

## 2.3 Kategorisierung von Performanz

Der in der *Public Administration*- und Verwaltungsforschung allgegenwärtige Begriff der *organisationalen Performanz* beschreibt ein gesellschaftlich konstruiertes Phänomen mit variablem Inhalt (Anspach 1991; Reiter et al. 2009). Ob seiner Breite und unterschiedlichen Füllung ist das Konzept nicht unumstritten (vgl. Talbot 2000: 64; Reichard 2004: 341; Nullmeier 2004: 47). Schon 1974 beschrieb Derlien in seinem Aufsatz „Theoretische und methodische Probleme der Beurteilung organisatorischer Effizienz der öffentlichen Verwaltung“ eindrücklich den bereits damals langen Weg zur Entwicklung eines multidimensionalen Performanzbegriffs in der *Public Administration* Forschung (Derlien 1974: 1). Reding (1981: 18) konstatierte schon bald zynisch, dass die Verwendung des Effizienzbegriffs – hier nur eines von mehreren Unterkonstrukten unter dem Dach des Performanzbegriffs – mangels einer akzeptierten Konvention immer ineffizienter werde.

Auch die bundesdeutsche staatswissenschaftliche Organisationsforschung trieb in der Phase der sog. *Steuerungseuphorie* in den 1960er und 1970er Jahren (vgl. Scharpf 1973) die Ausdifferenzierung des Performanzbegriffs voran. Die systematische empirische Überprüfung der hier getroffenen Annahmen fand jedoch kaum statt. Dies änderte sich erst in den 1980er Jahren mit einer neuen Welle performanzorientierter Organisationsreformen. Dieses New Public Management (NPM) erfasste ausgehend von den konservativen Regierungen der angelsächsischen Staaten in diversen Transformationen die meisten OECD-Staaten (vgl. Christensen/Læg Reid 2001a). Mit der Zielsetzung, den öffentlichen Sektor zu reformieren und leistungsfähiger zu machen, wuchs die Notwendigkeit präziser Performanzdefinitionen. Ein konsequent umgesetztes, von betriebswirtschaftlichen Konzepten abgeleitetes Steuerungsmodell erfordert ein *performance management* mit klar definierten politischen Zielen, Produkten der Verwaltungstätigkeit und Kennzahlen über Kosten und Zielerreichung: Was sind öffentliche Leistungen? In welche einzelnen Leistungsdimensionen lassen sich diese zerlegen und welche davon verdienen Berücksichtigung oder gar Priorisierung? Und schließlich, aber alles andere als unwichtig: Wie können die entsprechenden Variablen operationalisiert und gemessen werden? Die Übersetzung politischer Zielsetzungen in tatsächlich steuerungsrelevante Indikatoren, die Entwicklung

reaching these goals. Employee self-reports have the virtue of being global measures of performance, and it is therefore reasonable to expect that such evaluations will include an appropriate mix of output- and outcome-based criteria. A similar employee-based measure of performance has been employed by others (Molnar and Rogers 1976; Mott 1972), most recently by Brewer and Selden (2000), who point out that while the organizational psychology literature argues that organizations and individuals are interdependent, the employee perspective on performance is underemployed.“

von Anreizsystemen und die anschließende Minimierung von nicht-intendierten Lerneffekten erwies sich als äußerst komplexe und teure Daueraufgabe (vgl. Power 1999; Talbot 2005; Jann/Jantz 2008). Auch scheinen die an dieser Schnittstelle zwischen wissenschaftlicher und praktischer Anwendung entstandenen Konzepte und Indikatoren häufig weniger theoretisch abgeleitet als vom Zugang zu Datenmaterial determiniert.

An Sammlungen von Performanzmerkmalen herrscht kein Mangel, sie finden sich unter den unterschiedlichsten Begrifflichkeiten: Beispielsweise spricht Ostrom (2011: 15f.) schlicht von „evaluative criteria“, Steiner (2002) von „Zieldimensionen“, De Peuter et al. (2011) von „government capabilities“, Boyne (2002) von „service performance“ und Kuhlmann et al. (2011) von „institutioneller Performanz“. Ebenso zahlreich und buntgemischt sind die Bewertungskriterien selbst. Einen frühen und bis heute beeindruckenden Versuch der umfassenden Definition von Verwaltungsleistung wählte Frido Wagener in seinem 1969 veröffentlichten Opus Magnum *Neubau der Verwaltung*. Er identifizierte in der damals schon umfangreichen Literatur 25 verschiedene, von ihm als *relevant* erachtete Bewertungskriterien der Verwaltungstätigkeit (Wagener 1969: 284ff.). Entsprechend des Prinzips *Form folgt Funktion* entwickelte er ausgehend von diesen Maßstäben und entlang theoretisch abgeleiteter Prämissen über optimale Vollzugsbedingungen die "richtige" Anlagerung für jede Aufgabe hinsichtlich Verwaltungstypus und Zuständigkeitsbereich. Als Ergebnis steht für jede Aufgabe die Vollzugsorganisation mit dem optimalen Ziel-Mittel-Verhältnis. Wagners theoriegeleitetes Konzept klammert explizit die Frage der Umsetzungschancen eines solchen *Neubaus* der Verwaltung aus. Ihm ging es darum, durch die Zusammenführung des verfügbaren Wissens über Ziele und Praxis des Verwaltungsvollzugs eine präzise Blaupause der idealen Verwaltung zu entwerfen. Zwar verschoben sich in den vergangenen gut 40 Jahren das Aufgabenportfolio der öffentlichen Verwaltung und die technischen Maßstäbe der *idealen* Aufgabenerledigung aufgrund sozio-ökonomischer und technischer Entwicklungen deutlich. Seinen Charme hat das Konzept aufgrund seiner radikalen Fokussierung auf die Aufgabenerledigung und seiner ausdifferenzierten Sammlungen von Performanzaspekten dennoch bewahrt.

Spätere Autoren wählten oft noch breiter gefasste Performanzkonstrukte, konkretisierten diese jedoch nicht so stark. So nennen Eichhorn und Siedentopf (1976: 31) als Ziele von Verwaltungsorganisation „Sparsamkeit“, „Arbeitszufriedenheit“, „Einfachheit und Übersichtlichkeit“, „Qualität der Aufgabenerfüllung“, „Verwaltungsexterne Ersparnisse“ sowie „Politische Integration“. Ellwein (1993: 159 f.) listet „Rechtmäßigkeit, Zweckmäßigkeit, Wirtschaftlichkeit, Bürgernähe, Dezentralisierung, Aufgabenbündelung sowie Willensbildung und Kontrolle“. Steiner (2002: 279 ff.) nennt mit Blick auf die Optimierung der Verwaltungsstruktur schweizer Gemeinden die ähnlichen Zieldimensionen „Verbesser-

„Verbesserung des Dienstleistungsangebots“, „Politische Integration“, „Gemeindeautonomie“, „Anpassungsfähigkeit“, „Führbarkeit/Begrenzung des Koordinationsaufwands“ und „Motivation der Mitarbeitenden“. Elinor Ostrom (2011: 15f.) fokussiert auf „(1) economic efficiency, (2) equity through fiscal equivalence, (3) redistributive equity, (4) accountability, (5) conformance to general morality, and (6) adaptability.“ Umfassende Sammlungen und Diskussionen dieser Bewertungskriterien finden sich in Heffron (1989) und Helbing (1998). Die vorausgehende kursorische Sichtung zeigt, dass sich viele der gewählten Konzepte stark überdecken. Darüber hinaus wird deutlich, dass häufig Aspekte der *operativen* Verwaltungsleistung mit politischen und strukturellen Aspekten vermischt werden. Letztere gehören jedoch bei der hier angestrebten Betrachtung der *Vollzugsleistung* zu den *erklärenden*, nicht den *erklärten* Variablen.

### 2.3.1. Performanzdimensionen

Da die Zahl der potenziell zu berücksichtigenden Einzelkriterien unüberschaubar ist, muss es zunächst um eine Grobgliederung potenzieller *Performanzdimensionen* gehen. Viele Autoren schlagen zur Beherrschung der Vielzahl an unterschiedlichen Leistungsaspekten die inhaltliche Gliederung in wenige grundlegende Performanzdimensionen vor, die lediglich die äußere Performanz der Verwaltungstätigkeit berücksichtigen. Diese Gliederungsmodelle orientieren sich an den bereits betrachteten Prozessstufen der Aufgabenerstellung – bspw. *Input*, *Output*, *Outcome* (vgl. Kapitel 2.2; Budäus/Buchholtz 1997; Kuhlmann et al. 2011: 29f.)<sup>24</sup>, an inhaltlichen Zielsetzungen oder Funktionen (vgl. Eichhorn 1993: 867; Hood 1991: 12ff.)<sup>25</sup> oder an Perspektiven und Interessen im Sinne der Befriedigung verschiedener *Stakeholder*-Gruppen (vgl. Kapitel 2.1).

Als grundlegendste Kategorisierung von Performanzdimensionen kann jedoch Scharpfs (2004) Differenzierung zwischen *Input*- und *Output-Legitimation* gelten.<sup>26</sup> Die *Input-Legitimation* beschreibt aus der Perspektive des in der Analyse von Delegationsbeziehungen dominanten *Principle-Agent Ansatzes* die intakte hierarchische Beziehung zwischen *Prinzipal* (Wahlvolk, Politiker, gewählte Re-

24 Kuhlmann et al. (2011) differenziert sechs Dimensionen institutioneller Performanz: (1) Im Bereich des Inputs „Demokratische Kontrolle und Verantwortlichkeit“, (2) prozessorientiert „Horizontale Koordination sowie der Vertikale Koordination“ und (3) für den Output „Effizienz, [...] Effektivität [...] Homogenität des Verwaltungshandelns.“

25 Eichhorn (1993: 867) identifiziert zwei grundsätzliche Zielsetzungen – Formalziele (Haushaltsausgleich, Liquidität, Sicherheit) und Sachziele (Rechtmäßigkeit, Leistungsfähigkeit, Sozial- und Umweltverträglichkeit).

26 Zur Problematik und Vielschichtigkeit des Legitimitätsbegriffs s. Hellmann (2006: 81ff.).

präsentanten) und der Verwaltung als beauftragtem *Agenten*. Erst durch die gegebene Steuerungsfähigkeit<sup>27</sup> verfüge die Bürokratie über jene Legitimation, welche sie in einer Demokratie zur Durchsetzung des staatlichen Machtmonopols berechtigt (Scharpf 2004: o.S.).<sup>28</sup> Die *Output-Legitimation* umfasst dabei eine „Problemlösungsperspektive“; die Verwaltung kann Legitimation erwerben, wenn es ihr gelingt, „Schaden abzuwenden und den gemeinsamen Nutzen des ‚Volkes‘ zu fördern“ (Scharpf 2004: o.S.).<sup>29</sup> Dieses auf den ersten Blick sehr klare Konzept erweist sich in der empirischen Praxis – insbesondere wenn nicht auf harte *Output*-Daten zurückgegriffen werden kann – als problematisch. So ist die trennscharfe Abgrenzung zwischen Aspekten der *Input*-Legitimation (bspw. der *Responsivität*, s.u.) und solchen der *Output*-Legitimation (bspw. der *Professionalität*, s.u.) nicht aufrecht zu halten.

Hood unterscheidet ähnlich grundsätzlich, aber mit anderer Stoßrichtung nach der *Funktion* einzelner Verwaltungscharakteristika (1991: 12ff.) zwischen „lamda“, „theta“ und „sigma-type values“, an denen sich öffentliche Verwaltungen messen lassen müssten. Während die *lamda-type* Werte die Beständigkeit der Verwaltung und damit den Verwaltungsvollzug auch in Krisenzeiten zuverlässig sichern, sind *theta-type* Werte Garanten für die faire und rechtschaffene Aufgabenwahrnehmung, die durch weberianische Mechanismen der Verwaltungsbindung Amtsmissbrauch unterbinden. Jenseits dieser beiden unverzichtbaren Grundanforderungen liegt bei den *sigma-type* Werten die Performanz im engeren Sinne: die Angemessenheit des Ressourceneinsatzes, Sparsamkeit und Zielgerichtetheit. So wichtig das erste Werteset zur Charakterisierung öffentlicher Aufgabenwahrnehmung und insbesondere für die Entscheidung für einen öffentlichen Vollzug sein mag<sup>30</sup>, im Normalbetrieb tritt es weitgehend hinter die

27 Vgl. die Definition von Benz (1994: 53): „Als Steuerung wird die Art und Weise bezeichnet, in der ein bestimmtes Ziel gegenüber einem Adressaten durchgesetzt bzw. in der ein Adressat staatlicher Steuerung zu einem angestrebten Verhalten veranlaßt wird“; mit Verweis auf Mayntz (1997b: 191).

28 Im Regelfall wird aufgrund der bloßen Existenz von harten *Institutionen* (freien, geheimen und gleichen Wahlen, unabhängigen Institutionen des Rechtsstaats) auf eine demokratische Legitimation geschlossen. Problematischer erscheint, dass erhöhte Beteiligungschancen (wie bspw. durch lokale Demokratieelemente im Zuge der kommunalen Selbstverwaltung, vgl. Stein (1807); Hendlar (1984); Andersen (1998); Wollmann (1998)) vereinfachend mit einer höheren *Leistungsfähigkeit* der nachgeordneten Strukturen gleichgesetzt werden. Vgl. kritisch Mittelinstanzenbericht (1973: 9); von Alemann/Gehne (2007); Reiners (2008a: 77 m.w.V). Dieser strukturelle Aspekt ist zwar keinesfalls unwichtig, allerdings ist er kein Performanzaspekt, sondern ein potenzieller Erklärungsfaktor von Verwaltungshandeln und wird deshalb in der Detailanalyse als unabhängige Variable in das Erklärungsmodell eingehen.

29 Diese Formulierung Scharpfs zeigt, dass er sich hier de facto auf die *Outcome* oder *Impact* Stufe des Erstellungsprozesses bezieht.

30 Vgl. ähnlich auch Goodin (1996: 39ff.).

zweite und dritte Dimension zurück. Diese beiden Gruppen stellen auch bei den meisten anderen, eher konventionellen Unterscheidungen den Kern der Betrachtungen. So spiegeln *theta-* und *sigma-type* Werte grundsätzlich Wagners einfachen Dualismus *politischer Maßstäbe* (Festigung der Demokratie und die Sicherung der Rechtmäßigkeit) und *technischer Maßstäbe* (Wirtschaftlichkeit und Leistungsfähigkeit) wider (Wagener 1969: 312f.; vgl. De Peuter et al. 2011: 547). Derivate dieses simplen Modells bilden Performanzdefinitionen, die versuchen, durch die Ausdifferenzierung in drei oder vier Dimensionen eine präzisere Abgrenzung zu erreichen und weitere Aspekte zu berücksichtigen. So prägen Bovens et al. 1996 (zitiert nach De Peuter et al. 2011: 547) das sog. *Quality Quartett* Rechtmäßigkeit, Demokratie, Legitimation und Effektivität/Effizienz als Bewertungskriterien. Nelissen (2002: 13ff.) führt weiterhin eine Dimension politischer Steuerung ein und kommt so zu den drei Perspektiven seines *JEP-Triangles*: „the jurisdictional perspective, the economic perspective and the political or societal perspective.“ Helbing (1998: 86-88) schlägt eine Dreiteilung in Maßstäbe mit organisatorisch-technokratischer Funktion, Maßstäbe ökonomischer Rationalität und Maßstäbe integrativer Funktion vor und führt deren mögliche Inhalte detailliert aus. Bogumil/Jann (2009: 55) listen mit Verweis auf Offe (1974: 344), Jann (1998: 50), Bogumil (2001: 26), König (2003) und Snellen (2006) vier Referenzsysteme, denen öffentliches Handeln unterworfen sei:

„[...] einem legalen Richtigkeitstest (juristische Rationalität), also die Frage der Legalität, der Gesetzmäßigkeit, der Gleichbehandlung und des Rechtsschutzes, einem Wirtschaftlichkeitstest (ökonomische Rationalität), also die Frage der Effizienz staatlichen Handelns, einem politischen Konsenstest (politische Rationalität im Sinne von politics), also die Frage der demokratischen Verantwortlichkeit und Kontrolle, d.h. der Legitimität der Verwaltung sowie einem funktionalen Wirksamkeitstest (politische Rationalität im Sinne von policy), also die Frage der Effektivität politischer und administrativer Maßnahmen [...].“

Zusammenfassend kann festgestellt werden, dass auch knapp vier Jahrzehnte nach der Veröffentlichung von Derliens Aufsatz (1974) die Literatur zu Organisationszielen und damit auch Bewertungskriterien von Verwaltungsleistung – immer noch oder mehr denn je – unüberschaubar ist. Heutige Versuche der Sichtung der für eine Performanzdefinition zu berücksichtigenden Dimensionen führen zu einer bunten Mischung an ökonomisch, demokratietheoretisch oder rechtsstaatlich abgeleiteten Ansätzen, die weder trennscharf abgrenzbar noch widerspruchsfrei sind. Die Differenzierungsversuche, sei es Wagners Dualismus *politischer* und *technischer Maßstäbe* (Wagener 1969: 312f.), das *Quality Quartett* aus Rechtmäßigkeit, Demokratie, Legitimation und Effektivität/Effizienz (De Peuter et al. 2011: 547) oder das *JEP-Triangle* rechtlicher, wirtschaftlicher und politischer Perspektiven (Nelissen 2002: 13ff.) verdeutlichen einerseits die Heterogenität der Interessen unterschiedlicher Akteursgruppen, deren Ziele je nach Bewertungslogik „komplementär, konfliktär oder indifferent“ (Steiner 2002: 274) sein können. Diese führen einen Kampf um die Deutungshoheit des Organi-

sationszwecks und die zur Zielerreichung zur Verfügung stehenden Ressourcen. Andererseits spiegeln diese Modelle durch die Kopplung ihrer Komponenten auch die Unmöglichkeit, die Einzelziele voneinander zu lösen, da eine Vernachlässigung einer Zieldimension zwangsläufig zum Scheitern der Organisation als Ganzes führen würde. Die Zieldimensionen müssen also trotz ihres Konkurrenzverhältnisses vereint betrachtet werden und lassen sich nicht separat, d.h. ohne Berücksichtigung der Auswirkungen auf die anderen Dimensionen optimieren. Diese mehrdimensionale Performanz theoretisch zu erfassen und zu bewerten gehört zu den großen Herausforderungen der vergleichenden Institutionenanalyse (vgl. Ostrom 2007: 35, 2011: 17). Zur Gliederung dieser Vielzahl an Aspekten bietet sich auf einer ersten Ebene eine Dreiteilung in Anlehnung an Wagener (1969) in *politische Maßstäbe*, *funktionale Maßstäbe* und *mitarbeiterzentrierte Maßstäbe* an, unter die die relevanten Einzelaspekte subsumiert werden sollen.

### 2.3.2. Konzepte der empirischen Erfassung

Im Vorgriff auf die später notwendige Operationalisierung der Performanzkonstrukte soll bereits hier, noch vor der Identifikation der Bewertungsmaßstäbe, eine Grundsatzentscheidung zur Art der verwendeten Indikatoren getroffen und begründet werden. Aufgrund der Individualität der Aufgabe nahezu jeder einzelnen Verwaltungseinheit und der oft kaum zu quantifizierenden *Outputs*, stellt auch die Entwicklung geeigneter Performanzindikatoren eine Herausforderung bei der Analyse administrativer Performanz dar (vgl. Fölster 1997: 131f.). Zur empirischen Erfassung der Performanzdimensionen in großzahligen Forschungsdesigns bieten sich theoretisch zwei Herangehensweisen an: Die Operationalisierung einzelner Ziele über verfügbare "harte" *Output*- oder *Outcome*-Daten, oder ihre Greifbarmachung über Perzeptionsdaten. Da die für die erste Methode notwendigen, vergleichbaren Daten in der Regel nicht verfügbar sind, hat sich der zweite, akteursbezogene Ansatzpunkt etabliert. Dieser setzt meist am zentralen Spannungsverhältnis zwischen politischen und fachlichen Zielen an und fragt zu dessen Erfassung die *Rollenbilder* der Verwaltungsmitarbeiter ab. Dieses Konzept wurde schon in den 1960er Jahren im Kontext großer Elitestudien entwickelt und soll stark abstrahierend die Wertvorstellungen und damit potenziell auch das Entscheidungsverhalten der Verwaltungsmitarbeiter abbilden (vgl. Putnam 1976b; Aberbach et al. 1981; Derlien 1994: 267; Fisch 2002: 463f.).<sup>31</sup> Während eine Spielart dabei abstrakt die Einstellungen zum öffentlichen Dienst und der Rolle von Politikern und Bürokraten abfragte, setzte eine zweite unmittelbar

31 Christensen (1991: 303) definiert die *Rolle* treffend als „the perception of bureaucratic norms and role enactment“.