der Regel Vergleiche entlang grundlegender Strukturparameter durchgeführt: staatlich vs. privat, staatlich vs. kommunal oder auf staatlicher Ebene bspw. zwischen Oberen Landesbehörden, Mittelbehörden, Unteren Landesbehörden). Problematisch an dieser Vorgehensweise ist, dass jede dieser Vergleichskonstruktionen unterschiedliche Kombinationen potenziell erklärender Merkmale in sich trägt. So sind bspw. Untere Landesbehörden gleichzeitig staatliche, spezialisierte und dekonzentrierte Einheiten. Eine in diesen äußeren Strukturen begründete Benennung potenzieller Erklärungsfaktoren liegt zwar häufig vor (z. B. Skalen- und Verbunderträge, Ortsnähe, Politiknähe, Wettbewerbsdruck). Eine belastbare Zuordnung ihres individuellen Einflusses einzelner struktureller Faktoren auf beobachtete Performanzeffekte kann jedoch auf dieser Basis nicht erfolgen. Dies gilt umso mehr, als davon ausgegangen werden muss, dass die einzelnen Erklärungsfaktoren gegenläufige, sich potenziell aufhebende Effekte entwickeln. Die mit dieser Tatsache einhergehende Komplexität kann dafür verantwortlich gemacht werden, dass auf eine empirische Überprüfung der theoretisch postulierten Effekte häufig gänzlich verzichtet wird (vgl. Gornas 1993; Hesse 2003, 2007).

https://doi.org/10.5771/9783845247137-130, am 09.10.2024, 02:22:58
Open Access – (cc) BY-NC-ND - https://www.nomos-elibrary.de/agb

Darüber hinaus dominieren qualitative Einzelfallstudien oder Vergleiche weniger Länder diesen angewandten Strang der Verwaltungsforschung. Querschnittsvergleiche weisen teilweise keine politikfeldspezifische Ausrichtung aus (vgl. Boyne 1995; Dollery et al. 2007, 2008). Dies führt dazu, dass letztlich im Nationalstaat zu verortende Besonderheiten (Verwaltungskultur, Pfadabhängigkeiten, Vetospieler und Machtallokation) in den Fokus der Analysen rücken und ihnen ein entsprechend großer Anteil der erklärten Varianz zugeschrieben wird (z.B. Jann 1983; Bothe 1986; Kuhlmann et al. 2011).

5.1.1.2 Ruhm und Elend des quasi-experimentellen Längsschnittvergleichs

Ein häufig gewählter Ansatz zur Überwindung der Beschränkungen einfacher Fallstudiendesigns sind an das Prinzip eines *Längsschnittvergleichs* angelehnte *Quasi-Experimente* oder *natürliche Experimente* (vgl. Shadish et al. 2002). Da Fragen der Staatsorganisation nur selten tatsächliche Experimente zulassen, werden hierbei einschneidende Veränderungen, sog. *Interventionen* in Umfeld, Auftrag, Organisation oder Arbeitsweise von Organisationen genutzt, um die Wirkung dieser Modifikationen isoliert betrachten zu können. Wird dieser diachrone Ansatz hypothesengeleitet und mit einer durchdachten Fallauswahl eingesetzt, lassen sich so die Probleme von Fallstudien in Generalisierbarkeit und Kausalitätszuweisung zumindest teilweise überwinden (vgl. Kuhlmann/Wollmann 2011: 484). Entsprechend häufig wird dieser Ansatz in Implementations- und Evaluationsstudien, insbesondere zur Dokumentation und ex-post Bewertung von Reformvorhaben und ihrer Performanzeffekte eingesetzt (vgl. bspw. Büchner/Franzke 2001; Bogumil/Ebinger 2005).

Dabei erweist sich diese Herangehensweise als nicht unproblematisch: Die Beobachtung der Veränderungen ist aufgrund ihrer Spezifität meist auf eine nur geringe Anzahl von Untersuchungseinheiten beschränkt. Zudem sind die analysierten Modifikationen extern vorgegeben und es treten auch hier regelmäßig sich gegenseitig beeinflussende Maßnahmenbündel auf, so dass Kausalitätsfragen oft nicht ausgeräumt werden können. Pollitt (2000) fasste in einer frühen Methodenkritik die Schwächen von Evaluationsstudien am Beispiel der zahlreichen Analysen der *Public Management* Reformen der 1990er Jahre zusammen. Die folgenden, von ihm identifizierten Defizite sind hierbei unmittelbar auf Schwierigkeiten bei der Verwendung quasi-experimenteller Designs zurückzuführen (Pollitt 2000: 187):

- Ein fehlender Messpunkt *vor* dem Eintreten der Veränderung, so dass kein tatsächliches *pretest-posttest design* vorliegt und über Veränderungen nur mehr oder weniger abgesichert spekuliert werden kann.

https://doi.org/10.5771/9783845247137-130, am 09.10.2024, 02:22:58
Open Access – (cc) BY-NC-ND – https://www.nomos-elibrary.de/agb

- Ein fehlender Abgleich mit nicht von den Veränderungen betroffenen Einheiten, so dass von Drittvariablen verursachte Veränderungen nicht identifiziert werden können.
- Die Vernachlässigung der Transaktionskosten.
- Die Vernachlässigung von Nebeneffekten wie bspw. Vertrauensverlust oder Konfusion hinsichtlich gültiger Wertvorstellungen.
- Die Vernachlässigung von spezifischen Kontextfaktoren, welche die Verallgemeinerbarkeit der Befunde beeinflussen könnten.
- Die Vernachlässigung der Parallelität von Reformmaßnahmen, so dass Attributionsprobleme nicht thematisiert werden.

Diese Aufzählung ist sehr instruktiv für die Einschätzung der Chancen und Risiken eines quasi-experimentellen Forschungsdesigns. Sie verdeutlicht, dass Ansätze, die auf die Beobachtung der Wirkung extern induzierter Veränderungen abzielen, sehr hohe methodische Anforderungen erfüllen müssen, um belastbare Ergebnisse zu liefern: Neben dem Vorliegen (a) eines eine *größere Zahl* von vergleichbaren Untersuchungseinheiten betreffenden und (b) *theoretisch relevanten* Veränderungsimpulses müssen (c) *zwei Messpunkte* – vor und nach der Reform – zur Verfügung stehen oder zumindest eine nicht schwerwiegend verfälschte Rekonstruktion des Status ante möglich sein. Zusätzlich müssen (d) nicht von der untersuchten Veränderung betroffene Vergleichsfälle zur *Kontrolle reformunabhängiger Dritteffekten* erfasst werden können. Darüber hinaus potenziert der Reformfokus (e) den Komplexitätsgrad der Betrachtung aufgrund der notwendigen Berücksichtigung von *Transformationskosten*, welche aufgrund ihres unterschiedlichen Charakters (von einmalig bis langfristig) nur schwer zu erfassen und darüber hinaus analytisch kaum von originären Performanzeffekten zu unterscheiden sind. Im Ergebnis verdeutlicht dieser Anforderungskatalog, dass die Beantwortung der hier gestellten deskriptiven und analytischen Forschungsfragen mit einem längsschnittorientierten, quasi-experimentellen Ansatz aus forschungspragmatischen Gründen nicht möglich sein wird.

5.1.1.3 Querschnittsvergleiche

Scheidet ein quasi-experimentelles Design aus, so könnte alternativ – insbesondere für einen quantitativen Hypothesentest – auf einen querschnittsorientierten Untersuchungsansatz im Sinne eines zeitgleichen (synchronen) Vergleichs unterschiedlich organisierter Untersuchungseinheiten zurückgegriffen werden. Aufgrund des im Verhältnis zu Längsschnittvergleichen deutlich geringeren zeitlichen und methodischen Aufwands, ist dieser Ansatz das dominierende Forschungsdesign in der komparativen Organisations- und Verwaltungsforschung.

https://doi.org/10.5771/9783845247137-130, am 09.10.2024, 02:22:58
Open Access – (cc) BY-NC-ND - https://www.nomos-elibrary.de/agb

Judge (1994) konstatiert, dass viele Untersuchungen zur Leistungsfähigkeit öffentlicher Verwaltungen eng spezifizierte und oft einseitig ökonomisch dominierte Performanzbegriffe verwenden. Diese Engführung wurde teilweise durch verpflichtende staatliche Programme zum *Performance Measurement* und *Performance Management* forciert (vgl. Lynn et al. 2001: 62; Talbot 2005; Hood 2006; Jann/Jantz 2008; Ritz/Sager 2010: 125f.). In der verwaltungswissenschaftlichen Organisationsforschung erwies sich die Verwendung derartiger „harter" *Output*-Daten aus vier Gründen als problematisch. Erstens verhindern selbst bei Untersuchungseinheiten mit sehr ähnlichen oder gar identischen Aufgaben unterschiedlich angewandte Methoden der Kosten- und Leistungsdefinition wie -erfassung häufig einen (validen) Vergleich. Tatsächlich vergleichbare *Output*-Daten über ein großzahliges Sample hinweg existieren äußerst selten.[97] Zweitens engt ein solcher ökonomischer Fokus den Performanzbegriff zu stark ein. Bogumil (2004: 393) erinnert daran,

> „[…] dass die meisten Ziele in öffentlichen Verwaltungen sich nicht in exakten Zahlen darstellen lassen, da sie immer mehrdimensional und häufig kompromisshaft entstanden sind. Leistungsmessungen […] neigen […] dazu, dem Mythos der quantitativen Messbarkeit zu verfallen."

Aus politikwissenschaftlicher Perspektive zentrale Performanzaspekte wie Zielerreichung, Leistungs- und Zugangsgerechtigkeit, Leistungsqualität und demokratische Zurechenbarkeit werden mit diesem Zugang ignoriert. Drittens kann angenommen werden, dass das Forschungsdesign derartiger Studien eher von der Datenverfügbarkeit als von theoretischen Erörterungen geprägt ist (vgl. Wollmann 2001: 26; Christensen/Lægreid 2001: 30, 2001a: 74). Damit könnte nicht mehr von einer theoriegeleiteten Auswahl der Vergleichsfälle gesprochen werden – womit eine der Grundannahmen politikwissenschaftlicher Querschnittsvergleiche verletzt wäre (vgl. Lehner/Widmaier 1995). Viertens gibt die alleinige Analyse von *Output*-Daten auch keine Auskunft über die Prozesse der Leistungserstellung. Die zu Erfolg oder Misserfolg führenden Mechanismen bleiben unentdeckt und fehlgeleitete Schlussfolgerungen aufgrund sich verstärkender oder gegenseitig aufhebender Erklärungsfaktoren wären wahrscheinlich.

Diese Schwierigkeiten führten dazu, dass andere Ansätze zur Entwicklung variablenzentrierter und hypothesentestender Forschungsdesigns gesucht wurden. Von besonderem Interesse ist in dieser Beziehung die bereits eingeführte Forschung zu Autonomie und Steuerung verselbständigter Behörden (Greve et al. 1999; Gilardi/Braun 2002; Bouckaert/Peters 2004; Pollitt et al. 2004; Verhoest et al. 2004). Hier finden sich innovative *large-n* Studien, die die statistische Ana-

97 Einen eindrucksvollen Versuch, eine derartige Datengrundlage durch weitgehende Kontrolle von Kontextfaktoren zu schaffen, stellen die vier Evaluationen der Experimentierklausel nach § 6c SGB II dar, s. zusammenfassend Deutscher Bundestag 2008.

https://doi.org/10.5771/9783845247137-130, am 09.10.2024, 02:22:58
Open Access – (cc) BY-NC-ND – https://www.nomos-elibrary.de/agb

lyse des Zusammenhangs zwischen Strukturen, Steuerungsformen und Verwaltungshandeln oder Performanz anstreben (vgl. bspw. Verhoest et al. 2010; Lægreid/Verhoest 2010; Ebinger/Schmitt 2010). Statt auf „harte" *Output*-Daten wird dabei auf *Perzeptionsdaten* zurückgegriffen, die über Surveys in den Behörden erfasst werden. Dies erlaubt nicht nur die Abdeckung eines breiten Spektrums an Performanzkonstrukten, sondern auch die Erfassung von Prozessdaten.

Die Erfahrung aus diesen Studien zeigt, dass diese Konzeption des Querschnittsvergleichs grundsätzlich zur Beantwortung der hier gestellten deskriptiven und analytischen Forschungsfragen geeignet ist. Allerdings bringt auch dieser Ansatz methodische Probleme mit sich. Wesentliche Voraussetzung für einen erfolgreichen Einsatz ist auch hier die Vergleichbarkeit der untersuchten Organisationen. Bei der Analyse von Verselbständigungseffekten wird aufgrund des Fokus auf möglichst idealtypische *Agenturen* die Unterschiedlichkeit der Rahmenbedingungen häufig ignoriert und oft länderübergreifend und/oder unabhängig von der konkret vollzogenen Aufgabe verglichen. Da angenommen werden kann, dass sowohl die politische Salienz der vollzogenen Aufgabe (vgl. Elder/Page 1998; Christensen/Lægreid 2003: 388; Gormley/Balla 2004; Egeberg/Trondal 2009; Ebinger/Schmitt 2010), als auch die in den Behörden gepflegte Verwaltungskultur (vgl. Jann 1983; Adler 1993; Hall/Taylor 1996; Howlett 2004; Yesilkagit 2004) Wahrnehmung und Handeln der Akteure beeinflussen, könnten derartige Vergleiche durch Unterschiede in nationaler oder aufgabenspezifischer Verwaltungskultur zu Verzerrungen führen, sollten sich die gelebten Rollenmodelle zwischen den Untersuchungseinheiten unterscheiden (vgl. Lægreid/Verhoest 2010: 9). Ein belastbares Forschungsdesign sollte diese Probleme thematisieren und möglichst abmildern.

5.1.1.4 Kritik der Führungskräftebefragung

Unmittelbar verknüpft mit dem im vorhergehenden Abschnitt diskutierten Einsatz von Perzeptionsdaten ist der dabei eingenommene Fokus: Untersuchungsgegenstand ist im Regelfall die Gesamtbehörde. Zur Gewinnung von Daten zu einzelnen Behörden wird häufig auf die Befragung einer Ebene (vgl. Lynn et al. 2000: 248), und zwar insbesondere der Behördenleitung zurückgegriffen (vgl. Meier/O'Toole 2001; Pautz 2008: 42). Die Wahl dieser Zielgruppe über diesen in der Elitenforschung entwickelten *Positionsansatz* (vgl. Herzog 1982: 98-106; Schwanke/Ebinger 2006: 228-232) wird damit begründet, dass Führungskräfte den besten Überblick über sämtliche Aspekte der ihnen unterstellten Behörde hätten (vgl. Snow/Hrebiniak 1980: 320). Die so gewonnenen Aussagen werden oft nicht als individuelle *Perzeption*, sondern als *Beschreibung* der organisationalen Realität auf einer Makro-Ebene übernommen. Diese Herangehensweise

https://doi.org/10.5771/9783845247137-130, am 09.10.2024, 02:22:58
Open Access – (cc) BY-NC-ND – https://www.nomos-elibrary.de/agb

birgt drei Problemlagen: Erstens verstärkt sie die Problematik der *vielen Variablen bei wenigen Fällen* (Scharpf 2000: 765f.). Indem lediglich ein Akteur pro Behörde befragt wird, ist aufgrund der geringen Fallzahl eine über einfache deskriptive Statistik hinausgehende quantitative Datenauswertung selten möglich. Zweitens ist die *Validität* der Aussagen von Führungskräften (oder mit dem Ausfüllen des Fragenbogens beauftragter Subalterner) zu hinterfragen. So arbeiten diese Akteure nicht selbst innerhalb der interessierenden Strukturen und Prozesse, sondern nehmen eine herausgehobene, durch Informationsasymmetrien gekennzeichnete Position ein. Gerade in größeren, multifunktional organisierten Behörden wie Mittelinstanzen, Kommunalverwaltungen und Sammelbehörden sind sie sehr weit von der praktischen Vollzugstätigkeit einzelner Bereiche entfernt und verfügen im Regelfall auch nicht über die zur *fachlichen* Einschätzung der Vollzugsleistung notwendige Expertise. In der Folge können ihre Wahrnehmungen stark von jenen nachgeordneter Verwaltungsmitarbeiter abweichen. Abgesehen davon kann ihnen ein starkes Interesse an der Kontrolle der nach außen gelangenden Informationen und einer möglichst positiven Darstellung der Behördentätigkeit unterstellt werden (vgl. Kumar 1993; Enticott et al. 2009: 231), da sie zu Recht oder zu Unrecht für die erreichte Vollzugsleistung verantwortlich gemacht werden. Vor diesem Hintergrund erscheint die Praxis, aus den individuellen Aussagen dieser speziellen Akteursebene vorbehaltlos eine Zustandsbeschreibung der Behörden auf einer Makro-Ebene herauszulesen, äußerst fragwürdig. Die Folgerung von Walker und Enticott (2004: 432), dass über derartige Elitesurveys gewonnene Daten mit Vorsicht zu genießen sind, wird hier geteilt.

5.1.2. Föderalismus als valides most similar systems design

Aus den genannten Problemen der häufig in der Organisationsforschung gewählter Ansätze lassen sich Anforderungen an ein geeignetes Untersuchungsdesign für die hier bearbeiteten Forschungsfragen ableiten:

- Größtmögliche Varianz auf den als theoretisch relevant erachteten strukturellen Erklärungsfaktoren,
- größtmögliche Homogenität oder Kontrolle relevanter Kontextvariablen,
- eine ausreichend große Fallzahl zur Anwendung multivariater Analysemethoden und
- Erfassung nicht nur der Performanz-, sondern auch von Prozessvariablen der internen Verwaltungsorganisation.

Um diese Anforderungen zu erfüllen, greift das gewählte Forschungsdesign auf eine simple, aber erfolgversprechende Strategie zurück: Den *subnationalen Institutionenvergleich*. Als Form des Intrasystem-Vergleichs wird hierbei eine Grup-

https://doi.org/10.5771/9783845247137-130, am 09.10.2024, 02:22:58
Open Access – (cc) BY-NC-ND – https://www.nomos-elibrary.de/agb