

Die Evaluationskultur im Fokus: Eine Untersuchung der kantonalen Verwaltungen Basel-Stadt, Bern und Luzern

Dissertation
der Kultur- und Sozialwissenschaftlichen Fakultät der
Universität Luzern¹

vorgelegt von
Walter Rohrbach
Spitalackerstrasse 61
3013 Bern
(Rüeggisberg/Bern)

Luzern, 2020

Angenommen am: 11. Dezember 2018 auf Antrag von
Prof. Dr. Andreas Balthasar, Erstgutachter
Prof. Dr. Frédéric Varone, Zweitgutachter

DOI: 10.5281/zenodo.3854846; Lucerne Open Repository [LORY]
CC BY-NC-ND

¹ Die Dissertation entstand im Rahmen des Forschungsprogramms „Policy Evaluation im schweizerischen Politiksystem – Roots and Fruits“ (SynE-val), unterstützt durch den Schweizerischen Nationalfonds zur Förderung der wissenschaftlichen Forschung (SNF).

Dank

Als erstes gilt ein grosser Dank meinen beiden Betreuern, Prof. Dr. Andreas Balthasar und Prof. Dr. Frédéric Varone, für die Unterstützung und die konstruktive Begleitung des Dissertationsvorhabens. Ein weiterer Dank richtet sich an meinen Mitdoktoranden Olivier Dolder, mit welchem ich eine sehr angenehme und kollegiale Zusammenarbeit hatte. Ein Dank ist zudem an das gesamte Team des Forschungsprogramms „SynEval“ gerichtet, von dem ich ebenfalls viele wertvolle Rückmeldungen erhalten habe.

Weiter danke ich den General- und Departementssekretariaten sowie den Personen, die das Dissertationsprojekt unterstützt und sich für mein Forschungsanliegen eingesetzt haben. Ein weiteres grosses Dankeschön gilt allen Verwaltungsmitarbeitenden, die sich Zeit für die Teilnahme an der Befragung genommen haben.

Des Weiteren ist es mir ein Anliegen, einen Dank an Valerie Berger, Ilana Bischof, Andrea Eichelberger, Oliver Hümbelin, Andrea Kretschmann, Lea Portmann, Ina Paukhovich, Sandra Schwander, Benjamin Spielmann, Jonas Wenger und Dominik Wyss zu richten. Schliesslich bedanke ich mich auch bei meinen Eltern, meinen Geschwistern und meinen Freunden für ihre Unterstützung.

Inhalt

1 Einleitung	9
1.1 Zielsetzung und Fragestellung	10
1.2 Untersuchungsdesign	12
1.3 Eingrenzung des Untersuchungsbereiches	13
1.3.1 Eingrenzung des Evaluationsbegriffes	14
1.3.2 Eingrenzung der Untersuchungseinheiten	16
1.4 Forschung über die Evaluationsentwicklung in der Schweiz	21
1.4.1 Evaluationsentwicklung in der Schweiz	21
1.4.2 Forschung über die Evaluationsaktivität in der Schweiz	22
1.4.3 Einordnung der Entwicklung der Evaluation in der Schweiz im internationalen Vergleich	27
1.5 Aufbau	28
A. Theoretischer Teil	29
2 Was ist die organisationale Evaluationskultur?	29
3 Theoretische Modelle	51
3.1 Theoretisches Modell für die Untersuchung der Evaluationspraxis	54
3.2 Theoretisches Modell für die Untersuchung der Einstellung zur Evaluation	69
B. Empirischer Teil	87
4 Messung der organisationalen Evaluationskultur	87
4.1 Operationalisierung der organisationalen Evaluationskultur	87
4.2 Einschränkungen	99
5 Welche Unterschiede bezüglich der Evaluationskultur lassen sich in den neun Departementen und Direktionen feststellen?	105
5.1 Luzern	105
5.1.1 Gesundheits- und Sozialdepartement	105
5.1.2 Bildungs- und Kulturdepartement	109
5.1.3 Bau-, Umwelt- und Wirtschaftsdepartement	114
5.2 Basel-Stadt	117
5.2.1 Gesundheitsdepartement	117
5.2.2 Erziehungsdepartement	121
5.3 Bern	126
5.3.1 Gesundheits- und Fürsorgedirektion	126

5.3.2	Erziehungsdirektion	129
5.3.3	Bau-, Verkehrs- und Energiedirektion	134
5.3.4	Volkswirtschaftsdirektion	137
5.4	Überblick der Departemente und Direktionen.....	140
6	Wie lassen sich die Unterschiede in der Evaluationskultur erklären?	150
6.1	Untersuchung der Unterschiede in der Evaluationspraxis	150
6.1.1	Bivariate Analyse	151
6.1.2	Multivariate Analysen	153
6.2	Untersuchung der Unterschiede in der Einstellung zur Evaluation	167
6.2.1	Bivariate Analyse	168
6.2.2	Multivariate Analyse	170
C.	Synthese und Schlussfolgerungen	177
7	Synthese..	177
7.1	Was ist die organisationale Evaluationskultur?	177
7.2	Welche Unterschiede bezüglich der Evaluationskultur lassen sich in den neun Departementen und Direktionen feststellen?	180
7.3	Wie lassen sich die Unterschiede in der Evaluationskultur erklären?	183
7.3.1	Untersuchung der Einflüsse auf die Evaluationspraxis.....	183
7.3.2	Untersuchung der Einflüsse auf die Einstellung zur Evaluation	185
7.4	Diskussion	187
8	Schlussfolgerungen	192
Anhang I	Literatur- und Quellenverzeichnis	197
Anhang II	Abkürzungsverzeichnis	220
Anhang III	Anmerkungen für die bivariate Analyse.....	222
Anhang IV	Operationalisierung, bivariate Zusammenhänge und Einebenenmodelle	224
Anhang V	Online-Fragebogen	233

Abbildungen

<i>Abbildung 1: Drei Dimensionen der organisationalen Evaluationskultur</i>	49
<i>Abbildung 2: Theoretisches Gesamtmodell der Untersuchung</i>	52
<i>Abbildung 3: Beziehungen der drei Dimensionen der organisationalen Evaluationskultur</i>	53
<i>Abbildung 4: Modell für die Untersuchung der Evaluationspraxis</i>	54
<i>Abbildung 5: Modell für die Untersuchung der Einstellung zur Evaluation</i>	69
<i>Abbildung 6: Histogramm der Einstellung zur Evaluation</i>	99
<i>Abbildung 7: Organigramm des Gesundheits- und Sozialdepartementes (Stand Juli 2014)</i>	106
<i>Abbildung 8: Evaluationspraxis im GSD Luzern</i>	107
<i>Abbildung 9: Organigramm des Bildungs- und Kulturdepartementes (Stand Juli 2014)</i>	110
<i>Abbildung 10: Evaluationspraxis im BKD Luzern</i>	112
<i>Abbildung 11: Organigramm des Bau-, Umwelt und Wirtschaftsdepartementes (Stand Juli 2014)</i>	114
<i>Abbildung 12: Evaluationspraxis im BUWD Luzern</i>	116
<i>Abbildung 13: Organigramm des Gesundheitsdepartementes Basel-Stadt (Stand August 2014)</i> ..	118
<i>Abbildung 14: Evaluationspraxis im GD Basel-Stadt</i>	120
<i>Abbildung 15: Organigramm des Erziehungsdepartementes Basel-Stadt (Stand August 2014)</i> ..	122
<i>Abbildung 16: Evaluationspraxis im ED Basel-Stadt</i>	124
<i>Abbildung 17: Organigramm der Gesundheits- und Fürsorgedirektion Bern (Stand August 2014)</i>	126
<i>Abbildung 18: Evaluationspraxis in der GEF Bern</i>	128
<i>Abbildung 19: Organigramm der Erziehungsdirektion Bern (Stand August 2014)</i>	130
<i>Abbildung 20: Evaluationspraxis in der ERZ Bern</i>	132
<i>Abbildung 21: Organigramm der Bau-, Verkehrs- und Energiedirektion Bern (Stand August 2014)</i>	134
<i>Abbildung 22: Evaluationspraxis in der BVE Bern</i>	136
<i>Abbildung 23: Organigramm der Volkswirtschaftsdirektion Bern (Stand August 2014)</i>	138
<i>Abbildung 24: Evaluationspraxis in der VOL Bern</i>	139
<i>Abbildung 25: Evaluationspraxis unterschieden nach Departementen und Direktionen</i>	143
<i>Abbildung 26: Einstellung zur Evaluation unterschieden nach Evaluationsaktivität</i>	148

<i>Abbildung 27: Modell für die Untersuchung der Evaluationspraxis</i>	150
<i>Abbildung 28: Modell für die Untersuchung der Einstellung zur Evaluation</i>	168

Tabellen

<i>Tabelle 1: Teilnehmende an der Online-Mitarbeitendenbefragung</i>	18
<i>Tabelle 2: Mitarbeitende in zentralen Verwaltungseinheiten</i>	20
<i>Tabelle 3: Untersuchungen der Evaluationsaktivität in der Schweiz</i>	23
<i>Tabelle 4: Explizite Beiträge zur Evaluationskultur</i>	39
<i>Tabelle 5: Weitere Beiträge mit Relevanz für die Konzeptualisierung der Evaluationskultur</i> ...	44
<i>Tabelle 6: Zu untersuchende Determinanten der Evaluationspraxis</i>	68
<i>Tabelle 7: Zu untersuchende Determinanten der Einstellung zur Evaluation</i>	86
<i>Tabelle 8: Indikatoren der Institutionalisierung der Evaluation</i>	88
<i>Tabelle 9: Indikatoren der Evaluationspraxis</i>	91
<i>Tabelle 10: Fragen bezüglich der Einstellung zur Evaluation</i>	95
<i>Tabelle 11: Deskription der Einstellung zur Evaluation</i>	98
<i>Tabelle 12: Datenquellen für die Deskription der Evaluationskultur</i>	100
<i>Tabelle 13: „Non-Response“-Analyse</i>	104
<i>Tabelle 14: Teilnehmende an der Online-Befragung</i>	106
<i>Tabelle 15: Einstellung zur Evaluation im GSD Luzern</i>	109
<i>Tabelle 16: Teilnehmende an der Online-Befragung</i>	111
<i>Tabelle 17: Einstellung zur Evaluation im BKD</i>	113
<i>Tabelle 18: Teilnehmende an der Online-Befragung</i>	115
<i>Tabelle 19: Einstellung zur Evaluation im BUWD Luzern</i>	117
<i>Tabelle 20: Teilnehmende an der Online-Befragung</i>	118
<i>Tabelle 21: Einstellung zur Evaluation im GD Basel-Stadt</i>	121
<i>Tabelle 22: Teilnehmende an der Online-Befragung</i>	123
<i>Tabelle 23: Einstellung zur Evaluation im ED Basel-Stadt</i>	125
<i>Tabelle 24: Teilnehmende an der Online-Befragung</i>	127
<i>Tabelle 25: Einstellung zur Evaluation in der GEF Bern</i>	129

<i>Tabelle 26: Teilnehmende an der Online-Befragung</i>	131
<i>Tabelle 27: Einstellung zur Evaluation in der ERZ Bern</i>	134
<i>Tabelle 28: Teilnehmende an der Online-Befragung</i>	135
<i>Tabelle 29: Einstellung zur Evaluation in der BVE Bern</i>	137
<i>Tabelle 30: Teilnehmende an der Online-Befragung</i>	138
<i>Tabelle 31: Einstellung zur Evaluation in der VOL Bern</i>	140
<i>Tabelle 32: Evaluationsinstitutionalisierung</i>	142
<i>Tabelle 33: Übersicht der Evaluationsaktivität</i>	146
<i>Tabelle 34: Bivariate Zusammenhänge</i>	151
<i>Tabelle 35: Nullmodelle</i>	155
<i>Tabelle 36: Negativ binomiale Regressionsanalyse der Evaluationsaktivität</i>	158
<i>Tabelle 37: Hypothesentest der Evaluationsaktivität</i>	162
<i>Tabelle 38: Logistische Regressionsanalyse der Evaluationsverwendung</i>	163
<i>Tabelle 39: Hypothesentest der Evaluationsverwendung</i>	166
<i>Tabelle 40: Bivariate Zusammenhänge</i>	169
<i>Tabelle 41: Nullmodelle</i>	171
<i>Tabelle 42: Lineare Regressionsanalyse der Einstellung zur Evaluation</i>	172
<i>Tabelle 43: Hypothesentest der Einstellung zur Evaluation</i>	175
<i>Tabelle 44: Gesamtübersicht der Hypothesen bezüglich der Evaluationspraxis</i>	184
<i>Tabelle 45: Hypothesentest der Einstellung zur Evaluation</i>	186
<i>Tabelle 46: Übersicht der verwendeten Korrelationskoeffizienten</i>	222
<i>Tabelle 47: Charakteristiken der verwendeten Korrelationskoeffizienten</i>	223
<i>Tabelle 48: Operationalisierung und Deskription der Variablen für die Analyse der Evaluationspraxis</i>	224
<i>Tabelle 49: Operationalisierung und Deskription der Variablen für die Analyse der Einstellung zur Evaluation</i>	226
<i>Tabelle 50: Bivariate Zusammenhänge der unabhängigen Variablen für die Untersuchung der Evaluationspraxis</i>	228
<i>Tabelle 51: Bivariate Zusammenhänge der unabhängigen Variablen für die Untersuchung der Einstellung zur Evaluation</i>	229
<i>Tabelle 52: Negativ binomiale Regressionsanalyse der Evaluationsaktivität (Einebenenmodell)</i>	230

Tabelle 53: Logistische Regressionsanalyse der Evaluationsverwendung (Einebenenmodell)....231

Tabelle 54: Lineare Regressionsanalyse der Einstellung zur Evaluation (Einebenenmodell)232

1 Einleitung

Der Begriff „Evaluation“ ist für viele längst kein Fremdwort mehr; gerade in Gesprächen mit Personen, die in einer Verwaltung tätig sind, zeigt sich, dass dieser Terminus nur noch selten fragende Blicke hervorruft. Stattdessen kann eine Vielzahl der Betroffenen mit eigenen Erfahrungen und Erlebnissen an die Evaluation anknüpfen. In den letzten Jahrzehnten haben Evaluationen stark an Bedeutung gewonnen und in der Schweiz wie auch in anderen Ländern eine starke Ausweitung erfahren (Balthasar 2007, 2015; Beywl/Widmer 2009; Frey et al. 2017; Frey/Ledermann 2017; Jacob et al. 2015; Widmer 2017).

Auch das Forschungsprogramm „SynEval“, in dessen Rahmen die vorliegende Dissertation entstanden ist, ist Ausdruck der zunehmenden Relevanz der Evaluation. Durch den Schweizerischen Nationalfond (SNF) finanziert, untersucht es in verschiedenen Teilprojekten die Beziehung zwischen dem politischen System der Schweiz und der Praxis sowie der Institutionalisierung der Politikevaluationen. Neben der Universität Luzern sind die Universitäten Zürich, Bern und Genf sowie das IDHEAP in Lausanne beteiligt. Die vorliegende Untersuchung ist dem Teilprojekt, an dem die Universitäten Luzern und Genf beteiligt sind, zuzuordnen. Ebenfalls im Rahmen dieses Teilprojektes wurde die Dissertation von Dolder (2018) „Evaluation und NPM: Gegenseitiger Einfluss von Evaluationskultur und New Public Management“ verfasst.²

Die zunehmende Verbreitung und Bedeutung der Evaluation wirft unweigerlich Fragen nach ihrem Einfluss auf. Um diesen ermessen zu können, ist es wichtig, mehr über die Erstellung, die Verbreitung und die Verwendung der Evaluation zu erfahren. Für die Verwaltung ist das Thema Evaluation zudem von besonderer Relevanz. Aus den Untersuchungen der Evaluationsintensität³ kann geschlossen werden, dass die Verwaltungen beachtliche Ressourcen für die Beurteilung ihrer Massnahmen, Projekte und Gesetze mittels Evaluati-

² Weitere Informationen des Programms „SynEval“ finden sich auf der Webseite: www.syneval.ch.

³ Die Untersuchung von Balthasar/Strotz (2017) erfasste für den Zeitraum 1999 bis 2015 insgesamt 1414 Evaluationen auf Bundesebene. Der Beitrag von Frey et al. (2017) verweist auf 657 Evaluationsberichte auf Bundesebene und der Kantone Bern, Genf und Zürich im Zeitraum 2000 bis 2012.

onen einsetzen. Dies legt es nahe, gerade diesen Bereich zu untersuchen und mehr über das Verhältnis zwischen der Institutionalisierung der Evaluation, der Evaluationspraxis und der allfälligen Widerstände der Mitarbeitenden gegenüber der Evaluation zu erfahren. Des Weiteren gibt es offene konzeptionelle Fragen zu den Elementen und Bestandteilen der Evaluationskultur. Die bisherige Forschungstätigkeit fokussiert sich vermehrt auf die Verbreitung und Verwendung der Evaluation, und es finden sich in den Beiträgen von Barbier (2012), De Peuter/Pattyn (2009) und Läubli Loud (2014) nur vereinzelt theoretische Auseinandersetzungen mit dem Konzept der Evaluationskultur.

Die vorliegende Dissertation wird sich den vorgestellten Themen widmen; konkret wird die organisationale Evaluationskultur in fünf Departementen (der Kantone Basel-Stadt und Luzern) und in vier Direktionen (des Kantons Bern) untersucht. Abgeleitet aus bisherigen Untersuchungen zur Evaluationsintensität in der Schweiz und den darin festgestellten Unterschieden zwischen Verwaltungseinheiten und Politikfeldern (Balthasar 2007; Balthasar/Rieder 2009; Frey et al. 2017; Frey/Ledermann 2017; Widmer 2008; Widmer/Beywl 2009) bieten die ausgewählten Verwaltungseinheiten gute Vergleichsfälle, um einen Einblick in Verbreitung und Einflussfaktoren der organisationalen Evaluationskultur zu gewinnen.

1.1 Zielsetzung und Fragestellung

Prioritäres Ziel der vorliegenden Arbeit ist es, die organisationale⁴ Evaluationskultur herzuleiten und zu konzeptualisieren, um diese dann anhand ausgewählter kantonaler Verwaltungseinheiten zu beschreiben. Weiter soll untersucht werden, mit welchen Faktoren die organisationale Evaluationskultur zusammenhängt. Für dieses Forschungsvorhaben werden erstmals die Angaben einer breit angelegten Mitarbeitendenbefragung, die Informationen über die jeweilige Einstellung zur Evaluation und zur Evaluationspraxis enthält, untersucht und in die Analyse der Evaluationskultur miteinbezogen.

⁴ Der Begriff organisational bezieht sich – um einige Beispiele zu nennen – auf Schulen, Universitäten, Nicht-Profit-Organisationen, Verwaltungseinheiten sowie Geschäftseinheiten und charakterisiert die Organisation als Analyseebene. Die organisationale Evaluationskultur betrachtet die Evaluationskultur in Organisationen und nicht, wie die Evaluationskultur in bestimmten Ländern (Länderebene) verbreitet ist.

Die Untersuchung orientiert sich an folgenden drei Forschungsfragen:

1. *Was ist die organisationale Evaluationskultur?*
2. *Welche Unterschiede in der organisationalen Evaluationskultur gibt es zwischen den Departementen und Direktionen?*
3. *Wie lassen sich Unterschiede in der organisationalen Evaluationskultur erklären?*

Zur Beantwortung der ersten Frage wird in dieser Untersuchung erstmals der Begriff der organisationalen Evaluationskultur vertieft theoretisch aufgearbeitet und diskutiert. Dafür werden Beiträge aus der Forschung über Evaluation als auch aus verwandten Forschungsdisziplinen zur Konzeptualisierung und Definition herangezogen. Eine weitere Zielsetzung der Dissertation besteht in der Beschreibung der bisher nahezu unbekanntem Verbreitung der Evaluation in kantonalen Verwaltungseinheiten. Die vorliegende Untersuchung beschränkt sich auf neun Verwaltungseinheiten und kann demzufolge lediglich punktuell Einblicke gewähren. Trotzdem bietet die folgende Deskription der Departemente und Direktionen – zusammen mit der im Rahmen des gleichen Teilprojektes des Forschungsprogrammes „SynEval“ verfassten Dissertation von Dolder (2018) – erstmalig vertiefte Einblicke in den Stand der Evaluation auf Kantonsebene.⁵ Die Untersuchung der Evaluationskultur selbst (dritte Frage) bietet die Möglichkeit, weiterführende wichtige Erkenntnisse für die Forschungsgemeinschaft zu gewinnen. Als Grundlage für die Untersuchung der organisationalen Evaluationskultur dienen zwei theoretische Modelle, die auch die zu untersuchenden Einflussfaktoren umfassen.⁶ Beispielsweise werden in einem der beiden Modelle die häufig festgestellten und bisher vorwiegend theoretisch diskutierten Unterschiede

⁵ Im Beitrag von Dolder et al. (2017) wurden erste Ergebnisse der vorliegenden Dissertation und der Dissertation von Dolder (2018) publiziert. Darin sind bereits Beschreibung und Analyse der Evaluationskultur von ausgewählten kantonalen Verwaltungseinheiten zu finden.

⁶ Diese Modelle adressieren nicht die organisationale Evaluationskultur als Ganzes pro Verwaltungseinheit, sondern berücksichtigen die einzelnen Dimensionen/Elemente der Evaluationskultur und die Faktoren, welche die Evaluationskultur beeinflussen (siehe dazu die Erläuterungen im dritten Kapitel).

der Evaluationspraxis in diversen Politikfeldern analysiert (siehe Balthasar 2007; Balthasar/Rieder 2009; Furubo/Sandahl 2002; Hense et al. 2013; Jacob et al. 2015; Pattyn 2014, 2015; Widmer 2008; Widmer/Beywl 2009). Die Untersuchung will damit aufzeigen, welche Faktoren die Evaluationskultur in der Verwaltung stärken und welche Faktoren eher hinderlich sind. Weiter können Aussagen darüber gemacht werden, mit welchen Massnahmen es gelingt, die Mitarbeitenden positiv gegenüber der Evaluation zu stimmen.

1.2 Untersuchungsdesign

Für die Beantwortung der drei Fragestellungen werden unterschiedliche methodische Vorgehen gewählt:

Die erste Forschungsfrage wird auf der Grundlage der theoretischen Aufarbeitung des Begriffes der organisationalen Evaluationskultur beantwortet. Die organisationale Evaluationskultur wird für diese Untersuchung als Konzept¹ angesehen. In Anlehnung an De Vaus (2014: 41ff.) wird folgendes Vorgehen für dessen Konkretisierung verfolgt: Zunächst geht es um die Klärung des Konzeptes; dabei wird mittels einer vertieften Literaturanalyse nach bestehenden Definitionen und Konzepten recherchiert. Anschliessend werden diese Definitionen auf gemeinsame, grundlegende Elemente hin analysiert, und basierend auf den daraus folgenden Erkenntnissen wird eine gültige Definition formuliert, welche die relevanten Dimensionen und Aspekte berücksichtigt. In einem weiteren Schritt werden für die identifizierten Dimensionen und Aspekte Indikatoren zur Messung gebildet.

Zur Beantwortung der zweiten Forschungsfrage wird die Evaluationskultur in den Departementen und Direktionen mithilfe von Methoden der deskriptiven Statistik dargestellt. Gemäss Schäfer (2010: 59ff.) vereint die deskriptive Statistik alle Methoden, mit denen empirische Daten zusammenfassend beschrieben werden können. Vor allem Masse der

¹ Konzepte werden von De Vaus (2014: 41) folgendermassen definiert: „Concepts are simply tools which fulfil a useful shorthand function: they are abstract summaries of a whole set of behaviours, attitudes and characteristics which we see as having something in common“. Ein typisches Beispiel eines Konzeptes ist für De Vaus das Sozialkapital.

univariaten Statistik, wie sie Benninghaus (2007: 29ff.) oder Weins (2010) einführen, werden mit Häufigkeitsverteilungen und Mittelwerten Anwendung finden.

Die dritte Forschungsfrage untersucht unterschiedliche Einflussfaktoren auf die Evaluationskultur. Dafür werden im dritten Kapitel zwei theoretische Modelle präsentiert, welche die zu untersuchenden Einflüsse aufzeigen. Die zu untersuchenden Daten basieren primär auf Befragungsdaten; zusätzlich werden erhobene Merkmale der Verwaltungseinheiten in die Analyse miteinbezogen. Lauth et al. (2015: 163ff.) präsentieren in ihrem Einführungsband ausführlich die verschiedenen Methoden der vergleichenden Politikwissenschaft; dabei wird auch die vergleichende Surveyforschung thematisiert. In diesem Zusammenhang verweisen sie für die Auswertung von Befragungsdaten auf Verfahren der quantitativen empirischen Datenanalyse, unter anderem auf die Regressionsanalyse. Diese wird in dieser Untersuchung, in Anlehnung an Jahn (2013: 358ff.), für die Analyse der Einflussfaktoren der Evaluationskultur mittels folgendem Vorgehen angewendet: Zuerst erfolgt die Formulierung und Operationalisierung der Hypothesen. Anschliessend werden dann die bivariaten und im nächsten Schritt die multivariaten Zusammenhänge analysiert und beurteilt.

Die Informationen dieser Untersuchung basieren auf einer in ausgewählten kantonalen Verwaltungseinheiten durchgeführten Online-Mitarbeitendenbefragung. Zusätzlich wird für die Ermittlung der Evaluationspraxis und der Institutionalisierung der Evaluation eine Onlinerecherche durchgeführt und darüber hinaus werden Auskünfte von Amtsvorstehern und den General- und Departementssekretariaten eingeholt.

1.3 Eingrenzung des Untersuchungsbereiches

Primärer Gegenstand der vorliegenden Untersuchung ist die organisationale Evaluationskultur. Diese wird im zweiten Kapitel hergeleitet und erläutert. Vorab soll jedoch bereits der Evaluationsbegriff erörtert und definiert werden, um danach die Untersuchungseinheiten zu beschreiben und einzugrenzen.

1.3.1 Eingrenzung des Evaluationsbegriffes

Widmer/De Rocchi (2012) führen das Wort „Evaluation“ auf das Französische „évaluer“ zurück. Dieses steht für die Tätigkeit der Schätzung, der Ermittlung oder der Wertbestimmung. Über das amerikanische Englisch hat der Begriff eine semantische Ausweitung erfahren und bezieht sich ferner auch auf die Auswertung und die Einschätzung. Die Evaluation wird im heutigen Sprachgebrauch häufig mit dem Begriff „Bewertung“ in Zusammenhang gebracht (Balthasar 2007; Stamm 2003; Stockmann 2000; Widmer/De Rocchi 2012).

In den letzten Jahren hat der Evaluationsterminus eine grosse Verbreitung erfahren und wird nunmehr in ganz unterschiedlichen Bereichen und Kontexten verwendet. Die von Widmer/De Rocchi (2012) als „fast inflationär“ bezeichnete Verbreitung dieses Begriffs benötigt deshalb eine nähere Umschreibung und eine Abgrenzung von anderen Begriffen. Stamm (2003: 34 ff.) weist darauf hin, dass für den Evaluationsbegriff in verschiedenen Anwendungsgebieten viele bereichsspezifische Definitionen existieren. Die von Widmer/De Rocchi (2012: 11 ff.) formulierte Arbeitsdefinition soll, trotz der von ihnen auferlegten beschränkten Eignung, dennoch Erwähnung finden. Sie beinhaltet grundlegende Elemente und dient einem ersten Verständnis des Begriffs: „Unter Evaluation wird eine wissenschaftliche Dienstleistung verstanden, die sich mit der systematischen und transparenten Bewertung eines Gegenstandes befasst“ (Widmer/De Rocchi 2012: 11). Somit sind es den beiden Autoren zufolge drei zentrale Komponenten, welche die Evaluation bestimmen: Wissenschaftlichkeit, Dienstleistungsgedanke und Bewertung.

Eine Begriffsdefinition, die weithin akzeptierte Begriffsbestimmungen aufnimmt, ist jene von Balthasar (2007: 71). Die Evaluation definiert er als eine: „[...] wissenschaftliche und empirisch gestützte retrospektive Beurteilung der Konzeption, des Vollzugs und/oder der Wirksamkeit staatlicher Aktivitäten, seien dies Massnahmen, Programme oder Projekte. Evaluationen bewerten staatliche Aktivitäten nach transparenten Kriterien und stellen Kausalzusammenhänge zwischen Aktivitäten und Wirkungen fest. Solche Angaben dienen der Entscheidungsfindung, der Rechenschaftsablage, der Kontrolle oder zur Optimierung der Aufgabenerfüllung“. Balthasar (2007) verweist auf fünf Dimensionen dieser Definition,

um Evaluationen von anderen gesellschaftspolitischen Tätigkeiten abzugrenzen. Erstens befassen sich Politikevaluationen mit staatlichem Handeln, die zweitens mittels einer systematischen Analyse und Bewertung vollzogen werden. Drittens wird die Evaluation mit wissenschaftlichen Methoden durchgeführt und viertens muss sie sich an Qualitätsstandards orientieren. Als fünfte Dimension berücksichtigt die Definition von Balthasar (2007) den Zeitpunkt der Betrachtung: In Anlehnung an Vedung (1999) rechnet Balthasar (2007) Ex-ante-Analysen nicht den Evaluationen zu und begründet dies mit den Unterschieden in Methode und Technik: „Ex-ante-Analysen beruhen auf Annahmen über mögliche Wirkungszusammenhänge, während Ex-post-Evaluationen real existierende Wirkungszusammenhänge empirisch zu untersuchen versuchen“ (siehe Balthasar 2007: 72).

Für die vorliegende Untersuchung wird eine Definition verwendet, die sich für die Untersuchung von Evaluationen in der Verwaltung eignet. Die Erläuterung von Widmer et al. (2001: 9) bezieht sich auf Untersuchungen auf der Ebene der Bundesverwaltung. Mit geringen Änderungen kann diese Definition des Evaluationsbegriffs übernommen werden; sie lautet in der leicht modifizierten Form folgendermassen:

Unter einer Evaluation wird eine Studie verstanden, welche die Wirkungen von Gesetzen, Programmen, Projekten oder Massnahmen systematisch untersucht, dabei ein methodisches Vorgehen wählt und bewertende Aussagen über Erfolg bzw. Misserfolg macht. Blosser Zustandsbeschreibungen einer Situation (beispielsweise im Rahmen eines Monitoring) oder Szenarien, die nicht auf empirischen Grundlagen beruhen, sind im Sinne dieser Definition keine Evaluationen.

Die folgende Verwendung des Evaluationsbegriffes beruht auf dieser Definition. Bei der durchgeführten Online-Mitarbeitendenbefragung findet ebenfalls diese Definition Anwendung.

1.3.2 Eingrenzung der Untersuchungseinheiten

Der Schweizerische Föderalismus (siehe Linder 2012a: 155ff.; Vatter 2014: 427ff.) ist eine der wichtigsten Charakteristiken des schweizerischen politischen Systems und hat unterschiedliche Ausgestaltungen der Kantonsverwaltungen zur Folge. Beispielsweise unterscheiden sich die Kantonsverwaltungen wesentlich in der Organisation der Verwaltungseinheiten als auch in der Anzahl an Angestellten (siehe Koller 2013). Die vorliegende Untersuchung beschränkt sich auf die Analyse von ausgewählten Verwaltungseinheiten der Kantone Basel-Stadt, Bern und Luzern. Die Auswahl der Kantone und Politikfelder orientiert sich einerseits an der erwarteten Varianz in der Evaluationskultur und der erwarteten Varianz in den zu untersuchenden Einflussfaktoren.⁸ Andererseits musste auch die Bereitschaft zur Teilnahme seitens der Verwaltungen bei der Auswahl berücksichtigt werden und führte zu den letztendlich ausgewählten Verwaltungseinheiten.

Da keine konkreten Daten zur Evaluationskultur in kantonalen Verwaltungseinheiten vorliegen, wird die Evaluationsaktivität als Indikator für das Vorhandensein einer Evaluationskultur herangezogen. Balthasar/Rieder (2009) untersuchen die Unterschiede der Evaluationsaktivität in den Kantonen. Dabei zeigt sich, dass zwischen 2003 und 2008 das Politikfeld Bildung mit 108 Evaluationen über eine hohe und das Politikfeld Gesundheit mit 23 Evaluationen über eine geringere Evaluationsaktivität verfügt. Für die vorliegende Untersuchung werden ebendiese beiden Politikfelder berücksichtigt. Für weitere Vergleiche werden Verwaltungseinheiten, welchen die Politikfelder Energie und Wirtschaft zuzuordnen sind, hinzugenommen; die Wahl dieser beiden Politikfelder orientiert sich an den erwarteten Unterschieden der Evaluationsaktivität und auch auf eine erwartete Varianz der Einflussfaktoren. Die Einflussfaktoren werden im dritten Kapitel näher erläutert. Beispiele sind die Komplexität und die Volatilität des Arbeitsbereiches.⁹ Auch für die Auswahl der

⁸ Jahn (2005: 57f.) verweist bei der Fallauswahl auf die Berücksichtigung von Fällen mit einer möglichst grossen Varianz der abhängigen und der unabhängigen Variable. In diesem Zusammenhang ist darauf zu verweisen, dass sich empirisch, mittels der hier zur Anwendung kommenden Regressionsanalyse, die Kausalität nicht beweisen lässt (siehe dazu die Diskussion der Kausalität von Opp 2010). Trotzdem wird in dieser Untersuchung von „Einfluss“ und „Einflussfaktoren“ gesprochen (siehe dazu auch die Anmerkungen im dritten Kapitel).

⁹ Konkrete empirische Untersuchungen für eine stärkere Volatilität oder Komplexität beispielsweise für das Politikfeld Energie in den Kantonen gibt es nicht. Die erwartete stärkere Volatilität wird von den starken Veränderungen in diesem Politikfeld abgeleitet, beispielsweise den Verände-

drei Kantone Basel-Stadt, Bern und Luzern weist die Studie von Balthasar/Rieder (2009) auf die Gewährleistung einer gewissen Varianz in der Evaluationsaktivität hin. Ihre Untersuchung zeigt, dass Bern eine sehr hohe Evaluationsaktivität aufweist, während Luzern und Basel-Stadt weiter unten im Kantonsrankings bezüglich der Evaluationsaktivität zu verorten sind. Die drei Kantone unterscheiden sich auch in ihrer Bevölkerungszahl und der Anzahl an Kantonsverwaltungsangestellten: Bern hat die zweitgrösste kantonale Bevölkerungszahl und die drittgrösste Kantonsverwaltung (gemessen an der Anzahl an Mitarbeitenden), Luzern die siebtgrösste Kantonsbevölkerung und die neuntgrösste Kantonsverwaltung. Basel-Stadt verfügt über die siebtgrösste Kantonsverwaltung und ist mit der Kantonsbevölkerung auf Rang 15.¹⁰

Blatter/Haverland (2014: 102) erwähnen, dass der Zugang zu Informationen ein wichtiges Kriterium für die Wahl der Untersuchungsfälle darstellt. Dies war auch für diese Untersuchung ein entscheidendes Kriterium, insofern bei der Auswahl der Verwaltungseinheiten die Bereitschaft zur Teilnahme an der Untersuchung berücksichtigt werden musste. Angefragt wurde neben den drei bereits genannten auch ein vierter Kanton, der die Teilnahme an der vorliegenden Studie allerdings wegen des erwarteten Zeitaufwandes für die Mitarbeitenden ablehnte. Auch eine weitere angefragte Verwaltungseinheit lehnte die Teilnahme mit der gleichen Begründung ab.

Aus den genannten Auswahlverfahren ergibt sich eine Untersuchung von Verwaltungseinheiten der drei Kantone Basel-Stadt, Bern und Luzern. Diese beinhalten unter anderem die Politikfelder Bildung, Energie, Gesundheit und Wirtschaft. Tabelle 1 listet die Departemente und Direktionen auf. In der Tabelle sind zudem die Kennzahlen der durchgeführten Online-Mitarbeitendenbefragung ersichtlich. Wie erwähnt wurde neben der Befragung

rungen aufgrund der Nuklearkatastrophe in Fukushima, wie sie im Beitrag von Hindmarsh (2013) aufgezeigt werden. Ebenso Widmer/Beywl (2009) verweisen auf die Energiepolitik als ein Politikfeld mit volatiler Themenkonjunktur und stellen für die Schweiz in diesem Politikfeld ein „gewisses Niveau an Evaluationsaktivität“ fest. Vertiefte empirische Untersuchungen zum Stand der Evaluation in den Politikfeldern Wirtschaft und Energie in den Kantonen fehlen aber. Die Nichterwähnung des Politikfeldes Wirtschaft in vielen Untersuchungen der Evaluationsaktivität deuten eher auf eine geringe Verbreitung der Evaluation in diesem Politikfeld auf Kantonsebene hin. Demgegenüber findet sich auf Bundesebene eine erhöhte Evaluationsaktivität für das Staatssekretariat für Wirtschaft (siehe Balthasar 2007 und Balthasar/Strotz 2017).

¹⁰Die Angaben zu den Verwaltungsgrössen (die Anzahl an Vollzeitstellen ohne diejenigen in den öffentlichen Unternehmungen) stammen aus Koller (2013a). Die Zahlen zur Kantonsbevölkerung kommen vom Bundesamt für Statistik: ständige Wohnbevölkerung nach Kantonen (BFS 2014).

auch eine Onlinerecherche nach Evaluationsberichten durchgeführt und es wurden zusätzlich Auskünfte von den Departementen und Direktionen eingeholt. Das Vorgehen für die Onlinerecherche wird im vierten Kapitel ausführlich erläutert werden. Im Folgenden wird der Selektionsprozess für die Online-Mitarbeitendenbefragung aufgezeigt.

Tabelle 1: Teilnehmende an der Online-Mitarbeitendenbefragung

Department/Direktion		Gesamtpersonalbestand	Angefragt	Antwortende	Rücklauf (in Prozent)
Luzern	Gesundheits- und Sozialdepartement	377	170	65	38.2
	Bildungs- und Kulturdepartement	3057	105	60	57.1
	Bau- Umwelt- und Wirtschaftsdepartement	387	125	76	60.8
Basel-Stadt	Gesundheitsdepartement	387	179	109	60.9
	Erziehungsdepartement	4859	64	36	56.3
Bern	Gesundheits- und Fürsorgedirektion	2960	234	125	53.4
	Erziehungsdirektion	1539	289	73	25.3
	Bau- Verkehrs- und Energiedirektion	873	410	209	51.0
	Volkswirtschaftsdirektion	875	213	154	72.3
Total		15314	1789	907	50.7

Quelle: Eigene Darstellung

Wie die Zahlen der angefragten Mitarbeitenden erkennen lassen, wurden nicht sämtliche Mitarbeitende der angegebenen Verwaltungen befragt. In den aufgeführten Departementen und Direktionen sind Verwaltungsmitarbeitende zu finden, die unterschiedliche Tätigkeiten ausführen, wie Hauswartinnen und Hauswarte, Försterinnen und Förster, Lebensmittelkontrolleurinnen und Lebensmittelkontrolleure, Sozialarbeitende, Berufsberatende, Krankenpflegefachleute, Ärztinnen und Ärzte. Um eine zu breite Befragung aller Kantonsmitarbeitenden zu vermeiden und um die Anzahl der Befragten auf einen Kreis zu beschränken, in dem die Befragung rund um Evaluationen auch Sinn ergibt, wurden folgende zwei Eingrenzungskriterien angewendet:

1. Kriterium: In Anlehnung an die „street-level bureaucracay“-Unterscheidung von Lipsky (2010) und in Anlehnung an Hupe/Bufat (2014) wurden Verwaltungsmitarbeitende, welche täglich individuellen Bürgerkontakt haben, nicht befragt.
2. Kriterium: Verwaltungsmitarbeitende, die mehrheitlich Unterhalts- und Instandhaltungsarbeiten unternehmen, wurden ebenfalls von der Befragung ausgeschlossen.

Zwei Gründe führten zum Entschluss zur Eingrenzung der Anzahl an angefragten Personen. Erstens macht es wenig Sinn, Berufsgruppen, für deren Arbeit Evaluationen keine Relevanz haben, in die Analyse miteinzubeziehen. Dem entspricht das zweite Kriterium, das solche Berufsgruppen ausschliessen soll. So ist es beispielsweise nicht zielführend, für die Analyse der Evaluationskultur Personen zu befragen, die für den Unterhalt der Strassen zuständig sind. Zweitens war die Zusicherung der Unterstützung der Vorstehenden der verschiedenen General- und Departementssekretariate wichtig. Eine solche konnte nur gewährleistet werden, wenn der dafür nötige Zeitaufwand geringgehalten werden konnte und die Befragung auf einen kleinen, für die Untersuchung relevanten Kreis von Befragten beschränkt wurde. Eine individuelle Vorselektion der zu Befragenden wäre anhand der beiden Kriterien zu kompliziert und zu aufwändig gewesen. In beinahe allen Verwaltungseinheiten finden sich sowohl Mitarbeitende, die den Kriterien entsprechen, als auch solche, die den Kriterien nicht entsprechen. Um trotzdem einen systematischen und effizienten Auswahlprozess zu erzielen, wurde folgendes Vorgehen gewählt: Alle Verwaltungseinheiten¹¹ wurden zuerst analysiert, und es wurde kontrolliert, inwieweit die Mehrheit der Mitarbeitenden den beiden Kriterien entsprechen. Bei Unsicherheiten wurde Rücksprache mit den Vorstehenden aus den betreffenden Verwaltungseinheiten gehalten. In einem zweiten Schritt wurde, wenn die Mehrheit der Verwaltungsmitarbeitenden beide Kriterien erfüllte, die gesamte Verwaltungseinheit in die Befragung miteinbezogen. Beispielsweise wurde eine Verwaltungseinheit befragt, wenn die Überzahl der dort tätigen Verwaltungsangestellten keinen täglichen individuellen Bürgerkontakt hat und mehrheitlich keine Unterhalts- und Instandhaltungsarbeiten übernimmt. Verwaltungseinheiten, die diesen Kriterien entsprechen, werden im Folgenden als „zentrale Verwaltungseinheiten“ bezeichnet und für die Online-Befragung berücksichtigt. Verwaltungseinheiten, welche die Kriterien nicht erfüllen, sind nicht Teil der Online-Befragung.

¹¹ In den untersuchten Verwaltungen finden sich folgende zwei Ebenen von Untereinheiten: Die Departemente (Basel-Stadt, Bern) und Direktionen (Luzern) lassen sich weiter unterteilen in Ämter (Bern), Bereiche (Basel-Stadt) und Dienststellen (Luzern).

Drei Beispiele sollen diesen Selektionsprozess näher erläutern. Erstens das Regionale Arbeitsvermittlungszentrum Bern (RAV), das arbeitslose Personen bei der Stellensuche unterstützt (siehe RAV 2015). Die Mehrheit der Personen in dieser Verwaltungseinheit verfügt täglich über individuellen Bürgerkontakt, weshalb keine Mitarbeitenden dieser Verwaltungseinheit befragt wurden. Die Erziehungsberatung der Erziehungsdirektion Bern – als zweites Beispiel – befasst sich mit Fragen rund um die Kindererziehung. Das angebotene Fachwissen dieser Verwaltungseinheit ist für die gesamte Bevölkerung zugänglich (siehe ERZB 2015). Die Mitarbeitenden dieser Verwaltungseinheit wurden ebenfalls nicht befragt, weil sie täglich und individuell mit Bürgern in Kontakt stehen. Das dritte Beispiel ist die Abteilung Volksschule des Amtes für Kindergarten, Volksschule und Beratung der Erziehungsdirektion Bern. Diese ist verantwortlich für die Entwicklung der Grundschule sowie die Koordination und Implementierung unterschiedlicher Massnahmen, beispielsweise für Fremdsprachen, Lehrmittel und interkulturelle Bildung (siehe AVD 2015). Die Abteilung entspricht den beiden Kriterien als zentrale Verwaltungseinheit, weshalb deren Mitarbeitenden für diese Untersuchung befragt wurden.

Der beschriebene Selektionsprozess resultierte in einer starken Reduktion der angefragten Mitarbeitenden, die in der nachfolgenden Tabelle 2, aufgeschlüsselt nach den Departementen und Direktionen, ersichtlich ist. Die Erziehungsdirektion des Kantons Bern beispielsweise zählt insgesamt 1519 Mitarbeitende. 289 Personen (rund 19 %) entsprachen den beiden Kriterien und wurden deshalb angefragt.

Tabelle 2: Mitarbeitende in zentralen Verwaltungseinheiten

Department/Direktion		Gesamt	Angefragt	Angefragt (in Prozent)
Luzern	Gesundheits- und Sozialdepartement	377	170	45.1
	Bildungs- und Kulturdepartement	3057	105	3.4
	Bau- Umwelt- und Wirtschaftsdepartement	387	125	32.3
Basel-Stadt	Gesundheitsdepartement	387	179	46.3
	Erziehungsdepartement	4859	64	1.3
Bern	Gesundheits- und Fürsorgedirektion	2960	234	7.9
	Erziehungsdirektion	1539	289	18.8
	Bau- Verkehrs- und Energiedirektion	873	410	49.0
	Volkswirtschaftsdirektion	875	213	24.9
Total		15314	1789	11.7

Anmerkungen: Die Daten bezüglich der Anzahl an Mitarbeitenden stammen aus unterschiedlichen Quellen. Diese werden im vierten Kapitel in der Beschreibung der Departemente und Direktionen aufgeführt.

Quelle: Eigene Darstellung

1.4 Forschung über die Evaluationsentwicklung in der Schweiz

Das vorliegende Unterkapitel wird die zunehmende Bedeutung der Evaluation in der Schweiz in den letzten Jahrzehnten aufzeigen, indem die einerseits für die Evaluation wichtigen Entwicklungsschritte dargestellt und andererseits bisher durchgeführte Studien über die Evaluationsaktivität in der Schweiz präsentiert werden.

Die bestehenden Forschungsergebnisse verweisen auf wesentliche Unterschiede in der Evaluationsaktivität innerhalb der verschiedenen Politikfelder und Verwaltungseinheiten und deuten damit auf eine unterschiedliche Evaluationskultur hin. Auf eine umfassende Darstellung der gesamten Forschungsaktivität über Evaluation in der Schweiz wird verzichtet. Hierfür gibt es mit dem Beitrag von Widmer et al. (2016) eine aktuelle Zusammenstellung der Forschungsbeiträge über Evaluation in der Schweiz.

1.4.1 Evaluationsentwicklung in der Schweiz

Zwischen 1970 und 1990 wurden die ersten Evaluationen auf kantonaler Ebene durchgeführt. Insbesondere das Politikfeld Bildung wurde häufig evaluiert (Stamm 2003). Auch Mader (1985) sowie Horber-Papazian/Thévoz (1990) verweisen auf die kantonale Ebene, auf der die ersten Vorläufer der Evaluation im Bildungsbereich zu finden sind. Widmer/De Rocchi (2012) betonen zwei wichtige Schritte bei der Entwicklung der Evaluation in der Schweiz. Auf Bundesebene setzte erstens das Eidgenössische Justiz- und Polizeidepartement (EJPD) die Arbeitsgruppe Gesetzesevaluation (AGEVAL) ein, welches die Aufgabe hatte, Gesetzesevaluationen zu untersuchen. Der zweite wichtige Schritt war das Nationale Forschungsprogramm „Wirksamkeit staatlicher Massnahmen“ (NFP 27), das 1987 lanciert wurde und die Untersuchung der Professionalisierung der schwach entwickelten Evaluationsaktivität in der Schweiz zum Ziel hatte. Durch dieses Programm wurden zwischen 1990 und 1995 einige Evaluationen angestossen (Bussmann 1997; Widmer 1996), und 1995 wurde in der Folge die Bildung eines Netzwerkes (Netzwerk Evaluation) in der Bundesverwaltung als auch die Gründung der Schweizerischen Evaluationsgesellschaft (SEVAL) initiiert (Bussmann 2007). Die SEVAL vereinigt Evaluationsexperten mit unterschiedli-

chen beruflichen Hintergründen, organisiert Konferenzen und gibt die wissenschaftliche Zeitschrift „LeGes – Gesetzgebung & Evaluation“ heraus (Widmer/De Rocchi 2012).

Ein weiterer Meilenstein in der Evaluationsentwicklung der Schweiz ist die 1999 eingeführte Evaluationsklausel in der Bundesverfassung. Artikel 170 (Wirksamkeit staatlicher Massnahmen) ist allerdings vage formuliert und lässt Interpretationsspielraum offen.¹² Wichtiger für die Evaluationspraxis sind gemäss Widmer/De Rocchi (2012) die Evaluationsklauseln in Gesetzen und Regulierungen, welche die Regierung oder die Verwaltung zur Durchführung von Evaluationen in spezifischen Fällen verpflichten. Eine umfassende Beschreibung der Entwicklung der Evaluation in der Schweiz bietet der Beitrag von Widmer (2017), der die historische Entwicklung der Evaluation in vier Phasen unterteilt: Angefangen bei der Pionierphase bis 1990 und nach der Aufbauphase von 1990 bis 2000 und der Reifephase von 2000 bis 2010 befindet sich gemäss Widmer (2017) die Evaluation seither in einer Bürokratisierungsphase, die sich durch stark standardisierte und normierte Evaluationsabläufe auszeichnet.

1.4.2 Forschung über die Evaluationsaktivität in der Schweiz

Es finden sich bislang einzelne Studien, welche die Verbreitung der Evaluation in der Schweiz untersuchen. Sehr selten dagegen wird die Evaluationsaktivität in den Kantonen thematisiert. Tabelle 3 führt die Untersuchungen zur Evaluationsaktivität in der Schweiz auf. Die Forschungsergebnisse dieser Untersuchungen werden im Folgenden präsentiert.

¹² Für eine ausführlichere Diskussion der Bedeutung dieses Gesetzestextes wird auf Mader (2009) sowie Widmer/De Rocchi (2012) verwiesen.

Tabelle 3: Untersuchungen der Evaluationsaktivität in der Schweiz

Autoren	Ergebnisse	Zeitraum	Ebene (Erhebung)
Zimmermann et al. (1989)	Insgesamt 168 Evaluationen: Wirtschaft (35 Evaluationen), Politisches System und Recht (19 Evaluationen), Raumplanung und Wohnungswesen (19 Evaluationen), Land- und Forstwirtschaft (18 Evaluationen), Umweltschutz (15 Evaluationen), Energie (13 Evaluationen), Verkehr (12 Evaluationen), Bildung und Gesellschaft (12 Evaluationen), Gesundheit (8 Evaluationen), Soziale Sicherheit (5 Evaluationen).	1978–1989	Bund (Recherche Datenbank, Befragung der Bundesämter und Befragung Evaluierende)
Balthasar (2007)	Insgesamt 278 Evaluationen: – In 14 der insgesamt 45 untersuchten Verwaltungseinheiten werden keine Evaluationen durchgeführt. – 22 Verwaltungseinheiten können zwischen einer und sieben Evaluationen zugeordnet werden. – In sieben Verwaltungseinheiten lassen sich zwischen elf und 49 Evaluationen erfassen.	1999–2002	Bund (Befragung in den Departementen)
Balthasar/Rieder (2009)	Insgesamt 242 Evaluationen: – Höchste Evaluationsaktivität in den Kantonen Zürich, Bern und Genf. – Seltene Evaluationsaktivität in den Kantonen Uri, Obwalden, Nidwalden, Zug, Basel-Landschaft, Tessin und Wallis. – Sehr seltene Evaluationsaktivität in den Kantonen Glarus, Freiburg, Schaffhausen, Appenzell Innerrhoden, Appenzell Ausserrhoden, Graubünden, Thurgau, Neuenburg und Jura. Ausserdem enorme Unterschiede zwischen den Themenfeldern: Bildung (108 Evaluationen), Gesundheit (23 Evaluationen) und Verwaltungsreformen (7 Evaluationen).	2003–2008	Kantone (Onlinerecherche und Telefoninterviews)
Balthasar (2015)	Insgesamt 234 Evaluationen (Zeitraum 2009–2012): Es finden sich erhebliche Unterschiede in der Evaluationsaktivität zwischen den Verwaltungseinheiten.	1999–2002 sowie 2009–2012	Bund (Einbezug Datenbank von Balthasar 2007, „ARAMIS“-Datenbank sowie Onlinerecherche)
Balthasar/Strotz (2017)	Insgesamt 1414 Evaluationen: – Elf Verwaltungseinheiten verfügen über eine tiefe Evaluationstätigkeit mit weniger als einer Evaluation im Jahr. – Acht Verwaltungseinheiten verfügen über eine mittlere Evaluationstätigkeit mit im Durchschnitt zwei bis drei Evaluationen im Jahr. – In neun Verwaltungseinheiten wird häufiger evaluiert (vier oder mehr Evaluationen im Jahr).	1999–2015	Bund (Einbezug Datenbank von Balthasar 2007, 2015, „ARAMIS“-Datenbank, die „SynEval“-Datenbank* sowie Onlinerecherche)
Frey et al. (2017)	Insgesamt 657 Evaluationen: – 364 Evaluationen (55.4 %) lassen sich der Bundesebene zuordnen. – 148 Evaluationen (22.5 %) lassen sich dem Kanton Zürich, 94 Evaluationen (14,3 %) dem Kanton Genf und 70 Evaluationen (10.7 %) dem Kanton Bern zuordnen.	2000–2012	Bund und drei Kantone („SynEval“-Datenbank)*

Anmerkungen: *Im Rahmen des „SynEval“-Programms wurde eine öffentlich zugängliche Datenbank erstellt, welche die Evaluationsberichte der Jahre 2000 bis 2012 beinhaltet (www.syneval.ch).

Quelle: Eigene Darstellung

Die erste Untersuchung der Evaluationsaktivität wurde von Zimmermann et al. (1989) durchgeführt. Für den Zeitraum zwischen 1978 und 1989 wurden insgesamt 168 Evaluationen identifiziert. Die meisten Evaluationsberichte können dem Politikfeld Wirtschaft (35) zugeordnet werden. Ebenso kommen in den Politikfeldern Politisches System/Recht (19), Raumplanung/Wohnungswesen (19), Land- und Forstwirtschaft (18) sowie Umweltschutz (15) Evaluationen häufig zur Anwendung. Eine geringere Evaluationsaktivität lässt sich in den Politikfeldern Gesundheit (8) und Soziale Sicherheit (5) ausmachen. Allerdings gehen Widmer/De Rocchi (2012) davon aus, dass es sich bei der Studie von Zimmermann et al. (1989) einerseits nicht um eine Vollerhebung der tatsächlichen Evaluationsaktivität handelt, andererseits die Untersuchung von einem breiten Verständnis von Evaluationen ausgeht und deshalb auch Grundlagenstudien in die Erhebung aufgenommen wurden.¹³

Balthasar (2007) untersucht die Verwendung der Evaluation in der Schweiz für den Zeitraum zwischen 1999 und 2002. Insgesamt kann die Untersuchung 278 Evaluationsberichte identifizieren und grosse Differenzen zwischen den Politikfeldern und Verwaltungseinheiten aufzeigen: Beinahe ein Drittel der untersuchten 45 Institutionen der Bundesverwaltung verfügen über keine Evaluationsaktivität. Im Gegensatz dazu wird in sieben Verwaltungseinheiten sehr viel evaluiert, beispielsweise im Staatssekretariat für Wirtschaft (SECO) mit 49 Evaluationen. Das Bundesamt für Gesundheit (BAG) und die Direktion für Entwicklung und Zusammenarbeit (DEZA) sind für je 25 und damit auch für zahlreiche Evaluationen verantwortlich. Dem Bundesamt für Kultur (BAK) und dem Bundesamt für Polizei (fedpol) kann je nur eine Evaluation zugewiesen werden. Zum Vergleich ermittelt Balthasar (2015) zehn Jahre später erneut Evaluationsberichte auf Bundesebene: Für den Zeitraum 2009 bis 2012 können insgesamt 234 Evaluationen erfasst werden. Auch in die-

¹³ Das Bundesamt für Justiz führte die Erhebung für den Zeitraum von 1991 bis 1995 weiter. Es wurden insgesamt 534 Evaluationen gefunden (Bundesamt für Justiz 1996). Bei dieser Studie kam ebenfalls eine breitere Definitionsauslegung von Evaluation zur Anwendung.

ser Untersuchung zeigen sich erhebliche Unterschiede in der Evaluationsaktivität zwischen den Verwaltungseinheiten des Bundes.

Balthasar/Rieder (2009) können durch Internetrecherchen und Telefoninterviews mit Experten für den Zeitraum 2003 bis 2008 insgesamt 242 kantonale Evaluationen ausfindig machen. Verglichen mit der oben erwähnten Studie von Balthasar (2007) kommen sie zu dem Schluss, dass fast doppelt so viele Evaluationen auf Bundesebene (77 Evaluationen im Jahr) durchgeführt werden wie auf Kantonsebene (insgesamt 40 Evaluationen im Jahr). Zudem stellen sie grosse Differenzen in der Evaluationsaktivität zwischen den Kantonen fest: In den Kantonen Zürich, Bern und Genf können die meisten Evaluationen erfasst werden. Dem stehen neun Kantone gegenüber, denen keine oder nur sehr wenige Evaluationen zugeordnet werden können: Glarus, Freiburg, Schaffhausen, Appenzell Innerrhoden, Appenzell Ausserrhoden, Graubünden, Thurgau, Neuenburg und Jura. In den sieben Kantonen Uri, Obwalden, Nidwalden, Zug, Basel-Landschaft, Tessin und Wallis wird selten evaluiert. Wie schon für die Bundesebene festgestellt, weist auch die Untersuchung in den Kantonen auf Unterschiede zwischen den verschiedenen Politikfeldern hin: Am meisten Evaluationen konnten dem Politikfeld Bildung (108 der 242 erfassten Evaluationen) zugeordnet werden. Auf das Politikfeld Gesundheit (23 Evaluationen) und den Themenbereich Verwaltungsreformen (7 Evaluationen) fallen erheblich weniger Evaluationen. Der Beitrag von Balthasar/Strotz (2017) untersucht die Evaluationsaktivität auf Bundesebene und erfasst insgesamt 1414 Evaluationen. Aus den erhobenen Daten lässt sich eine leichte Zunahme der Anzahl an durchgeführten Evaluationen im Zeitraum 1999 bis 2015 ablesen. Zudem offenbaren sich starke Unterschiede der Aktivität in den einzelnen Jahren: 2001 ist das Spitzenjahr mit 112 Evaluationen gegenüber dem Jahr 2003 mit der niedrigsten Aktivität von 51 Evaluationen. Des Weiteren stellen die beiden Autoren fest, dass diejenigen Verwaltungseinheiten, die bereits früher evaluiert hatten, dies auch heute noch tun; beispielsweise lassen sich dem Staatssekretariat für Wirtschaft (SECO) 131 Evaluationen zuweisen. Auch im Bundesamt für Gesundheit (BAG) wird mit 169 und in der Direktion

für Entwicklung und Zusammenarbeit (DEZA) mit 163 Evaluationen sehr viel evaluiert. Bereits Balthasar (2007) weist diesen Verwaltungseinheiten eine hohe Evaluationsaktivität zu. Zudem lässt sich aus der Gesamtanzahl der von Balthasar/Strotz (2017) erfassten Evaluationsberichte ableiten, dass in den Politikbereichen Gesundheit, Wirtschaft, Soziales, Entwicklungszusammenarbeit und Umwelt am meisten evaluiert wird.

Frey et al. (2017) werten die im Rahmen des Programmes „SynEval“ erfassten Evaluationsberichte aus. Erhoben werden die Evaluationen der Bundesebene und der drei Kantone Bern, Genf und Zürich in den Politikfeldern Bildung, Energie und Gesundheit. Unterschiede in der Evaluationsaktivität ergeben sich für die verschiedenen Politikbereiche: Am meisten Berichte können der Gesundheit (44.1 %) zugewiesen werden, gefolgt von der Bildung (39.3 %) und der Energie (19.2 %). Des Weiteren lässt sich auch in den „SynEval“-Daten insgesamt eine Zunahme der erfassten Evaluationen feststellen. Zudem werden kantonale Politiken im Vergleich zu nationalen Politiken weniger häufig evaluiert; von den erfassten Evaluationen lassen sich 55.4 % der Bundesebene zuordnen. In den Kantonen Zürich (22.5 %), Genf (14.3 %) und Bern (10.7 %) wird weniger häufig evaluiert.

Die vorgestellten Beiträge zeigen, dass es keine kontinuierliche Untersuchung der Verbreitung der Evaluation über einen längeren Zeitraum gibt. Die verschiedenen Studien orientieren sich zudem an unterschiedlichen Definitionen und Methoden.¹⁴ In jüngerer Zeit lässt sich allerdings mit den Beiträgen von Balthasar (2007, 2015), Balthasar/Strotz (2017) sowie Frey et al. (2017) eine zunehmende Forschungstätigkeit im Bereich der Untersuchung der Evaluationsaktivität feststellen. Diese Beiträge fokussieren hauptsächlich auf die Bundesebene, mehrheitlich unerforscht bleibt die Situation der Evaluation in den Kantonen. Die fehlende Übersicht über die Evaluationsaktivität hängt sicherlich damit zu-

¹⁴ Siehe dazu die Ausführungen von Widmer/De Rocchi (2012).

sammen, dass es bisher keine öffentlich zugänglichen Datenbanken oder Archive gibt, in denen die Evaluationsberichte zentral gesammelt werden. Eine Zusammenstellung und Recherche der Evaluationsaktivität ist aufgrund der vielen involvierten Akteure mit erheblichem Aufwand verbunden (siehe dazu die Ausführungen von Frey et al. 2017: 83f.). Dessen ungeachtet weist die bisherige Forschungstätigkeit auf erhebliche Unterschiede in der Evaluationsaktivität bei den Politikfeldern und den untersuchten Verwaltungseinheiten hin. Aufschlussreich ist es auch, im Folgenden die Entwicklung und den Stand der Evaluation in der Schweiz im internationalen Vergleich einordnen.

1.4.3 Einordnung der Entwicklung der Evaluation in der Schweiz im internationalen Vergleich

Gemäss Widmer/Neuenschwander (2004) gehört die Schweiz nicht zu den Vorreitern in der Durchführung von Evaluationen. Die beiden Autoren beziehen sich auf Leeuw (2000), der Schweden, Deutschland und Grossbritannien eine Pionierrolle in der Etablierung der Evaluationskultur in Europa zuweist. Ebenso attestiert die bekannteste vergleichende Studie zur Evaluationskultur von Furubo et al. (2002) der Schweiz keine lange Tradition der Evaluationskultur im Gegensatz zu Kanada, Schweden, Deutschland und Grossbritannien, wo durch den Einfluss des Evaluationsdiskurses in den USA der 60er und 70er Jahren allmählich eine Evaluationskultur entstand. Weitere Beiträge teilen diese Einschätzung der verspäten Evaluationsentwicklung in der Schweiz (siehe Horber-Papazian/Thévoz 1990; Spinatsch 2002).

Widmer/Beywl (2009: 511f.) diskutieren anhand der Einschätzung von zehn Politikfeldern die Entwicklung der Evaluation in den Ländern Deutschland, Schweiz und Österreich. Die Autoren argumentieren, dass die Schweiz im Vergleich zu Deutschland und Österreich durchaus eine Vorreiterposition einnimmt, da in den beiden anderen Ländern in den meisten Themenfeldern keine stark verbreitete Evaluation auf „hohem Reflexionsstand und mit ausgeprägter Institutionalisierung“ zu finden sei. Die bekannte Studie von

Furubo et al. (2002) vergleicht mithilfe von neun Kriterien den Entwicklungsstand der Evaluation in 21 Ländern. Dabei wird die Schweiz mit dem 14. Rang im unteren Mittelfeld eingestuft. Später wurde dieser Vergleich von Jacob et al. (2015) anhand der gleichen neun Kriterien in 19 OECD-Ländern wiederholt und die Schweiz nun nach Finnland auf dem zweiten Platz eingestuft. Der Beitrag von Jacob/Varone (2003) vergleicht den Stand der Institutionalisierung der Evaluation anhand von sechs Kriterien in 24 Ländern. Die Schweiz teilt sich in diesem Ranking zusammen mit Australien den ersten Platz. Einige Jahre später hat Varone (2007) diese Studie aktualisiert und damit einhergehend die Schweiz in die Gruppe von Ländern mit hoher Institutionalisierung kategorisiert.

Die vorgestellten Beiträge konnten einerseits aufzeigen, dass die Schweiz zwar nicht zu den Wegbereitern der Evaluation gezählt werden kann und die Evaluationstätigkeit verspätet einsetzte. Andererseits attestieren die aktuellen internationalen Untersuchungen der Schweiz Spitzenplätze in der Institutionalisierung der Evaluation.

1.5 Aufbau

Das nachfolgende zweite Kapitel widmet sich der Definition und Herleitung der organisationalen Evaluationskultur. Darauf aufbauend werden im dritten Kapitel die beiden theoretischen Modelle erläutert, welche die zu untersuchenden Einflüsse auf die Evaluationskultur aufzeigen. Ausserdem werden die konkreten Forschungshypothesen formuliert.

Die Kapitel vier, fünf und sechs markieren die empirischen Teile der Untersuchung: Im vierten Kapitel wird die organisationale Evaluationskultur operationalisiert, Kapitel fünf beschreibt die Evaluationskultur in den neun Departementen und Direktionen, während Kapitel sechs sich der Untersuchung der Einflüsse auf die Evaluationskultur, wie sie im theoretischen Modell in Kapitel drei dargestellt wurden, widmet. In den beiden letzten Kapiteln der Untersuchung werden die empirischen Befunde zusammenfassend dargestellt, die Ergebnisse diskutiert und die zentralen Schlussfolgerungen präsentiert.

A. Theoretischer Teil

2 Was ist die organisationale Evaluationskultur?

In der Einleitung wurde darauf hingewiesen, dass eine vertiefende, theoretische Auseinandersetzung mit der organisationalen Evaluationskultur in der Forschungsliteratur nur ansatzweise zu finden ist. Dieses Kapitel wird sich dieser Forschungslücke widmen.

Folgende Gründe sprechen für eine vertiefende Aufarbeitung der organisationalen Evaluationskultur. Beispielsweise erfordert die breite Verwendung des Kulturbegriffs eine Auseinandersetzung mit Messung und Definition von Kultur. Des Weiteren verfügt Kultur über eine amorphe Eigenschaft. Diese Eigenschaft bringt die folgende Aussage von Lee/Lawrence (2013: 114) treffend zum Ausdruck: „Culture has something in common with atoms: you can see where they have been, but you cannot actually see them“. Ausserdem besteht keine einheitliche Kulturdefinition. Dadurch besteht Unklarheit betreffend der Bestandteile von Kultur. Beispielsweise ist dies in der Forschung über die Organisationskultur festzustellen, die in diversen Beiträgen Unterschiede der Definitionen und auch der Messung der Organisationskultur thematisiert (siehe Martin 2002; Smircich 1983). Das Zitat von Lee/Lawrence drückt zudem das Vorhandensein von latenten, nicht direkt messbaren Merkmalen bei der Untersuchung von Kultur aus.¹⁵ In der Literatur wird oft auf die Multidimensionalität der Kultur verwiesen, was bedeutet, dass Kultur sich nicht auf eine Dimension (Eigenschaft) beschränkt, sondern sich in mehreren Dimensionen ausprägt (siehe Lee/Lawrence 2013; Martin 2002; Schein 2010; Smircich 1983). Auch dies spricht für eine vertiefende Auseinandersetzung mit der Evaluationskultur. Ein weiterer Grund für eine theoretische Erörterung ist, dass die Untersuchung der Evaluationskultur

¹⁵ Siehe für eine ausführlichere Beschreibung des Begriffes der latenten Variable den Beitrag von Kopp/Lois (2012: 80ff.).

ein neuer Forschungsbereich ist und noch viele Konzeptionalisierungsfragen offen sind. Insofern sollen die folgenden Überlegungen einen Beitrag zu den möglichen und relevanten Elementen für die Messung der Evaluationskultur leisten. Diese Aufarbeitung erfolgt im vorliegenden Kapitel, das sich der theoretischen Herleitung und Definition der organisationalen Evaluationskultur widmet.

Da sich die Forschung über Evaluation bislang nur begrenzt mit der Evaluationskultur befasst, werden auch Forschungsbeiträge aus verwandten Disziplinen zur Herleitung der Definition herangezogen. Durch den Einbezug der Forschungsliteratur aus der politischen Kulturforschung, der Organisationskulturforschung sowie aus der Forschung über Evaluation können wichtige Erkenntnisse für die theoretische Herleitung der organisationalen Evaluationskultur gewonnen werden.

Aus diesem Grund wird zuerst das Konzept der politischen Kultur erläutert und darauf folgend wichtige Beiträge der Organisationskulturforschung vorgestellt. Zudem werden ausgewählte Konzepte der Forschung über Evaluation mit einbezogen, um am Schluss eine anwendbare Definition zu erhalten.

Politische Kultur

Die Untersuchung von Almond/Verba (1963) erlangte in der Politikwissenschaft grosse Bekanntheit und gilt als Geburtsstunde der systematischen Untersuchung der politischen Kultur. Diese wird von den beiden Autoren folgendermassen definiert: „[...] political system as internalised in the cognitions, feelings, and evaluations of its population” (Almond/Verba 1963: 14). Demzufolge resultiert die politische Kultur aus der Verteilung von individuellen Einstellungen, Beurteilungen und Wahrnehmungen gegenüber politischen Objekten. Dieses Konzept erlaubt es gemäss den Autoren, die Stabilität eines politi-

schen Systems zu bestimmen:¹⁶ Ein System gilt dann als stabil, wenn die politische Kultur und das politische System kongruent sind. Das von Almond/Verba (1963) entwickelte Konzept der politischen Kultur wurde seither von verschiedenen Forschenden weiterentwickelt.¹⁷

Das Konzept der politischen Kultur bezieht sich in der aktuellen Forschung ebenso auf die gesammelten Überzeugungen der Bürgerinnen und Bürger gegenüber dem politischen System (für eine Einführung siehe Pickel/Pickel 2006). Wichtig ist der Vermerk auf das Zusammenspiel zwischen der politischen Kultur mit der vorherrschenden politischen Struktur: „Die politische Kultur benötigt eine politische Struktur, zu der ihre Träger Bewertungen aufbauen können“ (Pickel/Pickel 2006: 56). Beispiele der Strukturebene sind Parlamente, Parteien, demokratische Spielregeln oder Elemente der Verfassung. Gemäss den beiden Autoren ist die politische Kultur zwischen der Mikro- und der Makroebene des politischen Systems anzusiedeln. Damit ist gemeint, dass sich die politische Kulturforschung mit der Untersuchung von Überzeugungen, Einstellungen und Werten von Individuen (Mikroebene) gegenüber der Politik und dem politischen System auseinandersetzt. Diese Individualdaten werden jedoch aggregiert betrachtet, um Aussagen auf Ebene der Gesamtbevölkerung (Makroebene) machen zu können.

Die politische Kulturforschung im „klassischen Sinne“¹⁸ unterscheidet zwischen dem politischen Verhalten und der politischen Einstellung. Hier wird für die alleinige Berücksichtigung der vorfindbaren Werte und Einstellungen gegenüber der Politik argumentiert. Das politische Verhalten von Personen, beispielsweise die politische Partizipation, ist gemäss dieser Auffassung nicht Bestandteil der politischen Kultur.

¹⁶ Dabei wurden die untersuchten Länder USA, Großbritannien, Italien, Bundesrepublik Deutschland und Mexiko in eine Grundtypisierung eingestuft

¹⁷ Beispielsweise von Easton (1965a, 1965b) mit einer differenzierteren Untersuchung der Beziehung, die er politische Unterstützung nannte, zwischen dem politischen System und dem Bürger. Oder von Lipset (1959, 1981), der eine Unterscheidung von langfristigen und kurzfristigen Einstellungen untersucht.

¹⁸ Diese Aussage bezieht sich auf die Konzeption und Argumentation gemäss der Forschungstradition von Almond/Verba (1963).

Für eine Ausweitung des politischen Kulturkonzeptes plädiert Rohe (1996). Dieser spricht sich für eine Erweiterung des politischen Kulturbegriffs von seiner ausschliesslichen Fokussierung auf die politische Einstellung aus. Konkret argumentiert Rohe (1996), neben der Einstellung zur Politik ebenso das politische Handeln zu erfassen, da diese die politische Kultur gemeinsam konstituieren. Im Gegensatz zu dem klassischen Ansatz der politischen Kulturforschung, geht es bei ihm darum, auch das politische Verhalten und Handeln zu berücksichtigen.

Das Konzept der politischen Kultur verweist somit auf wichtige Eigenschaften, die bei dieser Untersuchung zu berücksichtigen sind: In diesem Abschnitt wurde aufgezeigt, dass die Messung der politischen Kultur auf aggregierten Individualdaten beruht und sich auf Einstellungen und Werte fokussiert. Zudem ist bei dem Konzept der politischen Kultur zwischen Verhalten und Einstellung zu unterscheiden.

Organisationskultur

Die Grundidee der Organisationskultur geht davon aus, dass sich durch Aufbau und Erlernung von kulturellen Wert- und Handlungsmustern in jeder Organisation eine spezifische Kultur herausbildet. Schein (2010), der als Wegbereiter dieses Konzeptes gilt, strukturiert die Organisationskultur in drei Ebenen: Die erste Ebene der Artefakte (englisch: artifacts) beinhaltet sichtbare Verhaltensweisen, Rituale und Mythen. Beispiele sind das Kommunikationsverhalten der Mitarbeitenden, die Bekleidung, die Büroeinrichtung oder die eingesetzte Technologie. Auf der zweiten Ebene finden sich die unterstützten Vorstellungen und Werte (englisch: espoused values). Diese geben das Gefühl wieder, wie die Dinge sein sollten und beinhalten kollektive Werte der Organisation; beispielsweise gegenüber der Technik oder die Offenheit gegenüber Neuerungen. Die dritte Ebene beinhaltet Grundannahmen (englisch: basic underlying assumptions): unbewusste, selbstverständliche Vorstellungen und Werte, die das Verhalten, die Wahrnehmung, das Denken und das Fühlen determinieren. Die drei genannten Ebenen unterscheiden sich zudem in

der Beobachtbarkeit. Während Artefakte (Ebene 1) gut beobachtbar sind, ist die Erfassung von Werten und Vorstellungen (Ebene 2) weitaus schwieriger. Noch schwieriger verhält es sich mit der Beobachtung der Grundannahmen (Ebene 3), diese sind selbstverständliche und unsichtbare Wahrnehmungen, die nur schwer zu erfassen sind. Damit argumentiert Schein (2010: 14ff.) für eine breite Konzeption der Organisationskultur, die an verschiedenen Aspekten sichtbar wird und über Normen, Verhaltensregeln und Meinungen bis hin zu gemeinsamen Werten reicht.

Ein weiterer und oft erwähnter Beitrag im Zusammenhang mit der Organisationskultur ist die Studie von Hofstede (1980). Diese vergleicht die Werte von Mitarbeitenden aus 40 verschiedenen Ländern entlang von vier Dimensionen⁹: Machtdistanz, Unsicherheitsvermeidung, Individualismus-Kollektivismus und Maskulinität-Femininität. Anhand dieser Dimensionen lassen sich unterschiedliche nationale Wertesysteme der Mitarbeitenden ableiten. So werden beispielsweise schwedische Angestellte von anderen Werten beeinflusst, als chinesische. Hofstede vertritt damit die Annahme, dass es eine einheitliche nationale Kultur gibt. Dieser Ansatz wird von verschiedener Seite allerdings in Frage gestellt (beispielsweise von Barbier 2012; Forss 2012).

Ein wichtiger Beitrag zur Erforschung der Organisationskultur stammt von Smircich (1983: 339). Der Beitrag weist auf die fehlende Einheitlichkeit bezüglich der Messung der Organisationskultur hin: "The culture concept has been borrowed from anthropology, where there is no consensus on its meaning. It should be no surprise that there is also a variety in its application to organization studies". In ihrem Artikel analysiert die Autorin die Beiträge von verschiedenen Organisationskulturforschenden und hebt unterschiedliche Konzeptionen von Kultur hervor. Die Differenzen in der Herangehensweise an die Beziehung zwischen Kultur und Organisation basieren auf unterschiedlichen Grundan-

⁹ In seinem Nachfolgewerk kommt mit der Langzeit- vs. Kurzzeitorientierung eine fünfte Dimension hinzu (Hofstede 2001).

nahmen bezüglich der beiden Begriffe. Beispielsweise zeigt Smircich (1983), dass Kultur einerseits als Organisationsvariable (sogenannte kritische Variable), andererseits von anderen Forschenden auch als „root metaphor“ benutzt wird. Demzufolge kann die Organisationskultur einerseits als eine Variable aufgefasst werden, die von einer Organisation verändert werden kann (Variablenansatz). Andererseits gibt es Forschende, welche die Kultur und die Organisation als identisch verstehen (root-metaphor-Ansatz). Bei diesem Ansatz wird die Kultur als „root metaphor“ angesehen, somit kann die Organisation nicht nur entlang von ökonomischen und materiellen Gesichtspunkten analysiert werden. Für eine umfassende Untersuchung sind auch symbolische, expressive und identitätsstiftende Aspekte zu berücksichtigen.

Martin (2002) weist auf weitere Aspekte und Eigenschaften der Organisationskultur hin. Bemerkenswert ist ihr Verweis, dass die Organisationskultur gleichzeitig integriert, differenziert und fragmentiert sein kann. Eine Organisationskultur ist bei der Betrachtung eines „breiten“ Konsenses integriert, differenziert bei der Betrachtung von unterschiedlichen Subkulturen und fragmentiert bei unklaren Situationen, bei welchen kein Konsens vorherrscht. Auch thematisiert Martin (2002) die unterschiedlichen Untersuchungsergebnisse, die aufgrund verschiedener Forschungszugänge und unterschiedlicher Untersuchungstiefe resultieren. Wird ein eher quantitativer Ansatz gewählt, der spezifische Aspekte in vielen Organisationen untersucht, können andere Schlussfolgerungen gezogen werden, als bei der Anwendung einer qualitativeren Methode. Des Weiteren zeigt Martin (2002: 21) die Debatte zwischen Forschenden unterschiedlicher Ausrichtung auf: „Cultural researchers have vehement methodological disputes. Quantitative culture researchers often disdain qualitative studies and vice versa. Qualitative researchers also quarrel among themselves, usually about whether a researcher has spent enough time to gain a deep understanding of a culture“.

Auf die unterschiedlichen Herangehensweisen bei der wissenschaftlichen Untersuchung von Kultur geht auch Forss (2012) ein. Laut ihm spiegelt die Definition von Kultur drei

epistemologische Herangehensweisen wider: Erstens kann sich dem Konzept der Kultur deduktiv genähert werden, bei der vorab die zu untersuchenden Dimensionen definiert werden: „Knowledge comes from constructing these dimensions, looking to see where organizations fall, and then revising the dimensions when previously overlooked phenomena are noticed” (Forss 2012: 183). Die zweite Herangehensweise ist der induktive Ansatz, die implizite Elemente identifiziert und beschreibt: „These tacit elements can sometimes be made explicit, but always shape the experience of specified constructs“ (Forss 2012: 184). Die dritte Herangehensweise wird als die „Radikale“ bezeichnet: Dem Forschenden wird eine unvoreingenommene und an der Richtigkeit interessierte Sichtweise abgesprochen. Vielmehr wird ein „Konstrukt“ verfolgt, das seinen eigenen Interessen und Erfahrungen entspricht. Zudem sei, argumentiert Forss (2012: 183), die Organisationskultur als ein komplexes Konzept zu verstehen, das sich aus verschiedenen Individuen zusammensetzt: “Clearly, any corporate culture will be complex it is, after all, the crucible in which many people, with widely different identities of gender, ethnicity, and so forth, come together and have to manage and be managed”.

Barbier (2012) diskutiert die universalistische und die kulturalistische Sichtweise von Kultur. Die kulturalistische Sichtweise betont die Determinierung der Menschen durch die Kultur: „They [human beings] are seen as the product of their societies, and, in each society prevails a set of different 'values'" (Barbier 2012: 4). Auf der Gegenposition sind die Universalisten anzusiedeln, welche die Wichtigkeit der Kultur herunterstufen. “Culture in their sense is merely one factor in the 'context'; and human/social phenomena are, by ontology, universal” (Barbier 2012: 5).

Die bestehende Forschungsliteratur diskutiert somit verschiedene Zugänge bei der Untersuchung von Kultur: Den eher quantitativ und deduktiv orientierten Zugang, gemäss den Universalisten, beziehungsweise den eher qualitativ und induktiv orientierten Zugang, in der Tradition der Kulturalisten.

Des Weiteren ist der Forschungsliteratur die Berücksichtigung mehrerer Dimensionen bei

der Konzeptualisierung der Organisationskultur zu entnehmen (siehe Lee/Lawrence 2013; Martin 2002).

Forschung über die Evaluationskultur

Dieser Absatz widmet sich dem Forschungsstand zur Evaluationskultur. Zudem werden Indikatoren und Konzepte, die im Zusammenhang mit der Evaluationskultur stehen, vorgestellt. Ausgangspunkt der Bestandsaufnahme sind Artikel mit Überblickscharakter (Hense/Widmer 2013; Nielsen et al. 2011; Taylor-Ritzler et al. 2013) sowie die eigene Literaturrecherche. Dazu wurden die Evaluationsfachzeitschriften *American Journal of Evaluation*, *Canadian Journal of Program Evaluation*, *Evaluation*, *Evaluation and Program Planning* und *New Directions for Evaluation* von 2000 bis 2013 systematisch ausgewertet.

Die nachfolgenden zwei Tabellen fassen die erfassten Beiträge zusammen. Tabelle 4 umfasst Beiträge, die explizit die Evaluationskultur adressieren, während in Tabelle 5 weitere, für die Evaluationskultur relevante, Untersuchungen präsentiert werden.

Zur Evaluationskultur lassen sich 18 Beiträge ermitteln (Tabelle 4). Aufgrund der hohen Anzahl an erfassten Beiträgen wird darauf verzichtet, auf alle Beiträge einzeln einzugehen. Vielmehr soll ein Überblick gegeben werden, der aufzeigen soll, wie und was bisher untersucht wurde. Auf einzelne besonders relevante Beiträge wird punktuell eingegangen werden.

Von den 18 Beiträgen befasst sich die Hälfte mit der Evaluationskultur auf Länderebene: Schweiz (Horber-Papazian/Jacot-Descombes 2012), Italien (Stame 2012), Spanien (Bustelo 2006), Rumänien (Mihalache 2010), Irland (Boyle et al. 2012; McNamara et al. 2009), Argentinien (Rodríguez-Bilella/Monterde-Díaz 2010), Europäische Länder (Toulemonde 2000) und der Vergleich von 21 Ländern (Furubo et al. 2002). Augenfällig bei diesen Beiträgen sind die Unterschiede hinsichtlich der untersuchten Indikatoren und

Elementen der Evaluationskultur, als auch in den unterschiedlichen Definitionen. Bei fast allen dieser Beiträge handelt es sich vorwiegend um die Beschreibung der Situation und der Entwicklungslinien der Evaluation des jeweiligen Landes. Dabei steht die Beschreibung der Evaluationsaktivität, der Evaluationsinstitutionalisierung und der Professionalisierung der Evaluation im Vordergrund. Die Untersuchung von Furubo et al. (2002) sticht in dieser Gruppe heraus. Ungewöhnlich ist hier der Vergleich zwischen Ländern und den viele Bereiche umfassenden Indikatoren. Die neun Indikatoren zur Messung der Evaluationskultur wurden als Ausgangspunkt von weiteren Studien übernommen (beispielsweise von Jacob 2005 oder Speer 2012). Allerdings basiert die Analyse der Evaluationskultur auf Experteneinstufungen in den entsprechenden Ländern. Auf diesen Nachteil der subjektiven Sichtweise von Experten wird auch bei der Studie von Jacob et al. (2015) hingewiesen, welche die Untersuchung von Furubo et al. (2002) wiederholten. Die Ermittlung einer nationalen, einheitlichen Evaluationskultur ist zudem mit einer gewissen Ungenauigkeit verbunden. Die Betrachtung der Evaluationskultur auf Länderebene wurde deshalb von verschiedener Seite kritisiert.²⁹ Erwähnenswert ist die Studie von Mihalache (2010), die in Anlehnung an die Studien von Patton (1998) und Trochim (2006) das Vorhandensein von sogenannten evaluativen Wertvorstellungen untersucht. Mit dem Einbezug von evaluativen Wertvorstellungen wird somit ein wichtiges Element der Evaluationskultur untersucht, das in den anderen Beiträgen nicht berücksichtigt wurde.

Sieben der 18 Beiträge können Studien zugeordnet werden, die Organisationen oder Sektoren untersuchen. Davon konzentrieren sich vier der sieben Beiträge auf die Beschreibung der Evaluationskultur einer einzelnen Organisation, respektive eines Sektors im Rahmen einer Fallanalyse: „Australian Volunteers International“, Australien

²⁹Beispielsweise von Barbier (2012) oder Forss (2012), die der Ermittlung einer nationalen Evaluationskultur skeptisch gegenüberstehen, da es innerhalb von Ländern, gerade zwischen Politikbereichen grosse Unterschiede gibt.

(Owen/McDonald 1999); „National Emergency Management“, Australien (Owen 2003); Bundesamt für Gesundheit, Schweiz (Läubli Loud 2014); Entwicklungszusammenarbeit in Deutschland (Meyer/Rech 2013). Der Beitrag von Scheerens (2004) adressiert nicht eine explizite Organisation, sondern diskutiert die Effekte der Selbstevaluation auf die Evaluationskultur in Schulen. Auch der Beitrag von Forss (2012) untersucht die Evaluationskultur auf Organisationsebene. Das Bemerkenswerte an diesem Buchbeitrag ist der Einbezug zahlreicher Indikatoren für die Untersuchung der Evaluationskultur. Neben der Untersuchung der Evaluationsprozesse analysiert der Beitrag von Forss (2012) ebenfalls den Formalisierungsgrad, also das Vorhandensein von Leitfäden und Evaluationsstandards sowie den Status, die Akzeptanz und die Haltung zur Evaluation. Für die Ermittlung der Haltung hatten die Befragten die Möglichkeit, unterschiedliche Attribute wie Respekt, Interesse, Furcht und Spass in eine bestimmte Reihenfolge zu bringen. Auch Speer (2012) berücksichtigt für die Untersuchung der Evaluationskultur, in den Politikfeldern Bildung und Arbeitsmarkt, eine Reihe von Indikatoren (siehe Tabelle 4). Das Bewusstsein des intrinsischen Wertes der Evaluation ist gemäss Speer (2012: 69f.) besonders wichtig für die Einführung der Evaluationssysteme und deren Institutionalisierung.

Tabelle 4: Explizite Beiträge zur Evaluationskultur

Autor/ <i>Gegenstand/</i> Untersuchungsebene	Inhalt der Untersuchung	Definition/Elemente der Evaluationskultur
1. Furubo et al. (2002) <i>21 Länder</i> Länderebene	Die Untersuchung überprüft die Entwicklung der Evaluationskultur in 21 Ländern anhand von Experteneinschätzungen. Die Evaluationskultur wird entlang von neun Indikatoren gemessen: 1. Evaluationen in vielen Politikfeldern, 2. Angebot an einheimischen Evaluatorinnen und Evaluatoren, 3. Bestehender nationaler Diskurs, 4. Professionalisierung mit Vereinigung, 5. Institutionalisierung der Evaluation in der Regierung, 6. Institutionalisierung im Parlament, 7. Pluralismus an Evaluatorinnen und Evaluatoren in jedem Politikfeld, 8. Evaluationsaktivität in der obersten Rechnungsbehörde, 9. Outcome Evaluationen.	Die Autoren verzichten auf eine einheitliche Definition.
2. Horber-Papazian/Jacot-Descombes (2012) <i>Schweiz</i> Länderebene	In diesem Beitrag wird der Einfluss des politischen Systems und der Multikulturalität auf die Evaluationskultur in der Schweiz anhand folgender Elemente untersucht. 1. Die Nachfrage nach Evaluation im schweizerischen politischen System, 2. Die Verteilung von Evaluationsbüros in der Schweiz, 3. Evaluationspraktizierende sowie deren Profil.	In Anlehnung an Barbier (2012) besteht die Evaluationskultur aus Werte, Praxis und Institutionalisierung.
3. Stame (2012) <i>Italien</i> Länderebene	Die Autorin beschreibt die „culture of evaluation“ in Italien und die Entwicklung zu einer darüber hinausgehenden Evaluationskultur. Zudem werden die Hauptcharakteristika der sogenannten „Italian Napoleonic-type politico-administrative tradition“ aufgezeigt sowie die „culture of evaluation“ anhand verschiedener öffentlicher Politiken analysiert.	In Anlehnung an Barbier (2012) wird die Kultur geformt durch Werte, Praxis und Institutionalisierung.
4. Bustelo (2006) <i>Spanien</i> Länderebene	Der Artikel diskutiert die potentielle Rolle von Evaluationsstandards und Evaluationsleitlinien bei der Entwicklung der Evaluationskultur in Spanien.	Die Evaluationskultur besteht aus der akademischen und professionellen Konsolidierung und Institutionalisierung der Evaluationsfunktion.
5. Mihalache (2010) Makro Rumänien	Die Untersuchung analysiert die Evaluationskultur in Rumänien. Dafür wurden Personen mit Evaluationspraxiserfahrung befragt, um folgende zwölf Werte einer idealtypischen Evaluationskultur zu ermitteln: 1. Aktionsorientiert, 2. Lernorientiert, 3. Inklusiv und partizipativ, 4. Reagierend und nichthierarchisch, 5. Innovations- und diversitätsorientiert, 6. Strikt wissenschaftlich, 7. Interdisziplinär, 8. Selbstkritisch, 9. Aufrichtig und objektiv, 10. Ethisch und demokratisch, 11. Vorwärts schauend, 12. Transparent.	Die Evaluationskultur besteht aus Werten von Personen in einem gegebenen organisationalem Kontext (angelehnt an Patton 1998 und Trochim 2006).
6. Toulemonde (2000) <i>Europa</i> Länderebene	Dieser Beitrag diskutiert und analysiert die Evaluationskultur in Europa anhand der Einführung und Entwicklung der Evaluation einerseits, und dem Einfluss der Evaluationen auf die Regeln des Regierens in verschiedenen europäischen Ländern andererseits.	Die Evaluationskultur wird anhand der Evaluationspraxis und der Evaluationsverwendung untersucht.
7. Boyle et al. (2012) <i>Irland</i> Länderebene	Untersuchung der nationalen Evaluationskultur in Irland anhand des öffentlichen Sektors und der Bildung. Der Beitrag lehnt sich an die Analyse von McNamara et al. (2009) an (siehe nächste Zeile) und untersucht die Einheitlichkeit einer nationalen oder organisationalen Evalu-	Die Autoren verweisen auf die Werte als das zentrale Element von Kultur.

	ationskultur.	
8. McNamara et al. (2009) <i>Irland</i> Länderebene	Fallstudie der Entstehung der Evaluationskultur in dem öffentlichen Sektor und besonders im Politikfeld Bildung in Irland. Berücksichtigt wird der Einfluss der EU und der OECD auf die Evaluationskultur.	Eine Evaluations- und Qualitätsgewissheitskultur. Fokussierung auf Evaluationspraxisaspekte.
9. Rodríguez-Bilella/Monterde-Díaz (2010) <i>Argentinien</i> Länderebene	Fallstudie des ländlichen Entwicklungsprogrammes in Argentinien. Untersuchung des Verhandlungsprozesses und der Konsensbildung zwischen Evaluatorinnen und Evaluatoren und den Evaluationsberichtsverantwortlichen.	Eine Atmosphäre des Lernens und ein Kapazitätsentwicklungsprozess, der die Beobachtung und das Verständnis, der Implementierung und der Verwendung von Evaluation erleichtert.
10. Forss (2012) <i>Vier Organisationen (Verwaltung, Firma, Forschungsinstitut, Non-Profit-Organisation)</i> Organisationsebene	Untersuchung von schwedischen Organisationen anhand der Ausarbeitung von vier Idealtypen in drei Schritten: 1. Untersuchung der Aktivität, 2. Untersuchung von Artefakten, 3. Untersuchung der Werte, Normen und Einstellung zur Evaluation. Forss nimmt Bezug auf Evaluationsprozesse, Evaluationsdurchführung, Bibliothek mit Evaluationsliteratur, Evaluationskapazität, Formalisierungsgrad der Evaluationen, Evaluationsnetzwerke und -gesellschaften, Status der Evaluation, Evaluationsapproach, Evaluationsbegründungen, Evaluationshaltung.	Die Evaluationskultur besteht aus Normen, Werten und Einstellungen sowie den Organisationsstrukturen und deren Prozesse.
11. Speer (2012) <i>Zwei Sektoren (Bildung und Arbeitsmarkt), Deutschland</i> Organisationsebene	Untersuchung der sektoralen Evaluationskultur durch den Vergleich der Bildung mit dem Arbeitsmarkt. Vergleich anhand von vierer von Furubo et al. (2002) übernommenen und zwei zusätzlichen Kriterien: 1. Angebot an Evaluatorinnen und Evaluatoren, 2. Institutionelle Arrangements, 3. Fokus auf den Outcome, 4. Pluralismus in jedem Politikfeld, 5. Bewusstsein des intrinsischen Wertes von Evaluationen, 6. Datensammelmechanismen.	Gemäss der Autorin sind geteilte Präferenzen für die Evaluationspraxis, für Evaluationsapproaches und für das Evaluationssystem die Bestandteile der Evaluationskultur.
12. Owen/McDonald (1999) <i>„Australian volunteers international“</i> Organisationsebene	Fallstudie der Evaluationskultur in der Nichtregierungsorganisation „Australian volunteers international“.	Evaluationskultur wird als ein Commitment für die Rolle von Evaluationen im Entscheidungsprozess einer Organisation gesehen.
13. Owen (2003) <i>„National emergency management“</i> , <i>Australien</i> Organisationsebene	Der Beitrag diskutiert die Relevanz und die Einflussfaktoren der Evaluationskultur anhand einer Fallstudie, welche die Entwicklung der internen Evaluationskultur beschreibt.	Die Evaluationskultur wird als Commitment für die Rolle der Evaluation im organisatorischen Entscheidungsprozess beschrieben.
14. Läubli Loud (2014) <i>Bundesamt für Gesundheit (BAG), Schweiz</i> Organisationsebene	Die Untersuchung diskutiert das Zusammenspiel zwischen der Evaluationskultur, der Evaluationskapazität und der Institutionalisierung der Evaluation. Des Weiteren werden diese Konzepte für das Bundesamt für Gesundheit (BAG) analysiert.	Die Autorin verweist auf geteilte Ideen, Werte und Überzeugungen auf Ebene der Organisation betreffen die Rolle, Funktion, Praxis und Verwendung der Evaluati-

Theoretisch		on.
15. Meyer/Rech (2013) <i>Deutsche Entwicklungszusammenarbeit</i> Organisationsebene	Der Beitrag untersucht die Determinanten der Evaluationskultur in der Deutschen Entwicklungszusammenarbeit. Die Untersuchung berücksichtigt die Kontexteinflüsse von nationalen, institutionellen, sektoralen und operationalen Charakteristiken auf die Evaluationskultur.	Die Evaluationskultur umfasst Traditionen und Gewohnheiten bezüglich des Umgangs und der Verwendung von Evaluationen in Organisationen, Staaten oder Netzwerken (in Anlehnung an eval-wiki*).
16. Scheerens (2004) <i>Schulen</i> Organisationsebene	In diesem Beitrag wird untersucht, ob die Selbstevaluation in Schulen zu einer Assimilierung der Schulkultur führt und demzufolge von einer Schulevaluationskultur gesprochen werden kann.	Die Evaluationskultur wird in Anlehnung als das Ebenenmodell von Schein (2010) beschrieben.
17. Barbier (2012) Theoretisch	Der Beitrag widmet sich der theoretischen Ableitung und Diskussion der politischen Kultur der Evaluation.	Die Evaluationskultur wird als Schnittmenge der Werte, der Praxis und der Institutionalisierung der Evaluation beschrieben.
18. De Peuter/Pattyn (2009) Theoretisch	Diskussion der Evaluationskapazität und der Evaluationskultur. Untersuchung von 16 Studien, die diese beiden Konzepte untersucht haben.	Die Evaluationskultur wird als geteilte Überzeugungen und Werte, welche Regeln für das Verhalten liefern und zu einer Evaluationspraxis führen definiert.

Anmerkung:*= siehe <http://eval-wiki.org/glossar/Evaluationskultur>

Quelle: Eigene Darstellung

Erwähnenswert der ebenen dargestellten Auflistung ist, dass die Beiträge von Barbier (2012), Boyle et al. (2012), Forss (2012), Horber-Papazian/Jacot-Descombes (2012), Speer (2012) und Stame (2012) alle aus dem Sammelband von Barbier/Hawkins (2012) stammen.

In der Einleitung dieses Sammelbandes nähert sich Barbier (2012) theoretisch dem Begriff der Evaluationskultur. Diese konzeptualisiert er als Schnittmenge der Dimensionen Praxis, Werte und Institutionen. Auch De Peuter/Pattyn befassen sich auf einer theoretischen Ebene mit Elementen und Indikatoren der Evaluationskultur, wobei sie die Evaluationskultur als „geteilte Werte und Überzeugungen“ definieren. Bei der Evaluationskapazität hingegen legen sie den Schwerpunkt auf operationale Aspekte: „[...] are deemed necessary for conducting an evaluation. In that sense, evaluation capacity is strongly linked to the evaluation practice itself“ (De Peuter/Pattyn 2009: 143).²¹ Auch Läubli Loud (2014) diskutiert die Konzepte Evaluationskapazität, Evaluationsinstitutionalisierung und Evaluationskultur. Dabei beziehe sich die Evaluationskultur auf gemeinsame Werte, Ideen und Überzeugungen gegenüber der Rolle, Funktion, Praxis und Verwendung der Evaluation. Die Kapazität dagegen fokussiere stärker auf die Instrumente und Strategien um eine „good practice“ einzuführen und die Evaluationsaktivität und die Evaluationsverwendung sicherzustellen. Damit sind sich die Autoren der beiden Beiträge (De Peuter/Pattyn 2009; Läubli Loud 2014) insofern einig, als dass das Konzept der Evaluationskapazität sich stärker auf die Evaluationspraxis konzentriert. Vollständigkeitshalber wird im Folgenden kurz auf das Konzept der Evaluationskapazitätsbildung (ECB) eingegangen, da dieses in der Forschung über Evaluationen ein häufig untersuchtes Konzept ist. Eine weitverbreitete Definition der ECB ist jene von Stockdill et al. (2002: 14): „ECB is the intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its uses routine“. Es gibt eine

²¹ Auch Mihalache (2010) weist auf den starken Bezug zur Evaluationspraxis beim Konzept der Evaluationskapazität hin und dass die Evaluationskultur, die Evaluationspraxis und die Evaluationskapazität austauschbare Begriffe sind, die teilweise dasselbe untersuchen.

Vielzahl an weiteren Definitionen, wie die Analyse bisheriger Beiträge der ECB von Preskill/Boyle (2008) aufzeigen konnte. Die Definitionsunterschiede sind dabei oft an die Betrachtung der Individual- oder die Organisationsebene gebunden. Preskill/Boyle (2008) finden aber auch Gemeinsamkeiten: beispielsweise verweisen alle Definitionen auf Aktivitäten, die über die bloße Evaluationsaktivität hinausgehen.

Nachdem die Beiträge, welche die Evaluationskultur thematisieren, vorgestellt wurden, widmet sich die nachfolgende Tabelle 5 der Auflistung von Beiträgen, die ähnliche Begriffe und Untersuchungsgegenstände der Evaluationskultur verwenden und für die weitere Konzeptualisierung von Bedeutung sind.

Tabelle 5: Weitere Beiträge mit Relevanz für die Konzeptualisierung der Evaluationskultur

Autor/ Begriff	Inhalt der Untersuchung	Untersuchungsgegenstand (Bezogen auf den Begriff)
1. Botcheva et al. (2002) „Learning culture and outcomes measurement practices“	Untersuchung der Lernkultur in 25 Non-Profit-Organisationen. Befragung von kommunalen Kinder- und Jugendarbeitenden in Nordkalifornien.	Eine Lernkultur beinhaltet Haltungen und Überzeugungen, welche die systematische und kontinuierliche Nutzung von Wissen und Informationen für Verbesserungen unterstützt (in Anlehnung an Preskill/Torres 1999).
2. Cousins et al. (2004) „Evaluative inquiry“	Eine Besprechung der aktuellen Literatur sowie eine Diskussion über die Verbindungen der Konzepte Evaluationsnutzung, Evaluationskapazität und organisationalem Lernen.	„Evaluative inquiry“ ist eine organisationale Unterstützungsstruktur, die organisationales Lernen potenziell fördern kann. Sie kann zudem durch andere Unterstützungsstrukturen, wie Schulung, berufliche Entwicklung und Infrastruktur gefördert werden. Zudem ist sie Teil der Organisationskultur.
3. Torres und Preskill (2001) „Evaluation and organizational learning“	Die Autoren untersuchen verschiedene Indikatoren der Evaluationsanwendung im Zusammenhang mit „organisationalem Lernen“.	Torres und Preskill verweisen auf einen kontinuierlichen Prozess des Wachstums und der Feedbackverwendung innerhalb Organisationen. Der Prozess beinhaltet Werte, Haltungen und Wahrnehmungen. Weiter beschreiben Torres und Preskill (2001) Bestandteile, die für die Entwicklung eines Evaluationsdienstes wichtig sind: 1. Rollen für die Evaluation, 2. Eine integrierte Evaluationsfunktion, 3. Entwicklung von Rahmenbedingungen für ähnliche Untersuchungsergebnisse, 4. Etablierung eines Spirits von „of ongoing inquiry“, 5. Zeit für die Reflektion, 6. Untersuchung von zugrundeliegenden Annahmen, 7. Dialog zwischen Evaluatoreninnen und Evaluatoren, Mitarbeitenden und Vorgesetzten.
4. Katz et al. (2002) „Developing an evaluation habit of mind“	Untersuchung der „evaluation habit of mind“ durch Interviews mit Schulleitenden in Kanada.	Eine prozessorientierte Denk- oder Kapazitätsart, um weise Entscheidungen zu treffen. Ein fundamentaler Wechsel im Denken darüber, welche Form von Information wichtig ist sowie die Nutzung dieser Information für die konkrete Planung und Umsetzung von Massnahmen.
5. Patton (1998) „Thinking evaluatively“	Patton beschreibt die individuellen Veränderungen, die sich bei der Durchführung von Evaluationen bei Personen ereignen. Der Evaluationsprozess führt dazu, dass man sich evaluationspezifischen Kriterien bewusst wird und „values of evaluation“ aufbaut.	Evaluatorinnen und Evaluatoren mit eigenen Werten, Denkweisen, Sprachen, Hierarchien und Belohnungssystemen. Beispiele von evaluativen Wertvorstellungen: Klarheit, Spezifität und Fokussierung, Systematik, Annahmen explizit machen, Operationalisierung von Programmkonzepten, Ideen und Ziele, Unterscheidung des Input und des Prozesses vom Outcome, Wertzuschreibung der empirischen Evidenz, Unterscheidung von Fakten, Interpretationen und Werturteilen.
6. Kaufman et al. (2006) „Evaluation Readiness“	Erarbeitung eines Instrumentes zur Selbsteinschätzung der Bereitschaft einer Organisation für den systematischen internen Einsatz von Evaluationen.	Bereitschaft der Organisation für den systematischen internen Einsatz von Evaluationen. Leitgedanke dieser Konzeption ist die Durchführung von praktikablen Evaluationen. Das pragmatisch angesetzte Messinstrument der Auto-

7. Sanders (2003, 2002) „Mainstreaming evaluation“	Theoretische Diskussion des Konzeptes „mainstreaming of evaluation“.	ren berücksichtigt vorwiegend organisationale Elemente. Mainstreaming bezieht sich auf den Prozess, welcher die Evaluation zu einem integralen Bestandteil der täglichen organisationalen Operationen macht. Evaluationen werden zur Routine in der Arbeitsethik der Organisation. Sie ist Teil der Kultur und ist Bestandteil auf allen Ebenen der Organisation.
8. Sutherland (2004) „Creating a culture of data use“	Fallstudie einer Schule in Kalifornien. Untersuchung, wie eine Schule evaluative Praktiken und die Datennutzung fördern kann.	Eine Kultur der Datennutzung.
9. Mayne (2011) „Evaluative culture“	Der Artikel diskutiert praktische Massnahmen, die eine Organisation bei Aufbau und Entwicklung einer evaluativen Kultur unternehmen kann.	Eine evaluative Kultur bezeichnet eine Organisationskultur, die empirische Informationen betreffend ihrer Leistungen sucht. Diese Informationen werden genutzt, um zu lernen, wie die Programme und Massnahmen erledigt werden können, welche die Leistungen verbessern sollen.

Anmerkung: Die Auflistung beinhaltet nicht einen umfassenden Überblick aller möglichen Konzepte, sondern führt die Beiträge auf, die im Rahmen der Recherche gefunden wurden. Die Auflistung hat somit exemplarischen Charakter um die Vielfalt der Konzeptualisierungen aufzuzeigen.

Quelle: Eigene Darstellung

Die aufgeführten Beiträge zeigen die Vielfalt der Untersuchungsgegenstände und Begriffe in der Forschung über Evaluation auf. Die Begriffe reichen von Pattons (1998) „thinking evaluatively“ bis zu „evaluation and organizational learning“ von Torres und Preskill (2001).

Die Beiträge eins bis drei analysieren die Evaluation im Zusammenhang mit organisationalem Lernen und der Lernkultur. Aufschlussreich sind zudem die Beiträge von Katz et al. (2002) und von Patton (1998), die Veränderungsprozesse bei Individuen thematisieren. Auf den Beitrag von Patton (1998) wird oft verwiesen, da dieser auf den Nutzen der Evaluation, weg von einer blossen Nutzung der Evaluationsergebnisse hin zu einer zusätzlichen Nutzung aufgrund des Evaluationsprozesses (Prozessnutzen), hinweist. Der Prozessnutzen berücksichtigt, was mit Personen geschieht, die am Evaluationsprozess beteiligt sind. Gemäss Patton (1998) findet durch den Evaluationsprozess ein Lernen statt, beispielsweise durch die Bewusstwerdung von evaluationsspezifischen Kriterien oder den Aufbau von sogenannten evaluativen Wertvorstellungen (diese sind in obiger Tabelle ersichtlich). Die Beiträge sechs bis neun stellen mit ihren Begriffen praktische Aspekte in den Vordergrund: Sutherland (2004) beschreibt anhand der Analyse einer Schule, wie evaluative Praktiken und eine Kultur der Datennutzung kreiert werden können. Sanders (2003, 2002) diskutiert das Konzept „mainstreaming evaluation“, bei dem es darum geht, die Evaluation in den Organisationsalltag fest zu integrieren. Auch Mayne (2011) diskutiert praktische Massnahmen zur Einführung einer evaluativen Kultur auf Basis von empirischen Informationen. Die Konzeption von Kaufman et al. (2006) – „evaluation readiness“ – orientiert sich am Leitgedanken der Durchführung von praktikablen Evaluationen in Organisationen und berücksichtigt vorwiegend organisationale Elemente.

Auf die Breite und Vielfalt der Bestandteile im Zusammenhang von Kultur wurde bereits in der Einleitung dieses Kapitels hingewiesen: dies trifft auch auf die Evaluationenkultur zu, wie dieser Abschnitt aufzeigen konnte. Der nächste Abschnitt widmet sich der Strukturierung dieser unterschiedlichen Bestandteile.

Synthese und Definition

Welche Folgerungen lassen sich aus den vorgestellten Forschungsbereichen für die Herleitung der organisationalen Evaluationskultur ziehen? Von der politischen Kultur-literatur wissen wir, dass der Begriff der politischen Kultur hauptsächlich die Werte und Einstellungen gegenüber der Politik und dem politischen System berücksichtigt und dass Kultur zwischen Mikro- und Makroebene anzusiedeln ist. Die politische Kultur wird aus individuellen Einstellungen aggregiert und auf einer Makroebene betrachtet. Auch aus den Beiträgen der Organisationskulturforschung können wichtige Erkenntnisse für die Konzeptualisierung abgeleitet werden. Dazu zählt die Debatte bezüglich der unterschiedlichen Herangehensweisen an die Kultur mittels eines eher qualitativen oder eines eher quantitativen Ansatzes. Eng verbunden damit ist die oben skizzierte Unterscheidung zwischen einer kulturalistischen und universalistischen Perspektive der Untersuchung. Dementsprechend ändert sich auch die Betrachtungstiefe der Kultur und wird der Definition der Organisationskultur angepasst: Einer breiten, viele Elemente beachtenden Definition, gegenüber einer Definition, die sich auf wenige Bestandteile beschränkt. Schein (2010) verweist auf die Problematik, die einer Bestimmung von Kategorien und Dimensionen der Organisationskultur im Voraus zugrunde liegt, da unbekannt ist, ob diese für die zu untersuchende Organisation relevant sind. Dieses Problem verkleinert sich bei der Ermittlung der Evaluationskultur, da der Begriff bereits spezifisch auf Evaluationen ausgerichtet ist und die Bandbreite der möglichen zu berücksichtigenden Aspekte kleiner wird. Die Evaluationskultur konzentriert sich auf jenen Teil der Organisationskultur, der auf Evaluationen gerichtet ist. Man kann von einer Teilmenge der Organisationskultur sprechen, die jene Einstellungen und Werte berücksichtigt, die sich auf die Evaluation beziehen. Zudem zeigt das Modell von Schein (2010: 23ff.) unterschiedliche Kulturebenen. Zwar sind die Einstellungen und Überzeugungen Kern der Kultur, es gibt allerdings auch weitere Bestandteile, die zu berücksichtigen sind. Tabelle 4 und Tabelle 5 stellen diese Vielfalt im Zusammenhang mit der Evaluationskultur dar. Bei der Strukturierung dieser Beiträge und Untersuchungsgegenstände hilft der Beitrag von Barbier (2012), der von der politi-

schen Kultur ausgehend die Evaluationskultur ableitet und drei verschiedenen Elemente – Praxis, Werte und Institutionalisierung – unterscheidet. In Anlehnung an diese Unterteilung wird für diese Untersuchung die Evaluationskultur in folgende drei Kategorien aufgeteilt:

1. Kategorie von Bestandteilen der Evaluationskultur, welche die organisationspezifischen Strukturen und Prozesse thematisieren.
2. Kategorie von Bestandteilen der Evaluationskultur, die sich auf die Evaluationspraxis (beispielsweise Evaluationsaktivität, Professionalisierung, Evaluationskapazität) beziehen.
3. Kategorie von Bestandteilen der Evaluationskultur, welche die Werte und Einstellungen zur Evaluation berücksichtigen.

Die erste Kategorie konzentriert sich auf die organisationale Ebene, analysiert die Institutionalisierung der Evaluation und fokussiert damit auf die Analyse der Strukturen und Prozesse in der Organisation. Die zweite Kategorie beinhaltet die Evaluationspraxis und analysiert, wie stark die Evaluation in den Arbeitsalltag der Mitarbeitenden Einzug gehalten hat. Die dritte Kategorie kann als Kern der Evaluationskultur angesehen werden. Diese berücksichtigt, was Personen über die Evaluation denken und ob sich evaluative Wertvorstellungen herausgebildet haben.

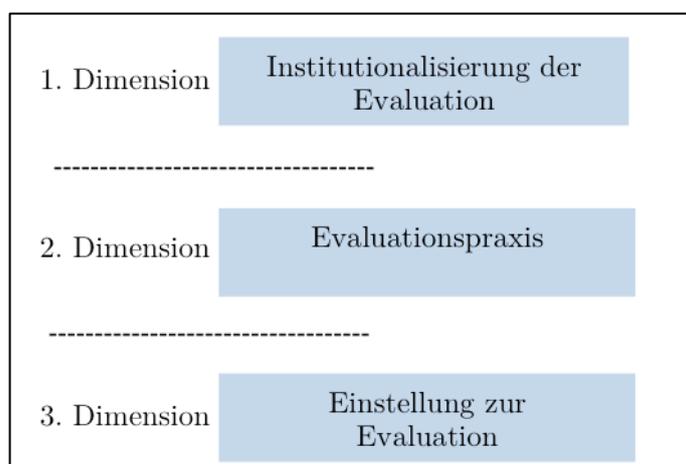
Zugunsten dieser Konzeptualisierung wird argumentiert, dass durch die Berücksichtigung der Einstellung, der Praxis und der Strukturen eine ganzheitliche Betrachtung der Evaluationskultur gewährleistet wird. Zudem ermöglicht diese Strukturierung eine Erhebung der organisationalen Evaluationskultur auf Individual- und auf Organisationsebene. Auf Individualebene können die Mitglieder einer Verwaltung auf deren Einstellungen und Werte gegenüber Evaluationen befragt werden. Die Strukturen und die Prozesse können auf Organisationsebene ermittelt werden. Bei der Ermittlung der Pra-

xis sind beide Ebenen denkbar.²² Abgeleitet aus diesen drei Kategorien wird die organisationale Evaluationskultur folgendermassen definiert:

Die organisationale Evaluationskultur umfasst die gemeinsamen Einstellungen und Werte der Mitarbeitenden zur Evaluation und findet Ausdruck im Verhalten (in der Evaluationspraxis) und in der Struktur (in der Institutionalisierung der Evaluation) der Organisation.

Diese Definition bringt den Charakter der Kultur als multidimensionales Konzept zum Ausdruck. In Abbildung 1 wird von drei Dimensionen gesprochen, die unterschiedliche, jedoch zusammengehörende Dimensionen der organisationalen Evaluationskultur umfassen.

Abbildung 1: Drei Dimensionen der organisationalen Evaluationskultur



Quelle: Eigene Darstellung

Diese Untersuchung verfolgt mit der Beschränkung auf wenige Dimensionen der Evaluationskultur eine quantitative Herangehensweise. Ein wichtiger Grund hierfür ist der

²² Damit verbunden ist eine Messung der Evaluationskultur durch verschiedene Zugänge. Beispielsweise nutzt Taut (2007) eine gemischte Strategie, um die Verwendung der Evaluation in der Organisation zu untersuchen. Neben der Befragung der Mitarbeitenden führt er Interviews und bezieht Unterlagen in seine Analyse ein. Ebenso übernehmen Taylor-Ritzler et al. (2013) in der Erarbeitung des „evaluation capacity assessment instrument (ECAD)“ die Kombination von individuellen und organisationalen Charakteristiken, um die Evaluationskapazität von Non-Profit-Organisationen zu messen.

bereits sehr spezifische Untersuchungsbereich der Evaluationskultur. Die Untersuchung beschränkt sich auf den Bereich der Organisationskultur, der auf die Evaluation gerichtet ist, was die Möglichkeit der zu untersuchenden Gegenstände bereits eingrenzt.²³ Gleichwohl berücksichtigt die vorliegende Konzeptualisierung eine Breite, die, trotz der Fokussierung auf wenige Aspekte, eine vertiefte Analyse der Evaluationskultur ermöglicht.

²³ Bei der Untersuchung der Organisationskultur, die sich nicht spezifisch auf die Evaluation bezieht würde sich ein qualitativer Zugang anbieten. Damit könnte zuerst untersucht werden, welche Aspekte für die schweizerischen Verwaltungseinheiten besonders relevant wären.

3 Theoretische Modelle

Nachdem im vorherigen Kapitel die organisationale Evaluationskultur in drei konstitutive Kulturdimensionen – Institutionalisierung der Evaluation, Evaluationspraxis und Einstellung zur Evaluation – strukturiert wurde, widmen sich die folgenden Abschnitte der Präsentation von zwei theoretischen Modellen. Diese bilden die Grundlage für die Untersuchung der Evaluationskultur. Das Gesamtmodell dient der Übersicht und beinhaltet diese beiden Modelle.

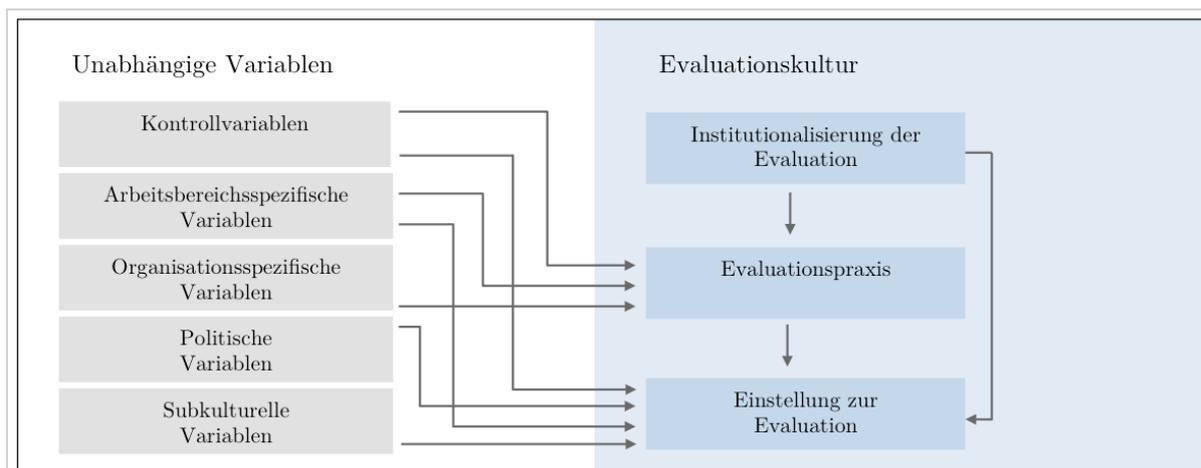
Im ersten Modell werden verschiedene Einflüsse auf die Evaluationspraxis untersucht, und es werden damit einhergehend die oft erwähnten Unterschiede der Evaluationspraxis zwischen den Politikfeldern analysiert, auf die in der Forschungsliteratur oft verwiesen wird (Balthasar 2007; Balthasar/Rieder 2009; Furubo/Sandahl 2002; Hense et al. 2013; Jacob et al. 2015; Pattyn 2014, 2015; Widmer 2008; Widmer/Beywl 2009). Die meisten bisher durchgeführten Studien basieren allerdings auf theoretischen Argumenten. Dieser Forschungslücke wird im ersten Modell begegnet, indem empirisch untersucht wird, welche Faktoren die Evaluationspraxis beeinflussen. Im zweiten Modell werden unterschiedliche Einflussfaktoren auf die Einstellung zur Evaluation analysiert. Wie im vorherigen Kapitel erwähnt, spielt die Einstellung bei der Untersuchung von Kulturen eine entscheidende Rolle.²⁴ Bei einer genaueren Betrachtung der Kultur muss diese Dimension zwingend berücksichtigt werden. Aus diesem Grund wird die Einstellung zur Evaluation in einem separaten Modell analysiert.

In der nachfolgenden Abbildung 2 wird das theoretische Gesamtmodell aufgezeigt. Die organisationale Evaluationskultur, wie sie auf der rechten Seite dargestellt ist, geht aus der Herleitung des vorherigen Kapitels hervor und unterscheidet drei konstitutiven Kulturdimensionen: Institutionalisierung der Evaluation, Evaluationspraxis und Einstellung zur Evaluation. Die zu untersuchenden Hypothesen zwischen den Dimensio-

²⁴ Für eine gute Einführung in das Konzept der politischen Kultur siehe Greiffenhagen (2009), Pickel/Pickel (2006: 49ff.), sowie Westle (2009). Für einen Überblick der Organisationskultur siehe Ashkanasy et al. (2011), Martin (2002), Lee/Lawrence (2013: 104ff.), sowie Alvesson (2013).

nen sind mithilfe von Pfeilen dargestellt. Auf der linken Seite der Abbildung 2 sind die unabhängigen Variablen und deren Einflüsse auf zwei der angeführten Kulturdimensionen aufgeführt.

Abbildung 2: Theoretisches Gesamtmodell der Untersuchung



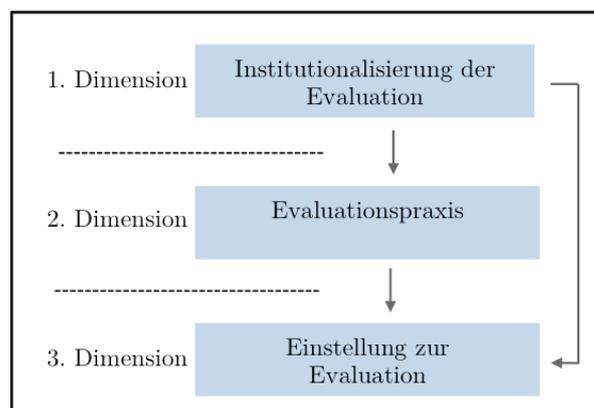
Anmerkungen: Die Pfeile geben die zu untersuchenden Einflussbeziehungen wieder.

Quelle: Eigene Darstellung

Betrachten wir zunächst die organisationale Evaluationskultur in Abbildung 3: In Anlehnung an Schein (2010: 23ff.) wird diese Untersuchung ausschliesslich die Einflussbeziehungen in Pfeilrichtung berücksichtigen. Schein (2010), der als Begründer der Organisationskultur gilt, schildert einen möglichen Vorgang innerhalb von Gruppen wie folgt: „If the manager convinces the group to act on her belief, the solution works, and the group has a shared perception of that success, then the perceived value ‘advertising is good’ gradually becomes transformed: first into a shared value or belief and ultimately into a shared assumption” (Schein 2010: 26). Für die organisationale Evaluationskultur wird daraus folgende Transformation abgeleitet (siehe Abbildung 3): Die Institutionalisierung der Evaluation (Dimension 1) kann als Signal für die Unterstützung der Evaluation seitens der Organisation verstanden werden. Sowohl mit der Institutionalisierung, als auch mit der Einführung einer Evaluationspraxis (Dimension 2) signalisiert die Organisation gegenüber den Mitarbeitenden, was die Evaluation angeht, Einverständnis und Unterstützung, was sich wiederum positiv auf die Einstellung zur Evaluation (Dimension 3) seitens der Mitarbeitenden auswirken kann. Die

Analyse dieser Richtung basiert damit auf dem Grundgedanken der Organisationskultur, der annimmt, dass die Organisation einen strukturierenden Einfluss auf die Mitarbeitenden nimmt.

Abbildung 3: Beziehungen der drei Dimensionen der organisationalen Evaluationskultur



Quelle: Eigene Darstellung

Es ist darauf hinzuweisen, dass durchaus Einflüsse in die umgekehrte Richtung möglich sind, insbesondere zwischen den drei Kulturdimensionen (Abbildung 3).²⁵

Varone (2007, 2015) unterscheidet zwei Pfade zur Evaluationskapazitätsbildung: Einerseits den „bottom-up“-Entwicklungsprozess, bei dem sich durch positive Erfahrungen eine Evaluationspraxis entwickelt, die in einem nächsten Schritt zu einer Institutionalisierung der Evaluation führen kann. Andererseits die „top-down“-Einflussrichtung, bei der die Institutionalisierung am Anfang steht und die Mitarbeitenden beeinflusst. Für diese Untersuchung wird die reziproke Wirkung der Organisationskultur nicht infrage gestellt.²⁶ Aufgrund der in den nächsten beiden Unterkapiteln folgenden theoretischen und empirischen Argumenten und in Anlehnung an den

²⁵ Ziel dieser Untersuchung ist es nicht, die Kausalität abschliessend zu klären; etwa die Frage, inwiefern die Evaluationspraxis die Institutionalisierung beeinflusst oder ob nicht vielmehr die Institutionalisierung die Evaluationspraxis beeinflusst, wird mittels dem vorliegenden Untersuchungsdesign nicht abschliessend beantwortet. Mittels der Regressionsanalyse kann die Kausalität nicht bewiesen werden (siehe dazu die Diskussion der Kausalität von Opp 2010). Trotzdem wird in dieser Untersuchung von „Einfluss“ gesprochen und die verwendeten Variablen, wie bei der Regressionsanalyse gebräuchlich, in abhängige und unabhängige Variablen unterschieden.

²⁶ Beispielsweise halten Dolder et al. (2017) Effekte, je nach Politikbereich (Bildung; Gesundheit) und Betrachtungsebene (Bund; Kantone), in beide Richtungen für möglich.

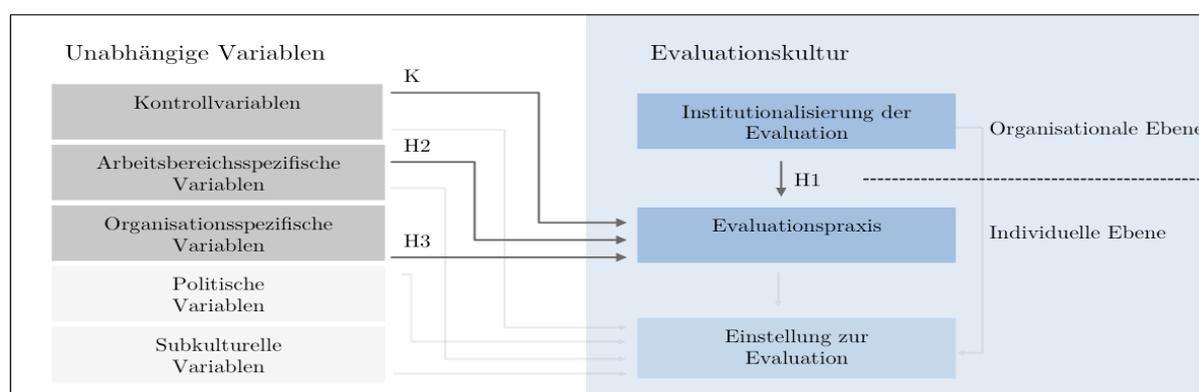
Grundgedanken der Organisationskultur, wird lediglich die dargestellte Einflussrichtung untersucht.

In den folgenden zwei Unterkapiteln werden die beiden theoretischen Modelle vorgestellt und die zu untersuchenden Hypothesen formuliert.

3.1 Theoretisches Modell für die Untersuchung der Evaluationspraxis

Die nachstehende Abbildung 4 zeigt das Modell zur Analyse der Evaluationspraxis (Dimension 2). Die erste Hypothese thematisiert den Einfluss der Institutionalisierung der Evaluation auf die Einstellung. Die Hypothesen zwei und drei berücksichtigen arbeitsbereichs- und organisationspezifische Variablen. Zusätzlich werden Kontrollvariablen in die Analyse einbezogen.

Abbildung 4: Modell für die Untersuchung der Evaluationspraxis



Anmerkungen: Die Pfeile geben die zu untersuchenden Einflussbeziehungen wieder.

Quelle: Eigene Darstellung

Im nächsten Unterkapitel wird ausführlich erläutert werden, dass die Evaluationspraxis mit den beiden Indikatoren Evaluationsaktivität und Evaluationsverwendung gemessen wird. Beide werden hinsichtlich der gleichen Hypothesen untersucht. Aufgrund der unterschiedlichen Verteilung und des unterschiedlichen Skalenniveaus der beiden abhängigen Variablen ergeben sich verschiedene Regressionsmodelle: ein negativ binomiales Regressionsmodell für die Evaluationsaktivität und ein binär logistisches Regressionsmodell für die Evaluationsverwendung. Dementsprechend ändert sich die Formulierung der Hypothesen. Während bei der linearen Einfachregression "Je-desto-

Hypothesen” formuliert werden, wird bei der logistischen Regression von der Eintrittswahrscheinlichkeit für das zu untersuchende Ereignis gesprochen (siehe dazu Backhaus et al. 2011: 257). Der Einfachheit halber werden die Annahmen als generell gültige formuliert und erst in den konkreten Modellen wird auf die spezifische Bedeutung, sowie Interpretation der Zusammenhänge eingegangen.

Institutionalisierung der Evaluation H1

Die erste Hypothese thematisiert den Einfluss der Institutionalisierung auf die individuelle Evaluationspraxis.

Damit wird der bereits geäußerte „top-down“-Einfluss von Dimension 1 auf Dimension 2 innerhalb der organisationalen Evaluationskultur empirisch überprüft.

Die Untersuchung dieser Hypothese nimmt den Grundgedanken der Organisationskultur auf; dieser schreibt der Organisation einen strukturierenden Einfluss auf die Mitarbeitenden zu. In Anlehnung an Schein (2010) wird überprüft, ob sich die Institutionalisierung der Evaluation (Dimension 1), die als Signal der Unterstützung der Evaluation seitens der Organisation verstanden werden kann, auch in einer erhöhten Evaluationspraxis (Dimension 2) der Mitarbeitenden finden lässt.²⁷ Des Weiteren ist anzunehmen, dass durch die bewusste Berücksichtigung der Evaluation in der Verwaltungseinheit mittels der Institutionalisierung und damit einhergehenden Spezialisierung, sowie Professionalisierung²⁸ der Mitarbeitenden bezüglich der Evaluation eine generell höhere Evaluationspraxis bei den Mitarbeitern zu beobachten ist, als bei nicht institutioneller Einbettung.

Ausserdem finden sich einige Beiträge aus der Forschung über Evaluation, die auf einen Zusammenhang zwischen der Institutionalisierung der Evaluation und der Ver-

²⁷ Auf die Untersuchung des Einflusses der Institutionalisierung der Evaluation auf die organisationale Evaluationspraxis (Evaluationspraxis aggregiert auf Organisationsebene) wird verzichtet. Grund dafür ist die erwähnte Fokussierung auf die Theorie der Organisationskultur und die Untersuchung des Einflusses der Organisation auf ihre Mitarbeitenden. Die bivariate Analyse zeigt, dass die Institutionalisierung der Evaluation mit der Höhe der Evaluationsaktivität auf Amts- und Bereichsebene signifikant positiv zusammenhängt (0.29 Gamma respektive 0.18 Kendalls-Tau-b).

²⁸ Für weitere Beiträge bezüglich der Professionalisierung der Evaluation siehe Jacob/Boisvert (2010), King et al. (2001), sowie Taylor-Ritzler et al. (2013).

breitung der Evaluation hindeuten. Dazu zählt der Beitrag von Horber-Papazian (2006), der auf die Wichtigkeit der Institutionalisierung der Evaluation für die Evaluationsaktivität in der Schweiz auf Bundesebene verweist. Ebenso erwähnen Varone et al. (2005: 254ff.) die Bedeutung der Institutionalisierung der Evaluation auf die Evaluationspraxis und postulieren, dass ein hoher Grad der Institutionalisierung bereits in der Phase der Policy-Formulierung eine Ausführung der Evaluationen garantiert. Zudem verweisen die Autoren auf einen Bericht der OECD (1997), der aufzeigt, dass ein gewisses Niveau der Institutionalisierung erreicht werden muss, damit die Evaluation ihre Rolle bei der „public governance“ einnehmen kann. Auch Leeuw/Rozendal (1994: 70) betonen die Wichtigkeit der Institutionalisierung für eine Evaluationspraxis, die nicht einfach nur sporadisch vollzogen wird: „Institutionalization is an instrument for promoting more efficient implementation of evaluation activities“. Zudem schlussfolgert Derlien (1990) bei der Analyse von acht Ländern, darunter auch der Schweiz, dass die Institutionalisierung einen wichtigen Beitrag zur Implementierung der Evaluation leistet.

Aufgrund der aufgezeigten theoretischen Überlegungen, sowie den empirischen Indizien wird davon ausgegangen, dass die Institutionalisierung auch in Organisationen einen positiven Einfluss auf die Evaluationspraxis ausübt. Dies wird mit folgender Hypothese überprüft:

H1: Wenn die Evaluation in der Verwaltungseinheit institutionalisiert ist, dann ist die Evaluationspraxis der Verwaltungsmitarbeitenden höher.

Im nächsten Kapitel wird für die Berücksichtigung von zwei Variablen bei der Messung der Institutionalisierung der Evaluation argumentiert: das Vorhandensein einer Evaluationsstelle, sowie die Existenz das Vorhandensein von Grundlagendokumenten und Leitfäden betreffend Evaluation. Allerdings zeigt die Beschreibung der Departemente und Direktionen im fünften Kapitel, dass die Verbreitung der beiden Variablen identisch ist. Denn Verwaltungseinheiten mit einer Evaluationsstelle verfügen auch über Dokumente betreffend Evaluation. Aus diesem Grund wird lediglich das Vorhan-

densein einer solchen Stelle in der Analyse berücksichtigt und als weiteres die unterschiedliche Ausprägung der Institutionalisierung der Evaluation für die Operationalisierung übernommen: Unterschieden wird in Verwaltungseinheiten mit keiner und mit einer ansatzweisen Institutionalisierung der Evaluation und Verwaltungseinheiten, in der die Evaluation institutionalisiert ist.²⁹

Arbeitsbereichsspezifische Variablen H2

In der Literatur existieren einige Beiträge, die auf unterschiedliche Evaluationspraxen in verschiedenen Politikfeldern hinweisen (Balthasar 2007; Balthasar/Rieder 2009; Furubo/Sandahl 2002; Hense et al. 2013; Jacob et al. 2015; Pattyn 2014, 2015; Widmer 2008; Widmer/Beywl 2009).

Die exakte Einteilung von ganzen Kantonsverwaltungseinheiten in unterschiedliche Politikfelder ist allerdings schwer zu vollziehen und für die Evaluationsaktivität häufig zu unpräzise, da sich auch innerhalb der Felder erhebliche Unterschiede in der Aktivität ausmachen lassen. Rhyn (2009) beispielsweise stellt bei der Beschreibung der Evaluationsaktivität des Bildungsbereichs der Schweiz erheblichen Unterschiede zwischen den verschiedenen Bildungsstufen (obligatorische Schule, Sekundarstufe 2 und Tertiärstufe) fest. Ebenso können die Tätigkeiten der Verwaltungsangestellten innerhalb von Politikfeldern stark voneinander abweichen. Uchtenhagen (2009) gruppiert in ihrer Übersicht der Evaluation des Gesundheitsbereichs in der Schweiz unterschiedliche Tätigkeitsfelder, in der Evaluationsberichte zu finden sind. Diese reichen von der Prävention über die Gesundheitsversorgung und -förderung, bis hin zu Evaluationen, die Behandlungsmethoden oder gesundheitsökonomische Aspekte untersuchen. Auch die hier untersuchten kantonalen Verwaltungseinheiten weisen innerhalb der Politikfelder sehr unterschiedliche Tätigkeitsfelder auf. Aus diesem Grund werden arbeitsbereichs-

²⁹ Diese Einteilung entspricht der Auswertung der Beschreibung der Departemente und Direktionen im fünften Kapitel. Bei dieser Beschreibung wird festgestellt, dass einige Verwaltungseinheiten eine Evaluationsstelle haben. Zudem gibt es Verwaltungseinheiten mit Ansätzen dieser und solche, die über keine verfügen.

spezifische Variablen in die Analyse miteinbezogen. Der Grundgedanke dahinter ist, dass arbeitsbereichsspezifische Variablen, welche die spezifischen Eigenschaften des Arbeitsbereiches der Mitarbeitenden berücksichtigen, Einfluss auf die Evaluationspraxis nehmen. In der vorliegenden Untersuchung werden vier solcher Variablen die Eigenschaften des Arbeitsbereiches charakterisieren. Diese berücksichtigen die Volatilität, Komplexität, sowie den Politikzyklus und die Politikmassnahmen.

Die erste Variable ist die Volatilität. Diese beschreibt in dieser Untersuchung, wie stark die Verwaltungstätigkeit von Wechseln und Änderungen betroffen ist und berücksichtigt nicht nur die Reformhäufigkeit.

Einen Verweis auf den Einfluss der Volatilität geben Widmer/Beywl (2009: 523) in der Analyse unterschiedlicher Evaluationsaktivität der untersuchten Politikfelder für die Schweiz, Österreich und Deutschland. Für die Erklärung der Unterschiede nennen sie verschiedene Faktoren: „Legitimationsdruck, Wertekonflikte und Kostendruck reichen jedoch nicht in jedem Fall aus, um Evaluationstätigkeiten auszulösen (Rentenversicherung)“. Weitere wichtige Faktoren sind nach Ansicht der beiden Autoren die Akteurskonstellation, sowie die Volatilität des Themenfeldes. Als Beispiel fügen sie die Energie- und Umweltpolitik als Themenfeld mit volatiler Themenkonjunktur an, in dem aufgrund des Ausmasses der Volatilität ein gewisses Niveau an Evaluationsaktivität festgestellt werden kann. Auch in der Literatur, die sich mit dem Einbezug forschungsbasierter Wirkungsinformationen³⁰ auseinandersetzt, wird der Einfluss des Politikwandels häufig erwähnt (siehe dazu Frey 2012: 67). Allerdings umfasst dieser Begriff nicht nur die Form der Evaluation, sondern berücksichtigt ein breiteres Verständnis von Informationen.

Ein weiteres Argument für die Relevanz der Volatilität kann von der durch Guba/Lincoln (1987) hervorgebrachten Unterscheidung der Evaluationstheorien in vier Generationen abgeleitet werden. Ausgehend von dieser Differenzierung, in der die

³⁰ Der Begriff „forschungsbasierte Wirkungsinformationen“ bezieht sich auf forschungsbasierte Informationen zur Wirksamkeit staatlicher Aktivitäten. Diese beinhalten systematische und transparente Messungen der Wirksamkeit staatlicher Aktivitäten (siehe Frey 2012: 21f.).

Evaluation der letzten Generation als Lernprozess angesehen wird, argumentieren Kissling-Näf/Knoepfel (1994: 111ff.), dass die Evaluation als Mittel für ein institutionalisiertes policy-orientiertes Lernen³¹ eingesetzt werden kann. In gleicher Weise erwähnt Van der Knaap (1995) diesen Paradigmenwechsel der Evaluation als eine Form von institutionalisiertem, policy-orientiertem Lernen³². Daraus wird abgeleitet, dass ein volatiler Arbeitsbereich angesichts der Veränderungen stärker policy-orientiertes Lernen, sprich Evaluationen, auslöst.

Aufgrund der geschilderten Beiträge wird der Einfluss der Volatilität auf die Evaluationspraxis mit folgender Hypothese untersucht:

H2a: Je volatiler der Arbeitsbereich der Mitarbeitenden ist, desto höher ist die Evaluationspraxis, über die sie verfügen.

Die Volatilität wird anhand der Selbsteinstufung der Mitarbeitenden über die Veränderungen ihres Tätigkeitsbereiches operationalisiert.

Eine weitere arbeitsbereichsspezifische Variable, die einen Erklärungsbeitrag für die Evaluationspraxis leisten kann, ist die Komplexität. Ein komplexer Arbeitsbereich zeichnet sich für diese Untersuchung dadurch aus, dass die Massnahmen und/oder die Zusammenhänge unklar sind, also etwa viel Technik- und oder Fachwissen voraussetzt. Ein weiteres Merkmal der Komplexität ist der Einbezug von empirischen Daten.

Zwei Argumente sprechen für einen Einfluss der Komplexität: Erstens generiert die Komplexität, um die ungewissen Zusammenhänge und Situationen zu verstehen, eine höhere Nachfrage nach Wissen in Form von Evaluationen. Zweitens bietet ein komplexer Arbeitsbereich mehr Handlungsspielräume für die involvierten Koalitionen.

³¹ Policy-orientiertes Lernen wird von Kissling-Näf/Knoepfel (1994: 110) mit Lernprozessen, die im Rahmen der Programmierung und Implementation von öffentlichen Politiken ablaufen, beschrieben.

³² Unter policy-orientiertem Lernen versteht Van der Knaap (1995: 190) einen Prozess, in dem Policy-Akteure nach Verbesserungen der öffentlichen Politiken streben.

Dem ersten Argument liegt die Überlegung zugrunde, dass einige Themenbereiche aufgrund inhärenter Komplexität mehr Evaluationen benötigen oder auslösen als andere. Es gibt beispielsweise Bereiche, in denen die Wirkungszusammenhänge, die mittels Evaluationen untersucht werden können, unklarer sind. In diesem Zusammenhang kann auf die bereits erwähnte Untersuchung der Verwendung von forschungsbasierten Wirkungsinformationen von Frey (2012) verwiesen werden. Als Grundlage für die Bedeutung der Komplexität dienen Frey (2012: 70f.) die Thesen von Radaelli (1999) und von Boswell (2009), wonach die Unsicherheit und die Komplexität von Problemen und deren Lösungen den Bedarf und die Verwendung von Forschungskennnissen moderieren. Demgemäss ist bei hochkomplexen und unsicheren Problemen der Einfluss von Forschungserkenntnissen höher, da die Entscheidungstragenden nicht über vorgefasste Präferenzen verfügen (siehe auch Barnes et al. 2003; Behrens et al. 2005). Des Weiteren kann als Begründung für den Einfluss der Komplexität die im oberen Abschnitt beschriebene Funktion der Evaluation als eine Form von institutionalisiertem, policy-orientiertem Lernen (Kissling-Näf/Knoepfel 1994; Van der Knaap 1995) herangezogen werden. Wie auch für volatile Arbeitsbereiche kann argumentiert werden, dass komplexe Arbeitsbereiche verstärkt Lernprozesse generieren, die unter anderem über Evaluationen erwirkt werden können.

Ein zweites Argument ist, dass Evaluationen von Koalitionen benutzt werden, um ihre Position zu stärken: „Several of the cases support the ACF’s argument that technical information and formal policy analysis are generally used in an advocacy fashion, that is, to buttress and support a predetermined position” (Sabatier/Jenkins-Smith 1993: 218). Dieses Argument bezieht sich auf den von Sabatier (1987) entwickelten Analyserahmen „ACF“ (Advocacy Coalition Framework, deutsch: Advocacy-Koalitionsansatz). Gemäss dem ACF werden politische Entscheidungen in politischen Subsystemen getroffen – darunter kann ein Netzwerk von Akteuren und Vermittlern verstanden werden, die sich mit einem spezifischen Problem befassen (siehe Bandelow 2015; Weible et al. 2011). Diesem Ansatz entsprechend werden in technischen und komplexen Bereichen vermehrt Evaluationen nachgefragt, da die beteiligten Akteure

die Evaluationsresultate verwenden, um ihre Position zu stärken. Auch Radaelli (1999) argumentiert, dass die Komplexität das politische und strategische Verhalten der Akteure beeinflusst und einen Faktor in der Verwendung von Forschungserkenntnissen darstellt. Demgemäss begünstigt ein Bereich mit geringer Komplexität den Wettbewerb um Macht und das Lernen wird weniger bedeutend.

Aufgrund der geschilderten theoretischen Überlegung wird der Einfluss der Komplexität auf die Evaluationspraxis mit folgender Hypothese untersucht:

H2b: Je komplexer der Arbeitsbereich der Mitarbeitenden ist, desto höher ist die Evaluationspraxis, über die sie verfügen.

Die Komplexität wird mittels der Selbsteinstufung der Mitarbeitenden über die Komplexität ihres Tätigkeitsbereiches operationalisiert.

Ein weiteres Unterscheidungsmerkmal der Arbeitstätigkeit der Verwaltungsmitarbeitenden ist die Einteilung in unterschiedliche Phasen des Politikzyklus. Im Folgenden wird argumentiert, dass für Mitarbeitende, die in der Phase der Formulierung und der Phase der Implementierung tätig sind, empirische Hinweise auf eine erhöhte Evaluationsaktivität gefunden werden können, im Vergleich zu Mitarbeitenden, die interne Dienstleistungen übernehmen.

Einen Hinweis für die Relevanz der Phasen des Politikzyklus für die Evaluationspraxis gibt der Beitrag von Frey/Ledermann (2017). Die beiden Autorinnen weisen auf eine unterschiedliche Nutzung der Evaluation zwischen der Politikformulierung und dem -vollzug hin und vermuten, dass Regierung und Verwaltung häufiger beim Vollzug als bei der Formulierung Evidenzen für die Entscheidung nutzen. Allerdings räumen sie ein, diese Aussage seien etwas gewagt, da bisher Studien, welche die Nutzung in den beiden Phasen direkt gegenüberstellen, fehlen. Auch die Datenbankauswertung³³ der erfassten Evaluationsberichte der Bundesebene und der drei Kantone Bern, Genf und

³³ Eine ausführlichere Beschreibung der Datenbank findet sich im Unterkapitel 1.4.2.

Zürich von Frey et al. (2017) lässt erhebliche Unterschiede in der Evaluationspraxis zwischen den Phasen des Politikzyklus vermuten. So zeigt das Ergebnis der zeitlichen Perspektive der Evaluationsberichte, dass begleitende, sowie Ex-post-Bewertungen deutlich häufiger durchgeführt werden. Rund 93 Prozent aller erfassten Studien haben begleitend oder mit einer Ex-post-Perspektive untersucht, während prospektive Untersuchungen deutlich seltener vollzogen wurden. Diese Zahlen lassen vermuten, dass die Evaluationspraxis bei dem Vollzug relevanter ist als in der Formulierungsphase. Ferner lässt sich auch im parlamentarischen Kontext ein weiteres Indiz für die Wichtigkeit der Evaluation in der Phase des Vollzugs ausmachen. Der Beitrag von Eberli/Bundi (2017) untersucht die Evaluationsnachfrage und die Evaluationsnutzung in Schweizer Parlamenten. Dabei verweisen die beiden Autoren auf die wichtige Rolle der Evaluation für die Parlamente, um die Aufsichtsfunktion über den Gesetzesvollzug der Regierung wahrzunehmen. Ihre empirische Untersuchung weist darauf hin, dass diese Aufsichtsfunktion Evaluationen generiert.

Im Gegensatz dazu gibt es ebenfalls Beiträge, die auf die Bedeutung der Evaluation in der Politikformulierung hinweisen. Frey/Ledermann (2017: 218ff.) thematisieren die Nutzung der Evaluation in Regierung und Verwaltung. Dabei verweisen sie auf Studien, die aufzeigen, dass bei der Bundes- und der Kantonalverwaltung oft systematische Evidenz erstellt und für den Gesetzgebungsprozess berücksichtigt wird. Bei diesen Studien (Balthasar/Müller 2014; Blatter et al. 2015; Klöti/Schneider 1989; Ledermann 2014) wird allerdings ein breiter Evidenzbegriff verwendet, der Evaluationen einschliesst.

Um die unterschiedlichen Phasen des Politikzyklus zu berücksichtigen, wird die Unterscheidung von Anderfuhren-Biget et al. (2013) beigezogen. Diese reduzieren in ihrer Untersuchung zur Motivation von Verwaltungsarbeitenden (Public Service Motivation; PSM) den Politikzyklus von ursprünglich fünf Phasen auf zwei: Formulierung

und Implementierung.³⁴ Dabei bezieht sich die Formulierung von öffentlichen Politiken auf das Agenda-Setting und die Formulierung von öffentlichen Politiken; wie zum Beispiel Verwaltungsmitarbeitende, die administrative Dienstleistungen übernehmen oder Mitarbeitende auf hierarchischen Stufen, wo politische Entscheidungen getroffen oder politische Dossiers vorbereitet werden. Die Implementierung³⁵ bezieht sich auf Massnahmen, die auf die Dienstleistungserbringung für Nutzer zielen. Zu den Beispielen zählen Subventionen, der Vollzug von Normen oder die Autorisierung von Bewilligungen. Des Weiteren unterteilen die Autoren die Verwaltungsmitarbeitenden, die für die internen Dienste (interne Abläufe und Funktionieren der Verwaltung) zuständig sind.

Aufgrund der vorgestellten Beiträge und unter Berücksichtigung der geschilderten Unterscheidung der Politikphasen wird argumentiert, dass die Phase des Politikzyklus relevant für die Evaluationspraxis ist. Aus den bisherigen Untersuchungen lassen sich sowohl für die Phase der Formulierung, als auch für die Phase der Implementierung Indizien, die auf eine erhöhte Evaluationspraxis hindeuten, ableiten. Diesen Beiträgen folgend wird argumentiert, dass Personen, die ihre Tätigkeit eher der Politikformulierung oder der -implementierung zuschreiben, im Gegensatz zu Personen, die interne Dienstleistungen erbringen, auch vermehrt mit der Evaluationspraxis konfrontiert werden. Daraus ergibt sich die folgende Hypothese:

H2c: Wenn die Verwaltungsmitarbeitenden in der Politikformulierung oder in der Politikimplementierung tätig sind, dann verfügen sie, im Vergleich zu Verwaltungsmitarbeitenden die interne Dienstleistungen tätigen, über eine höhere Evaluationspraxis.

³⁴ Die Autoren argumentieren, dass gerade diese beiden Phasen für die Untersuchung relevant sind und zudem die Zuteilung der Verwaltungsmitarbeitenden geeigneter ist. Beispielsweise sind Verwaltungsmitarbeitende, die im Agenda-Setting beteiligt sind, oft auch in die Politikformulierung involviert.

³⁵ Die Implementierung wird für diese Untersuchung synonym verwendet wie der Vollzug.

Die Einteilung der Mitarbeitenden entlang der Tätigkeitsfelder (Politikformulierung, Politikimplementierung und interne Dienste) wird mithilfe der Selbsteinstufung der Mitarbeitenden operationalisiert.

Eine andere Unterscheidung der Tätigkeiten von Verwaltungsmitarbeitenden kann bei Lowi (1972) gefunden werden. Dieser unterteilt die Politikfelder anhand individueller oder kollektiver Impacts, die diese generieren. Gemäss seiner Kategorisierung können Politikmassnahmen in distributive, redistributive und regulative Politiken unterschieden werden. Die unterschiedlichen Massnahmen haben Kosten, beziehungsweise Nutzen, für die involvierten Akteure zur Folge. Die Ungleichheit des Nutzens der Massnahmen, beziehungsweise ihrer Kosten, lösen wiederum Reaktionen bei den Betroffenen aus. Distributive Policy-Bereiche verfügen über wenig Konfliktpotenzial, da die Tangierten dieser Massnahmen den gleichen Nutzen zu keinen Kosten erwarten können. Anders sind die Folgen von redistributiven Massnahmen einzuordnen, wo der Nutzen nicht gleichmässig verteilt wird. Dementsprechend resultiert aus Massnahmen dieser Art ein hohes Konfliktpotenzial zwischen den Betroffenen. Zum Dritten zeichnen sich die regulativen Massnahmen durch in Konflikt stehende Koalitionen aus, die versuchen die Effekte der Regulierung zu ihren Gunsten zu beeinflussen (Windhoff-Héritier 1987). Aufgrund der unterschiedlichen Wirkungen wird angenommen, dass regulative und redistributive Massnahmen mit hohem Konfliktpotenzial zu einer höheren Evaluationspraxis führt. Massnahmen dieser Art fördern aufgrund stärkerer Konflikte zusätzliche wissenschaftliche Erkenntnisse in Form von Evaluationen, mit denen sich Kosten und Nutzen von Massnahmen besser einschätzen lassen.

Ein Indiz, das auf einen Einfluss von regulativen Massnahmen hindeutet, ist der Vermerk von Jann/Wegrich (2006: 54), demgemäss Evaluationen ihren Ursprung im Kontext regulativer sozialpolitischer Reformprogramme (Great Society) der 60er Jahre in den USA haben, die eine Reihe von Evaluationen generierten.

Während die eben geschilderten Einflüsse der Politikmassnahmen bisher noch nicht explizit untersucht wurden, weisen einige Beiträge auf die Bedeutung der Konfliktivität hin. Nach der Theorie von Lowi (1972) beinhalten die drei Massnahmen (distribu-

tive, redistributive und regulative Politiken) ein unterschiedliches Konfliktpotenzial. Damit wird die Konfliktivität mit der Unterscheidung der drei Politikmassnahmen in dieser Untersuchung indirekt berücksichtigt. Der Beitrag von Frey (2012) ist eine Untersuchung, welche die Konfliktivität in den Politikbereichen berücksichtigt. Mit Verweis auf andere Autoren (beispielsweise Sabatier/Jenkins-Smith 1993; Schrefler 2010; Weible 2008; Weible et al. 2009) argumentiert Frey (2012: 100ff.), dass in intensiven Konflikten forschungsbasierte Wirkungsinformationen genutzt werden, um die eigene Position zu legitimieren und sich in der Debatte Vorteile zu verschaffen. Des Weiteren legt Frey (2012: 73) dar, dass die Konfliktivität ein bedeutsamer Faktor ist, da diese die Rolle oder den Gestaltungsspielraum der Verwaltung und der Expertenkommissionen mitbestimmt. Bei einer hohen Konfliktivität kann dementsprechend die Evidenz dazu dienen, die Politiklösung zu legitimieren.

Aus den angeführten Erläuterungen wird abgeleitet, dass der Impact von Politikmassnahmen für die Evaluationspraxis relevant ist. Konkret wird argumentiert, dass regulative und redistributive Politikmassnahmen aufgrund eines höheren Konfliktpotenzials eine höhere Evaluationspraxis fördern. Dies soll mit folgender Hypothese empirisch überprüft werden:

H2d: Wenn die Verwaltungsmitarbeitenden Regulierungs- oder Redistributionsmassnahmen umsetzen, dann verfügen sie, im Vergleich zu Verwaltungsmitarbeitenden mit distributiven Politiken, über eine höhere Evaluationspraxis.

Die Einteilung der Mitarbeitenden anhand der unterschiedlichen Politikmassnahmen (Regulierungs-, Redistributiv- und Distributionsmassnahmen) wird mittels der Selbsteinstufung der Mitarbeitenden operationalisiert.

Organisationsspezifische Variablen H3

Einige Untersuchungen weisen auf den Einfluss von ausreichend vorhandenen Ressourcen auf die Verbreitung von NPM-Reformen hin (Kübler/Widmer 2007; Rieder/Widmer 2007). Horber-Papazian (2006) erwähnt die Mitverantwortung von insti-

tionellen Charakteristiken, wie finanzielle oder personelle Ressourcen in Verwaltungseinheiten, auf die Implementierung von Politiken. In Zusammenhang mit dem Konzept der Evaluationskapazitätsbildung³⁶ (ECB) wird häufig die Bedeutung von ausreichend vorhandenen Ressourcen (finanziellen, personellen und materiellen Ressourcen) für die Evaluationspraxis erwähnt (Arnold 2006; Preskill/Boyle 2008; Stockdill et al. 2002; Taut 2007; Volkov/King 2007).³⁷ Eine Studie konnte zudem aufzeigen, dass bei der Untersuchung von Entwicklungsorganisationen in OECD-Mitgliedsstaaten die Organisationsgrösse, sowie die finanziellen Ressourcen positiv mit der Anzahl an Evaluationen korrelieren (OECD 2010). Eine andere Untersuchung weist auf die Tatsache hin, dass Evaluationen im Bereich der Entwicklungszusammenarbeit häufig aufgrund fehlender Ressourcen nicht verwirklicht werden (Forss/Bandstein 2008).

Aufgrund der angeführten Hinweise wird angenommen, dass die Evaluationspraxis einen gewissen Umfang an Ressourcen erfordert. Nur wenn dieser verfügbar ist, kann sich eine Evaluationspraxis entwickeln. Damit orientiert sich diese Untersuchung an der von Howlett (2009) gemachten Äusserung, dass die Kapazität die wichtigste Vorbedingung für eine evidenzbasierte Politikgestaltung ist.

Empirische Evidenzen, die auf einen Zusammenhang in dieser Richtung deuten, lassen sich auch für die Schweiz finden. Balthasar/Rieder (2009) zeigen in der Untersuchung der Evaluationsaktivität in den Kantonen einen Zusammenhang zwischen der Kantoneinwohnerzahl und der Evaluationsaktivität der Verwaltungen auf. Wenn gefolgert wird, dass bevölkerungsreiche Kantone auch über grössere Verwaltungen verfügen, kann vermutet werden, dass die Verwaltungsgrösse die Evaluationspraxis positiv beeinflusst. Auch Balthasar/Müller (2014) können den Zusammenhang der relativen Grösse der Kantonsverwaltung auf die Verwendung von evidenzbasierten und gleichstellungssensitiven Informationen in den Kantonen empirisch nachweisen. Allerdings

³⁶ Der Begriff wurde im zweiten Kapitel ausführlicher beschrieben. Kurz zusammengefasst geht es bei diesem Konzept um die Verbesserung und Aufrechterhaltung der Evaluationspraxis (siehe Stockdill et al. 2002).

³⁷ Allerdings beziehen sich die genannten Ressourcen explizit auf solche für die Evaluationspraxis, beispielsweise genügend Mittel für die Evaluationsaktivität.

berücksichtigt die Studie von Balthasar/Müller (2014) Evidenz nicht nur in Form der Evaluation, sondern arbeitet mit einem breiteren Begriffsverständnis von Evidenz. Trotz dieser Einschränkung weisen beide Studien auf die Bedeutung der Ressourcenausstattung der Kantonsverwaltung auf die Evaluationspraxis hin. Aus diesem Grund wird die Ressourcenstärke mit folgender Hypothese in die Analyse miteinbezogen:

H3: Je ressourcenstärker die Verwaltung ist, desto höher ist die Evaluationspraxis der Verwaltungsmitarbeitenden.

Die Ressourcenstärke wird mit der Gesamtanzahl der in der Online-Mitarbeitendenbefragung angefragter Mitarbeitenden der Departemente und Direktionen operationalisiert.

Kontrollvariablen K

Folgende vier sozioökonomischen Variablen und eine geographische Variable werden als Kontrollvariablen in die Analyse miteinbezogen:

- Geschlecht, Alter, Bildungsniveau, Hierarchieposition und Kanton

Die Variablen Geschlecht, Alter und Bildungsniveau werden ebenfalls bei der Untersuchung der Evaluationsnachfrage von Parlamentsangehörigen bei der Studie von Bundi (2016) als Kontrolle verwendet und gelten zudem als robuste und oft verwendete Kontrollvariablen in der Untersuchung von Verwaltungsangestellten (für eine ausführliche Diskussion des Gebrauchs dieser Variablen für die Untersuchung der Arbeitsmotivation siehe Pandey/Stazyk 2008: 101ff.). Die Hierarchieposition ist zur Kontrolle hinzuzuziehen, da diese die Tätigkeit und die Einstellung von Mitarbeitenden beeinflusst (siehe Schein 2010: 63ff.). Auch bestehende Untersuchungen von Schweizer Verwaltungsangestellten beziehen die Hierarchieposition in die Analyse mit ein (siehe Anderfuhren-Biget et al. 2013; Giaque et al. 2010). Zudem wird mit der

fünften Kontrollvariable, die auf Unterschiede zwischen den Kantonen prüft, die Eigenheit des Schweizerischen Föderalismus bedacht.³⁸

Die soeben geschilderten Zusammenhänge sind in der folgenden Tabelle 6 zur Übersicht nochmals aufgelistet und werden im sechsten Kapitel, zuerst bivariat und danach multivariat, untersucht. Die exakte Operationalisierung der Merkmale kann der Tabelle 48 im Anhang entnommen werden.

Tabelle 6: Zu untersuchende Determinanten der Evaluationspraxis

Hypothese	Bezug
H1: Wenn die Evaluation in der Verwaltungseinheit institutionalisiert ist, dann ist die Evaluationspraxis der Verwaltungsmitarbeitenden höher.	Derlien (1990), Horber-Papazian (2006), Leeuw/Rozendal (1994), Varone et al. (2005),
H2a: Je volatiler der Arbeitsbereich der Mitarbeitenden ist, desto höher ist die Evaluationspraxis, über die sie verfügen.	Widmer/Beywl (2009)
H2b: Je komplexer der Arbeitsbereich der Mitarbeitenden ist, desto höher ist die Evaluationspraxis, über die sie verfügen.	Boswell (2009), Frey (2012), Radaelli (1999), Sabatier/Jenkins-Smith (1993)
H2c: Wenn die Verwaltungsmitarbeitenden in der Politikformulierung oder in der Politikimplementierung tätig sind, dann verfügen sie im Vergleich zu Verwaltungsmitarbeitenden, die interne Dienstleistungen übernehmen, über eine höhere Evaluationspraxis.	Balthasar/Müller (2014), Blatter et al. (2015), Frey/Ledermann (2017), Ledermann (2014)
H2d: Wenn die Verwaltungsmitarbeitenden Regulierungs- oder Redistributionsmassnahmen umsetzen, dann verfügen sie, im Vergleich zu Verwaltungsmitarbeitenden mit distributiven Politiken, über eine höhere Evaluationspraxis.	Frey (2012), Jann/Wegrich (2006), Sabatier/Jenkins-Smith (1993), Schrefler (2010), Weible (2008), Weible et al. (2009)
H3: Je ressourcenstärker die Verwaltung ist, desto höher ist die Evaluationspraxis der Verwaltungsmitarbeitenden.	Balthasar/Müller (2014), Balthasar/Rieder (2009), Forss/Bandstein (2008), Preskill/Boyle (2008), Taut (2007)
Kontrollvariablen: Geschlecht, Alter, Hierarchieposition, Bildungsniveau, Kanton	

Quelle: Eigene Darstellung

Die beiden abhängigen Variablen der Evaluationspraxis werden folgendermassen operationalisiert: Die Evaluationsaktivität wird mit der Häufigkeit der Mitarbeitenden an der Beteiligung der Evaluationsaktivität (Einschätzung der individuellen Evaluationsaktivität im Jahresdurchschnitt), die Verwendung mit der Häufigkeit der Verwendung von Evaluationen in die Arbeit ermittelt. Beide Angaben stammen aus der Online-

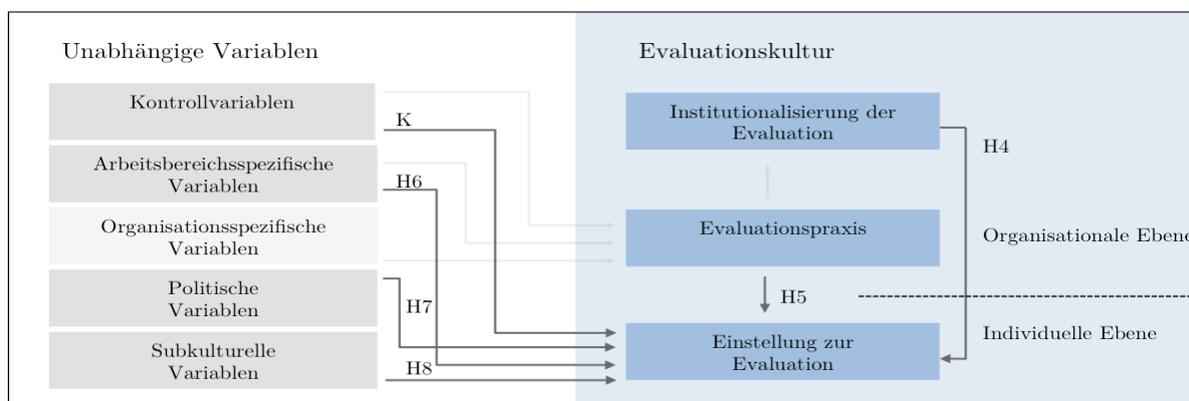
³⁸ Der Schweizerische Föderalismus zeichnet sich durch wirtschaftliche, kulturelle und politische Eigenheiten der Kantone aus (Freitag 2005; Vatter 2002: 17f.), diese manifestieren sich auch in unterschiedlichen Ausgestaltungen der kantonalen Verwaltungen (siehe Koller 2013).

Mitarbeitendenbefragung (siehe für die konkrete Operationalisierung die Tabelle 48 im Anhang).

3.2 Theoretisches Modell für die Untersuchung der Einstellung zur Evaluation

Das zweite Modell widmet sich der Untersuchung der Einstellung zur Evaluation, ersichtlich in der nachfolgenden Abbildung 5. Die beiden Hypothesen vier und fünf erörtern Einflüsse der Evaluationspraxis und der Institutionalisierung der Evaluation auf die Einstellung. Die Hypothesen sechs bis acht berücksichtigen arbeitsbereichsspezifische, politische und subkulturelle Variablen. Zudem werden Kontrollvariablen in der Analyse berücksichtigt.

Abbildung 5: Modell für die Untersuchung der Einstellung zur Evaluation



Anmerkungen: Die Pfeile geben die zu untersuchenden Einflussbeziehungen wieder.

Quelle: Eigene Darstellung

Institutionalisierung der Evaluation H4

Mit der vierten Hypothese wird der Einfluss der Institutionalisierung der Evaluation auf die Einstellung hinsichtlich dieser untersucht.

Bemerkenswert bei dieser Hypothese ist der Bezug zur politischen Kulturforschung. Gemäss Salzborn (2009: 7) ist die Vermittlung zwischen Individuum und Struktur ein häufiger Untersuchungsgegenstand der politischen Kulturforschung. In diesem Zusammenhang wird oft auf die Kongruenz von Struktur und Politik hingewiesen (siehe

Greiffenhagen 2009: 7ff.; Pickel/Pickel 2006: 49ff.). Diese verweist auf die Konsistenz der Werte und Einstellungen einer Bevölkerung mit den Institutionen und wurde bereits in der Studie von Almond/Verba (1963) erwähnt.³⁹ Dementsprechend untersucht auch Hypothese H4 das Zusammengehen der Institutionalisierung (Struktur) mit dem Individuum.

Gemäss Dolder et al. (2017) in Anlehnung an Varone (2007, 2015) lassen sich zwei Entwicklungsprozesse im Zusammenhang mit der Evaluation unterscheiden: Einerseits der „bottom-up“-Entwicklungsprozess, bei dem sich durch positive Erfahrungen eine Evaluationspraxis entwickelt, die in einem nächsten Schritt zu einer Institutionalisierung der Evaluation führen kann, und andererseits die „top-down“-Entwicklung, bei der die Institutionalisierung am Anfang steht, und eine Evaluationstätigkeit initiiert, die bei den Mitarbeitenden Lernprozesse auslöst und einen positiven Einfluss auf die Einstellung nimmt. Die vorliegende Untersuchung wird sich, wie bereits erläutert wurde, auf die Analyse der „top-down“-Einflussrichtung beschränken.

Ein weiteres Argument, das für einen Einfluss entsprechend der “top-down“-Einflussrichtung spricht, kann aus der Forschung über die Organisationskultur abgeleitet werden. Der Grundgedanke der Organisationskultur ist, dass die Organisation in Bezug auf ihre Mitglieder strukturierend wirkt. Schein (2010) beschreibt die Transformation von Werten und Vorstellungen der Mitarbeitenden einer Organisation. Beispielsweise mithilfe von Gruppenlernprozessen können Vorgesetzte Einfluss auf die Einstellungen der Mitarbeitenden ausüben: „If the manager convinces the group to act on her belief, the solution works, and the group has a shared perception of that success, then the perceived value [...] gradually becomes transformed: first into a shared value or belief and ultimately into a shared assumption” (Schein 2010: 26). Die Institutionalisierung (Dimension 1) kann als Signal zur Unterstützung von Evaluation seitens der Organisation (Vorgesetzten) verstanden werden. Sowohl mit der Institutionalisierung,

³⁹ Im zweiten Kapitel wurde die Studie von Almond/Verba (1963) bereits eingeführt. Die Untersuchung der beiden Autoren gilt als Geburtsstunde der systematischen Untersuchung der politischen Kultur.

als auch mit der Einführung einer Evaluationspraxis (Dimension 2) signalisiert die Organisation gegenüber den Mitarbeitenden in Bezug auf die Evaluation ihr Einverständnis und ihre Unterstützung, was sich wiederum positiv auf die Einstellung zur Evaluation (Dimension 3) der Mitarbeitenden auswirken kann. Des Weiteren impliziert die Definition der Organisationskultur selbst einen Einfluss der Organisation auf die Mitarbeitenden: „As such, organizational culture refers to a set of shared values, belief, assumptions, and practices that shape and guide members’ attitudes and behavior in the organization“.⁴⁰ Sollte die Organisationskultur auch im Bereich der Evaluation auf die Mitarbeitenden wirken, kann in diesem Zusammenhang argumentiert werden, dass wenn die Evaluation in den Organisationen institutionalisiert ist, sich das auch auf die Einstellung der Mitarbeitenden positiv auswirkt.

Empirische Untersuchungen hinsichtlich der Wirkung der Institutionalisierung der Evaluation auf die Einstellung zur Evaluation konnten keine gefunden werden. Im Gegensatz dazu gibt es Untersuchungen, die aufzeigen, dass Organisationen einen Einfluss auf die Einstellung ihrer Mitarbeitenden ausüben. Als Beispiel dient die Studie von Abdul Rashid et al. (2004), die nachweisen kann, dass unterschiedliche Organisationskulturtypen einen Einfluss auf die Einstellung der Mitarbeitenden bezüglich organisationaler Veränderungen nehmen.

Aufgrund der geschilderten Beiträge wird angenommen, dass die Institutionalisierung der Evaluation die Einstellung zur Evaluation beeinflusst. Dies wird mit folgender Hypothese überprüft:

H4: Wenn die Evaluation in der Verwaltungseinheit institutionalisiert ist, dann ist die Einstellung zur Evaluation der Mitarbeitenden positiver.

Die Institutionalisierung der Evaluation wird wie bei der vorherigen Untersuchung der Evaluationspraxis operationalisiert. Dabei werden Verwaltungseinheiten in solche mit

⁴⁰ Definition von Abdul Rashid et al. (2004: 165f.).

keiner oder mit einer ansatzweisen Institutionalisierung der Evaluation und andere, in denen die Evaluation institutionalisiert ist, unterteilt.

Evaluationspraxis H5

Mit der Untersuchung des Einflusses der Evaluationspraxis auf die Einstellung zur Evaluation wird ein zentraler Zusammenhang analysiert, der in der psychologischen Forschung oft Erwähnung findet. Bereits Ende der 90er Jahren nannte Mummendey (1988: 1) die Untersuchung der Zusammenhänge zwischen Verhalten und Einstellung als eine der Grundfragen der Psychologie. Allerdings interessieren sich diese Forschungsarbeiten häufig dafür, inwiefern das konkrete Handeln der dazu analogen Gesinnung entspricht und wieso diesbezüglich Diskrepanzen ausgemacht werden können. Im Gegensatz dazu wird in der vorliegenden Untersuchung der Einfluss des organisationalen Verhaltens auf die individuelle Einstellung analysiert und es wird erforscht, ob Mitarbeitende, die in einer Verwaltung mit einer höheren Evaluationspraxis tätig sind, über eine positivere Einstellung zur Evaluation verfügen.

Deshalb ändert sich die Analyseebene für die Evaluationspraxis: Diese wird für das zweite Modell aggregiert auf Organisationsebene betrachtet. Grund für diese Betrachtungsebene ist der Einbezug des sozialen/organisationalen Einflusses¹ auf die Untersuchung der Einstellung. Theoretischer Hintergrund ist die sogenannte „Kontakthypothese“ und der „Effekt der blossen Darbietung“ (englisch: mere exposure effect).² Grundgedanke dieser beiden Theorien ist, dass durch den Kontakt mit Menschen oder auch Objekten die Vorurteile gegenüber diesen abgebaut werden. Die

¹ Für eine ausführliche Diskussion des sozialen Einflusses auf die Einstellung siehe den Beitrag von Wood (2000). Ausserdem ist zu berücksichtigen, dass für die Analyse der Einstellung zur Evaluation nur Personen, die angeben, über Evaluationsaktivität zu verfügen (aufgrund der „Non-Attitude-Problematik“), berücksichtigt werden (siehe dazu das nächste Kapitel).

² Neben theoretischen Gründen ergeben sich auch methodische Vorteile dieser Betrachtungsebene. Durch die Berücksichtigung verschiedener Quellen (Online-Mitarbeitendenbefragung und der Recherche nach Evaluationsberichten) wird verhindert, dass die Befragten die Quelle für die exogene und auch die endogenen Variablen sind und kann eine Verzerrung in Form des sogenannten „common method bias“ ausgeschlossen werden (siehe dazu Spector 2006 und auch die Ausführungen im nächsten Kapitel). Des Weiteren können Verzerrungen, die in Online-Befragungen auftreten können, für die Evaluationsaktivität beispielsweise durch die Selbstselektion (englisch: self-selection-bias) und die Möglichkeit von Falschangaben (englisch: misreporting), vermieden werden (siehe auch dazu die Ausführungen im nächsten Kapitel).

Kontakthypothese, eingeführt von Allport (1954), geht davon aus, dass der Kontakt zu Personen aus einer Fremdgruppe die Vorurteile gegenüber dieser verringern vermag. Unter Vorurteile sind meist negative Einstellungen gegenüber etwas zu verstehen. Die Kontakthypothese wurde seither in zahlreichen Studien angewendet um die Vorurteilsreduktion zu erklären (für einen Überblick siehe Pettigrew/Tropp 2006). Beispielsweise kann die Untersuchung von Wagner et al. (2016) nachweisen, dass in Bezirken mit höheren Anteilen an Ausländern die Vorurteile gegenüber diesen niedriger sind. Während sich die meisten Untersuchungen der Kontakthypothese auf die Einstellung gegenüber Fremdgruppen beschränken, gibt es einige Beiträge, die einen Einfluss auf die Einstellung gegenüber weiteren Objekten nachweisen. Ausgangspunkt dieser Untersuchungen ist der Beitrag von Zajonc (1968), der auf einen positiven Einfluss auf die Einstellung durch wiederholte Wahrnehmung und Exposition zu einem Objekt verweist. Entscheidend ist dabei die Darbietungshäufigkeit. Gemäss der Studie von Zajonc (1968) wird sich durch eine erhöhte Darbietungshäufigkeit eine verbesserte Einstellung gegenüber des Reizes (wahrgenommenes Objekt) einstellen. Der „Effekt der blossen Darbietung“ konnte auf Objekte verschiedener Art nachgewiesen werden. Diese reichen von Gemälden bis hin zu Photographien und Musik (für einen Überblick bisheriger Studien siehe Bornstein 1989; Montoya et al. 2017).⁴³ Abgeleitet aus den eben vorgestellten Effekten ist anzunehmen, dass diese auch auf die Einstellung zur Evaluation angewendet werden können und dass demzufolge Mitarbeitende, die in einer Verwaltungseinheit mit einer erhöhten Evaluationspraxis arbeiten häufiger in Kontakt mit Evaluationen kommen und dadurch ihre Vorurteile abbauen.

Zwei weitere Argumente, die für einen Einfluss der organisationalen Evaluationspraxis auf die Einstellung sprechen, wurden bereits bei der vorhergehenden Hypothese H4 vorgestellt. Das erste Argument bezieht sich auf die von Dolder et al. (2017) in Anleh-

⁴³ Des Weiteren zeigen die beiden Meta-Studien (Bornstein 1989; Montoya et al. 2017), dass weitere Elemente die Beziehung zwischen Darbietungshäufigkeit und Einstellung moderieren können, wie beispielsweise der „erste Eindruck“. Zudem verweisen diese Beiträge, dass je nach Stimulus der Zusammenhang zwischen Darbietungshäufigkeit und Einstellung nicht immer linear ist. Beispielsweise geht die positive Einstellung gegenüber einem Lied ab einer gewissen Darbietungshäufigkeit wieder zurück.

nung an Varone (2007, 2015) geschilderten „top-down“-Entwicklungsprozess. Bei diesem stellt sich aufgrund der Evaluationspraxis eine positive Erfahrung bei den Mitarbeitenden ein. Daraus lässt sich auch ein Einfluss in Form der Etablierung einer positiven Einstellung ableiten. Das zweite Argument lehnt an die von Schein (2010: 26 ff.) geschilderte Transformation der verschiedenen Ebenen der Evaluationskultur an. Ein Teil dieser Transformation wird mit der vorliegenden Hypothese überprüft. Wie bereits bei der vorherigen Hypothese argumentiert wurde, kann die Einführung einer Evaluationspraxis als Signal der Organisation für die Unterstützung gegenüber der Evaluation gewertet werden. Diese Unterstützung seitens der Organisation sollte wiederum eine positivere Einstellung zur Evaluation hervorbringen.

Ein weiteres Argument für den Einfluss in die postulierte Richtung kann bei Beiträgen gefunden werden, welche individuelle Veränderungsprozesse aufgrund der Evaluationspraxis beschreiben. Der bekannteste Beitrag zu diesem Thema stammt von Patton (1998). Dieser Beitrag wurde bereits im zweiten Kapitel ausführlich beschrieben. Der Autor konnte Veränderungen bei Personen, die an Evaluationen beteiligt sind, beobachten. Insbesondere verweist Patton (1998: 225f.) auf den Aufbau von bestimmten Sichtweisen und sogenannten evaluativen Wertvorstellungen bei Evaluatoreninnen und Evaluatoren aufgrund ihrer Evaluationserfahrung: „We, as evaluators, have our own values, our own ways of thinking, our own language, our own hierarchy, and our own reward system“. In diesem Zusammenhang erwähnt Patton zudem die Erlernung eines evaluativen Denkens (englisch: thinking evaluatively). Diese Denkweise geht mit spezifischen Wertvorstellungen⁴ einher, die Evaluatoreninnen und Evaluatoren aufbauen. Auch der Beitrag von Katz et al. (2002) erwähnt eine Veränderung in Richtung einer evaluativen Denkgewohnheit (englisch: evaluation habit of mind) aufgrund der Evaluationsaktivität. Eine evaluative Denkgewohnheit beschreibt einen Wechsel im

⁴ Beispiele von evaluativen Wertvorstellungen sind gemäss Patton (1998: 226): Klarheit; Spezifität und Fokussierung; Systematik; Annahmen explizit machen; Operationalisierung von Programmkonzepten; Ideen und Ziele; Unterscheidung Input und Prozess von dem Outcome; Wertzuschreibung der empirischen Evidenz; Unterscheidung von Fakten, Interpretationen und Werturteilen.

Denken darüber, welche Form von Information für die Planung und zu treffende Maßnahmen wichtig sind. Zudem basiert diese Denkgewohnheit auf einer evidenzbasierten Grundlage, die vorhandene Daten zur Entscheidungsfindung einbezieht. In der Untersuchung von Katz et al. (2002), in der Schulleitende diesbezüglich interviewt wurden, konnte ein Einfluss in Richtung der evaluativen Denkgewohnheit infolge der Evaluationspraxis aufgezeigt werden.

Hinsichtlich der geschilderten Beiträge wird angenommen, dass die organisationale Evaluationspraxis die Einstellung zur Evaluation beeinflusst. In Anlehnung an den Mechanismus der Kontakthypothese und den Effekt der blossen Darbietung ist anzunehmen, dass in Verwaltungseinheiten mit einer stärkeren Evaluationsaktivität und einer stärker verbreiteten Evaluationsverwendung die Mitarbeitenden häufiger in Kontakt mit Evaluationen kommen und sich damit die Ablehnung gegen die Evaluation abbaut. Ferner werden durch die Erfahrungen der Evaluationspraxis evaluative Wertvorstellungen in einer Organisation aufgebaut, die sich über den geschilderten „top-down“-Einflussprozesses auf die Mitarbeitenden übertragen. Diese Annahmen werden mit den folgenden zwei Hypothesen getestet:

H5a: Je höher die Evaluationsaktivität in der Verwaltungseinheit ist, desto positiver ist die Einstellung zur Evaluation.

Für die Ermittlung der Evaluationsaktivität der Departemente und Direktionen wurde für den Zeitraum 2005 bis 2013 online nach Evaluationen gesucht. Diese wird mit der Gesamtanzahl an Evaluationsberichten auf Stufe der Departemente und Direktionen operationalisiert.⁴⁵

⁴⁵ Die Evaluationsaktivität kann nicht auf Stufe der Ämter und Bereiche operationalisiert werden. Grund dafür sind die General- und Departementssekretariate, denen häufig keine Evaluationsberichte zugeordnet werden konnten, die aber aufgrund ihrer Stabs- und Koordinationsfunktion über Evaluationsaktivität verfügten. Siehe dazu Kapitel fünf, wo die organisationale Evaluationskultur der Departemente und Direktionen beschrieben wird.

H5b: Je höher die Anteile der Evaluationsverwendung in der Verwaltungseinheit sind, desto positiver ist die Einstellung zur Evaluation.

Die Häufigkeit der Evaluationsverwendung der Evaluation wurde in der Online-Mitarbeitendenbefragung erhoben. Die Evaluationsverwendung wird auf Stufe der Ämter und Bereiche als Prozentanteile von Mitarbeitenden mit mindestens seltener Verwendung der Evaluation in die Arbeit operationalisiert.

Arbeitsbereichsspezifische Variablen H6

Bereits bei den Hypothesen zur Analyse der Evaluationspraxis wird zwischen den verschiedenen Phasen des Politikzyklus unterschieden. Analog zur dort gemachten Unterteilung der Verwaltungsmitarbeitende wird auch hier, in Anlehnung an Anderfuhren-Biget et al. (2013), unterteilt: in jene, die in der Politikformulierung arbeiten, solche, die in der Implementierung von öffentlichen Politiken tätig sind, sowie in Mitarbeitende, die interne Dienstleistungen übernehmen.⁴⁶

Im Folgenden wird argumentiert, dass die Verwaltungsmitarbeitenden aufgrund ihrer unterschiedlichen Position und Rolle, die sie im Entscheidungsprozess entlang des Politikzyklus einnehmen, eine unterschiedliche Einstellung zur Evaluation einnehmen.

Bisher wurde der Zusammenhang zwischen der Position im Politikzyklus und der Arbeitsmotivation empirisch untersucht. Der Artikel von Anderfuhren-Biget et al. (2013) analysiert die Arbeitsmotivation von Schweizer Kantons- und Gemeindeverwaltungsangestellten. Die Arbeitsmotivation (englisch: Public Service Motivation, PSM) wird an vier verschiedenen Dimensionen gemessen: „attraction to politics, commitment to the public interest, compassion and self-sacrifice“. Anderfuhren-Biget et al. (2013) weisen darauf hin, dass Verwaltungsmitarbeitende, die in der Formulierungsphase von öffentlichen Politiken tätig sind, über ein politisches Interesse verfügen, da sie in ei-

⁴⁶ Die konkrete Unterscheidung wurde bereits im vorherigen Unterkapitel in der zweiten Hypothese (H2c) erläutert und kann dort nachvollzogen werden. Des Weiteren ist der genaue Wortlaut für die Einstufung sowie die Operationalisierung im Anhang aufgeführt.

nem Umfeld arbeiten, wo politische Entscheidungen getroffen werden und sie gegenüber gewählten Politikerinnen und Politikern direkte Verantwortung übernehmen. Ausserdem haben diese Mitarbeitenden die Absicht, die politische Agenda zu kontrollieren und verfügen über klar definierte politische Werte und Ideologien, die sie über die Konstruktion oder mittels der Definition der Ziele dieser öffentlichen Politiken verfolgen. Analog dazu lassen sich auch für Verwaltungsmitarbeitende, die in der Implementierungsphase arbeiten, Unterschiede in der Arbeitsmotivation ableiten. Anderfuhren-Biget et al. (2013) argumentieren in Anlehnung an Pandey/Stazyk (2008), dass Mitarbeitende, die in der Implementierungsphase tätig sind, eher in Kontakt mit den Zielgruppen stehen und dadurch die Möglichkeit haben, Empathie mit diesen aufzubauen. Da diese Mitarbeitenden in Kontakt zu den Nutzniessern der öffentlichen Politik stehen, verfügen sie über eine stärker ausgeprägte Verpflichtungsbereitschaft gegenüber dem Gemeinwohl.

Anderfuhren-Biget et al. (2013) finden nachweisbare, positive Effekte für in der Politikformulierung und in der -implementierung tätige Verwaltungsangestellten auf zwei Dimensionen der Arbeitsmotivation: „attraction to politics“ und „commitment to the public interest“.⁴⁷ Damit belegt die Untersuchung eine höhere Arbeitsmotivation bei Personen, die in der Formulierung oder in der Implementierung tätig sind, als bei jenen, die in den internen Diensten arbeiten. Die „attraction to politics“ charakterisiert, wie stark sich die Mitarbeitenden für das politische Geschehen interessieren und wie sehr sie durch eine mögliche Einflussnahme auf die Politikgestaltung motiviert werden. Das „commitment to the public interest“ adressiert die Orientierung der Mitarbeitenden am Gemeinwohl und an ihrer gesellschaftlichen Verantwortung.

Der Beitrag von Anderfuhren-Biget et al. (2013) kann somit einen Einfluss des Politikzyklus auf die Werte von Verwaltungsangestellten in der Schweiz empirisch nachweisen. Im Zusammenhang mit der Einstellung zur Evaluation wird argumentiert, dass

⁴⁷ Auch die Untersuchung von Castaing (2006) stellt für die in der Politikformulierung tätigen Verwaltungsmitarbeitenden, im Vergleich zu Mitarbeitenden, die interne Dienstleistungen übernehmen, ein höheres Level in der Dimension „commitment to the public interest“ fest.

gerade die Dimension „attraction to politics“ eine wichtige Rolle einnimmt. Diese Dimension berücksichtigt das Interesse für das politische Geschehen und die Motivation für die Einflussnahme auf die Politikgestaltung (siehe für die konkrete Messung Giauque et al. 2011) und weist damit wichtige Elemente auf, die von Schwandt/Dahler-Larsen (2006) als Gründe für die Ablehnung gegenüber der Evaluation angeführt werden.⁴⁸ Die Begründung lautet, dass die Ablehnung der Evaluation in Gesellschaften in der die Auffassung vorherrscht, dass man Sachverhalte beeinflussen und steuern kann, schwächer ist.

Aufgrund des geäußerten Zusammenhanges wird angenommen, dass Mitarbeitende der Implementierungs- und auch der Formulierungsphase, die über eine höhere „attraction to politics“ verfügen, aufgrund ihrer Gestaltungsmotivation gewillt sind, Politikinhalt zu steuern und zu beeinflussen, demzufolge eine positivere Einstellung zur Evaluation aufweisen. Dieser Zusammenhang wird mit der nachfolgenden Hypothese untersucht.

H6: Wenn die Verwaltungsmitarbeitenden in der Politikformulierung oder in der Politikimplementierung tätig sind, dann verfügen sie im Vergleich zu Mitarbeitenden, die interne Dienstleistungen tätigen, über eine positivere Einstellung zur Evaluation.

Die Einteilung der Mitarbeitenden entlang der Tätigkeitsfelder (Politikformulierung, Politikimplementierung und interne Dienstleistungen) wird mithilfe der Selbsteinstufung der Mitarbeitenden operationalisiert.

Politische Variablen H7

Milic (2008) verweist auf die Links-Rechts-Selbsteinschätzung (englisch: left-right self-identification) als einen der zuverlässigsten Prädiktoren für den Abstimmungsentcheid bei schweizerischen Sachabstimmungen. Hinweise dafür, argumentiert Milic

⁴⁸ Auf den Beitrag von Schwandt/Dahler-Larsen (2006) wird in der übernächsten Hypothese (H7b) vertiefter eingegangen.

(2008), würden sich in zahlreichen bivariaten Vox-Analysen⁴⁹ und ebenso in der multivariaten Studie von Buri/Schneider (1993) finden. Auch für die Eidgenössischen Wahlen lassen sich erkenntliche Tendenzen entlang der Links-Rechts Positionierung und dem Wahlentscheid nachweisen (siehe Lutz 2016: 22ff.). Freire (2006: 360) bemerkt, dass die Links-Rechts-Selbsteinstufung für viele andere Länder den Hauptprädiktoren des Wahlverhaltens darstellt und folgert: „Consequently, there is considerable evidence to suggest that there is little empirical support for the ‘end of ideology’ thesis, and that, particularly at the individual level, the left-right divide is still a very important information-economizing device, enabling electors to cope with political complexities, at least in western Europe“. In diesem Zusammenhang ist darauf zu verweisen, dass in den USA anstelle der Links-Rechts-Unterscheidung vermehrt die Liberal-Konservatismus-Unterscheidung zur Messung der politischen Ideologie angewendet wird. Diese wird häufig auch synonym verwendet (für eine ausführliche Diskussion siehe Jost et al. 2009: 310ff.).⁵⁰

Ausserdem geht man davon aus, dass die politische Ideologie über den Prädiktator des Wahlentscheids hinaus ein Glaubenssystem (englisch: belief system) und eine Gruppenidentität darstellt (siehe Gauchat 2015; Martin/Desmond 2010; Perrin et al. 2014). Jost et al. (2009) weisen in ihrem umfassenden Beitrag der Literatur der politischen Ideologie darauf hin, dass diese sich aus Überzeugungen, Bewertungen und Ontologien zusammensetzt. Der Beitrag weist auf eine Vielzahl an Untersuchungen hin, die signifikante Unterschiede entlang der politischen Ideologie nachweisen können. Die Vielfalt der Differenzen wird im Beitrag von Jost et al. (2008) ersichtlich. So sind beispielsweise Liberale positiver gegenüber ausländischen Filmen, grossen Städten, Tätowierungen, Gedichten und Auslandsreisen eingestellt. Weitere Studien zeigen, dass

⁴⁹ Im Nachgang an die eidgenössischen Volksabstimmungen wird das Stimmverhalten der Schweizer Bevölkerung mittels einer repräsentativen Befragung erhoben und analysiert. Siehe dazu http://www.ipw.unibe.ch/forschung/vox_analysen/index_ger.html oder <http://www.gfsbern.ch/de-ch/Publikationen/VOX-Analysen>.

⁵⁰ Als weiteres ist zu bedenken, dass die politische Ideologie (Links-Rechts) in unterschiedlichen Ländern und Kontinenten verschiedene Bedeutungen hat. Beispielsweise kann die Untersuchung von Freire/Kivistik (2013) aufzeigen, dass sogenannte westliche Werte in den untersuchten dreizehn Ländern (auf vier Kontinenten) unterschiedlich mit der Links-Rechts-Selbsteinschätzung in Beziehung stehen.

die Links-Rechts-Selbsteinstufung mit politischen Überzeugungen und Werten korreliert (Freire/Kivistik 2013; Knutsen 1995; De Vries et al. 2013). Die Studie von De Vries et al. (2013) beispielsweise kann den Zusammenhang zwischen der Links-Rechts-Selbsteinschätzung, sowie der ökonomischen und kulturellen Einstellungen empirisch nachweisen. Die Untersuchung von Blank/Shaw (2015) analysiert den Einfluss der politischen Ideologie auf die Bereitschaft der Bürgerinnen und Bürger, wissenschaftliche Empfehlungen zu politischen Themen anzunehmen. Die Studie zeigt empirische Evidenz für den Zusammenhang zwischen der Liberalismus-Konservatismus-Selbsteinstufung und der Einstellung gegenüber wissenschaftlichen Informationen. In allen sechzehn untersuchten Themenfeldern und im Gesamtmodell sind Personen, die sich liberaler einstufen, positiver gegenüber wissenschaftlichen Experten eingestellt. Damit kommt die Studie zu den gleichen Ergebnissen wie weitere Untersuchungen, die den Einfluss der politischen Ideologie auf einzelne Themenbereiche nachweisen können: beispielsweise auf die Einstellung gegenüber der Klimawissenschaften (siehe Kahan et al. 2012; McCright/Dunlap 2011). Auch bei diesen Untersuchungen geht eine linke respektive liberale Selbsteinstufung mit einer Sichtweise, die mit der Klimawissenschaft übereinstimmt, einher.

Aufgrund der vorgestellten Beiträge konnte aufgezeigt werden, dass erstens die Links-Rechts-Selbsteinschätzung für die Schweiz nach wie vor von Relevanz ist und zweitens diese einen Einfluss auf ökonomische, kulturelle und politische Einstellungen ausübt. Ferner ist aufgrund der nachgewiesenen Einflüsse der politischen Ideologie auf die Einstellung gegenüber wissenschaftlichen Informationen anzunehmen, dass die Links-Rechts-Selbsteinstufung der Mitarbeitenden auch für die Einstellung zur Evaluation relevant ist. Infolgedessen wird die Links-Rechts-Selbsteinstufung mit nachfolgender Hypothese in die Analyse miteinbezogen:

H7a: Je weiter links sich die Mitarbeitenden in der Links-Rechts-Skala einstufen, desto positiver ist die Einstellung zur Evaluation.

Die Links-Rechts-Einstufung wird anhand der Selbsteinstufung der Mitarbeitenden entlang der Links-Rechts-Skala operationalisiert.

Auf einen weiteren Einfluss auf die Einstellung zur Evaluation wurde bereits in der Hypothese H6 Bezug genommen. Thomas Schwandt und Peter Dahler-Larsen diskutieren basierend auf den Erfahrungen von zwei Evaluationsprojekten die Gründe für eine Ablehnung gegenüber der Evaluation.⁵¹ Dabei stellen sie fest, dass eine Ablehnung gegenüber der Evaluation in Gemeinschaften mit traditionellen Werten stärker verbreitet sind, als in Gemeinschaften mit modernen Werten (Schwandt/Dahler-Larsen 2006). In Entgegnung zu Peter Dahler-Larsen stellt Thomas Schwandt fest, dass die Evaluation als ein modernistisches Unterfangen (englisch: modernist undertaking) betrachtet werden kann, das in einem Gegensatz steht zu den traditionellen Werten und bemerkt: „Contemporary societies are marked by their Enlightenment, modernist belief in the power of reason and evidence to ‘fix things’, as you say. In the self-understanding of modernity, no set of circumstances, no problem, is beyond the reach of reason and evidence” (Schwandt/Dahler-Larsen 2006: 500).

Diese Feststellung soll in dieser Untersuchung berücksichtigt werden. Konkret wird überprüft, ob modernistische, beziehungsweise traditionalistische Werte tatsächlich Einfluss auf die Einstellung zur Evaluation nehmen. Gemäss der Überlegungen von Schwandt/Dahler-Larsen (2006) wird die Hypothese in folgende Einflussrichtung formuliert:

H7b: Je traditionalistischer sich die Mitarbeitenden in der Modernismus-Traditionalismus-Skala einstufen, desto negativer ist die Einstellung zur Evaluation.

⁵¹ Der Beitrag von Schwandt/Dahler-Larsen (2006) folgt nicht dem Aufbau eines herkömmlichen wissenschaftlichen Artikels. Beide Forscher kommen in dem Beitrag abwechselnd zu Wort, ähnlich einer mündlichen Diskussion.

Der Traditionalismus wird anhand der Selbsteinstufung der Mitarbeitenden entlang der Modernismus-Traditionalismus-Skala operationalisiert.⁵²

Subkulturelle Variablen H8

In der Organisationskulturliteratur finden sich einige Beiträge, die auf den vorherrschenden Unterschied in der Organisationskultur zwischen den Angestellten und den Vorgesetzten verweisen.

Schein (2010: 56) beschreibt eine Subkultur, die sich zwischen Angestellten der gleichen Position finden lässt: „Elders will teach newly promoted supervisors how to perform their roles, and this mentoring will be more powerful than any formal training they might be given. In the same way, middle management and higher levels will develop their own shared assumptions, and at each level, those assumptions will be taught to newcomers as they get promoted to that level“. Des Weiteren beschreibt Schein (2010: 66ff.) das Vorhandensein einer sogenannten Vorgesetzten-Subkultur (englisch: executive subculture). Der Grund für diese Subkultur kann darin gefunden werden, dass Personen in einer Vorgesetztenposition Finanzierungs- und Überwachungsfunktionen wahrnehmen. In dieser Position ist es schwierig, verlässliche und vertrauenswürdige Informationen zu erhalten, da die ihnen unterstellten Mitarbeitenden den Anreiz haben, diese zu ihren Gunsten zu beeinflussen. Die Schwierigkeit, verlässliche Informationen zu erhalten, steigt ausserdem mit der Höhe der Hierarchieposition. Gleichzeitig nimmt die Notwendigkeit, verlässliche Informationen zu erhalten, in höheren Positionen zu, da entlang der Hierarchiestufen die Rechenschaftspflicht der Führungspersonen zunimmt. Gemäss Schein (2010) hängt diese Veränderung der Tätigkeit und der Rechenschaftspflicht auch mit einer höheren Wertschätzung und einer

⁵² In einem Pretest wurde die Messung des Traditionalismus-Modernismus mithilfe von vier Fragen in Anlehnung an Epple (2013) und Stolz (2001) getestet. Diese vier Fragen zeigten für die Befragung der Verwaltungsmitarbeitenden ein tiefes Cronbachs Alpha von 0.541 auf. Aus diesem Grund wird die Variable Traditionalismus direkt mit der Traditionalismus-Modernismus-Skala gemessen (siehe den Fragebogen im Anhang).

grösseren Wichtigkeit von Informationen, Informationssystemen und Evidenzen zusammen.

Auch die in Hypothese H7a vorgestellte Untersuchung der Arbeitsmotivation von Schweizer Verwaltungsmitarbeitenden, durchgeführt von Anderfuhren-Biget et al. (2013), zeigt Unterschiede zwischen Mitarbeitenden in verschiedenen Hierarchiepositionen auf. Die Arbeitsmotivation (englisch: Public Service Motivation, PSM) wurde wie erwähnt in vier verschiedenen Dimensionen gemessen. Die Untersuchung konnte Unterschiede zwischen Mitarbeitenden unterschiedlicher Hierarchiepositionen in den Dimensionen „commitment“ und „compassion“ im multivariaten Modell nachweisen. Auch die Analysen von Camilleri (2007) und von Moynihan/Pandey (2007) konnte einen positiven Zusammenhang zwischen hierarchischer Position und der PSM aufzeigen.

Aufgrund des Zusammenhanges der Hierarchieposition mit den Werten und der Einstellung zu Informationen und Evidenzen wird angenommen, dass die Hierarchieposition die Einstellung zur Evaluation positiv beeinflusst. Dies wird mit folgender Hypothese empirisch überprüft:

H8a: Je höher die Position der Mitarbeitenden in der Verwaltungshierarchie ist, desto positiver ist die Einstellung zur Evaluation.

Die Hierarchieposition wird anhand der Selbsteinstufung der Mitarbeitenden operationalisiert.

Auch die nächste Hypothese thematisiert die Bildung einer Subkultur. Die Hypothese H8b geht von einer Subkultur unter Mitarbeitenden mit dem gleichen Studienhintergrund aus. Schein (2010: 60ff.) weist auf das Vorhandensein einer Ingenieur-/Design-Subkultur in Organisationen hin und argumentiert, dass die Ingenieure gewisse Annahmen untereinander teilen, die sie von anderen Personen unterscheiden. Grund für diese Differenzen sind die gemeinsame Ausbildung, die gleichen gesammelten Erfahrungen bei der Arbeit, sowie identische Arbeitsanforderungen. Gemäss Schein (2010) sind Ingenieure und Mitarbeitende mit einem technischen Studienhintergrund der

Überzeugung, dass Lösungen auf Wissenschaft und verfügbaren Technologien basieren sollten.

Auch für die Studienrichtung der Naturwissenschaften lässt sich aufgrund der wissenschaftlichen Messbarkeit der Probleme auf eine positivere Einstellung gegenüber Forschungserkenntnissen schliessen. Der Beitrag von Frey (2012: 71), der die Verwendung von forschungsbasierten Wirkungsinformationen⁵³ untersucht, argumentiert, dass in Politikbereichen, die soziale Phänomene betreffen und primär durch die Sozialwissenschaften erfasst werden können, die Produktion und Nähe von wissenschaftsbasierten Wirkungsinformation geringer ausfällt, als in Politikbereichen mit einer grossen Nähe zu den Naturwissenschaften.

Abgeleitet aus den präsentierten Beiträgen sollten Mitarbeitende mit einem technischen, exakten oder naturwissenschaftlichen Studienhintergrund⁵⁴ aufgrund des besseren Zuganges zu wissenschaftlichen Erkenntnissen und der besseren Messbarkeit der Phänomene eine positivere Einstellung zur Evaluation aufweisen. Durch ihre Ausbildung verfügen sie über andere Erfahrungen und einen anderen Wissensstand im Umgang mit wissenschaftlichen Informationen, was sich in ihrer Einstellung zur Evaluation abbilden sollte. Dementsprechend formuliert sich die Hypothese folgendermassen:

H8b: Mitarbeitende, die ein Studium technischer, exakter oder naturwissenschaftlicher Richtung absolviert haben, verfügen im Vergleich zu den anderen Mitarbeitenden über eine positivere Einstellung zur Evaluation.

Der Studienhintergrund wird anhand der Selbsteinstufung der Mitarbeitenden operationalisiert.

⁵³ Auf den Beitrag von Frey (2012) wurde bereits im vorherigen Unterkapitel verwiesen. Weitere Informationen zu dieser Studie sind dort zu entnehmen.

⁵⁴ In der Online-Mitarbeitendenbefragung bilden die Exakten- und die Naturwissenschaften eine gemeinsame Kategorie. Aus diesem Grund werden die Mitarbeitenden mit einem Studienhintergrund der exakten Wissenschaften ebenfalls in der Hypothese berücksichtigt. Es ist aber anzunehmen, dass der Studienhintergrund der exakten Wissenschaften über eine bessere Messbarkeit der Probleme gemäss der geschilderten Argumentation von Frey (2012: 71) verfügt, im Vergleich zu den anderen Studienrichtungen.

Kontrollvariablen K

Zusätzlich werden folgende zwei sozioökonomische Variablen und eine geographische Variable zur Kontrolle in die Analyse miteinbezogen:

- Geschlecht, Alter und Kanton

Die Variablen Geschlecht und Alter gelten als robuste und oft verwendete Kontrollvariablen in der Untersuchung der Einstellung von Verwaltungsangestellten (für eine ausführliche Diskussion siehe Pandey/Stazyk 2008: 101ff.). Ausserdem werden die drei unterschiedlichen Kantone in der Analyse berücksichtigt.⁵⁵

In der nachfolgenden Tabelle 7 sind die Determinanten zusammenfassend dargestellt. Die genaue Operationalisierung der Merkmale ist in der Tabelle 49 im Anhang aufgeführt.

⁵⁵ Auf die Gründe des Einbezugs des Kantons in die Analyse wurde bereits beim vorherigen Modell eingegangen. Die Unterscheidung der Kantone berücksichtigt die vorherrschenden Eigenheiten des Schweizerischen Föderalismus. Der Föderalismus zeichnet sich aus durch wirtschaftliche, kulturelle und politische Eigenheiten der Kantone (Freitag 2005; Vatter 2002: 17f.) und auch in den unterschiedlichen Ausgestaltungen der kantonalen Verwaltungen (siehe Koller 2013) und wird aus diesem Grund in der Analyse berücksichtigt.

Tabelle 7: Zu untersuchende Determinanten der Einstellung zur Evaluation

Hypothese	Bezug
H4: Wenn die Evaluation in der Verwaltungseinheit institutionalisiert ist, dann ist die Einstellung zur Evaluation der Mitarbeitenden positiver.	Schein (2010), Varone (2007, 2015)
H5a: Je höher die Evaluationsaktivität in der Verwaltungseinheit ist, desto positiver ist die Einstellung zur Evaluation.	Allport (1954), Katz et al. (2002), Patton (1998), Schein (2010), Varone (2007, 2015) Zajonc (1968)
H5b: Je höher die Anteile der Evaluationsverwendung in der Verwaltungseinheit sind, desto positiver ist die Einstellung zur Evaluation.	Allport (1954), Katz et al. (2002), Patton (1998), Schein (2010), Varone (2007, 2015) Zajonc (1968)
H6: Wenn die Verwaltungsmitarbeitenden in der Politikformulierung oder in der Politikimplementierung tätig sind, dann verfügen sie im Vergleich zu Mitarbeitenden, die interne Dienstleistungen tätigen, über eine positivere Einstellung zur Evaluation.	Anderfuhren-Biget et al. (2013), Schwandt/Dahler-Larsen (2006)
H7a: Je weiter links sich die Mitarbeitenden in der Links-Rechts-Skala einstuft, desto positiver ist die Einstellung zur Evaluation.	Blank/Shaw (2015), De Vries et al. (2013), Freire (2006) Freire/Kivistik (2013), Kahan et al. (2012), McCright/Dunlap (2011)
H7b: Je traditionalistischer sich die Mitarbeitenden in der Modernismus-Traditionalismus-Skala einstuft, desto negativer ist die Einstellung zur Evaluation.	Schwandt/Dahler-Larsen (2006)
H8a: Je höher die Position der Mitarbeitenden in der Verwaltungshierarchie ist, desto positiver ist die Einstellung zur Evaluation.	Anderfuhren-Biget et al. (2013), Camilleri (2007), Moynihan/Pandey (2007), Schein (2010)
H8b: Mitarbeitende, die ein Studium technischer, exakter oder naturwissenschaftlicher Richtung absolviert haben, verfügen im Vergleich zu den anderen Mitarbeitenden über eine positivere Einstellung zur Evaluation.	Frey (2012), Schein (2010)
Kontrollvariablen: Geschlecht, Alter, Kanton	

Quelle: Eigene Darstellung

Wie die abhängige Variable Einstellung zur Evaluation konkret gemessen wird, wird im nächsten Kapitel ausführlich beschrieben.

B. Empirischer Teil

4 Messung der organisationalen Evaluationskultur

4.1 Operationalisierung der organisationalen Evaluationskultur

Nachdem im zweiten Kapitel die organisationale Evaluationskultur definiert wurde, behandelt dieses Kapitel, wie diese konkret gemessen werden kann. Dazu wird auf die drei Dimensionen – Institutionalisierung der Evaluation, Evaluationspraxis und die Einstellung zur Evaluation – eingegangen und diskutiert, anhand welcher Indikatoren diese Dimensionen gemessen werden können. Allerdings sollte immer bedacht werden, dass die in dieser Arbeit untersuchten Dimensionen und Indikatoren der Evaluationskultur nicht auf alle Organisationen gleich gut anzuwenden sind. Zweckmäßiger ist es, die zu untersuchenden Bestandteile der Evaluationskultur spezifisch auf die zu untersuchende Organisation anzupassen.⁵⁶

Institutionalisierung der Evaluation

Die erste Dimension der Evaluationskultur ist die Institutionalisierung der Evaluation. Hierzu gibt es bereits einige Beiträge: Jacob/Varone (2003, 2004) untersuchen den Stand der Institutionalisierung in verschiedenen Ländern. Dabei analysieren sie, ob sogenannte Evaluationsinstanzen in Parlament, Verwaltung und Rechnungshof zu finden sind. Ausserdem befassen sie sich mit der Regulierung und Steuerung der Evaluation, wie sie durch Standards, Gesellschaften oder Fachzeitschriften geschehen. Jacob (2005: 23) bemerkt in seiner Dissertation, in der er die Institutionalisierung der Evaluation in Belgien, Frankreich, Holland und der Schweiz vergleicht, dass Institutionen mit Regeln, Normen und Prozeduren das Handeln der Individuen beeinflussen. Auch Preskill/Boyle (2008: 454) verweisen in ihrem Artikel über die Evaluationskapazität auf „policies“ und Prozeduren, um die Evaluation in einer Organisation zu institutiona-

⁵⁶ Siehe dazu den Beitrag von Martin (2009: 205ff.).

lisieren. Dies könne, so die beiden Autoren, durch das Vorhandensein einer speziellen Evaluationseinheit oder eines Evaluationsteams geschehen. Auch Compton et al. 2002, Stufflebeam 2002, Valéry/Shakir 2005 und Volkov/King 2007 verweisen auf die Institutionalisierung der Evaluation durch die Einführung einer Evaluationsstelle. Preskill/Torres (1999) erwähnen die Implementierung von internen Prozeduren, die Evaluationen in den täglichen Arbeitsprozess einbetten, als eine Form der Institutionalisierung der Evaluation. Forss 2012, Jacob/Varone 2003, 2004, Läubli Loud 2014, Stufflebeam 2002 und Volkov/King 2007 untersuchen das Vorhandensein von Unterlagen, Evaluationsstandards oder Verbreitungsstrategien. Andere Studien betrachten, wie der Umgang mit der Evaluation in der Organisation vonstatten geht (Forss 2012; Speer 2012; Stufflebeam 2002; Widmer et al. 2001): Beispielsweise die Unterscheidung zwischen einer zentralisierten gegenüber einer dezentralisierten Organisation der Evaluationstätigkeit, die berücksichtigt, wie breit die Zuständigkeit für die Evaluation innerhalb der Organisation gestreut ist (siehe Forss 2012; Widmer et al. 2001). In der nachfolgenden Tabelle 8 sind die erwähnten Indikatoren zur Übersicht nochmals aufgeführt.

Tabelle 8: Indikatoren der Institutionalisierung der Evaluation

Indikatoren	Bezug
Evaluationsstelle/Evaluationsspezialisten	Compton et al. (2002), Stufflebeam (2002), Valéry/Shakir (2005), Volkov/King (2007)
Evaluationsunterlagen (Leitfäden, Vorgaben für Berichte, Methoden)	Forss (2012), Stufflebeam (2002)
Evaluationsverbreitungsstrategie	Läubli Loud (2014), Stufflebeam (2002)
Evaluationsstandards	Jacob/Varone (2003, 2004)
Formalisierte Evaluationsprozesse	Läubli Loud (2014), Preskill/Torres (1999), Stufflebeam (2002)
Zentralisierung/Dezentralisierung	Forss (2012), Widmer et al. (2001)

Quelle: Eigene Darstellung

Die vorliegende Untersuchung wird sich auf zwei Indikatoren der Institutionalisierung der Evaluation beschränken. Einerseits dient die Beschränkung dazu, die Analyse in einem überschaubaren Rahmen zu halten, andererseits wird damit der Stand der Institutionalisierung in den untersuchten neun Verwaltungseinheiten berücksichtigt. Diese werden weiter unten vorgestellt. Im nächsten Kapitel wird diesbezüglich aufgezeigt, dass die Berücksichtigung der beiden Indikatoren die Institutionalisierung in den De-

partementen und Direktionen genügend zu beschreiben vermag und aufgrund der geringen Verbreitung der Institutionalisierung nicht noch weitere Indikatoren berücksichtigt werden müssen.

Für die vorliegende Untersuchung wird die Institutionalisierung der Evaluation durch das Vorhandensein einer Evaluationsstelle und das Vorhandensein von spezifischen Unterlagen, betreffend Evaluation, berücksichtigt.

Inwiefern die Verwaltungseinheiten über eine Institutionalisierung der Evaluation verfügen, wurde mit der Onlinerecherche sowie über Auskünfte von Amtsvorstehern und den Generalsekretariaten in Erfahrung gebracht. Zusätzlich wurden die Mitarbeitenden zu diesen Indikatoren befragt.

Evaluationspraxis

Die Evaluationspraxis bildet die zweite Dimension der Evaluationskultur und soll aufzeigen, wie stark die Evaluation in den Arbeitsalltag der Mitarbeitenden Einzug gehalten hat. Damit orientiert sich diese Untersuchung an Kallemeyn et al. (2015: 341f.), die in Anlehnung an Shadish et al. (1991) und Stern (2006) folgendes Statement abgegeben haben: „When evaluators use the term evaluation practice, they typically refer to the everyday work of doing evaluation, such as dealing with stakeholders, developing an evaluation plan, collecting evidence, communicating findings, and so on“. Stern (2006: 293) verweist in seinem Buchkapitel, das die Evaluationspraxis von Evaluatoreninnen und Evaluatoren thematisiert, auf die Unterscheidung zwischen „Evaluationen zu machen“ und „Evaluationen zu studieren“. Auch der Beitrag von Cousins et al. (2008) unterscheidet bei der Konzeptualisierung der Evaluationskapazitätsbildung (ECB) zwischen der Evaluationsaktivität und der Evaluationsverwendung (englisch: do and use). Diese Einteilung findet sich auch in weiteren Beiträgen der ECB und bildet somit Kernelemente dieses Konzeptes ab (siehe Hueftle Stockdill et al. 2002: 8f.; Labin et al. 2012: 2; Nielsen et al. 2011; Stufflebeam 2001: 325). Die beiden Merkma-

le sind die offensichtlichsten Charakteristika der Evaluationspraxis. Konkret geht es bei der Untersuchung der Erarbeitung von Evaluationen um die Evaluationsaktivität, wie sie beispielsweise von (Balthasar 2007, 2015; Balthasar/Rieder 2009; Frey/Ledermann 2017; Zimmermann et al. 1989)⁵⁷ für Schweizer Verwaltungseinheiten untersucht wurde oder die sogenannte „evaluation regularity“, wie sie Pattyn (2015) analysierte. Die „evaluation regularity“ untersucht die Regelmässigkeit, in der Evaluationen durchgeführt werden und fokussiert auf die Frequenz der Evaluationspraxis (Pattyn 2015: 1477).⁵⁸

Für die Analyse der Evaluationsverwendung können drei verschiedene Gegenstände untersucht werden. Erstens die Betrachtung der vorhandenen Kapazität der Evaluationsnutzung, wie von Cousins et al. (2014) erwähnt. Diese beachtet die Fähigkeit, Nutzen aus den Evaluationsergebnissen und aus dem Evaluationsprozess zu ziehen. Zweitens die Betrachtung der Verwendung anhand der Rolle von Evaluationen im Entscheidungsprozess, wie sie von Owen (2003), Owen/McDonald (1999) oder Läubli Loud (2014) erwähnt wird. Hierbei geht es darum, dass Evaluationsresultate in den Entscheidungsprozess einfließen und auch berücksichtigt werden.⁵⁹ Die dritte mögliche Form der Untersuchung der Verwendung beinhaltet eine vertiefte inhaltliche Untersuchung der Art der Verwendung von Evaluationen gemäss Johnson et al. (2009: 378): „[...] of the application of evaluation processes, products, or findings to produce an effect“. Dieser fokussiert auf den stark erforschten Bereich und auf die oft angewendete Unterscheidung in instrumentale, konzeptionelle, symbolische Verwendung und die spätere Erweiterung der Prozessverwendung (Balthasar 2007: 23ff.; Henry/Mark 2003; Kirkhart 2000).

Bei der Diskussion um die relevanten Aspekte der Evaluationspraxis ist die Berücksichtigung der Debatte dienlich, die sich in der Forschung über die Evaluationskapazi-

⁵⁷ Diese Beiträge erheben die Anzahl der gemachten oder gefundenen Evaluationsberichte für eine spezifische Zeitperiode.

⁵⁸ Pattyn (2015) operationalisiert die „evaluation regularity“ mit der Anzahl im Durchschnitt durchgeführten Evaluationen der im öffentlichen Sektor tätigen Organisationen.

⁵⁹ Diese Form des Einbezugs von Evaluationen berücksichtigt die formative Einbindung der Evaluation in der Organisation und ist daher eher der Dimension der Institutionalisierung der Evaluation zuzuordnen.

tätsbildung finden lässt. In diesem Forschungsfeld sprechen sich einige Autoren für eine breitere Sichtweise der Evaluationskapazitätsbildung, respektive der Evaluationspraxis, über die Elemente „do and use“ hinaus aus (beispielsweise Julnes 2006; Mackay 2002; Mayne/Rist 2006; Nielsen/Ejler 2008; Rist 2006).⁶⁰ Mackay (2002) schlägt den Einbezug von evaluativem Wissen vor. Dementsprechend ist das Ziel der Evaluationskapazität nicht alleine die Erzielung qualitativ guter Programmevaluationen, sondern beinhaltet ein Engagement in einem breiteren Feld der Evaluationspraxis, welches evaluatives Wissen im ganzen Politikzyklus produziert und anwendet. Weitere Indikatoren der Evaluationspraxis sind für deren Messung denkbar. Beispielsweise ist die Abdeckung der evaluierten Programme und Aktivitäten einer Organisation, wie sie Speer (2012) untersucht hat, ein weiterer geeigneter Indikator für die Evaluationspraxis. Auch die Schulung der und das Wissen über Evaluation können als Elemente der Evaluationspraxis angesehen werden. Beispielsweise die Professionalisierung der Evaluatorinnen und Evaluatoren (siehe Jacob/Boisvert 2010; King et al. 2001; Taylor-Ritzler et al. 2013).

In der nachfolgenden Tabelle 9 sind die hergeleiteten Aspekte der Evaluationspraxis zur besseren Übersicht nochmals aufgeführt.

Tabelle 9: Indikatoren der Evaluationspraxis

Indikatoren	Bezug
Evaluationsaktivität	Balthasar (2007, 2015), Balthasar/Rieder (2009), Balthasar/Strotz (2017), Pattyn (2015), Zimmermann et al. (1989)
Evaluationsverwendung	Cousins et al. (2014), Owen (2003), Owen/McDonald (1999), Läubli Loud (2014), Balthasar (2007: 23ff.), Henry (2003), Kirkhart (2000)
Abdeckung	Speer (2012)
Evaluationswissen	Mackay (2002)
Professionalisierung	Jacob/Boisvert (2010), King et al. (2001), Taylor-Ritzler et al. (2013)

Quelle: Eigene Darstellung

Aus den Beobachtungen lässt sich zusammenfassen, dass die Evaluationsaktivität und

⁶⁰ Bei der Diskussion der Aspekte der Evaluationspraxis wird häufig Bezug auf die Evaluationskapazität genommen. Die beiden Konzepte stehen in enger Verbindung, weil die Evaluationskapazität eine nachhaltige Evaluationspraxis aufrechterhält oder (überhaupt) ermöglicht (siehe Hueftle Stockdill et al. 2002; Preskill/Boyle 2008: 444).

die Evaluationsverwendung die beiden wichtigsten Indikatoren der Evaluationspraxis sind, sich diese Liste jedoch erweitern lässt. Für eine geeignete Fokussierung wird sich die Analyse in dieser Untersuchung jedoch auf diese beschränken. Es kann davon ausgegangen werden, dass diese beiden Indikatoren die Evaluationspraxis in den Verwaltungen genügend zu beschreiben vermögen und gleichzeitig einer Fokussierung auf zentrale Elemente, für eine übersichtlichere Analyse, dienlich sind.

Für diese Untersuchung werden folgende Indikatoren der Evaluationspraxis berücksichtigt: *Die Evaluationsaktivität und die Evaluationsverwendung.*

Die Evaluationsaktivität gibt die Häufigkeit an, an der die Mitarbeitenden an der Erstellung an Evaluationen im Jahr im Durchschnitt beteiligt sind. Die Evaluationsverwendung gibt die Häufigkeit der Verwendung von Evaluationen bei der Arbeit an. Die Evaluationsaktivität und die Evaluationsverwendung der Mitarbeitenden werden anhand der Online-Befragung ermittelt. Ausserdem wird für den Vergleich der Departemente und Direktionen im fünften Kapitel die Evaluationsaktivität zusätzlich durch die Evaluationsrecherche, also der Anzahl an gefundenen Evaluationsberichten in den Verwaltungseinheiten, ergänzt.

Einstellung zur Evaluation und evaluative Wertvorstellungen

Die dritte Dimension der Evaluationskultur berücksichtigt die Einstellung der Mitarbeitenden zur Evaluation und evaluative Wertvorstellungen. Hierbei stellt sich die Frage der Messbarkeit besonders, da weder die Einstellung zur Evaluation, noch die evaluativen Wertvorstellungen an eindeutigen Kriterien beobachtbar sind und damit latente Messkonzepte darstellen (siehe Kopp/Lois 2012: 80f.; Kuckartz et al. 2010: 217).

Der Begriff evaluative Wertevorstellung (englisch: values of evaluation) findet bei der Untersuchung von Patton (1998: 226) Erwähnung. Dieser spricht von spezifischen Werten, die sich bei am Evaluationsprozess beteiligten Personen bilden, und um-

schreibt diese folgendermassen: „[...] clarity, specificity and focusing; being systematic and making assumptions explicit; operationalizing program concepts; ideas and goals; distinguishing inputs and processes from outcomes; valuing empirical evidence; and separating statements of fact from interpretations and judgements”. Forss (2012) untersucht die Emotionen Respekt, Angst, Misstrauen, Genuss, Innovation und Spass gegenüber der Evaluation und wie der Nutzen der Evaluation eingeschätzt wird. Auch Mihalache (2010) analysiert in seiner Untersuchung der Evaluationskultur zwölf Werte.⁶¹

Für die Ermittlung von evaluativen Wertvorstellungen muss vor der Befragung bestimmt werden, welche Charakteristika darunter zu verstehen sind. Beispielsweise steht eine spezifische oder eher eine systematische Vorgehensweise für Wertvorstellungen solcher Art oder sind gar andere Charakteristiken wichtiger? Um diese Schwierigkeit zu umgehen und um die Untersuchung in einem vernünftigen und durchführbaren Rahmen zu halten, wird nachfolgend lediglich die Einstellung gegenüber der Evaluation ermittelt. Dabei bestehen Einstellungen in Anlehnung an Ajzen (2001) aus Bewertungen gegenüber unterschiedlichen Objekten (beispielsweise Personen, Situationen, Objekten). Demzufolge bezieht sich die Einstellung zur Evaluation darauf, wie Mitarbeitende die Evaluation bewerten.

Die Messung der Einstellung zur Evaluation basiert auf den Daten einer Online-Mitarbeitendenbefragung. Für die Messung von Einstellungen mittels Befragung empfiehlt De Vaus (2014: 47) folgende Vorgehensweise: „Typically attitudes and opinions are complex and are best measured with a number of questions to capture the scope of the concept”. Gerade bei Messobjekten, die empirisch noch zu wenig getestet sind, wie es bei der Einstellung gegenüber Evaluationen der Fall ist, empfiehlt De Vaus, diese mit mehreren Fragen zu ermitteln. Die Studie von Brandon/Higa (2004) hat bereits in

⁶¹ Mihalache (2010) untersucht folgende zwölf Werte einer idealtypischen Evaluationskultur: 1. Aktionsorientiert, 2. Lernorientiert, 3. Inklusiv und partizipativ, 4. Reagierend und nichthierarchisch, 5. Innovations- und diversitätsorientiert, 6. Strikt wissenschaftlich, 7. Interdisziplinär, 8. Selbstkritisch, 9. Aufrichtig und objektiv, 10. Ethisch und demokratisch, 11. Vorwärts schauend, 12. Transparent.

Anlehnung an die Untersuchung von Cousins/Earl (1995) die Einstellung zur Evaluation gemessen. Mit fünf Fragen wurde untersucht, wie angenehm, wichtig und nützlich Evaluationen empfunden werden und wie die Evaluationshaltung und die Durchführung von Evaluationen eingestuft werden.⁶² Diese Untersuchung wird sich, mit einigen Anpassungen, an dem Fragenkatalog von Brandon/Higa (2004) orientieren.

De Vaus (2014: 95) verweist auf die Berücksichtigung von folgender Unterscheidung bei der Erhebung von Einstellungen in Befragungen: „It is important to distinguish the direction of a person’s attitude from both the extremity of their position and the intensity with which they hold that position”. Porst (2011: 93ff.) empfiehlt für die Erfragung von Einstellungen endpunkt-basierte Skalen (sogenannte Likert-Skalen) und ein Set von Aussagen (Item-Batterien). Tabelle 10 zeigt die fünf Fragen, mit denen die Einstellung zur Evaluation für diese Untersuchung gemessen wird. Diese Aussagen wurden mithilfe von Likert-Skalen gemessen und orientieren sich an den eben genannten Unterscheidungen: Die grundsätzliche Haltung wird mittels der Frage bezüglich der positiven oder negativen Einstellung erfragt (Frage F1). Mit der Zustimmungserfragung zu den Aussagen (F3 und F4) wird die Extremität der Einstellung ermittelt. Mit den Aussagen (F2 und F5) wird indirekt die Intensität ermittelt. Allerdings ist diese Unterscheidung nicht immer klar zu ziehen und die Fragen können beispielsweise sowohl die Position, als auch die Extremität beinhalten (siehe De Vaus 2014: 95).

⁶² Die Cronbachs Alpha-Werte zeigten (0.87 Pretest und 0.90) hohe Beziehungswerte zwischen den Fragen auf.

Tabelle 10: Fragen bezüglich der Einstellung zur Evaluation

Frage	Min	Max	Mittel	Schiefe	SA (Varianz)	N
F1 Haltung: Wie ist Ihre grundsätzliche Haltung gegenüber Evaluationen?	0	10	7.19	-0.84	1.76 (3.11)	397
F2 Ausweitung: Wie stark stimmen Sie der folgenden Aussage zu? Die Evaluationsaktivität und der Evaluationseinbezug in der Verwaltung sollte ausgeweitet werden.	0	10	5.50	-0.31	2.41 (5.80)	392
F3 Sinnvoll: Finden Sie die Durchführung oder den Einbezug von Evaluationen in ihren Arbeitsalltag sinnvoll?	0	10	6.70	-0.81	2.20 (4.84)	375
F4 Nutzen: Nützen Evaluationen (oder würden helfen) Ihrer Meinung nach, um bessere Produkte, Programme, Prozesse und Dienstleistungen zu entwickeln?	0	10	6.92	-0.91	1.98 (3.93)	378
F5 Wert: Wie stark stimmen Sie der folgenden Aussage zu? Das Geld und die Zeit, die in Evaluationen investiert wird, sind es wert?	0	10	6.11	-0.57	2.21 (4.89)	385

Anmerkungen: Genau genommen ist bei diesen Fragen ein ordinales Skalenniveau zutreffend. Aufgrund der hier vorhandenen Likert-Skala mit elf verschiedenen Abstufungen werden auch für diese Merkmale die Skala als intervallskaliert eingestuft (siehe dazu Urban/Mayerl 2011: 276). N=Anzahl Verwaltungsmitarbeitende.

Quelle: Eigene Darstellung

Die Antwortwerte der fünf Fragen werden addiert und bilden die Variable Einstellung zur Evaluation (siehe Tabelle 11).⁴³ Diese Vorgehensweise entspricht somit der Bildung einer Skala, wie sie in der Literatur häufig erwähnt wird (siehe Kopp/Lois 2012: 80ff.; Kuckartz et al. 2010: 215ff.).

Die Skala wird aus mehreren Items (Fragen) gebildet, die meist die Form von Aussagen haben und den Grad der Zustimmung erfragen. Daraus kann jedem einzelnen Befragten ein Skalenwert zugeordnet werden. Diese Skalierungsverfahren werden in den Sozialwissenschaften überwiegend eingesetzt, um Einstellungen zu messen (Kopp/Lois 2012: 80).

In Anlehnung an De Vaus (2014: 179f.) werden fünf Argumente für die Skalenbildung an dieser Stelle nochmals wiedergegeben. Für die Skalenbildung spricht: 1.) die Berücksichtigung der Komplexität: die Einstellung zur Evaluation kann durch die Berücksichtigung verschiedener Items gemessen werden und wird nicht nur durch ein

⁴³ Es wird keine Gewichtung und keine Transformation durchgeführt. De Vaus (2014: 192) erwähnt die Durchführung einer Z-Transformation bei der Bildung von Skalen aus Fragen mit unterschiedlicher Anzahl an Antwortkategorien und bei Fragen die sehr unterschiedliche Verteilungen aufweisen. Die Beschreibung der Einstellungsfragen in Tabelle 10 zeigt, dass alle Fragen die gleiche Antwortskala haben (Werte null bis zehn) und zudem eine sehr ähnliche Verteilung aufweisen (Mittelwerte zwischen 5.50 und 7.19 und Schiefe zwischen -0.91 und -0.31). Deshalb wird auf eine Gewichtung und eine Z-Transformation verzichtet.

Item beschrieben. 2.) Validere Messresultate durch mehrere Items: häufig ist eine Beobachtung mangelhaft und die Interpretation im Kontext von anderen Beobachtungen hilft die Gefahr von Fehlinterpretationen einzuschränken. 3.) Mehrere Items erhöhen die Reliabilität: Die Wortwahl der Fragen kann das Antwortverhalten der Befragten stark beeinflussen. Der Einbezug von mehreren Fragen minimiert die Stärke dieses Effektes. 4.) Mehrere Items erhöhen die Präzision: eine einzelne Frage erlaubt nur eine beschränkte Differenzierung zwischen den Befragten. Durch mehrere Fragen kann eine präzisere Rangfolge in der Einstellung zur Evaluation zwischen den verschiedenen Befragten gemacht werden. 5.) Durch die Summierung der Information der verschiedenen Items in eine Variable wird die Analyse stark vereinfacht.

Eine wichtige Kenngrösse für die Beurteilung von Skalen ist das Cronbachs Alpha. Dieses ist ein Mass für die interne Konsistenz und errechnet die durchschnittliche Korrelation zwischen den Items (Fragen). Nachfolgend ist die Formel aufgeführt. Hierbei ist n die Anzahl der Items und \bar{r} die mittlere Interkorrelation der Items.

$$\alpha = \frac{n \bar{r}}{1 + \bar{r} (n - 1)}^{64}$$

Cronbachs Alpha misst die Reliabilität einer Skala, indem bewertet wird, wie konsistent die Befragten geantwortet haben.⁶⁵ Es wird hierbei davon ausgegangen, dass jeder der fünf Fragen das gleiche messen soll. Demzufolge sollten die Befragten diese Fragen konsistent beantworten. Konkret wird angenommen, dass, wenn eine Person bei allen anderen Fragen eine positive Einstellung gegenüber Evaluationen hat, diese Person auch bei dieser Frage eine positive Einstellung gegenüber Evaluationen haben sollte. Die vorgestellte Skala wurde in einem Pretest⁶⁶ getestet und weist mit einem

⁶⁴ Formel aus Kopp/Lois (2012: 95).

⁶⁵ Die Reliabilität oder Zuverlässigkeit beschreibt, wie stark Messwiederholungen mit einem Messinstrument die gleichen Werte liefern (siehe (Kopp/Lois 2012: 95ff.).

⁶⁶ Für den Pretest wurden Mitarbeitende aus dem Departement Bildung, Kultur und Sport, Aargau sowie Personen aus dem Bildungsdepartement des Kantons St. Gallens angefragt. Ausgewählt wurden die beiden Verwaltungseinheiten, weil es sich dabei um Verwaltungen auf gleicher Ebene (Kantonsebene) handelt und eines der für diese Untersuchung relevanter Politikfelder umfasst (Bildung). Insgesamt nahmen, von den 68 angefragten Mitarbeitenden, 34 Personen an der Umfrage teil, was einer Rücklaufquote von 50.0 % entspricht.

Cronbachs Alpha von 0,846 mit allen fünf Einstellungsfragen die höchsten Werte auf. Damit können die fünf Fragen als reliabel angesehen werden und wurden für die Befragung übernommen.⁶⁷

Im Zusammenhang mit der Messung von Einstellungen ist das Problem von „Non-Attitudes“ zu berücksichtigen. Hierbei zeigt ein Anteil der Befragten keine stabile Disposition bei der Beantwortung der Fragen (siehe Best/McDermott 2007; Cobb/Nance 2013; Norpoth/Lodge 1985). Die Befragten werden bei einem Thema zu einer Antwort gezwungen, zu dem keine gefestigte Meinung existiert. In diesem Zusammenhang werden oft auch die Begriffe „pseudo-opinions“ oder „doorsteep opinions“ verwendet. Viele Studien untersuchen „Non-Attitudes“ von „Normalbürgern“ gegenüber spezifischen Themen: Beispielsweise fanden Cobb und Nance (2013) Effekte dieser Art bei Bürgern aus North Carolina in Bezug auf Themen des Freihandels. Auch fanden Best und McDermott (2007) Effekte in Bezug von US-Bürgern auf den „USA Patriot Act“. Um dieser Problematik genügend Rechnung zu tragen, wird folgendes Vorgehen gewählt. Es wird zwischen der Untersuchung mit allen Mitarbeitenden und der Analyse, die lediglich die Mitarbeitenden mit Evaluationsaktivität berücksichtigt, unterschieden. Diese Aufteilung stützt sich auf die Annahme, dass die Wahrscheinlichkeit für „Non-Attitudes“ bei Personen mit Evaluationsaktivität geringer ist und davon auszugehen ist, dass sich diese Stichprobe demzufolge besser als diejenige mit sämtlichen Mitarbeitenden eignet, um die Einstellung zur Evaluation zu messen. Aus diesem Grund wird folgendes Vorgehen gewählt: Für die Beschreibung der Einstellung zur Evaluation in den Departementen und Direktion im fünften Kapitel wird auf beide Grössen Bezug genommen. Während die Stichprobe mit den Evaluationsaktiven weniger anfällig für die „Non-Attitudes“ ist, bietet die Untersuchung mit allen Befragten den Vorteil, dass diese die Einstellung einer höheren Anzahl an Befragten abbildet. Für die Untersuchung der Zusammenhänge im sechsten Kapitel, in welchem

⁶⁷ Gemäss Bortz/Döring (2006: 199) sollte Cronbachs Alpha grösser als 0.8 sein. Werte über 0.9 weisen auf eine sehr hohe Reliabilität hin. Für De Vaus (2014: 84) ist ein Cronbachs Alpha Wert von mindestens 0.7 bereits eine reliable Skala.

keine Aussagen auf Departements- und Direktionsstufe gemacht werden, wird die Stichprobe mit den Evaluationsaktiven verwendet.

Tabelle 11 bildet die Kennwerte der Einstellung zur Evaluation ab. Das Cronbach Alpha weist mit 0.886 eine hohe Konsistenz auf. Der Mittelwert der Einstellung zur Evaluation zwischen dem Minimum null und Maximum 50 liegt bei 30.6. Ein hoher Einstellungswert steht für die Zuschreibung von Sinn und Nutzen der Evaluation und beinhaltet zudem eine positive Haltung gegenüber der Evaluation. Im Folgenden wird von einer positiven Einstellung zur Evaluation bei Werten über 34 und von einer negativen Einstellung bei Werten unter 17 gesprochen.⁶⁸ Dazwischen liegende Werte entsprechen einer neutralen Einstellung.⁶⁹

Tabelle 11: Deskription der Einstellung zur Evaluation

	Min	Max	Mittel	SK	SA	Varianz	Cronbachs Apha	N
Alle Mitarbeitende	0	50	30.6	-0.59	9.25	85.63	0.886	823
Personen mit Evaluationsaktivität	0	50	32.6	-0.85	8.92	79.57	0.894	369

Anmerkungen: Die Angaben der Mitarbeitenden mit Evaluationsaktivität berücksichtigt nur die Antworten der Mitarbeitenden, die angegeben haben, an mindestens einer Evaluation im Durchschnitt pro Jahr mitzuarbeiten. N=Anzahl Verwaltungsangestellte. SK=Schiefe. SA=Standardabweichung. Skala=Skalenniveau. Mittel=Mittelwert.

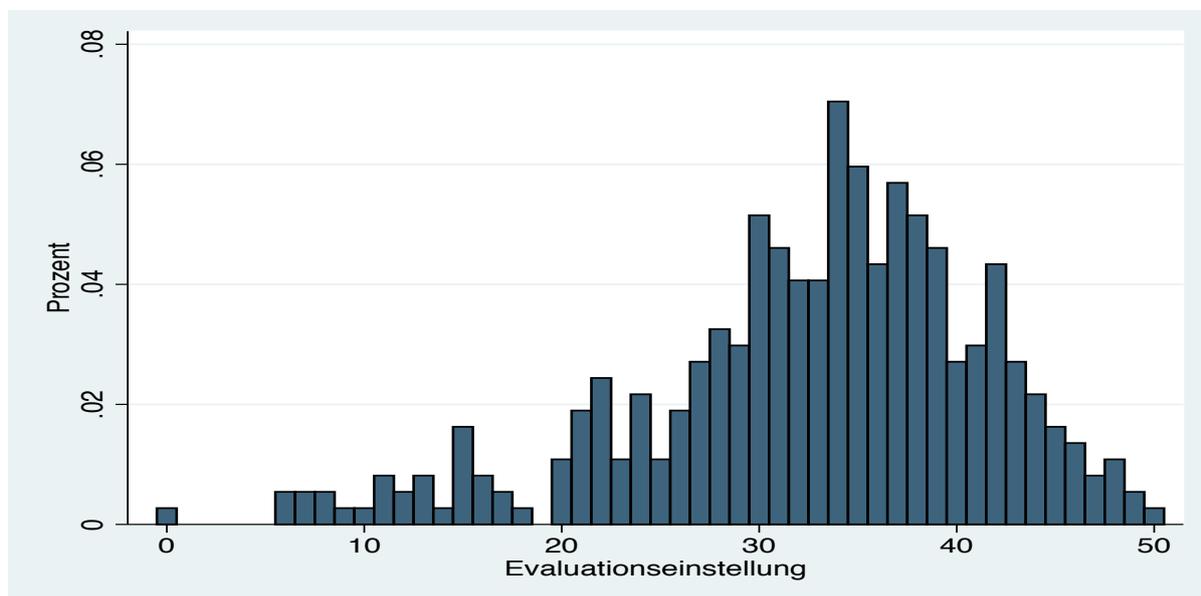
Quelle: Eigene Darstellung

Einen Überblick über die Verteilung der Skalenwerte illustriert die nachfolgende Abbildung 6 mit der prozentualen Häufigkeitsverteilung für jeden Wert. Am meisten Mitarbeitende (26 Befragte) mit Evaluationsaktivität haben einen Wert von 34. Ebenso ersichtlich ist die linksschiefe Verteilung, die auf eine mehrheitliche Einstellung von einem Wert über 25 hindeutet.

⁶⁸ Der Ausdruck der negativen oder der positiven Einstellung zur Evaluation ist nicht als Wertung zu verstehen. Inhalt dieser Untersuchung ist nicht die Analyse der Nützlichkeit oder der Wichtigkeit von positiver oder negativer Einstellung zur Evaluation.

⁶⁹ Die vorhandene Skala von null bis 50 wurde in möglichst drei gleiche grosse Wertebereiche eingeteilt. Zudem ermöglicht diese Aufteilung, aufgrund der Verteilung in den Departementen und Direktionen (ersichtlich in dem fünften Kapitel), die Verwaltungseinheiten in die verschiedenen Kategorien (negativ, neutral und positiv) einzustufen. Damit folgt die Aufteilung der Skalenwerte in gleich grosse Wertebereiche dem Vorgehen wie es De Vaus (2014: 163f.) unter dem Begriff „substantiv approach“ beschreibt.

Abbildung 6: Histogramm der Einstellung zur Evaluation



Quelle: Eigene Darstellung; N: 369

Einige Punkte sind im Gebrauch und bei der Interpretation von Skalen zu berücksichtigen. Bei der Reduktion auf den Gesamtskalenwert, wie dies im vorliegenden Fall gemacht wurde, sind keine Rückschlüsse zu den einzelnen Fragen möglich. Auch muss berücksichtigt werden, dass die gleichen Werte durch unterschiedliches Antwortverhalten zustande kommen können. Ein Wert von 25 kann einerseits durch konstante Wahl der Mitteposition (Wert 5) zustande kommen, andererseits führen jedoch auch andere Kombinationen von Antwortverhalten zu diesem Skalenwert. Zudem sind Skalenwerte eher relativ, als absolut zu interpretieren (siehe De Vaus 2014: 191ff.).

4.2 Einschränkungen

Nachdem die organisationale Evaluationskultur hergeleitet und definiert wurde, widmet sich das nächste Kapitel der konkreten Beschreibung der Evaluationskultur in den neun Departementen und Direktionen. Vorab wird in diesem Unterkapitel auf die Daten, die dieser Beschreibung zugrunde liegen, eingegangen.

Tabelle 12 zeigt, dass die organisationale Evaluationskultur mittels einer Online-Mitarbeitendenbefragung, der Recherche nach Evaluationsberichten sowie von den

Auskünften von Amtsvorstehern und den General- und Departementssekretariaten erhoben wird.

Tabelle 12: Datenquellen für die Deskription der Evaluationskultur

	Quellen
Organisation und Tätigkeitsbereich der Verwaltung	- Onlinerecherche sowie Auskünfte von Amtsvorstehern und den General- und Departementssekretariaten, statistischen Ämtern und Personalämtern
Institutionalisierung der Evaluation	- Onlinerecherche und Auskünfte von Amtsvorstehern und den General- und den Departementssekretariaten
	- Online-Mitarbeitendenbefragung
Evaluationspraxis	- Onlinerecherche der Evaluationsberichte für die neun Verwaltungseinheiten (sowie Bezug auf die Recherche von Dolder 2017)
	- Online-Mitarbeitendenbefragung
Einstellung zur Evaluation	- Online-Mitarbeitendenbefragung

Anmerkungen: Für zusätzliche Informationen über die Online-Mitarbeitendenbefragung siehe Abschnitt 1.3.2 in der Einleitung.

Quelle: Eigene Darstellung

Hierbei sei nochmals erwähnt, dass sich die Online-Mitarbeitendenbefragung lediglich auf Verwaltungseinheiten bezieht, deren Mehrheit der Mitarbeitenden keinen täglichen, individuellen Bürgerkontakt hat und mehrheitlich keine Unterhalts- oder Instandhaltungsarbeiten ausführt. Damit berücksichtigt diese Befragung nur jene Mitarbeitenden der Verwaltung, die potentiell mit Evaluationen in Kontakt kommen (siehe Abschnitt 1.3.2 in der Einleitung).

Die Beschreibung der Evaluationspraxis basiert auf zwei unterschiedlichen Quellen: der Evaluationsrecherche und der Mitarbeitendenbefragung. Für die Recherche der Evaluationspraxis wurde für den Zeitraum 2005 bis 2013 online nach Evaluationen geforscht, welche die Verwaltungseinheiten durchgeführt oder in Auftrag gegeben haben.⁷⁰ Als Evaluation gilt jeder Evaluationsbericht, wobei bei einer wiederholten Evaluation des gleichen Projektes die einzelnen Berichte als eigenständige Evaluation gezählt werden. Ebenso werden Zwischen- und Schlussberichte hinzugerechnet. Die Berücksichtigung aller Evaluationsberichte soll eine angemessene Abbildung der tatsäch-

⁷⁰ Unter einer Evaluation wird die in der Einleitung formulierte Definition verwendet. Wichtig ist dabei der bewertende Charakter der Studie oder des Berichtes. Massgebend für die Einteilung der Jahre ist das Publikationsdatum. Die Recherche nach Evaluationsberichten wurde mit Olivier Dolder (2018) abgeglichen, da es Überschneidungen von gemeinsam untersuchten Verwaltungseinheiten gab. Allerdings kann es zu geringen Abweichungen der Anzahl erfasster Evaluationsberichte kommen, da Dolder (2018) die Definition restriktiver anwendet.

lichen Evaluationstätigkeit einer Verwaltungseinheit sicherstellen. Bei mehreren Jahren dauernden Projekten würde die alleinige Darstellung des Schlussberichtes die Evaluationstätigkeit der vorherigen Jahre nicht adäquat abbilden. Für die Evaluationsrecherche wurden einerseits die Internetplattformen und -archive (Webseiten) der betreffenden Verwaltungseinheiten nach Evaluationsberichten durchsucht, andererseits wurde mittels der Suche bei „Google“ nach entsprechenden Stichworten (beispielsweise Evaluation oder Wirkungsanalyse), verbunden mit den Namen der jeweiligen Verwaltungseinheiten, im Internet nach Evaluationsberichten gesucht. Die Recherche erhebt somit keinen Anspruch auf Vollständigkeit, soll aber als Anhaltspunkt für die tatsächliche Evaluationstätigkeit dienen. Separat für die Analyse berücksichtigt werden die Schulevaluationen. Diese sind von den Politikevaluationen zu unterscheiden. Schulevaluationen werten die Qualität von Schulen aus und sind stark auf die Qualitätssicherung von Schulen ausgerichtet. Eine undifferenzierte Berücksichtigung dieser Berichte würde zudem die Analyse verzerren.⁷¹ In diesem Zusammenhang gilt es zu beachten, dass bei den Antworten der Online-Mitarbeitendenbefragung teilweise Angaben bezüglich der Schulevaluationen gemacht worden sind. Bei der Umfrage wurde zwar die Definition von Politikevaluation angeführt, die Schulevaluation wurde aber nicht explizit als „keine Politikevaluation“ definiert. Weiter ist zu berücksichtigen, dass die Online-Mitarbeitendenbefragung und die Onlinerecherche jeweils unterschiedliche Charakteristiken wiedergeben. Die Befragung basiert auf der Selbsteinschätzung der individuellen Evaluationsaktivität (im Jahresdurchschnitt) und dient dazu die Anteile der Mitarbeitenden mit Evaluationsaktivität pro Verwaltungseinheit darzustellen, während die Evaluationsrecherche die gesamte Anzahl in Auftrag gegebener oder durchgeführter Evaluationen pro Verwaltungseinheit abbildet. Deshalb ist es möglich, dass aufgrund einer starken Zentralisierung der Evaluationsaktivität, eine hohe Anzahl an Evaluationsberichten einem geringen Anteil an Mitarbeitenden mit

⁷¹ In Luzern beispielsweise wurden für den Zeitraum 2005 bis 2013 insgesamt 339 Schulevaluationen durchgeführt. Würde die Schulevaluation und die Politikevaluation undifferenziert berücksichtigt ergäbe es eine sehr hohe Evaluationspraxis.

Evaluationsaktivität gegenübersteht. Aus diesem Grund eignet sich die Evaluationsrecherche nur bedingt für die Validierung der Befragungsdaten.

Für die Beschreibung der Institutionalisierung wird ebenso auf diese zwei Quellen zurückgegriffen: Die Befragungsdaten der Mitarbeitenden (subjektive Wahrnehmung der Institutionalisierung) wird den recherchierten Angaben von Amtsvorstehern, Generalsekretariaten und Onlineinformationen gegenübergestellt. Mittels des Vergleiches dieser beiden Quellen kann die Wahrnehmung der Mitarbeitenden mit der tatsächlich vorhandenen Institutionalisierung der Evaluation verglichen werden.

Beide Quellen sind allerdings mit Einschränkungen verbunden. Die Befragungsdaten basieren auf den Angaben der Mitarbeitenden, weshalb Verzerrungen sowie inkonsistente Antworten vorliegen können (für einen Überblick siehe Bethlehem 2010). Beispielsweise bezeichnet die Methodenverzerrung (englisch: common-method bias) eine Verfälschung der Messergebnisse, die dadurch entsteht, dass die Befragten die Quelle für die exogene und die endogenen Variablen sind. Damit könnten die Befragten aufgrund des Fragebogens Schlussfolgerungen auf die Hypothesen ziehen und ihr Antwortverhalten entsprechend anpassen. Konkret wird angenommen, dass die Tendenz der Befragten, während einer Befragung konsistent zu antworten, steigt. Allerdings ist umstritten, ob eine solche Verzerrung überhaupt existiert (siehe Spector 2006). Weitere Probleme sind die Selbstselektion (englisch: self-selection-bias) und die Möglichkeit von Falschangaben (englisch: misreporting), die auf soziale Erwünschtheit zurückzuführen ist. Eine Selbstselektion kann zu verzerrten Ergebnissen führen, wenn gewisse Personen häufiger an der Befragung teilnehmen als andere: „Self-selection means that it is completely left to individuals to select themselves for the survey (Bethlehem 2010: 162)“. Bei der vorliegenden Untersuchung ist eine überdurchschnittliche Teilnahme von Personen mit einer höheren Evaluationsaktivität oder höherem Evaluationsinteresse möglich. Bei einer Verzerrung aufgrund sozialer Erwünschtheit orientieren sich die Antworten, aus Furcht vor sozialer Verurteilung, an gängigen Normen und Erwartungen (siehe Bortz/Döring 2006: 232ff.). Für die hier verwendete Befragung ist eine Verzerrung dieser Art denkbar, insofern die Verwaltungsangestellten Falschanga-

ben bezüglich ihrer Evaluationspraxis und bezüglich ihrer Einstellung zur Evaluation aufgrund von sozialer Erwünschtheit machen könnten. In diesem Zusammenhang konnten Bundi et al. (2016) Falschangaben betreffend der Evaluation bei der Untersuchung der Befragungsangaben von Schweizer Parlamentsangehörigen nachweisen und zeigen, dass die Einstellung zur Evaluation hierbei eine wichtige Rolle spielt. Politiker mit einer positiveren Einstellung zur Evaluation übertrieben bei den Angaben über ihre Evaluationsnachfrage, während Politiker mit einer negativeren Einstellung zur Evaluation ihre Nachfrage untertrieben haben. Diese Befunde sind allerdings nur beschränkt auf die vorliegende Untersuchung übertragbar, da Parlamentsangehörige nicht direkt mit Verwaltungsmitarbeitenden verglichen werden können. Bundi et al. (2016: 3) erwähnen die Spezifität von Parlamentsbefragungen, die folgendermassen betrachtet werden kann: „As a kind of elite surveys, they are shaped by special circumstances“. Zudem wurden bei der hier zur Anwendung kommenden Online-Befragung die Verwaltungsmitarbeitenden ausdrücklich darauf hingewiesen, dass die Befragung anonym erfolgt. Auch bei der Evaluationsrecherche ist auf Einschränkungen hinzuweisen. Beispielsweise besteht die Möglichkeit, dass nicht alle Evaluationsberichte erfasst wurden. So sind zum Beispiel nicht alle Berichte öffentlich zugänglich beziehungsweise nicht in den Onlinearchiven und -listen der Verwaltungen aufgeführt. Auch ist anzunehmen, dass vielmehr jüngere Evaluationen in den Online-Archiven zu finden sind und ältere gelöscht oder nicht aufgenommen wurden.

Den genannten Verzerrungen wird mit folgendem Vorgehen begegnet: Einerseits werden die Selbstangaben bezüglich der Evaluationsaktivität den erfassten Evaluationen der entsprechenden Verwaltungen gegenübergestellt. Einige offensichtliche Falschangaben können mit diesem Vorgehen identifiziert und korrigiert werden.⁷² Andererseits wird im nächsten Kapitel auf die Möglichkeit der Selbstselektion eingegangen.⁷³ Des

⁷² Die Gegenüberstellung erlaubte die Korrektur von Falschangaben lediglich in Verwaltungseinheiten, in welchen keine Evaluationsaktivität ausgemacht werden können und die Mitarbeitenden trotzdem angeben, über eine Evaluationsaktivität zu verfügen.

⁷³ Die Gegenüberstellung der Rücklaufquote mit der Evaluationsaktivität und der Einstellung zur Evaluation zeigt keine Hinweise in Richtung einer Selbstselektion (siehe dazu das Unterkapitel 5.4).

Weiteren wird anhand der nachfolgenden „Non-Response-Analyse“ beurteilt, ob Anzeichen einer Selbstselektion ausgemacht werden können.

Wie erwähnt besteht die Möglichkeit, dass Personen mit Evaluationserfahrung verglichen mit Personen ohne Evaluationserfahrung eine höhere Teilnahmemotivation haben (Selbstselektion). Um eine bessere Einschätzung der Befragungsdaten zu erhalten, wurden im Nachgang zur Befragung 13 Mitarbeitende, die nicht teilgenommen hatten, zufällig ausgewählt und telefonisch befragt. Acht dieser Personen gaben Auskunft über ihre Gründe zur Nichtteilnahme sowie über deren Einstellung zur Evaluation und deren Evaluationspraxis. Die nachfolgende Tabelle 13 zeigt die Angaben dieser Mitarbeitenden, und für den Vergleich die Angaben der Teilnehmenden, auf.

Tabelle 13: „Non-Response“-Analyse

	Teilnehmende	Nichtteilnehmende
Anteile mit Evaluationsaktivität	38.3	37.5
Anteile mit Evaluationsverwendung	60.3	71.4
Einstellung zur Evaluation (Mittelwert)	30.6	30.4

Anmerkungen: Die Einstellung zur Evaluation ergibt sich aus der Summierung von fünf Fragen und umfasst den Wertebereich null bis 50. Ein hoher Wert entspricht einer positiveren Einstellung gegenüber der Evaluation. Evaluationsaktivität = Prozentualer Anteil an Mitarbeitenden mit Evaluationsaktivität (mindestens an einer Evaluation im Durchschnitt im Jahr beteiligt). Evaluationsverwendung = Prozentualer Anteil an Mitarbeitenden mit Evaluationsverwendung (mindestens seltener Miteinbezug von Evaluationen in die Arbeit).

Quelle: Eigene Darstellung

Der Vergleich zeigt, dass die Angaben der acht Personen, die nicht teilgenommen hatten, den Angaben der Teilnehmenden stark gleichen. Die Anteile der Personen mit Evaluationsaktivität und die Anteile jener Personen mit Evaluationsverwendung sind sich ähnlich. Die Einstellungswerte zur Evaluation sind beinahe identisch. Damit lassen sich keine Anhaltspunkte für eine Selbstselektion aufgrund von unterschiedlichen Evaluationserfahrungen in Form von Selbstselektion erkennen. Die Gründe für die Nichtteilnahme weisen ebenfalls nicht in Richtung einer starken Selbstselektion hin; lediglich zwei Personen geben an, aufgrund von mangelndem Interesse am Thema Evaluation nicht an der Befragung teilgenommen zu haben. Zwei Personen haben die Teilnahme übersehen beziehungsweise vergessen und drei haben aufgrund von Zeitmangel nicht an der Befragung teilgenommen.

5 Welche Unterschiede bezüglich der Evaluationskultur lassen sich in den neun Departementen und Direktionen feststellen?

In diesem Kapitel wird die hergeleitete Unterteilung der Evaluationskultur in drei Dimensionen empirisch angewendet, indem für die fünf Departemente (Kanton Basel-Stadt und Kanton Luzern) und vier Direktionen (Kanton Bern) die Evaluationsinstitutionalisierung, die Evaluationspraxis und die Einstellung zur Evaluation beschrieben werden. Eine kurze Vorstellung der Verwaltungen soll einen Überblick über deren Tätigkeitsbereich, den Personalbestand sowie den Organisationsaufbau geben. Am Ende des Kapitels wird der Stand der Evaluationskultur in den Direktionen und Departementen nochmals zusammenfassend dargestellt.

5.1 Luzern

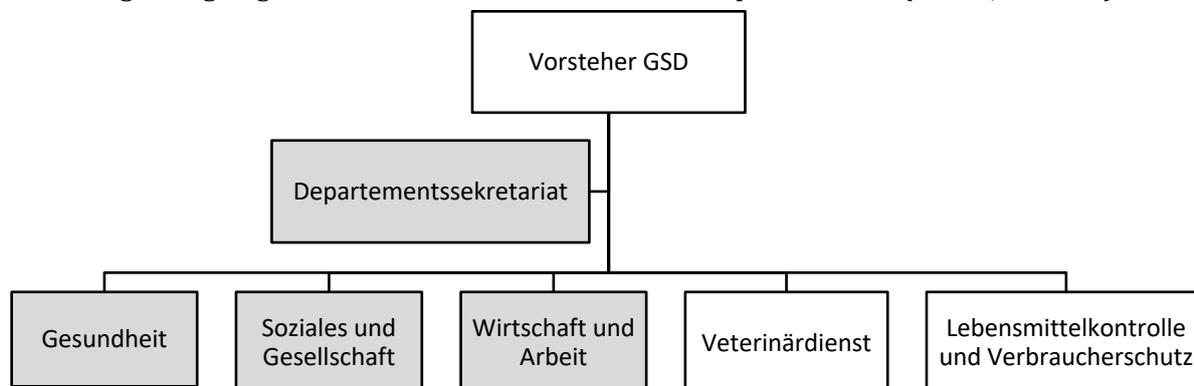
5.1.1 Gesundheits- und Sozialdepartement

Der Aufgabenbereich des Gesundheits- und Sozialdepartementes (GSD) des Kantons Luzern umfasst neben den Themenbereichen Gesundheit und Soziales, das Veterinärwesen, die Lebensmittelkontrolle, den Verbraucherschutz, die Heilmittel-, Betäubungsmittel- und Chemikalienkontrolle sowie Gesellschaftsthemen in den Bereichen Jugend, Alter, Familie, Gleichstellung und Integration. Ebenso ist dieses Departement verantwortlich für das Asylwesen und den Arbeitsmarkt (siehe GSD 2016a).

Das GSD setzt sich zusammen aus dem Departementssekretariat sowie den folgenden fünf Einheiten: Dienststelle Gesundheit, Dienststelle Soziales und Gesellschaft, Dienststelle Wirtschaft und Arbeit, Dienststelle Veterinärdienst und Dienststelle Lebensmittelkontrolle und Verbraucherschutz (siehe nachfolgende Abbildung 7).⁷⁴

⁷⁴ In den Luzerner Kantonsverwaltungen finden sich zwei Hierarchiestufen: Jedes Departement lässt sich in Dienststellen (Ebene 1) aufteilen, die wiederum in Bereiche (Ebene 2) unterteilt werden.

Abbildung 7: Organigramm des Gesundheits- und Sozialdepartementes (Stand Juli 2014)



Anmerkungen: Die grau schattierten Dienststellen sind Gegenstand der Untersuchung.

Quelle: Eigene Darstellung in Anlehnung an GSD (2014)

Aufgrund der oben erwähnten Kriterien (kein täglicher, individueller Bürgerkontakt und keine Unterhalts- oder Instandhaltungsarbeiten) beschränkt sich die Untersuchung auf die in Abbildung 7 grau schattierten Dienststellen. Zudem wurde nur jener Bereich der Dienststelle Wirtschaft und Arbeit berücksichtigt, der nicht in direktem Kontakt mit Bürgerinnen und Bürgern steht.

Die nachfolgende Tabelle 14 zeigt die Charakteristiken der Online-Befragung im GSD Luzern auf: Von den insgesamt 377 Mitarbeitenden wurden schliesslich 170 angefragt, wovon 65 Mitarbeitende den Fragebogen ausfüllen. Die folgende Tabelle 14 zeigt die Anzahl angefragter Personen sowie die Anzahl der Personen, die geantwortet haben. Die Rücklaufquote des GSD fällt mit 38.2 % – im Vergleich zum Rücklauf in den anderen untersuchten Departementen – deutlich tiefer aus.

Tabelle 14: Teilnehmende an der Online-Befragung

Dienststelle	Mitarbeitende	Angefragt	Antworten	Rücklauf (in Prozent)
Departementssekretariat	14	14	7	50.0
Dienststelle Gesundheit*	38	43	19	44.2
Lebensmittelkontrolle und Verbraucherschutz	31	-	-	-
Veterinärdienst	34	-	-	-
Soziales und Gesellschaft*	44	54	14	25.9
Wirtschaft und Arbeit	216	59	11	18.6
Keine Angaben	-	-	14	-
Total	377	170	65	38.2

Anmerkungen: Die Daten bezüglich der Anzahl an Mitarbeitenden stammen von der Website des Gesundheits- und Sozialdepartementes (GSD 2016b) und geben den Gesamtpersonalbestand per 1.7.2015 wieder. Nicht einbezogen ist das Personal des Kantonsspitals, der Psychiatrie, der Ausgleichskasse und der IV-Stelle.

*: Die Diskrepanz zwischen der Anzahl Mitarbeitenden und der Anzahl angefragter Personen ergibt sich aufgrund unterschiedlicher Quellen. Die Anzahl an Mitarbeitenden in den Dienststellen ist der kantonalen Webseite entnommen (GSD 2016b). Angefragt wurden die Mitarbeitenden aufgrund eines E-Mail Verteilers, der vom Departementssekretariat zur Verfügung gestellt wurde.

Quelle: Eigene Darstellung

Institutionalisierung der Evaluation

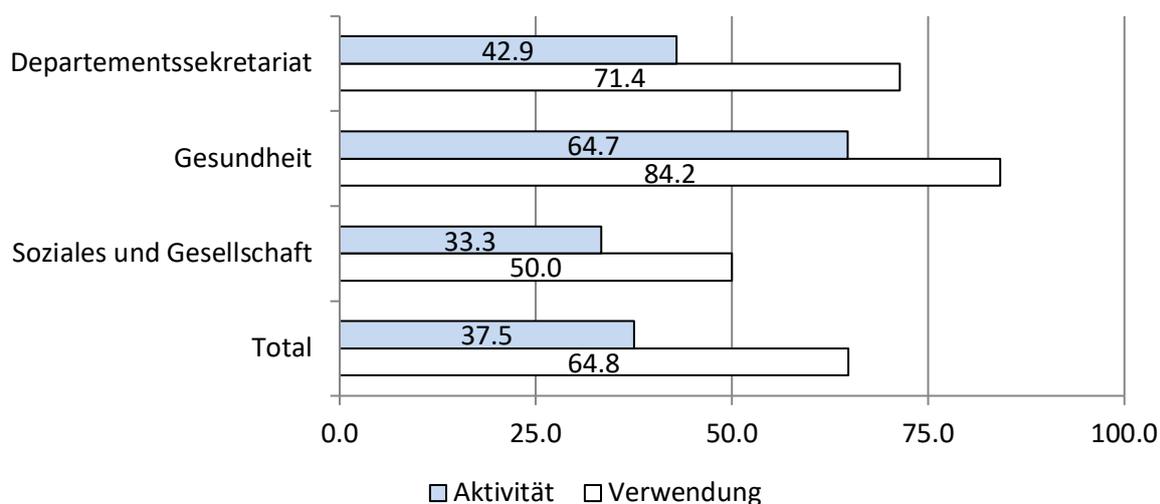
Die Evaluation ist im GSD Luzern nicht institutionalisiert. Es gibt im GSD weder eine Evaluationsstelle noch spezifische Grundlegendokumente betreffend Evaluation.

Die Gegenüberstellung der Einschätzung der Mitarbeitenden zeigt, dass sechs der 61 Mitarbeitenden, die diese Frage beantwortet haben, angeben, dass in ihrem Departement eine Evaluationsstelle bestehe. Diese Diskrepanz lässt sich durch das Dienstleistungszentrum Arbeitsmarktliche Massnahmen erklären, die Metaevaluations durchführt. Allerdings handelt es sich bei diesen Evaluationen nicht um eine wissenschaftliche Bewertung und demzufolge auch um keine Institutionalisierung in der hier definierten Form (siehe GSD WIRA DLZ 2016).

Evaluationspraxis

In unserer Online-Befragung haben wir ebenfalls die Mitarbeitenden nach der Evaluationspraxis gefragt. Ermittelt wurden die Häufigkeit der Beteiligung an Evaluationsaktivitäten und die Häufigkeit der Verwendung von Evaluationen. Die nachfolgende Abbildung 8 fasst die Ergebnisse zusammen.

Abbildung 8: Evaluationspraxis im GSD Luzern



Anmerkungen: Aktivität = Prozentualer Anteil an Mitarbeitende mit Evaluationsaktivität (mindestens an einer Evaluation im Durchschnitt im Jahr beteiligt). Verwendung = Prozentualer Anteil an Mitarbeitende mit Evaluationsverwendung (mindestens seltener Miteinbezug von Evaluationen in die Arbeit). Total Antwortende: Aktivität = 48; Verwendung = 54.

Quelle: Eigene Darstellung

Betrachtet man die Selbsteinstufung der Evaluationsaktivität und der Evaluationsverwendung, lassen sich für die Dienststelle Gesundheit die höchsten und für das Departementssekretariat die zweithöchsten Anteile nachweisen. Die Angaben der Mitarbeitenden der Dienststelle Wirtschaft und Arbeit können nicht berücksichtigt werden und sind deshalb nicht in der Abbildung aufgeführt. Der Grund sind die soeben erwähnten Metaevaluationen, die in der Dienststelle Arbeitsmarktliche Massnahmen durchgeführt werden und zu Falschangaben der Mitarbeitenden führten.⁷⁵

Aufgrund der Evaluationsrecherche können für den Zeitraum 2005 bis 2013 der Dienststelle Gesundheit elf Evaluationsberichte zugeordnet werden. Für die Dienststelle Soziales und Gesellschaft finden sich zwölf Evaluationsberichte. Dem Departementssekretariat kann keine Evaluation direkt zugeordnet werden. Allerdings ist denkbar, dass einige Mitarbeitende des Departementssekretariats dennoch an der Erarbeitung von Evaluationen beteiligt waren, da dieses auch Koordinationsaufgaben wahrnimmt (GSD DS 2016).

Mit total 24 Evaluationen⁷⁶ verfügt die GSD Luzern über eine mittlere Evaluationsaktivität und ist mit einem Anteil von 37.5 % der Mitarbeitenden mit Evaluationsaktivität im unteren Mittelfeld anzusiedeln.

Einstellung zur Evaluation

Die Variable Einstellung zur Evaluation drückt die Grundstimmung gegenüber Evaluationen in einer Dienststelle aus. Ein hoher Wert steht für die Zuschreibung einer hohen Nützlichkeit von Evaluationen und für die Tatsache, dass dem Thema Evaluation eine hohe Wichtigkeit zugesprochen wird. Die genaue Ermittlung und Interpretation des Wertes wurde im vierten Kapitel erläutert.

Um die Problematik von sogenannten „non-attitudes“ zu berücksichtigen (siehe die ausführliche Diskussion im vierten Kapitel) werden die Mittelwerte von Mitarbeiten-

⁷⁵ Bei den Metaevaluationen handelt es sich nicht um Evaluationen in der hier definierten Form. Anhand der Rückfragen ist davon auszugehen, dass eine Person (entspricht 11.1 %) in dieser Dienststelle über tatsächliche Evaluationsaktivität verfügt.

⁷⁶ Eine Evaluation konnte zwei Dienststellen zugeordnet werden.

den mit und ohne Evaluationsaktivität angegeben. Für die Einstufung der Einstellungsmittelwerte der Direktionen und Departemente in die Bereiche negativ, neutral und positiv⁷⁷ werden im Folgenden die Werte der evaluationsaktiven Mitarbeitenden („non-attitude“-bereinigte Stichprobe) Berücksichtigung finden.

Die nachfolgende Tabelle 15 listet diese Einstellungswerte auf und zeigt für die GSD Luzern insgesamt eine positive Einstellung zur Evaluation. Ausserdem sind Personen mit Evaluationsaktivität statistisch nachweisbar positiver gegenüber der Evaluation eingestellt.

Tabelle 15: Einstellung zur Evaluation im GSD Luzern

	Mittelwert	Min	Max	SA	N
Personen mit Evaluationsaktivität	34.0	16	44	9.8	18
Personen ohne Evaluationsaktivität	24.7	9	41	8.8	30
T-Test: $t(46) = -3.823^{**}$					
Alle befragte Personen	27.7	6	44	9.8	56

Anmerkung: Die Einstellung zur Evaluation ergibt sich aus der Summierung von fünf Fragen und umfasst den Wertebereich 0 (minimale Einstellung zur Evaluation) bis 50 (maximale Einstellung zur Evaluation). Ein hoher Wert entspricht einer positiveren Einstellung gegenüber der Evaluation. T-Test zum Vergleich der Mittelwerte.

+ = $p < 0.10$; * = $p < 0,05$; ** = $p < 0,01$; 2-seitig.

Quelle: Eigene Darstellung

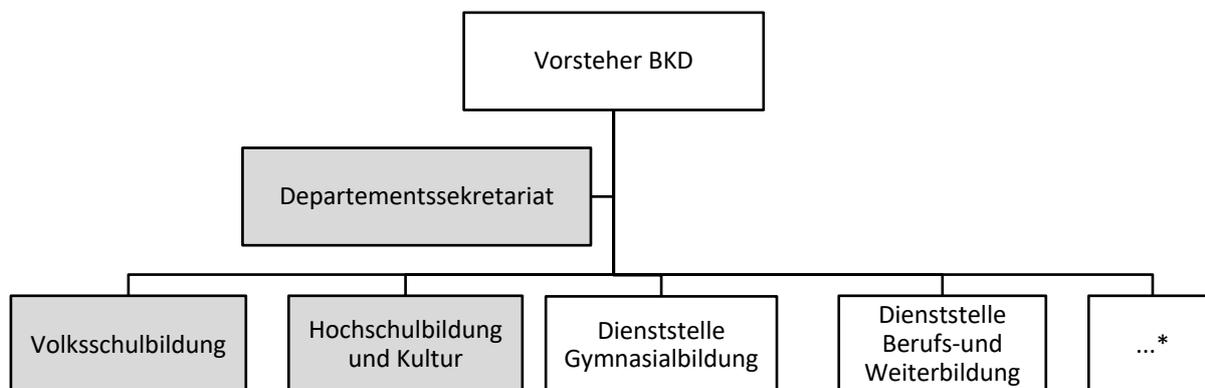
5.1.2 Bildungs- und Kulturdepartement

Zum Zuständigkeitsbereich des Bildungs- und Kulturdepartementes (BKD) des Kantons Luzern gehören, nebst der Verwaltung von Bildung und Kultur, zwei Heilpädagogische Zentren, sechs Berufsbildungszentren und acht Gymnasien. Auch die Zentral- und Hochschulbibliothek, die Hochschule Luzern, die Pädagogische Hochschule Luzern, die Universität Luzern, zwei Museen sowie die Archäologie und die Denkmalpflege sind dem BKD Luzern zugeordnet (BKD 2016).

Das BKD Luzern ist in vier verschiedene Einheiten und ein Departementssekretariat gegliedert: Die Dienststelle Volksschulbildung, die Dienststelle Gymnasialbildung, Berufs- und Weiterbildung sowie die Dienststelle Hochschulbildung und Kultur (siehe Abbildung 9).

⁷⁷ Im vierten Kapitel wurden die Einstellungswerte in folgende Bereiche aufgeteilt: Werte bis 17.0 sind negative Einstellungswerte; Werte ab 34.0 sind positive Einstellungswerte; Werte dazwischen sind neutrale Einstellungswerte.

Abbildung 9: Organigramm des Bildungs- und Kulturdepartementes (Stand Juli 2014)



Anmerkungen: Die grau schattierten Dienststellen sind Gegenstand der Untersuchung.

*: Selbstständige Anstalten (beispielsweise die Universität Luzern). Diese sind nicht Gegenstand dieser Untersuchung.

Quelle: Eigene Darstellung in Anlehnung an BKD (2014)

Von den insgesamt 3057 Mitarbeitenden des Bildungs- und Kulturdepartementes wurden Mitarbeitende aus der Dienststelle Volksschulbildung und der Dienststelle Hochschulbildung und Kultur, sowie dem Departementssekretariat befragt.⁷⁸ Die Zahl der angefragten Personen reduziert sich auf 105, wovon schliesslich 60 Personen die Fragen beantworteten. Die Höhe der Rücklaufquote des BKD Luzern ist mit 57.1 % im Vergleich mit dem Rücklauf aus den anderen Departementen im oberen Mittelfeld anzusiedeln.

Nicht befragt wurden die Dienststelle Gymnasialbildung und die Dienststelle Berufs- und Weiterbildung. Die Dienststelle Berufs- und Weiterbildung verfügt mehrheitlich über individuellen Bürgerkontakt, wie etwa die Berufs-, Studien- und Laufbahnberatung und das Stipendienwesen, und wurde deshalb von der Befragung ausgeschlossen. Die Dienststelle Gymnasialbildung besteht vorwiegend aus den Schulen (Gymnasien)

⁷⁸ Die Gesamtzahl von 3057 Mitarbeitenden umfasst auch Lehrpersonen und Mitarbeitende der selbstständigen Anstalten, beispielsweise Museums- oder Bibliotheksangestellte, die nicht befragt wurden und die erhebliche Reduktion der Befragten erklären.

und den zugehörigen Rektoraten, weshalb sie ebenfalls von der Befragung ausgeschlossen wurde.

Tabelle 16: Teilnehmende an der Online-Befragung

Dienststelle	Mitarbeitende	Angefragt*	Antworten	Rücklauf (in Prozent)
Departementssekretariat	20	-	15	-
Volksschulbildung	897	-	34	-
Hochschulbildung und Kultur	175	-	3	-
Keine Angaben		-	8	-
Total	3057	105	60	57.1

Anmerkungen: Die Daten bezüglich der Anzahl an Mitarbeitenden stammen von dem statistischen Amt des Kantons Luzern. Diese wurden auf Anfrage bekanntgegeben und geben den Personalbestand des gesamten Bildungs- und Kulturdepartementes per Dezember 2014 wieder. Diese Zahlen schliessen viele Mitarbeitenden ein, beispielsweise Lehrpersonen, die nicht befragt wurden.

*: Die Anzahl angefragter Mitarbeitenden in den einzelnen Dienststellen kann nicht nachvollzogen werden, da die erhaltenen Mail-Adressen nicht den einzelnen Dienststellen zugeordnet werden können.

Quelle: Eigene Darstellung

Institutionalisierung der Evaluation

Innerhalb der Dienststelle Volksschulbildung gibt es den Bereich Bildungsplanung mit einer verantwortlichen Person, die sich mit System- und Projektevaluation⁷⁹ beschäftigt (für weitere Informationen siehe BKD VB 2016). Die Evaluationsstelle ist in alle Evaluationsprozesse der Dienststelle involviert und verfügt über ein Dokument, welches die Leitlinien der Politikevaluation in der Dienststelle Volksschulbildung aufführt (siehe Dolder 2017). In den beiden anderen hier betrachteten Verwaltungseinheiten gibt es weder eine Stelle noch Grundlegendokumente betreffend Evaluation. Erwähnenswert ist die Abteilung für externe Schulevaluation (Dienststelle Volksschulbildung), welche die Schulqualität der Schulen des Kantons Luzern alle sechs Jahre evaluiert. Für die Schulevaluation existieren klar definierte Vorgaben betreffend Inhalt und Vorgehen der Evaluation.⁸⁰

Die Gegenüberstellung der Einschätzung der Mitarbeitenden zeigt keine Diskrepanz zu diesen Angaben: 32 von 34 Personen der Dienststelle Volksschulbildung geben an,

⁷⁹ Mit Systemevaluationen wird die Wirksamkeit des Bildungssystems überprüft. Projektevaluationen beurteilen kantonale Projekte (siehe dazu die Angaben der Webseite der Volksschulbildung VB 2016).

⁸⁰ Für weitere Informationen siehe die entsprechende Webseite der Volksschulbildung (BKD VB Evaluation 2016).

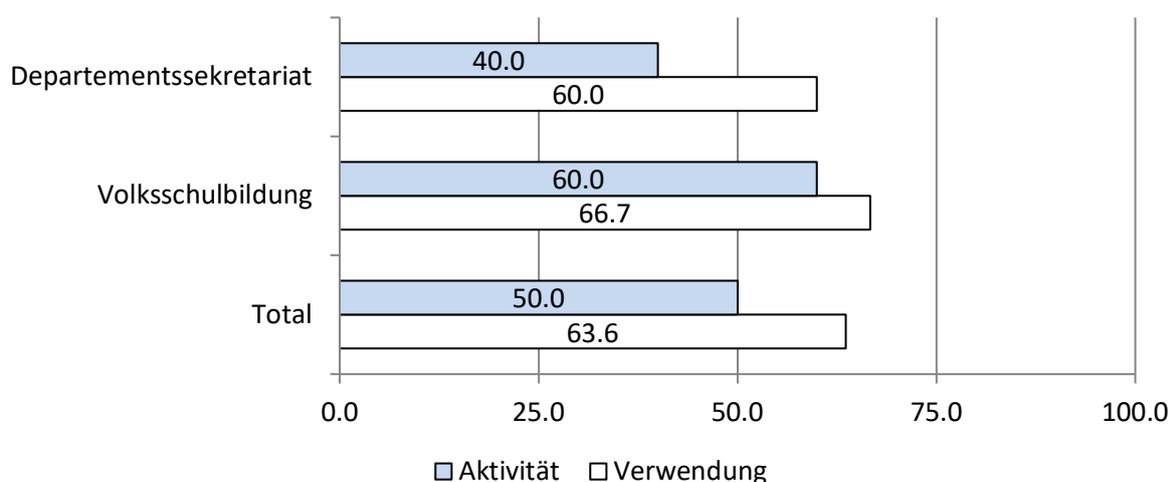
über eine Evaluationsstelle zu verfügen, während alle 18 Personen aus den beiden anderen Dienststellen angeben, über keine Evaluationsstelle zu verfügen.

Evaluationspraxis

Die nachfolgende Abbildung 10 illustriert die Evaluationspraxis von zwei Verwaltungseinheiten des BKD. Aufgrund der geringen Anzahl an Antwortenden, kann nicht einzeln auf die Dienststelle Hochschulbildung und Kultur eingegangen werden.⁸¹

Die Dienststelle Volksschulbildung verfügt über die höchsten Anteile an Mitarbeitenden mit Evaluationsaktivität (60.0 %) und Evaluationsverwendung (66.7 %). Ebenso wird gezeigt, dass in der Volksschulbildung, die im Vergleich zum Departementssekretariat über eine starke Institutionalisierung der Evaluation verfügt, eine höhere Evaluationspraxis vorliegt.

Abbildung 10: Evaluationspraxis im BKD Luzern



Anmerkungen: Aktivität = Prozentualer Anteil an Mitarbeitenden mit Evaluationsaktivität (mindestens an einer Evaluation im Durchschnitt im Jahr beteiligt). Verwendung = Prozentualer Anteil an Mitarbeitenden mit Evaluationsverwendung (mindestens seltener Miteinbezug von Evaluationen in die Arbeit). Total Antwortende: Aktivität = 52; Verwendung = 55.

Quelle: Eigene Darstellung

Der Dienststelle Volksschulbildung können insgesamt 13, dem Departementssekretariat dagegen keine Evaluationsberichte zugeordnet werden. Daneben führte die Abtei-

⁸¹ Im Total der Abbildung 10 sind auch die Mitarbeitenden der Dienststelle Hochschulbildung und Kultur enthalten.

lung für externe Schulevaluationen im Zeitraum 2005 bis 2013 insgesamt 339 Evaluationen durch.⁸² Allerdings kann kein direkter Vergleich zwischen den beiden Einheiten gezogen werden da das Departementssekretariat eine Stabsstelle ist.

Betrachtet man die Anzahl recherchierter Politikevaluationen für das BKD Luzern ist mit total 13 Evaluationen eine tiefe Evaluationsaktivität zu finden. Dem steht mit 50% ein hoher Anteil von Mitarbeitenden mit Evaluationsaktivität gegenüber. Eine Erklärung kann in der hohen Anzahl an Schulevaluationen gefunden werden, die in dem BKD durchgeführt wird.

Einstellung zur Evaluation

Die nachfolgende Tabelle 17 listet die Mittelwerte der Einstellung zur Evaluation auf. Insgesamt ist die Einstellung der Mitarbeitenden des BKD für die „non-attitude“-bereinigte Stichprobe (Mitarbeitende mit Evaluationsaktivität) mit einem Wert von 30.7 neutral.

Tabelle 17: Einstellung zur Evaluation im BKD

	Mittelwert	Min	Max	SA	N
Mitarbeitende mit Evaluationsaktivität	30.7	6	46	8.8	26
Mitarbeitende ohne Evaluationsaktivität	27.8	15	43	7.4	26
T-Test: $t(50) = -1.227$					
Volksschulbildung (Evaluationsstelle)	30.8	13	46	8.9	34
Departementssekretariat (keine Evaluationsstelle)	26.7	5	36	7.9	15
T-Test: $t(50) = -1.959+$					
Alle befragten Mitarbeitenden	30.0	6	46	8.6	58

Anmerkung: Die Einstellung zur Evaluation ergibt sich aus der Summierung von fünf Fragen und umfasst den Wertebereich 0 (minimale Einstellung zur Evaluation) bis 50 (maximale Einstellung zur Evaluation). Ein hoher Wert entspricht einer positiveren Einstellung gegenüber der Evaluation. T-Test zum Vergleich der Mittelwerte.

+ = $p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$; 2-seitig.

Quelle: Eigene Darstellung

Die Dienststelle Volksschulbildung verfügt über eine Evaluationsstelle und ist gegenüber der Evaluation auch positiver eingestellt (Mittelwert 30.8) als das Departementssekretariat, bei dem der Mittelwert bei 26.7 liegt.

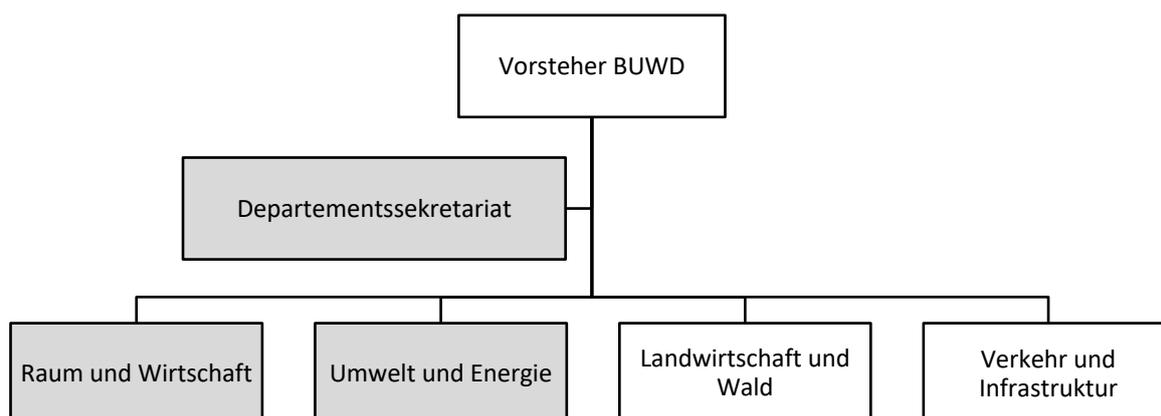
⁸² Die Angaben der Anzahl an Schulevaluationen stammen von der Leitung der Abteilung Externe Schulevaluation.

5.1.3 Bau-, Umwelt- und Wirtschaftsdepartement

Das Bau-, Umwelt- und Wirtschaftsdepartement (BUWD) des Kantons Luzern umfasst viele unterschiedliche Aufgabenbereiche: Die Wirtschaft, das Bauwesen, den Verkehr, die Energie, die Umwelt sowie die Land- und die Forstwirtschaft. Dazu kommen die Bereiche Fischerei und Jagd sowie der Schutz von Menschen und Sachwerten gegen Naturgefahren (siehe BUWD 2016).

Organisiert werden die verschiedenen Themen in vier Einheiten und dem Departementssekretariat.

Abbildung 11: Organigramm des Bau-, Umwelt und Wirtschaftsdepartementes (Stand Juli 2014)



Anmerkungen: Die grau schattierten Dienststellen sind Gegenstand der Untersuchung.
Quelle: Eigene Darstellung in Anlehnung an BUWD (2014)

Insgesamt wurden 125 Personen aus den Dienststellen Raum und Wirtschaft, der Dienststelle Umwelt und Energie sowie dem Departementssekretariat um die Teilnahme an der Befragung angefragt. Mit 76 Teilnehmerinnen und Teilnehmern resultiert eine hohe Rücklaufquote von 60.8 %.

Aufgrund der beiden angewandten Kriterien wurde die Dienststelle Landwirtschaft und Wald sowie die Dienststelle Verkehr und Infrastruktur nicht befragt. Diese neh-

men mehrheitlich betriebliche Arbeiten sowie Unterhalts- und Instandhaltungsarbeiten wahr.⁸³

Tabelle 18: Teilnehmende an der Online-Befragung

Dienststelle	Mitarbeitende	Angefragt	Antworten	Rücklauf (in Prozent)
Departementssekretariat		4	4	100.0
Raum und Wirtschaft		57	28	44.1
Umwelt und Energie		64	43	67.2
Keine Angaben		-	5	-
Total	387	125	76	60.8

Anmerkungen: Die Daten bezüglich der Anzahl an Mitarbeitenden stammen vom statistischen Amt des Kantons Luzern. Diese wurden nach Anfrage bekanntgegeben und geben den Personalbestand des gesamten Bau-, Umwelt- und Wirtschaftsdepartementes per Dezember 2014 wieder. Der Personalbestand in den einzelnen Dienststellen konnte nicht ermittelt werden.

Quelle: Eigene Darstellung

Institutionalisierung der Evaluation

Beim BUWD Luzern ist die Evaluation nicht institutionalisiert. Es gibt in diesem Departement keine Evaluationsstelle und auch keine spezifischen Dokumente, die den Umgang oder die Ausführung von Evaluationen thematisieren.

Die Gegenüberstellung der Selbstangaben der Mitarbeitenden zeigen keine wesentlichen Widersprüche. Einzig drei Mitarbeitende geben an, dass in ihrer Verwaltungseinheit eine spezifische Evaluationsstelle zu finden sei. Die grosse Mehrheit (56 von 76 Mitarbeitende) geben an, über keine solche zu verfügen (17 wissen es nicht).

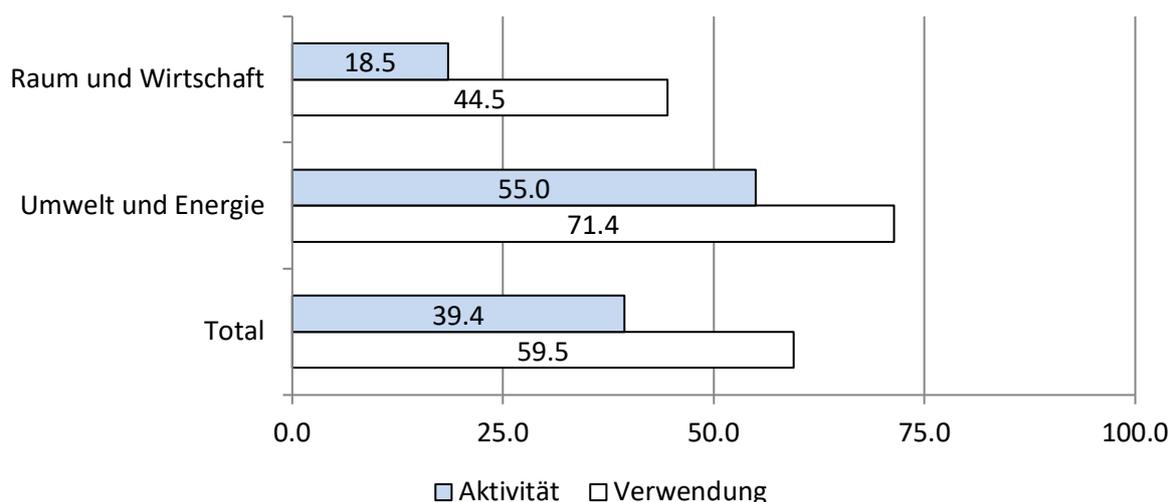
Evaluationspraxis

Die nachfolgende Abbildung 12 zeigt Unterschiede der Evaluationspraxis zwischen der Dienststelle Raum und Wirtschaft und der Dienststelle Umwelt und Energie auf. Die Dienststelle Umwelt und Energie verfügt über wesentlich höhere Anteile an Mitarbeitende mit Evaluationsaktivität (55.0 %) als die Dienststelle Raum und Wirtschaft

⁸³Die Dienststelle Landwirtschaft und Wald fördert die Land- und Forstwirtschaftsbetriebe und schafft bessere Rahmenbedingungen für die primäre Produktion. Sie strebt artenreiche, naturnahe ländliche Räume an und stellt die nachhaltige Nutzung des staatlichen Fischerei- und Jagdregals sicher (BUWD LAWA 2016). Die Dienststelle Verkehr und Infrastruktur ist verantwortlich für Planung, Bau, Betrieb und Werterhaltung der Kantonsstrassen und stellt ausserdem den Betrieb und Unterhalt der Nationalstrassen sowie die Verbesserung der Infrastruktur des öffentlichen Verkehrs sicher (BUWD VIF 2016).

(18.5 %). Das Departementssekretariat wird aufgrund der tiefen Anzahl an Befragten nicht dargestellt.

Abbildung 12: Evaluationspraxis im BUWD Luzern



Anmerkungen: Aktivität = Prozentualer Anteil an Mitarbeitende mit Evaluationsaktivität (mindestens an einer Evaluation im Durchschnitt im Jahr beteiligt). Verwendung = Prozentualer Anteil an Mitarbeitende mit Evaluationsverwendung (mindestens seltener Miteinbezug von Evaluationen in die Arbeit). Total Antwortende: Aktivität = 71; Verwendung = 74.
Quelle: Eigene Darstellung

Eine Differenz ergibt sich auch in der erfassten Anzahl an Evaluationsberichten. Der Dienststelle Raum und Wirtschaft lassen sich zwei, der Dienststelle Umwelt und Energie acht Evaluationsberichte zuordnen.

Mit total zehn gesichteten Evaluationen zeigt das BUWD Luzern insgesamt eine tiefe Evaluationsaktivität auf. Gleichzeitig gibt ein Anteil von 39.4 % an, bei der Erarbeitung von Evaluationen beteiligt zu sein. Ein möglicher Grund für diese Diskrepanz könnte sein, dass beim Amt für Umwelt und Energie sechs der acht gesichteten Evaluationen im Zeitraum zwischen 2011 und 2013 gemacht wurden – dass also in kurzer Zeit sehr viel evaluiert wurde.

Einstellung zur Evaluation

Die nachfolgende Tabelle 19 zeigt die Mittelwerte der Einstellung zur Evaluation. Insgesamt verfügt das BUWD Luzern bei Mitarbeitenden mit Evaluationsaktivität über neutrale Einstellungswerte. Analog zu den beiden vorherigen Departementen haben

auch die Mitarbeitenden des BUWD Luzern mit Evaluationsaktivität eine positivere Einstellung. Die geringen Mittelwertunterschiede dieser beiden Gruppen sind allerdings statistisch nicht belegbar.

Tabelle 19: Einstellung zur Evaluation im BUWD Luzern

	Mittelwert	Min	Max	SA	N
Mitarbeitende mit Evaluationsaktivität	33.2	15	49	9.8	28
Mitarbeitende ohne Evaluationsaktivität	30.5	7	43	7.2	42
T-Test: $t(68) = -1.343$					
Alle befragten Mitarbeitenden	31.4	9	49	8.4	74

Anmerkung: Die Einstellung zur Evaluation ergibt sich aus der Summierung von fünf Fragen und umfasst den Wertebereich 0 (minimale Einstellung zur Evaluation) bis 50 (maximale Einstellung zur Evaluation). Ein hoher Wert entspricht einer positiveren Einstellung gegenüber der Evaluation. T-Test zum Vergleich der Mittelwerte.

+ = $p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$; 2-seitig.

Quelle: Eigene Darstellung

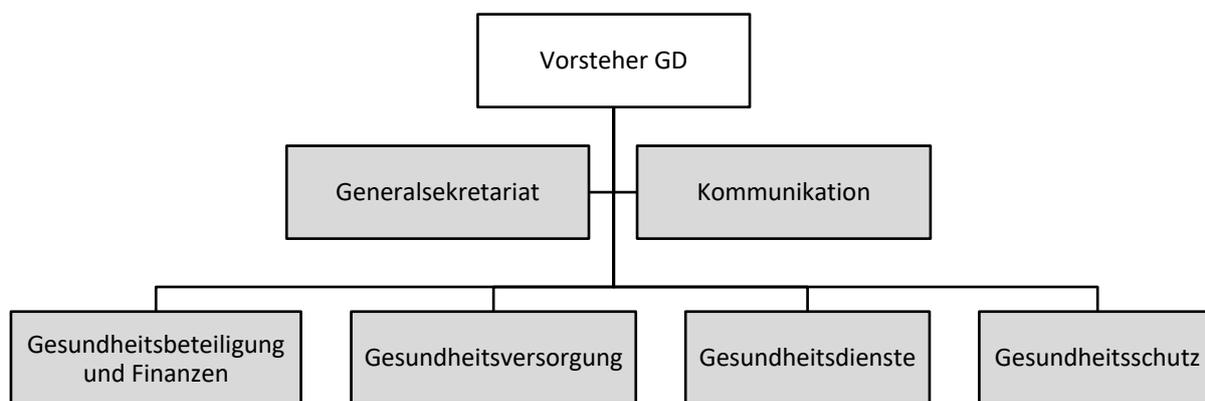
5.2 Basel-Stadt

5.2.1 Gesundheitsdepartement

Das Gesundheitsdepartement (GD) des Kantons Basel-Stadt hat die Förderung und den Schutz der psychischen und physischen Gesundheit zum Ziel und gliedert sich in folgende Einheiten: das Generalsekretariat, die Kommunikation, den Bereich Gesundheitsbeteiligung und Finanzen, den Bereich Gesundheitsversorgung, den Bereich Gesundheitsdienste sowie den Bereich Gesundheitsschutz (GD 2016).⁸⁴

⁸⁴ In den Kantonsverwaltungen Basel-Stadt finden sich zwei Hierarchiestufen: Jedes Departement lässt sich in Bereiche (Ebene 1) aufteilen, die wiederum in Abteilungen (Ebene 2) unterteilt sind.

Abbildung 13: Organigramm des Gesundheitsdepartementes Basel-Stadt (Stand August 2014)



Anmerkungen: Die grau schattierten Bereiche sind Gegenstand der Untersuchung.
Quelle: Eigene Darstellung in Anlehnung an GD (2014)

Befragt wurden alle Bereiche des Gesundheitsdepartementes. Im Bereich Gesundheitsschutz wurden einzelne Abteilungen von der Befragung ausgeschlossen, wie das Veterinäramt, das nicht angefragt wurde, weil es sich dabei um eine Abteilung mit täglichem individuellen Bürgerkontakt handelt.

Die Reduktion der Mitarbeitenden ist in der nachfolgenden Tabelle 20 dargestellt: Insgesamt wurden 179 Personen angefragt, wovon 109 Personen den Fragebogen beantworteten. Die Rücklaufquote ist mit 60.9 % aussergewöhnlich hoch.

Tabelle 20: Teilnehmende an der Online-Befragung

Bereich	Mitarbeitende	Angefragt	Antworten	Rücklauf (in Prozent)
Generalsekretariat	34*	34*	16	-
Kommunikation	34*	34*	2	-
Gesundheitsbeteiligungen und Finanzen	34*	34*	5	-
Gesundheitsversorgung	36	36	24	66.7
Gesundheitsdienste	87	98	46	46.9
Gesundheitsschutz	105	11	5	45.5
Keine Angaben	-	-	11	-
Total	387	179	109	60.9

Anmerkung: Die Daten bezüglich der Anzahl an Mitarbeitenden stammen von der Webseite und Auskünften des statistischen Amtes Basel-Stadt (Statistisches Amt 2014) und geben den Gesamtpersonalbestand des Gesundheitsdepartementes für das Jahr 2014 wieder.

*: Die Anzahl angefragter Mitarbeitende in den einzelnen Bereichen (Generalsekretariat, Kommunikation und Gesundheitsbeteiligungen und Finanzen) kann nicht nachvollzogen werden. Insgesamt können 34 Mitarbeitende diesen drei Bereichen zugeordnet werden.

Quelle: Eigene Darstellung

Institutionalisierung der Evaluation

Der Bereich Gesundheitsdienste verfügt über die Steuergruppe Evaluation. Diese unterstützt die Projektleiterinnen und Projektleiter und kann bei Fragen bezüglich der Evaluation angefragt werden. Daneben muss in den Projektskizzen die Methodik zur Projektauswertung angegeben werden. Die anderen Bereiche verfügen über keine Evaluationsstelle und keine spezifischen Grundlagendokumente betreffend Evaluation.

Die Gegenüberstellung der Selbstangaben der Mitarbeitenden zur Institutionalisierung der Evaluation zeigen ein ähnliches Bild. Die Mitarbeitenden aus dem Bereich Gesundheitsdienste sind sich betreffend dem Vorhandensein einer Evaluationsstelle in ihrem Bereich uneinig. Diese Stelle wird als solche nicht einheitlich wahrgenommen: 16 Mitarbeitende geben an, über eine solche zu verfügen, 18 Mitarbeitende antworten, sie würden über keine Evaluationsstelle in ihrem Bereich verfügen (zwölf wissen es nicht).

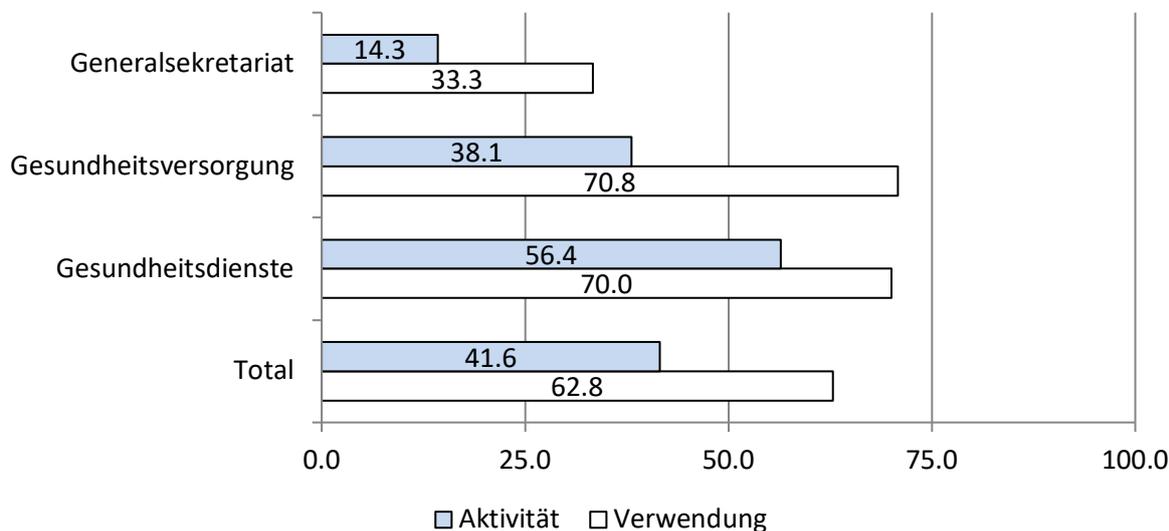
Evaluationspraxis

Unterschiede zwischen den Verwaltungseinheiten lassen sich bei den Angaben der Evaluationspraxis ausmachen. Die nachfolgende Abbildung 14 zeigt eine Diskrepanz der Evaluationspraxisangaben zwischen dem Generalsekretariat und den beiden Bereichen Gesundheitsversorgung und Gesundheitsdienste.⁸⁵ Im Bereich Gesundheitsdienste, der über die Steuergruppe Evaluation verfügt, ist auch der höchste Anteil an Mitarbeitenden mit Evaluationsaktivität (56.4 %) zu finden.

Besonders ausgeprägt sind zudem die Differenzen der Angaben hinsichtlich der Evaluationsverwendung: Mit 33.3 % ist der Anteil an Mitarbeitenden mit Evaluationsverwendung im Generalsekretariat erheblich tiefer als in den Bereichen Gesundheitsversorgung (70.8 %) und Gesundheitsdienste (70.0 %).

⁸⁵ Aufgrund der geringen Anzahl an angefragten und antwortenden Mitarbeitenden werden die Bereiche Kommunikation, Gesundheitsbeteiligungen und Finanzen sowie der Bereich Gesundheitsschutz nicht einzeln dargestellt.

Abbildung 14: Evaluationspraxis im GD Basel-Stadt



Anmerkungen: Aktivität = Prozentualer Anteil an Mitarbeitende mit Evaluationsaktivität (mindestens an einer Evaluation im Durchschnitt im Jahr beteiligt). Verwendung = Prozentualer Anteil an Mitarbeitende mit Evaluationsverwendung (mindestens seltener Miteinbezug von Evaluationen in die Arbeit). Total Antwortende: Aktivität = 89; Verwendung = 102.

Quelle: Eigene Darstellung

Aufgrund der Evaluationsrecherche können drei Evaluationsberichte dem Bereich Gesundheitsversorgung und zehn Evaluationen dem Bereich Gesundheitsdienste zugeordnet werden.

Für das GD Basel-Stadt ergibt sich mit total zehn Evaluationsberichten die tiefste Evaluationsaktivität. Der Anteil der Mitarbeitenden mit Evaluationsaktivität ist trotzdem im Mittelfeld (41.6 %) anzusiedeln. Grund sind die höheren Anteile im Bereich Gesundheitsdienste (56.4 %), der über die Steuergruppe Evaluation verfügt.

Einstellung zur Evaluation

Zusammengefasst ist die Einstellung zur Evaluation der evaluationsaktiven Mitarbeitenden („non-attitude“-bereinigt) des GD Basel-Stadt positiv und im Vergleich zu den anderen Verwaltungseinheiten ausserordentlich hoch. Zudem sind die Unterschiede der Mittelwerte zwischen evaluationsaktiven Mitarbeitenden und Mitarbeitenden ohne Evaluationsaktivität statistisch nachweisbar.

Tabelle 21: Einstellung zur Evaluation im GD Basel-Stadt

	Mittelwert	Min	Max	SA	N
Mitarbeitende mit Evaluationsaktivität	34.6	12	48	9.0	37
Mitarbeitende ohne Evaluationsaktivität	30.1	7	45	8.0	48
T-Test: $t(83) = -2.426^*$					
Gesundheitsdienste	32.9	7	49	9.1	44
Alle anderen Bereiche ohne die Gesundheitsdienste	30.4	7	48	9.1	59
T-Test: $t(101) = -1.360$					
Alle befragten Mitarbeitenden	31.4	7	49	9.2	103

Anmerkung: Die Einstellung zur Evaluation ergibt sich aus der Summierung von fünf Fragen und umfasst den Wertebereich 0 (minimale Einstellung zur Evaluation) bis 50 (maximale Einstellung zur Evaluation). Ein hoher Wert entspricht einer positiveren Einstellung gegenüber der Evaluation. T-Test zum Vergleich der Mittelwerte.

+ = $p < 0.10$; * = $p < 0,05$; ** = $p < 0,01$; 2-seitig.

Quelle: Eigene Darstellung

Der Mittelwert der Einstellung für den Bereich Gesundheitsdienste, der über die Steuergruppe Evaluation verfügt, ist mit 32.9 nur geringfügig positiver als der Mittelwert in den anderen Bereichen (30.4).

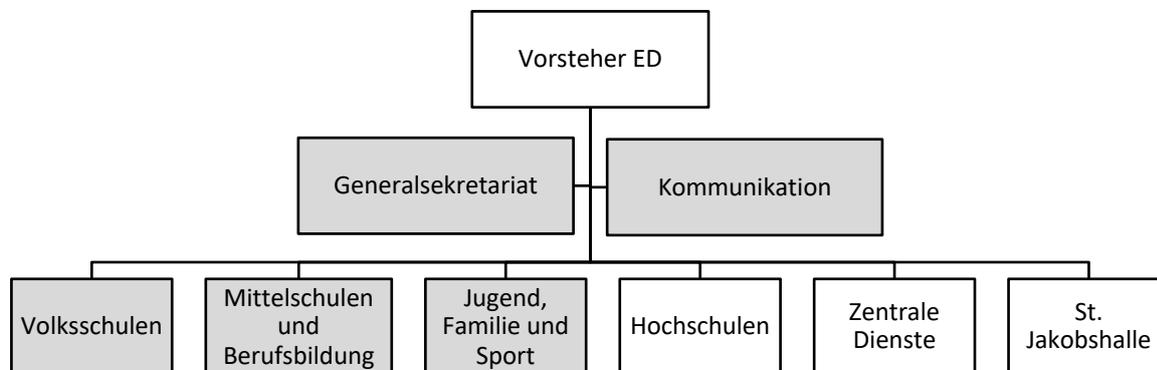
5.2.2 Erziehungsdepartement

Das Erziehungsdepartement (ED) des Kantons Basel-Stadt ist von der Volksschule bis zur Hochschule zuständig für das Bildungsangebot, und umfasst die Berufsfachschule sowie die Erwachsenen- und Weiterbildung. Das Erziehungsdepartement ist zudem verantwortlich für die Tagesbetreuung und die Themen Bewegung und Sport. Weiter übernimmt das ED Basel-Stadt Aufgaben zur Förderung von Kindern, Jugendlichen und Familien und kümmert sich auch um Kinder- und Jugendschutz (siehe ED 2016).

Verwaltet wird dieses breite Themenspektrum von sechs verschiedenen Einheiten sowie dem Generalsekretariat und der Kommunikation.

Zu beachten ist, dass in Basel-Stadt die kommunalen Geschäfte vom Kanton verwaltet und geführt werden und es diesbezüglich keine kommunale Ebene gibt. Anders als bei den Kantonen Bern und Luzern, wo die Gemeinden für die kommunalen Schulen zuständig sind, ist in Basel-Stadt der Kanton dafür verantwortlich.

Abbildung 15: Organigramm des Erziehungsdepartementes Basel-Stadt (Stand August 2014)



Anmerkungen: Die grau schattierten Bereiche sind Gegenstand der Untersuchung.
Quelle: Eigene Darstellung in Anlehnung an ED (2014)

Aufgrund der beiden Ausschlusskriterien wurden die Mitarbeitenden aus dem Bereich Hochschulen, dem Bereich Zentrale Dienste und dem Bereich St. Jakobshalle nicht befragt.⁸⁶ Im Bereich Mittelschulen und Berufsbildung wurden Personen aus Abteilungen mit individuellem Bürgerkontakt aus der Befragung ausgeschlossen. Ebenso in den Bereichen Volksschulen sowie Jugend, Familie und Sport gab es eine erhebliche Reduktion der angefragten Abteilungen. Beispielsweise wurde die Abteilung Schulpsychologischer Dienst aus der Befragung ausgeschlossen.⁸⁷

Von den insgesamt 4859 angestellten Personen wurden lediglich 64 Mitarbeitende befragt. Analog zur Reduktion der Personen des Bildungs- und Kulturdepartementes Luzern erklärt sich der grösste Anteil dieser Reduktion durch den Ausschluss der Schulen und der Hochschulen. Schlussendlich beantworteten von 64 angefragten Mitarbeitenden 36 Personen den Fragebogen.

⁸⁶ Die zentralen Dienste beispielsweise bestehen aus 200 Hauswartinnen und Hauswarten, dem Informatiksupport sowie dem Telefondienst und können nicht der Zentralverwaltung zugeordnet werden. Ebenso wenig entspricht der Bereich Hochschule, mit dem Amt für Ausbildungsbeiträge und dem Pädagogischen Zentrum, nicht unseren Kriterien der Zentralverwaltung (siehe ED HS 2016).

⁸⁷ Der Schulpsychologische Dienst (SPD) verfügt als öffentliche Beratungsstelle für Kinder und Jugendliche, die in Basel zur Schule gehen, über tägliche individuellen Bürgerkontakt (siehe ED SPD 2016).

Tabelle 22: Teilnehmende an der Online-Befragung

Bereich	Mitarbeitende	Angefragt	Antworten	Rücklauf (in Prozent)
Generalsekretariat	3	3	1	33.3
Kommunikation	6	6	4	66.7
Volksschulen	2435	14	7	50.0
Mittelschulen und Berufsbildung	723	13	5	38.5
Jugend, Familie und Sport	327	28	16	57.1
Keine Angaben	-	-	3	-
Total	4859	64	36	56.3

Anmerkungen: Die Daten bezüglich der Anzahl an Mitarbeitenden stammen von der Webseite und Auskünften des statistischen Amtes Basel-Stadt (Statistisches Amt 2014) und geben den Personalbestand des gesamten Erziehungsdepartementes für das Jahr 2014 wieder. Diese Zahlen schliessen viele Mitarbeitenden ein, beispielsweise Lehrpersonen und Personen aus dem Gebäudeunterhalt, die nicht befragt wurden.
Quelle: Eigene Darstellung

Institutionalisierung der Evaluation

Dem Bereich Jugend, Familie und Sport untersteht die Abteilung Planung, die unter anderem Planungs- und Evaluationstätigkeiten durchführt (ED JFF 2016). Die Abteilung Planung ist aber keine institutionalisierte Evaluationsstelle. Diese verfügt zwar über Evaluationsknowhow, beschäftigt sich aber lediglich punktuell mit Evaluationen. Die anderen Bereiche verfügen über keine Evaluationsstelle. Spezifische Dokumente betreffend Evaluation gibt es in keinem der befragten Bereiche des Erziehungsdepartementes. Zum Zeitpunkt der Durchführung der Befragung gab es im ED Basel-Stadt keine Abteilung für externe Schulevaluation.

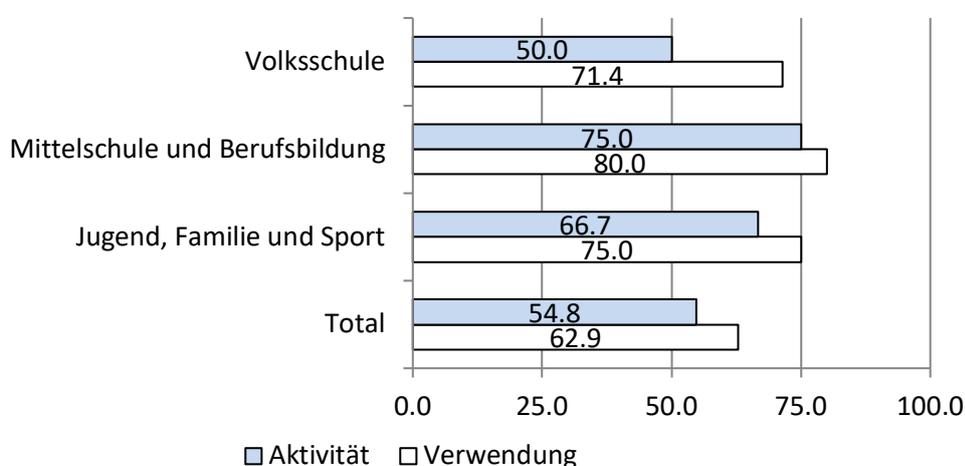
Die Gegenüberstellung der Selbstangaben der Mitarbeitenden aus dem Bereich Jugend, Familie und Sport bezüglich des Vorhandenseins einer Evaluationsstelle, zeigt eine andere Wahrnehmung: Hier geben 15 von 16 Befragten an, dass dieser Bereich über keine solche verfügt, was veranschaulicht, dass die Abteilung Planung von den Mitarbeitenden nicht als Evaluationsstelle wahrgenommen wird.

Evaluationspraxis

Die folgende Abbildung 16 illustriert die Evaluationspraxis im Erziehungsdepartement und zeigt, dass ein hoher Anteil an Mitarbeitenden über Evaluationsaktivität verfügt und viele Mitarbeitenden die Evaluationsberichte auch verwenden. Beim Bereich Mittelschule und Berufsbildung geben beispielsweise 75.0 % an, im Durchschnitt jährlich an der Erarbeitung von Evaluationen beteiligt zu sein. Allerdings sind nur wenige Be-

fragte den einzelnen Bereichen zuzuordnen und die Prozentwerte vorsichtig zu interpretieren. Dem Bereich Mittelschulen und Berufsbildung sind lediglich vier und dem Bereich Volksschule sechs Befragte zuzuordnen. Aufgrund der tiefen Anzahl an Befragten werden die Bereiche Generalsekretariat und Kommunikation nicht einzeln dargestellt.

Abbildung 16: Evaluationspraxis im ED Basel-Stadt



Anmerkungen: Aktivität = Prozentualer Anteil an Mitarbeitende mit Evaluationsaktivität (mindestens an einer Evaluation im Durchschnitt im Jahr beteiligt). Verwendung = Prozentualer Anteil an Mitarbeitende mit Evaluationsverwendung (mindestens seltener Miteinbezug von Evaluationen in die Arbeit). Total Antwortende: Aktivität = 31; Verwendung = 35.

Quelle: Eigene Darstellung

Aufgrund der Evaluationsrecherche können dem Bereich Volksschulen 17, dem Bereich Mittelschulen und Berufsbildung fünf sowie dem Bereich Jugend, Familie und Sport drei Evaluationsberichte zugeordnet werden. Ausserdem wurden für den Bereich Mittelschulen und Berufsbildung zehn externe IFES-Schulevaluationen⁸⁸ im Zeitraum 2005 bis 2013 durchgeführt. Diese wurden von dem Bereich in Auftrag gegeben und finanziert.⁸⁹ Für die Volksschulen wurden keine externen Schulevaluationen in diesem Zeitraum gemacht.

⁸⁸ Das IFES (Institut für Externe Schulevaluation auf der Sekundarstufe II) ist eine Institution der EDK (Schweizerische Konferenz der kantonalen Erziehungsdirektoren) und führt externe Schulevaluationen auf Sekundarstufe II durch (siehe: www.ifes-ipes.ch).

⁸⁹ Die Angaben stammen von der Leitung des Bereichs Mittelschulen und Berufsbildung.

Insgesamt finden sich damit 25 Evaluationsberichte (ohne die Schulevaluationen), was im Vergleich zu den anderen Verwaltungseinheiten einer mittleren Evaluationsaktivität entspricht. Gleichzeitig gibt im ED Basel-Stadt insgesamt der höchste Anteil an Mitarbeitenden (54.8 %) an, pro Jahr mindestens an einer Evaluation beteiligt zu sein. Auch hier kann eine wichtige Erklärung in den externen Schulevaluationen gefunden werden.

Einstellung zur Evaluation

Im Vergleich zu den anderen Departementen und Direktionen haben die evaluationsaktiven Mitarbeitenden des ED Basel-Stadt den höchsten Einstellungsmittelwert und somit die positivste Einstellung zur Evaluation.

Die nachfolgende Tabelle 23 listet die Mittelwerte der Einstellung zur Evaluation auf. Daraus ersichtlich werden einerseits die Werte aller Befragten des ED Basel-Stadt und andererseits jener, die bereits über Evaluationserfahrung verfügen. Die Mittelwertunterschiede dieser beiden Gruppen sind hoch (Wert 6.0), statistisch aber lediglich auf dem Signifikanzniveau von 90 % belegbar.

Tabelle 23: Einstellung zur Evaluation im ED Basel-Stadt

	Mittelwert	Min	Max	SA	N
Mitarbeitende mit Evaluationsaktivität	35.7	14	47	12.8	17
Mitarbeitende ohne Evaluationsaktivität	29.7	7	48	12.8	14
T-Test: t(29) = -1.701+					
Jugend, Familie und Sport	35.8	25	47	7.1	17
Alle anderen Bereiche	29.7	7	48	9.5	19
T-Test: t(15) = -1.795+					
Alle befragten Mitarbeitenden	32.5	7	48	11.4	35

Anmerkung: Die Einstellung zur Evaluation ergibt sich aus der Summierung von fünf Fragen und umfasst den Wertebereich 0 (minimale Einstellung zur Evaluation) bis 50 (maximale Einstellung zur Evaluation). Ein hoher Wert entspricht einer positiveren Einstellung gegenüber der Evaluation. T-Test zum Vergleich der Mittelwerte.

+ = p < 0.10; * = p < 0.05; ** = p < 0.01; 2-seitig.

Quelle: Eigene Darstellung

Der Mittelwert der Einstellung für Mitarbeitende aus dem Bereich Jugend, Familie und Sport, der über die Abteilung Planung verfügt, ist mit 35.8 klar positiver als der Mittelwert der Mitarbeitenden der anderen Bereichen des ED (29.7). Auch diese Differenz lässt sich lediglich mit einer Irrtumswahrscheinlichkeit von 10 % (Signifikanzniveau von 90 %) nachweisen.

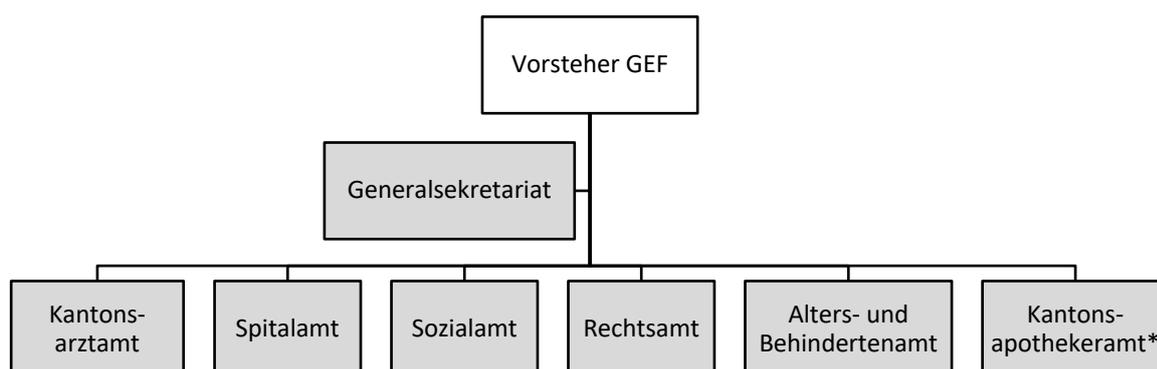
5.3 Bern

5.3.1 Gesundheits- und Fürsorgedirektion

Die Gesundheits- und Fürsorgedirektion (GEF) des Kantons Bern ist zuständig für das Gesundheits- und das Sozialwesen. Darunter fallen die Spitalversorgung, die Sozialhilfe, der Vollzug der Heilmittel-, der Lebensmittel- und der Giftgesetzgebung sowie die Strahlenschutzgesetzgebung. Ebenso ist die Direktion verantwortlich für die Gesundheitsvorsorge und die Bekämpfung übertragbarer Krankheiten (siehe GEF 2016).

In der nachfolgenden Abbildung 17 sind die Einheiten und das Generalsekretariat zu erkennen, in denen die oben erwähnten Themenbereiche verwaltet werden.

Abbildung 17: Organigramm der Gesundheits- und Fürsorgedirektion Bern (Stand August 2014)



Anmerkungen: Die grau schattierten Ämter sind Gegenstand der Untersuchung.

*: Das kantonale Laboratorium ist auch Bestandteil der GEF Bern. Dieses Amt wurde nicht befragt und wird deshalb nicht in der Abbildung dargestellt.

Quelle: Eigene Darstellung in Anlehnung an (GEF 2014)

Das kantonale Laboratorium wurde als Ganzes aus der Befragung ausgeschlossen, weil die Mitarbeitenden dieses Amtes mehrheitlich in direktem Bürgerkontakt stehen. Auch das Personal der administrativ angegliederten Organisationseinheiten, wie die Universitären Psychiatrischen Dienste, die Schulheime, das Psychiatriezentrum Münsingen und die Psychiatrischen Dienste, wurden nicht befragt.

Dementsprechend reduziert sich die Anzahl angefragter Personen von 2960 auf 234. Mit 125 Antwortenden resultiert für die GEF Bern eine Rücklaufquote von 53.4 %.

Tabelle 24: Teilnehmende an der Online-Befragung

Ämter	Mitarbeitende	Angefragt	Antworten	Rücklauf (in Prozent)
Generalsekretariat	50	44	22	50.0
Kantonsapothekeramt	10	5	2	40.0
Kantonsarztamt	15	20	8	32.0
Spitalamt	34	36	17	47.2
Sozialamt	61	65	36	55.4
Rechtsamt	13	13	6	46.1
Alters- und Behindertenamt	45	51	19	37.3
Keine Angaben	-	-	15	-
Total	2960	234	125	53.4

Anmerkungen: Die Daten bezüglich der Anzahl an Mitarbeitenden stammen von der Rückmeldung und der Webseite des Personalamtes des Kantons Bern (Personalamt 2014) und geben den Gesamtpersonalbestand der Gesundheits- und Fürsorgedirektion per Dezember 2014 wieder. Einbezogen sind die administrativ angeordneten Einheiten: Universitäre Psychiatrische Dienste, die Schulheime, das Psychiatriezentrums Münsingen und die psychiatrischen Dienste.

Quelle: Eigene Darstellung

Institutionalisierung der Evaluation

Im Generalsekretariat gibt es die Abteilung Grundlagen, welche in den Schwerpunktthemen Gesundheit, Soziales und Familie Statistiken erhebt, auswertet und Bericht erstattet (GEF GS 2016). Die Abteilung Grundlagen verfügt über Knowhow betreffend Evaluation, wird aber nicht für jede Evaluation einbezogen. Dolder (2018) konnte aufzeigen, dass im Zuge dem Programmende von ERKOS⁹⁰ eine Deinstitutionalisierung der Evaluation in der Abteilung Grundlagen stattgefunden hat. Während der Zeit des Programmes ERKOS gab es eine Evaluationsstelle, die seit der Beendigung dieses Programmes massiv eingeschränkt wurde. Die GEF Bern verfügt seither weder über eine Evaluationsstelle noch über definierte Evaluationsprozesse.

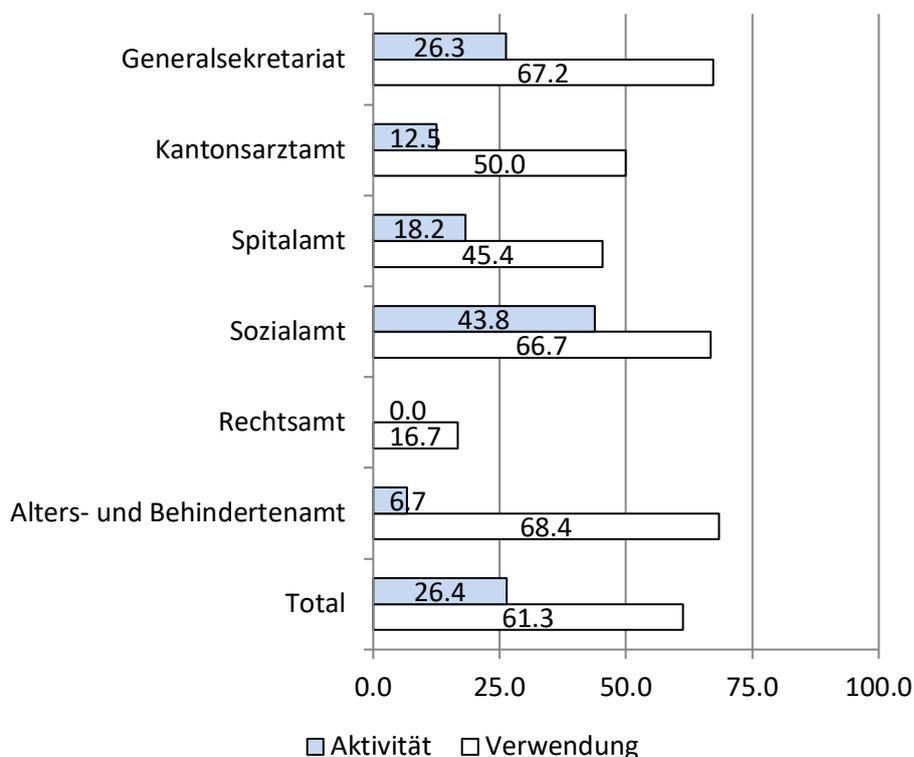
Dies wird auch von den Mitarbeitenden des Generalsekretariates so wahrgenommen; von den 22 Befragten geben lediglich zwei Befragte an, dass ihr Amt über eine Evaluationsstelle verfügt. Zwölf geben an, es gebe keine Evaluationsstelle und Acht wissen es nicht.

⁹⁰ ERKOS steht für Erfolgskontrolle von Staatsbeiträgen im Kanton Bern und hat zum Ziel, die bernischen Staatsbeiträge mindestens einmal innerhalb von zehn Jahren einer Erfolgskontrolle zu unterziehen. Das Programm wurde 1998 eingeführt und 2004 wieder beendet (siehe (Dolder 2017)).

Evaluationspraxis

Die untenstehende Abbildung 18 der Evaluationspraxis der GEF Bern zeigt erhebliche Unterschiede zwischen den Ämtern. Auffallend sind die Anteile im Rechtsamt. Hier gibt niemand an, an der Evaluationserstellung beteiligt zu sein. Auch die Verwendung der Evaluation in die Arbeit wird lediglich von einem kleinen Anteil bestätigt (einer von sechs Befragten). Die Evaluationsaktivität ist im Generalsekretariat trotz der Abteilung Grundlagen, die über Evaluations-Knowhow verfügt, wenig verbreitet.

Abbildung 18: Evaluationspraxis in der GEF Bern



Anmerkungen: Aktivität = Prozentualer Anteil an Mitarbeitenden mit Evaluationsaktivität (mindestens an einer Evaluation im Durchschnitt im Jahr beteiligt). Verwendung = Prozentualer Anteil an Mitarbeitenden mit Evaluationsverwendung (mindestens seltener Miteinbezug von Evaluationen in die Arbeit). Total Antwortende: Aktivität = 104; Verwendung = 117.
Quelle: Eigene Darstellung

Die grösste Anzahl Evaluationsberichte (26 Evaluationen) finden sich für das Sozialamt. Fünf Evaluationen können dem Alters- und Behindertenamt zugeordnet werden. Dem Spitalamt können zwei, dem Kantonsarztamt eine Evaluation zugeschrieben werden und für das Rechtsamt lässt sich kein Evaluationsbericht finden.

Insgesamt ergibt sich für die GEF Bern mit 34 Evaluationen eine hohe Evaluationsaktivität. Gleichzeitig ist mit einem Anteil von 26.4 % der tiefste Anteil an Mitarbeitenden mit Evaluationsaktivität zu finden. Grund dafür ist die sehr heterogene Verteilung der Mitarbeitenden mit Evaluationsaktivität zwischen den Ämtern.

Einstellung zur Evaluation

Mitarbeitende mit Evaluationsaktivität der GEF Bern verfügen insgesamt über eine neutrale Einstellung zur Evaluation.

Die Unterschiede der Mittelwerte in der nachfolgenden Tabelle 25 aufgeführten Gruppen mit und ohne Evaluationsaktivität fallen klein aus und sind nicht signifikant.

Tabelle 25: Einstellung zur Evaluation in der GEF Bern

	Mittelwert	Min	Max	SA	N
Mitarbeitende mit Evaluationsaktivität	33.1	9	48	9.5	28
Mitarbeitende ohne Evaluationsaktivität	31.0	6	50	8.4	73
T-Test: $t(46) = -1.00$					
Alle befragten Mitarbeitenden	31.4	6	50	8.8	117

Anmerkung: Die Einstellung zur Evaluation ergibt sich aus der Summierung von fünf Fragen und umfasst den Wertebereich 0 (minimale Einstellung zur Evaluation) bis 50 (maximale Einstellung zur Evaluation). Ein hoher Wert entspricht einer positiveren Einstellung gegenüber der Evaluation. T-Test zum Vergleich der Mittelwerte.

+ = $p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$; 2-seitig.

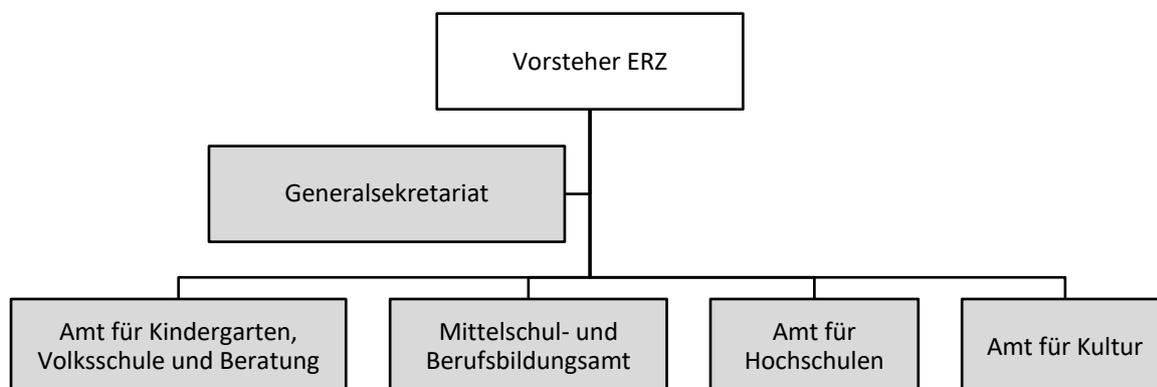
Quelle: Eigene Darstellung

5.3.2 Erziehungsdirektion

Die Erziehungsdirektion (ERZ) des Kantons Bern verwaltet die beiden Themenbereiche Bildung und Kultur, überwacht die Volks- und Mittelschulen ebenso wie die Berufsbildung, die Berufsfachschulen, die Pädagogische Hochschule und die Universität (ERZ 2016).

Die Erziehungsdirektion besteht aus fünf Einheiten und dem Generalsekretariat (siehe nachfolgende Abbildung 19).

Abbildung 19: Organigramm der Erziehungsdirektion Bern (Stand August 2014)



Anmerkungen: Die grau schattierten Ämter sind Gegenstand der Untersuchung. Das Amt für zentrale Dienste wird aus Platzgründen nicht in dem Organigramm aufgeführt.

Quelle: Eigene Darstellung in Anlehnung an (ERZ 2014)

Von der grossen Anzahl an Mitarbeitenden (1539) wurden schlussendlich 289 angefragt, wovon 73 an der Befragung teilnahmen. Das Amt für zentrale Dienste wurde, aufgrund vorwiegender Übernahme von Unterhalts- und Instandhaltungsarbeiten (technische Infrastruktur und Hausdienst) und individuellem Bürgerkontakt (Stipendienwesen), nicht befragt. Die nachfolgende Tabelle 26 zeigt eine tiefe Rücklaufquote von 25.3 % für die ERZ Bern auf. Zudem fällt die sehr ungleiche Beteiligung der verschiedenen Ämter auf. Ein möglicher Grund kann im unterschiedlichen Befragungsvorgehen bei der ERZ Bern gefunden werden: Anders als bei den anderen acht Verwaltungseinheiten wurden die Befragten nicht direkt angeschrieben. Auf Wunsch der ERZ Bern übernahm das Generalsekretariat die Anfrage zur Befragungs-Teilnahme an die Mitarbeitenden. Insgesamt wurde der Link zur Befragung an 289 Mitarbeitende gesendet. Die nachfolgende Tabelle 26 zeigt, dass lediglich vier Mitarbeitende aus dem Amt für Kindergarten, Volksschule, und Beratung teilnahmen und nur ein Mitarbeiter dem Amt für Hochschulen zugeordnet werden kann. Entsprechend wurde auf eine vertiefte Analyse dieser beiden Ämter verzichtet.

Tabelle 26: Teilnehmende an der Online-Befragung

Ämter	Mitarbeitende	Angefragt*	Antworten	Rücklauf (in Prozent)
Generalsekretariat	69	73	41	-
Amt für Kindergarten, Volksschule und Beratung	227	11	4	-
Mittelschul- und Berufsbildungsamt	958	23	8	-
Amt für Hochschulen	30	22	1	-
Amt für Kultur	175	22	16	-
Keine Angaben	-	-	3	-
Total	1539	289	73	25.3

Anmerkungen: Die Daten bezüglich der Anzahl an Mitarbeitenden stammen von der Webseite und den Rückmeldungen des Personalamtes des Kantons Bern (Personalamt 2014) und geben den Gesamtpersonalbestand der Erziehungsdirektion per Dezember 2014 wieder.

*: Die Anzahl angefragter Mitarbeitende in den einzelnen Ämtern lässt sich nicht exakt nachvollziehen. Der Link zur Befragung wurde durch das Generalsekretariat direkt an die Mitarbeitenden versendet. Die aufgeführte Anzahl ist dem Staatskalender entnommen und stellt ein Näherungswert dar (Staatskanzlei 2016).

Quelle: Eigene Darstellung

Institutionalisierung der Evaluation

Im Generalsekretariat der ERZ Bern gibt es die Abteilung Bildungsplanung und Evaluation, die über den Fachbereich Evaluation verfügt. Dieser führt interne Evaluationen durch und begleitet externe Evaluationsaufträge (ERZ B&E 2016). Der Fachbereich Evaluation ist die Evaluationsstelle für die gesamte Erziehungsdirektion. Zudem existieren in der Erziehungsdirektion spezifische Vorlagen betreffend Evaluation: Antragsbeurteilung, Grobkonzept (für interne Evaluationen), Offertenerstellung und -beurteilung, Pflichtenheft (externe Evaluationen), Auftragsvergabe (externe Evaluationen) sowie Berichtsvorlagen. Demgegenüber gibt es im Kanton Bern keine flächendeckende externe Schulevaluation (siehe ARGEV 2016).

Die Gegenüberstellung der Selbstangaben der Mitarbeitenden aus dem Generalsekretariat bezüglich des Vorhandenseins einer Evaluationsstelle zeigt, dass die überwiegende Mehrheit diese Stelle auch als solche wahrnimmt. Insgesamt geben 33 Befragte aus dem Generalsekretariat an, dass ihre Einheit über eine Evaluationsstelle verfügt. Lediglich drei Befragte verneinten eine solche Stelle und fünf wissen es nicht.

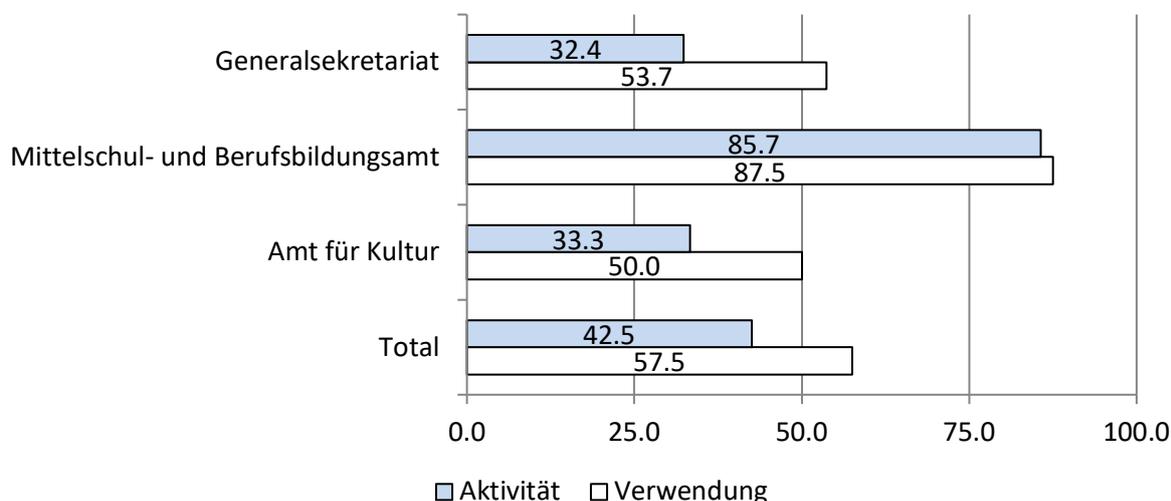
Evaluationspraxis

Die nachstehende Abbildung 20 illustriert die Evaluationspraxis von drei Verwaltungseinheiten der ERZ Bern und zeigt erhebliche Unterschiede zwischen den Ämtern auf. Hohe Anteile im Mittelschul- und Berufsbildungsamt mit einem Anteil an Evalua-

tionsaktivität von 85.7 % und einen Evaluationsverwendungsanteil von 87.5 %. Wesentlich tiefere Anteile sind für das Generalsekretariat, mit 32.4 % für die Evaluationsaktivität und 53.7 % für die Evaluationsverwendung, zu beobachten. Das Amt für Hochschulen sowie das Amt für Kindergarten, Volksschule und Beratung werden aufgrund der geringen Anzahl an Antwortenden nicht dargestellt.

Dabei ist nochmals auf die tiefe Rücklaufquote für die ERZ Bern zu verweisen und die damit verbundene eingeschränkte Aussagekraft. Die grossen Unterschiede der Evaluationspraxisanteile zwischen den Verwaltungseinheiten könnten zudem teilweise auf einem Selbstselektionseffekt beruhen. Beispielsweise indem vor allem die evaluationsaktiven Mitarbeitenden aus dem Mittelschul- und Berufsbildungsamt überdurchschnittlich häufig an der Befragung teilgenommen haben. In Anbetracht der hohen Rücklaufquote, ist diese Art der Verzerrung im Generalsekretariat weniger bedeutend.⁹¹

Abbildung 20: Evaluationspraxis in der ERZ Bern



Anmerkungen: Aktivität = Prozentualer Anteil an Mitarbeitende mit Evaluationsaktivität (mindestens an einer Evaluation im Durchschnitt im Jahr beteiligt). Verwendung = Prozentualer Anteil an Mitarbeitende mit Evaluationsverwendung (mindestens seltener Miteinbezug von Evaluationen in die Arbeit). Total Antwortende: Aktivität = 63; Verwendung = 73.

Quelle: Eigene Darstellung

⁹¹ Allerdings kann die Rücklaufquote der einzelnen Ämter nicht exakt berechnet und verglichen werden, da der Link für die Befragung indirekt über das Generalsekretariat versendet wurde. Es ist aber anzunehmen, dass bei dem Generalsekretariat ein sehr hoher Anteil an Mitarbeitenden an der Befragung teilgenommen hat (69 Mitarbeitende und 41 Antwortende).

Die Evaluationsrecherche zeigt, dass im Generalsekretariat weitaus am meisten evaluiert wird (32 Evaluationsberichte). Dem Amt für Kultur lassen sich vier und dem Amt für Mittelschul- und Berufsbildung fünf Evaluationsberichte zuordnen. Die Evaluationsaktivität ist demzufolge stark innerhalb des Generalsekretariats, in der Abteilung Bildungsplanung und Evaluation, zentralisiert. Dort werden alle Themenbereiche der ERZ Bern evaluiert. Allerdings kommt der Impuls für die Evaluation häufig von den einzelnen Ämtern. Deshalb ist es möglich, dass die Evaluation zwar vom Generalsekretariat gemacht und demzufolge auch dem Generalsekretariat zugeordnet wird, trotzdem aber auch Mitarbeitende aus den anderen Ämtern daran beteiligt sind. Ausserdem wurden für das Mittelschul- und Berufsbildungsamt 30 externe IFES-Schulevaluationen⁹² im Zeitraum 2005 bis 2013 durchgeführt. Diese Schulevaluationen wurden von dem Amt in Auftrag gegeben und finanziert.⁹³ Zudem können diese externen Schulevaluationen die sehr hohen Anteile an Mitarbeitenden mit Evaluationsaktivität im Mittelschul- und Berufsbildungsamt erklären.

Insgesamt lassen sich damit 41 Evaluationsberichte (ohne die externen Schulevaluationen) der ERZ Bern zuordnen. Dies entspricht einer hohen Aktivität. Gleichzeitig findet sich mit 42.5 % ein mittlerer Anteil an Mitarbeitenden mit Evaluationsaktivität.

Einstellung zur Evaluation

Die evaluationsaktiven Mitarbeitenden („non-attitude“-bereinigt) der ERZ Bern haben insgesamt eine neutrale Einstellung zur Evaluation. Auch weisen diese Mitarbeitenden im Schnitt eine deutlich positivere Einstellung gegenüber der Evaluation auf als Mitarbeitende ohne Evaluationsaktivität. Diese Unterschiede sind aber lediglich auf dem Signifikanzniveau von 90 % belegbar.

⁹² Das IFES (Institut für Externe Schulevaluation auf der Sekundarstufe II) ist eine Institution der EDK (Schweizerische Konferenz der kantonalen Erziehungsdirektoren) und führt externe Schulevaluationen auf Sekundarstufe II durch (siehe: www.ifes-ipes.ch).

⁹³ Die Angaben stammen von der Leitung des Bereichs Mittelschulen und Berufsbildung.

Tabelle 27: Einstellung zur Evaluation in der ERZ Bern

	Mittelwert	Min	Max	SA	N
Mitarbeitende mit Evaluationsaktivität	32.8	11	47	8.5	26
Mitarbeitende ohne Evaluationsaktivität	27.7	3	44	10.7	37
T-Test: $t(61) = -1.995+$					
Alle befragten Mitarbeitenden	30.6	3	48	10.3	72

Anmerkung: Die Einstellung zur Evaluation ergibt sich aus der Summierung von fünf Fragen und umfasst den Wertebereich 0 (minimale Einstellung zur Evaluation) bis 50 (maximale Einstellung zur Evaluation). Ein hoher Wert entspricht einer positiveren Einstellung gegenüber der Evaluation. T-Test zum Vergleich der Mittelwerte.

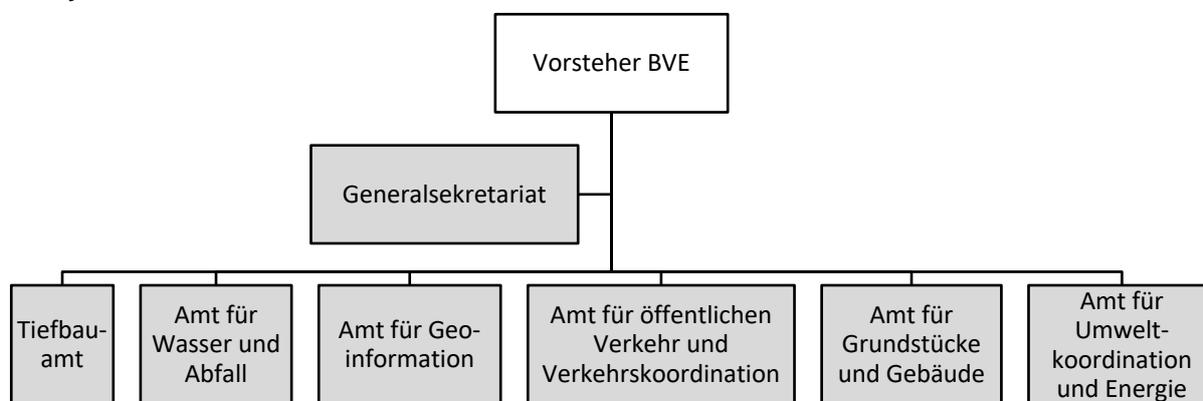
+ = $p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$; 2-seitig.

Quelle: Eigene Darstellung

5.3.3 Bau-, Verkehrs- und Energiedirektion

Der sorgfältige Umgang mit der Landschaft, dem Boden, der Luft und mit dem Wasser stehen im Zentrum der Tätigkeiten der Bau-, Verkehrs- und Energiedirektion (BVE) des Kantons Bern (BVE 2016). Diese Themen werden von sieben verschiedene Einheiten und dem Generalsekretariat verwaltet.

Abbildung 21: Organigramm der Bau-, Verkehrs- und Energiedirektion Bern (Stand August 2014)



Anmerkungen: Die grau schattierten Ämter sind Gegenstand der Untersuchung. Das Rechtsamt wird aus Platzgründen nicht in dem Organigramm aufgeführt.

Quelle: Eigene Darstellung in Anlehnung an BVE (2014)

Die nachfolgende Tabelle 28 zeigt die Charakteristiken der Online-Befragung für die BVE Bern auf. Insgesamt wurden 410 Personen angefragt. Der Rücklauf ist mit 51.0 % im Mittelfeld anzusiedeln.

Von der Befragung ausgeschlossen wurde das Rechtsamt, da die Mitarbeitenden dieses Amtes mehrheitlich über individuellen Bürgerkontakt verfügen. Ebenfalls musste im Tiefbauamt eine erhebliche Reduktion der befragten Mitarbeitenden vorgenommen werden; die Oberingenieurkreise wurden beispielsweise nicht in die Befragung aufge-

nommen. Diese stellen unter anderem den Betrieb und den Unterhalt der Kantonsstrassen sicher.

Tabelle 28: Teilnehmende an der Online-Befragung

Ämter	Mitarbeitende	Angefragt	Antworten	Rücklauf (in Prozent)
Generalsekretariat	69	88	36	40.9
Tiefbauamt	521	72	37	51.4
Amt für Geoinformation	33	40	15	37.5
Amt für Wasser und Abfall	127	117	48	41.0
Amt für öffentlichen Verkehr und Verkehrskoordination	16	17	12	70.6
Amt für Grundstücke und Gebäude	72	59	24	40.7
Amt für Umweltkoordination und Energie	17	18	10	55.6
Keine Angaben	-	-	27	-
Total	873	410	209	51.0

Anmerkungen: Die Daten bezüglich der Anzahl an Mitarbeitenden stammen von der Webseite und den Rückmeldungen des Personalamtes des Kantons Bern (Personalamt 2014) und geben den Gesamtpersonalbestand der Bau-, Verkehrs- und Energiedirektion per Dezember 2014 wieder.

Quelle: Eigene Darstellung

Institutionalisierung der Evaluation

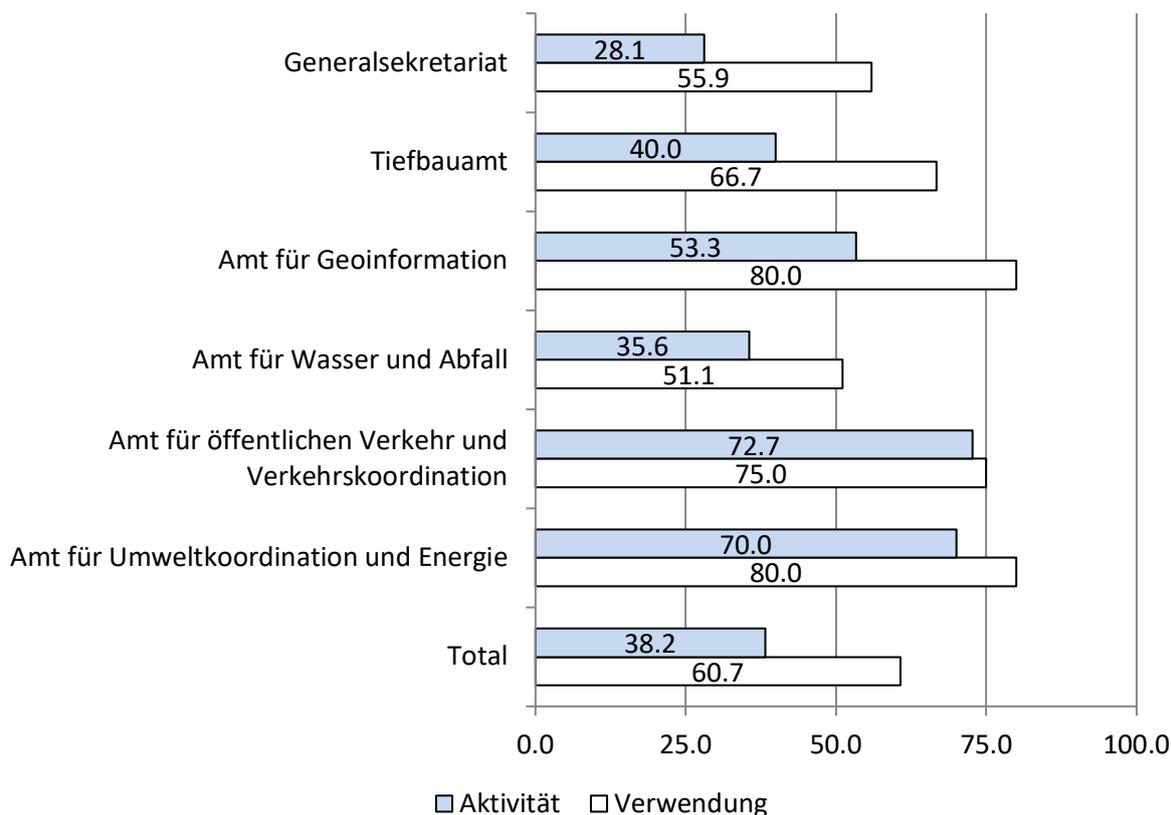
Die Evaluation ist in der BVE Bern nicht institutionalisiert. Es gibt keine Evaluationsstelle und keine spezifischen Grundlegendokumente betreffend Evaluation. Dies wird auch von einer Mehrheit der Mitarbeitenden so wahrgenommen: Von den 186 Befragten aus dem Generalsekretariat geben lediglich 14 Befragte an (entspricht 7.5 %), dass ihr Amt über eine Evaluationsstelle verfügt (83 geben an, es gebe keine Evaluationsstelle und 89 wissen es nicht).

Evaluationspraxis

Die Selbstangaben der Mitarbeitenden über ihre Evaluationspraxis sind in der nachfolgenden Abbildung 22 ersichtlich. Auffallend sind die hohen Anteile im Amt für öffentlichen Verkehr und Verkehrskoordination und im Amt für Umweltkoordination und Energie. Die Mitarbeitenden aus dem Amt für Wasser und Abfall und dem Generalsekretariat geben zu wesentlich kleineren Anteilen an, über Evaluationsaktivität und Evaluationsverwendung zu verfügen. Die Angaben der Mitarbeitenden aus dem Amt

für Grundstücke und Gebäude werden aufgrund von Falschangaben für die Deskription nicht berücksichtigt. Nach Rücksprache mit der Amtsleitung davon ausgegangen werden, dass in diesem Amt keine Evaluation durchgeführt wurde.⁹⁴

Abbildung 22: Evaluationspraxis in der BVE Bern



Anmerkungen: Aktivität = Prozentualer Anteil an Mitarbeitenden mit Evaluationsaktivität (mindestens an einer Evaluation im Durchschnitt im Jahr beteiligt). Verwendung = Prozentualer Anteil an Mitarbeitenden mit Evaluationsverwendung (mindestens seltener Miteinbezug von Evaluationen in die Arbeit). Total Antwortende: Aktivität = 165; Verwendung = 173.

Quelle: Eigene Darstellung

Die Evaluationsrecherche zeigt, dass in der BVE Bern sehr viele Evaluationen in Auftrag gegeben oder durchgeführt werden: Dem Tiefbauamt können neun Evaluationen, dem Amt für Wasser und Abfall zehn, dem Amt für öffentlichen Verkehr 20, dem Amt

⁹⁴ Der Grund für die Angaben der Evaluationsaktivität könnte sein, dass in diesem Amt sogenannte Planer- und Unternehmerevaluation durchgeführt werden. Dabei geht es aber nicht um Evaluationen im hier untersuchten Sinn, sondern um die Beurteilung der Vergabe von Dienstleistungen. Aus diesem Grund werden die Angaben zur Evaluationspraxis der Mitarbeitenden aus dem Amt für Grundstücke und Gebäude, die angegeben haben über Evaluationsaktivität zu verfügen, für die weiteren Analysen nicht berücksichtigt.

für Geoinformation drei und dem Amt für Umweltkoordination und Energie neun Evaluationen zugeordnet werden. Keine Evaluationsberichte lassen sich dem Generalsekretariat zuordnen.

Auch bei diesem Generalsekretariat handelt es sich um eine Stabstelle, in der die Mitarbeitenden bei der Erstellung einer Evaluation beteiligt sein können, obwohl dieser Verwaltungseinheit direkt keine Evaluation zugeordnet werden kann.

Mit insgesamt 51 recherchierten Evaluationsberichten kann der BVE Bern die höchste Evaluationsaktivität zugeordnet werden. Gleichzeitig gibt es mit 38.2 % einen tiefen Anteil an Mitarbeitenden, die über Evaluationsaktivität verfügen.

Einstellung zur Evaluation

Der Mittelwert der Einstellung zur Evaluation ist bei Mitarbeitenden mit Evaluationsaktivität mit einem Wert von 34.0 im positiven Bereich. Damit ist die BVE die Berner Direktion mit der positivsten Einstellung zur Evaluation.

Auffallend ist zudem die statistisch nachweisbare Differenz der Einstellungswerte zwischen Mitarbeitenden mit Evaluationsaktivität und Mitarbeitenden ohne Evaluationsaktivität. Diese fällt mit einem Unterschied von 5.5 Punkten grösser aus als in den meisten anderen Departementen und Direktionen.

Tabelle 29: Einstellung zur Evaluation in der BVE Bern

	Mittelwert	Min	Max	SA	N
Mitarbeitende mit Evaluationsaktivität	34.0	15	48	6.5	62
Mitarbeitende ohne Evaluationsaktivität	28.5	3	50	9.8	98
T-Test: $t(158) = -3.911^{**}$					
Alle befragten Mitarbeitenden	30.9	3	50	8.9	174

Anmerkung: Die Einstellung zur Evaluation ergibt sich aus der Summierung von fünf Fragen und umfasst den Wertebereich 0 (minimale Einstellung zur Evaluation) bis 50 (maximale Einstellung zur Evaluation). Ein hoher Wert entspricht einer positiveren Einstellung gegenüber der Evaluation. T-Test zum Vergleich der Mittelwerte.

+ = $p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$; 2-seitig.

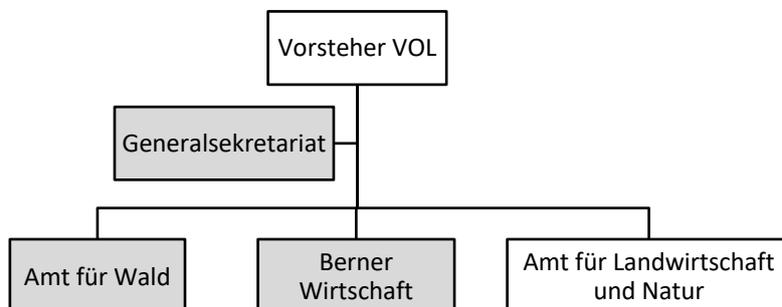
Quelle: Eigene Darstellung

5.3.4 Volkswirtschaftsdirektion

Die Volkswirtschaftsdirektion (VOL) Bern umfasst die Themenbereiche Wirtschaft, Landwirtschaft und Wald (siehe VOL 2016).

Wie der nachfolgenden Abbildung 23 entnommen werden kann, gliedern sich entlang dieser Themenbereiche drei Einheiten und das Generalsekretariat.

Abbildung 23: Organigramm der Volkswirtschaftsdirektion Bern (Stand August 2014)



Anmerkungen: Die grau schattierten Ämter sind Gegenstand der Untersuchung.
Quelle: Eigene Darstellung in Anlehnung an (VOL 2014)

Die nachfolgende Tabelle 30 zeigt für die VOL Bern eine ausserordentlich hohe Rücklaufquote von 72.3 %.

Das Amt für Landwirtschaft und Natur wurde als Ganzes ausgeschlossen, da die grosse Mehrheit der Mitarbeitenden dieses Amtes direkten Bürgerkontakt haben und nicht primär Verwaltungstätigkeiten nachgehen. Beispielsweise stehen Mitarbeitende der Abteilung für Naturgefahren vorwiegend im individuellen Bürgerkontakt und werden aus diesem Grund aus der Befragung ausgeschlossen. Ebenso trifft dies bei einigen Ämtern des Amts für Wald und der Berner Wirtschaft zu, entsprechend reduzierte sich die Zahl der angefragten Personen.

Tabelle 30: Teilnehmende an der Online-Befragung

Ämter	Mitarbeitende	Angefragt	Antworten	Rücklauf (in Prozent)
Generalsekretariat	45	42	31	73.8
Amt für Wald	176	30	24	80.0
Berner Wirtschaft	173	141	86	61.0
Keine Angaben	-	-	12	-
Total	875	213	154	72.3

Anmerkungen: Die Daten bezüglich der Anzahl an Mitarbeitenden stammen von der Webseite und den Rückmeldungen des Personalamtes des Kantons Bern (Personalamt 2014) und geben den Gesamtpersonalbestand der Volkswirtschaftsdirektion per Dezember 2014 wieder.
Quelle: Eigene Darstellung

Institutionalisierung der Evaluation

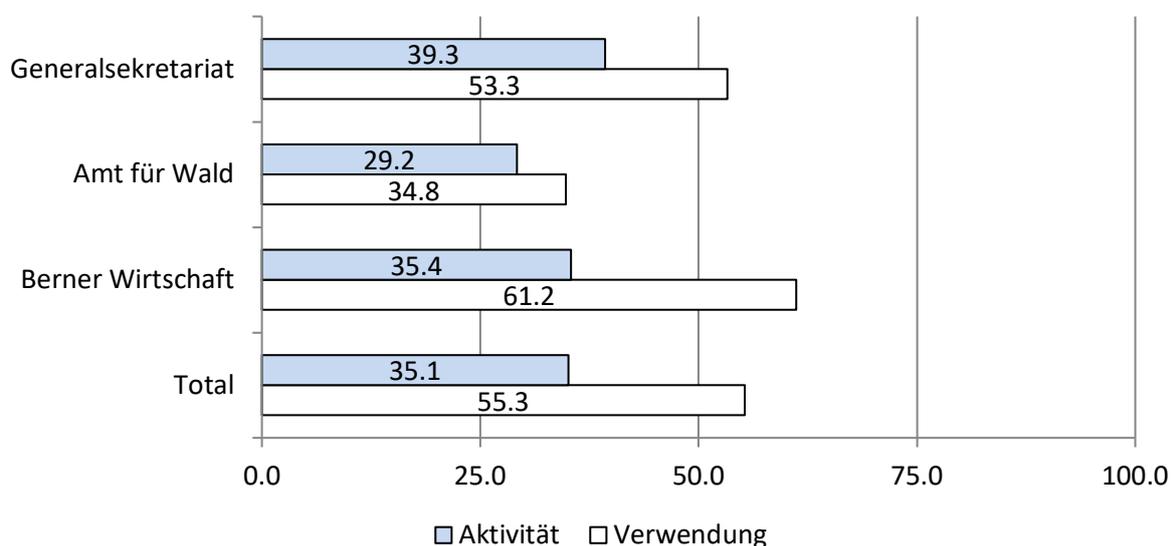
Die VOL Bern hat keine Evaluationsstelle und auch keine spezifischen Dokumente, die den Umgang oder die Erstellung von Evaluation regeln.

Die Gegenüberstellung der Selbstangaben der Mitarbeitenden der Volkswirtschaftsdi- rektion bezüglich des Vorhandenseins einer Evaluationsstelle zeigt, dass die überwie- gende Mehrheit aussagt, dass keine solche existiert (lediglich 4.1 % bejahen diese Fra- ge). Fünf der sechs Befragten, die angegeben haben in ihrer Verwaltungseinheit über eine Evaluationsstelle zu verfügen, können dem Amt Berner Wirtschaft zugeordnet werden.⁹⁵

Evaluationspraxis

Bei der Betrachtung der nachstehenden Abbildung 24 der Evaluationspraxis fällt auf, dass ein hoher Anteil (61.2 %) der Mitarbeitenden aus dem Amt Berner Wirtschaft angegeben hat, Evaluationen in ihre Arbeit einzubeziehen. Die Anteile an Mitarbeiten- den mit Evaluationsaktivität sind in den drei Verwaltungseinheiten, mit Werten zwi- schen 29.2 % und 39.3 %, ähnlich hoch.

Abbildung 24: Evaluationspraxis in der VOL Bern



Anmerkungen: Aktivität = Prozentualer Anteil an Mitarbeitenden mit Evaluationsaktivität (mindestens an einer Evaluation im Durchschnitt im Jahr beteiligt). Verwendung = Prozentualer Anteil an Mitarbeitenden mit Evaluationsverwendung (mindestens seltener Miteinbezug von Evaluationen in die Arbeit). Total Antwortende: Aktivität = 134; Verwendung = 141.

Quelle: Eigene Darstellung

* Ein möglicher Grund der Wahrnehmung einer Evaluationsstelle der Mitarbeitenden könnten die in der Standortförderung gemachten "Evaluationsbesuche" sein. Dabei geht es um die Unterstützung bei der Standortwahl von Unternehmungen durch Spezialistinnen und Spezialisten und um keine Evaluation in der hier untersuchten Form (siehe dazu VOL beco SF 2016).

Dem Amt für Wald können insgesamt fünf, der Berner Wirtschaft 20 in Auftrag gegebene oder durchgeführte Evaluationen zugeordnet werden. Dem Generalsekretariat lässt sich dagegen keine Evaluation zuordnen. Trotzdem ist es denkbar, dass einige Mitarbeitende des Generalsekretariats an Evaluationen beteiligt sind, da dieses unter anderem Koordinationsaufgaben wahrnimmt (VOL GS 2016).

Mit gesamthaft 25 Evaluationen lässt sich bei der VOL Bern eine mittlere Evaluationsaktivität nachweisen. Der Anteil an Mitarbeitenden mit Evaluationsaktivität ist mit 35.1 % im Vergleich zu den anderen Verwaltungseinheiten sehr tief.

Einstellung zur Evaluation

Zusammengefasst kommen die Einstellungsmittelwerte der evaluationsaktiven Mitarbeitenden der VOL Bern dem neutralen Bereich am nächsten.

Vergleicht man die Mittelwerte der Einstellung zur Evaluation mit den Werten in den anderen Direktionen in Bern, verfügt die VOL Bern über die negativste Einstellung zur Evaluation.⁹⁶

Tabelle 31: Einstellung zur Evaluation in der VOL Bern

	Mittelwert	Min	Max	SA	N
Mitarbeitende mit Evaluationsaktivität	31.4	0	46	10.0	45
Mitarbeitende ohne Evaluationsaktivität	28.7	2	50	9.9	86
T-Test: $t(46) = -1.529$					
Alle befragten Mitarbeitenden	29.5	0	50	9.8	141

Anmerkung: Die Einstellung zur Evaluation ergibt sich aus der Summierung von fünf Fragen und umfasst den Wertebereich 0 (minimale Einstellung zur Evaluation) bis 50 (maximale Einstellung zur Evaluation). Ein hoher Wert entspricht einer positiveren Einstellung gegenüber der Evaluation. T-Test zum Vergleich der Mittelwerte.

+ = $p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$; 2-seitig.

Quelle: Eigene Darstellung

5.4 Überblick der Departemente und Direktionen

Nach der Vorstellung der einzelnen Departemente und Direktionen, werden die Verwaltungseinheiten in diesem Unterkapitel zusammenfassend nochmals dargestellt. Auf

⁹⁶ Bei der VOL lässt sich der negativste Einstellungswert bei der höchsten Rücklaufquote finden. Die Verzerrung aufgrund der Selbstselektion wurde zudem im Unterkapitel 4.2 besprochen.

eine Beurteilung und Ursachenerklärung für allfällige Unterschiede wird in diesem Überblick bewusst verzichtet. Auch ein systematischer Vergleich der Evaluationskultur zwischen unterschiedlichen Politikfeldern oder zwischen den drei Kantonen wird unterlassen und lediglich punktuell auf besonders offensichtliche Unterschiede hingewiesen. Ein Vergleich dieser Art wurde bereits in dem Beitrag von Dolder et al. (2017), der teilweise auch auf Daten der vorliegenden Untersuchung basiert vollzogen.⁹⁷

Institutionalisierung der Evaluation

Die Institutionalisierung der Evaluation in den neun Departementen und Direktionen ist in nachfolgender Tabelle 32 dargestellt. Die Tabelle zeigt, dass die Evaluation in den untersuchten Verwaltungseinheiten unterschiedlich stark institutionalisiert ist. Ausserdem ist die ganze Bandbreite an verschiedenen Ausprägungen der Institutionalisierung zu finden.

Lediglich in einer der neun Verwaltungseinheiten ist die Evaluation auf Departementsstufe institutionalisiert: In der Erziehungsdirektion (ERZ) des Kantons Bern bestehen eine Evaluationsstelle sowie spezifische Grundlegendokumente betreffend Evaluation. Beim Bildungs- und Kulturdepartement (BKD) des Kantons Luzern kann von einer teilweisen Institutionalisierung gesprochen werden, wo die Evaluation innerhalb der Dienststelle Volksschulbildung⁹⁸ institutionalisiert ist. Die Dienststelle Volksschulbildung verfügt über eine Evaluationsstelle sowie über Dokumente betreffend Evaluation.

⁹⁷ Der Vergleich berücksichtigt die gleichen drei Kantone wie die vorliegende Untersuchung und untersucht die beiden Politikfelder Gesundheit und Bildung. Anhand der Auswertung der Entwicklung der Periode 2000 bis 2013 kann aufgezeigt werden, dass es einen politikfeldspezifischen Entwicklungspfad der Evaluationskultur gibt. Eine kantonsspezifische Entwicklung lässt sich für die sechs Verwaltungseinheiten hingegen nicht ausmachen.

⁹⁸ Ebenso ist die externe Schulevaluation, die mit einer spezifischen Einheit lediglich im Bildungs- und Kulturdepartement des Kantons Luzern zu finden ist, im Bereich Volksschule des BKD institutionalisiert.

Tabelle 32: Evaluationsinstitutionalisierung

	Department/Direktion	Grundlagendokumente	Evaluationsstelle
Luzern	Gesundheits- und Sozialdepartement	Nein	Nein
	Bildungs- und Kulturdepartement	Teilweise	Teilweise
	Bau- Umwelt- und Wirtschaftsdepartement	Nein	Nein
Basel-Stadt	Gesundheitsdepartement	Nein	Ansatzweise
	Erziehungsdepartement	Nein	Ansatzweise
Bern	Gesundheits- und Fürsorgedirektion	Nein	Nein
	Erziehungsdirektion	Ja	Ja
	Bau- Verkehrs- und Energiedirektion	Nein	Nein
	Volkswirtschaftsdirektion	Nein	Nein

Anmerkungen: Dunkelgraue Schattierung = Institutionalisierung. Mittelgraue Schattierung = teilweise Institutionalisierung. Hellgraue Schattierung = Ansätze einer Institutionalisierung. Keine Schattierung = keine Institutionalisierung.

Quelle: Eigene Darstellung

Bei den Departementen des Kantons Basel-Stadt kann aufgrund gewisser Ansätze einer Evaluationsstelle und fehlender Dokumente betreffend Evaluation von einer ansatzweisen Institutionalisierung gesprochen werden. Die Steuergruppe Evaluation im Gesundheitsdepartement (GD) des Kantons Basel-Stadt stellt eine leichte Form der Institutionalisierung dar und ist nur für den Bereich Gesundheitsdienste zuständig. Gleich verhält es sich mit dem Erziehungsdepartement (ED) des Kantons Basel-Stadt, bei dem es Ansätze der Institutionalisierung im Bereich Jugend, Familie und Sport gibt. Dieser Bereich verfügt über ein kleines Team, das Evaluationsfragen formuliert und kleinere Wirkungsanalysen und Evaluationen selber durchführt.

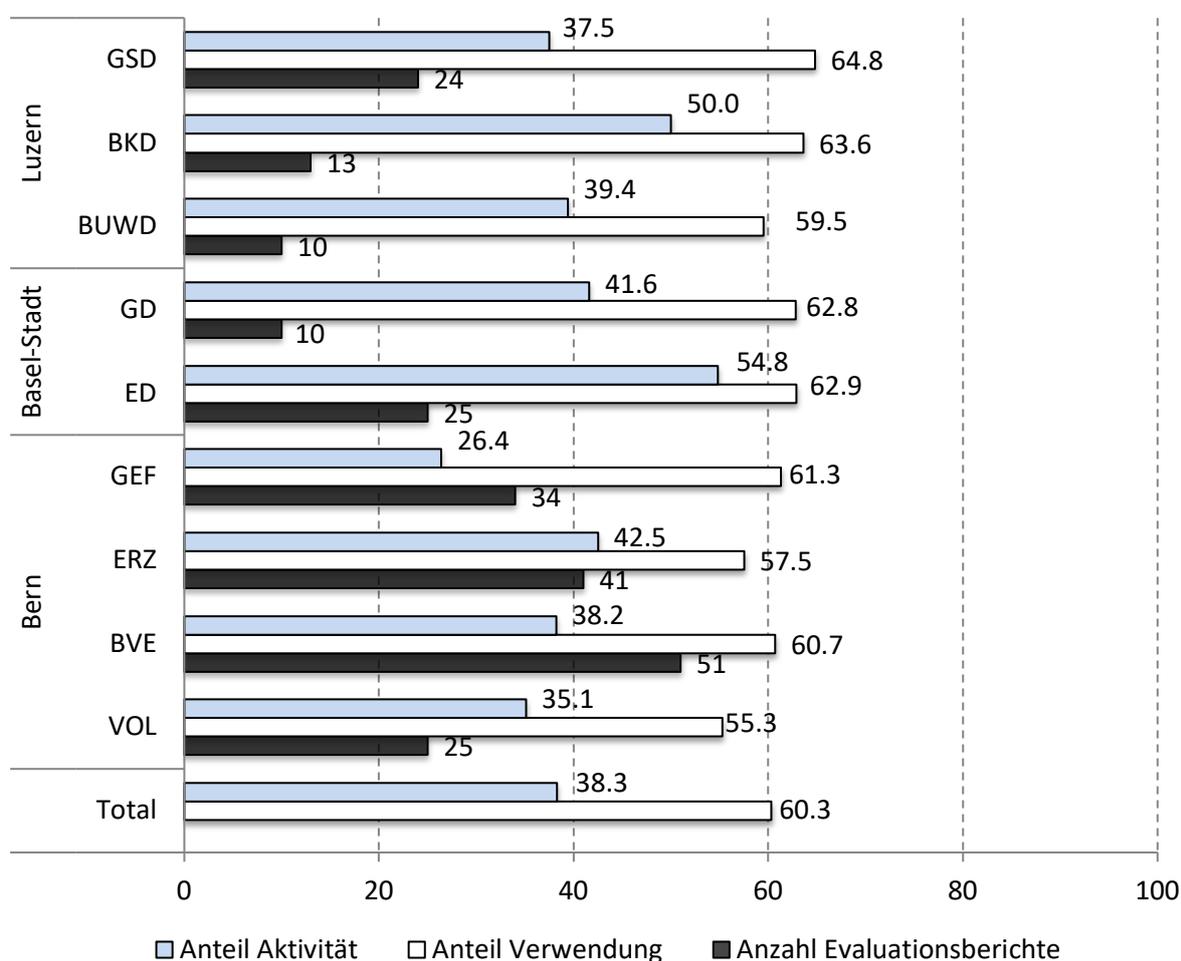
In den anderen fünf Verwaltungseinheiten ist die Evaluation nicht institutionalisiert: Im Kanton Luzern ist die Evaluation im Gesundheits- und Sozialdepartement (GSD) und im Bau- Umwelt- und Wirtschaftsdepartement (BUWD) nicht institutionalisiert. Im Kanton Bern ist die Evaluation weder in der Volkswirtschaftsdirektion (VOL), noch in der Bau- Verkehrs- und Energiedirektion (BVE) oder der Gesundheits- und Fürsorgedirektion (GEF) institutionalisiert. Zu erwähnen ist die Situation in der GEF Bern. Hier war die Evaluation bis 2004 institutionalisiert, danach wurde sie im Zuge des ERKOS-Programmendes deinstitutionalisiert.

Evaluationspraxis

In der Einleitung dieses Kapitels wurde darauf hingewiesen, dass die Evaluationsaktivität mittels zwei unterschiedlicher Quellen ermittelt wird. Die folgende Abbildung 25

stellt die Ergebnisse der Online-Mitarbeitendenbefragung mit der Onlinerecherche nach Evaluationsberichten gegenüber. Der schwarze Balken gibt die Höhe der Evaluationsaktivität wieder (Evaluationsrecherche für den Zeitraum 2005 bis 2013) und zeigt die Anzahl, der von den Verwaltungseinheiten durchgeführten oder in Auftrag gegebenen Evaluationen. Der blaue Balken gibt die prozentualen Anteile von Mitarbeitenden mit Evaluationsaktivität (Selbstangaben) an. Als Drittes sind ebenfalls die Anteile der Mitarbeitenden mit Evaluationsverwendung aufgeführt.

Abbildung 25: Evaluationspraxis unterschieden nach Departementen und Direktionen



Anmerkungen: Anteil Aktivität = Prozentualer Anteil an Mitarbeitenden mit Evaluationsaktivität (Selbstangaben). Anteil Verwendung = Prozentualer Anteil an Mitarbeitenden mit Evaluationsverwendung (mindestens seltener Miteinbezug von Evaluationen in die Arbeit). Anzahl Evaluationsberichte = Anzahl der Evaluationen aufgrund der Evaluationsrecherche. Dargestellt wird die Anzahl an Politikevaluationen. Schulevaluationen werden nicht dargestellt. Ebenso ist die Gesamtanzahl der Evaluationsberichte aus Darstellungsgründen nicht aufgeführt.
Quelle: Eigene Darstellung

Die Evaluationsrecherche (schwarzer Balken) zeigt, dass auf Ebene der Direktionen und Departemente keine „weissen Flecken“ zu finden sind. Eine gewisse Evaluationsaktivität ist bei allen neun Verwaltungseinheiten vorhanden. Wie die vorangehenden Unterkapitel zeigen, ergeben sich innerhalb der Verwaltungseinheiten teilweise grössere Differenzen. Gemessen an der Anzahl der Evaluationsberichte verfügen die Berner Verwaltungseinheiten über die höchste Evaluationsaktivität. Am meisten wird in der BVE Bern evaluiert, gefolgt von der ERZ Bern und der GEF Bern. Über eine mittlere Evaluationsaktivität verfügen die VOL Bern, das ED Basel-Stadt und das GSD Luzern. Die kleinste Evaluationsaktivität haben das BKD Luzern, das BUWD Luzern und das GD Basel-Stadt.

Betrachten wir als nächstes die Anteile der Mitarbeitenden, die Evaluationsaktivitäten angegeben haben (blauer Balken in Abbildung 25). Wie im vierten Kapitel erwähnt, sind diese Angaben individuelle Einschätzung der Mitarbeitenden und geben einen Anhaltspunkt wie verbreitet die Evaluationsaktivität in der Verwaltung ist. Die höchsten Anteile sind in den drei Verwaltungseinheiten, welche die Bildung enthalten zu finden: für das ED Basel-Stadt 54.8 %, das BKD Luzern 50.0 % und die ERZ Bern 42.5 %. An dieser Stelle ist nochmals darauf hinzuweisen, dass diese Verwaltungseinheiten neben der Bildung auch für andere Themenbereiche zuständig sind. Die tiefsten Anteile an evaluationsaktiven Mitarbeitenden sind der GEF Bern, der BVE Bern, der VOL Bern und dem GSD Luzern zuzuordnen.

Wie sind nun aber die Unterschiede zwischen den beiden soeben beschriebenen Charakteristiken zu interpretieren? Erstens ist darauf hinzuweisen, dass die Online-Mitarbeitendenbefragung und die Onlinerecherche jeweils unterschiedliche Charakteristiken der Evaluationsaktivität wiedergeben: Die Befragung gibt die Anteile der Mitarbeitenden mit Evaluationsaktivität pro Verwaltungseinheit wieder, während die Evaluationsrecherche die gesamte Anzahl in Auftrag gegebener oder durchgeführter Evaluationen pro Verwaltungseinheit abbildet. Als Zweites fällt auf, dass gerade in den Berner Verwaltungen trotz hoher Evaluationsaktivität, diese nicht verbreiteter ist als in den anderen Verwaltungseinheiten. Ein wichtiger Grund für diese Situation ist wahr-

scheinlich die Grösse der Verwaltung. Die Berner Verwaltungseinheiten in unserem Vergleich über die mitarbeiterstärksten Verwaltungen (siehe Tabelle 33). Bei diesen führt, die zwar hohe Anzahl an Evaluationsberichte, aufgrund der hohen Mitarbeiterzahlen nicht sehr stark ins Gewicht. Beispielsweise können der BVE Bern sehr viele Evaluationsberichte zugeordnet werden und es gibt lediglich geringe Anteile von Mitarbeitenden mit Evaluationsaktivität und damit anteilmässig nicht viele die in den Evaluationsprozess involviert sind. Die gleiche Interpretation lässt sich für die Situation der GEF Bern, der VOL Bern und der ERZ Bern machen. Auch die Werte des BUWD Luzern und des GD Basel-Stadt entsprechen für einen Zusammenhang in dieser Richtung – beides Departemente mit wenig Mitarbeitenden, die über eine breitere Einbindung (hohen Anteil) der Mitarbeitenden in den Evaluationsprozess verfügen und über eine geringe Anzahl an Evaluationsberichten verfügt. Dies könnte auch daran liegen, dass die Mitarbeitenden in kleineren Verwaltungseinheiten eher eine „Generalistenfunktion“ ausüben und häufiger in Kontakt mit Evaluationen kommen. Als Drittes ist bei den Werten für das BKD Luzern und ED Basel-Stadt die Schulevaluation zu berücksichtigen. Bei diesen Verwaltungseinheiten gaben sehr viele Mitarbeitende an, über Evaluationsaktivität zu verfügen. Gleichzeitig lassen sich diesen beiden Verwaltungseinheiten nur eine geringe Anzahl an Berichten zuordnen. Es ist davon auszugehen, dass sich diese Differenz durch die externen Schulevaluationen erklären lässt, die bei der Anzahl Evaluationsberichte Abbildung 25 nicht hinzugerechnet wurden.⁹⁹

Bei der Betrachtung der Evaluationsaktivität der Verwaltungseinheiten, wie sie in diesem Abschnitt dargestellt wird, ist man geneigt Schlüsse auf die Politikbereiche zu ziehen. In diesem Zusammenhang ist aber stets zu bedenken, dass sich die Departemente und Direktionen sehr unterschiedlich zusammensetzen und ein präziser Vergleich nur erschwert zu vollziehen ist. Beispielsweise ist die ERZ Bern, neben der Bil-

⁹⁹ Wie bereits erwähnt, wurde in der Befragung zwar die Definition von Politikevaluation angeführt, die Schulevaluation wurde aber nicht explizit als „keine Politikevaluation“ definiert. Es ist anzunehmen, dass einige Befragte ihre Angaben auf die Schulevaluation gemacht haben.

dung, auch für die Kultur zuständig. In Basel-Stadt dagegen wird im ED die Kultur nicht betreut, dafür aber die Themenbereiche Jugend und Sport.

Bereits im vierten Kapitel wurde auf eine Verzerrung der Befragungsergebnisse aufgrund einer selektiven Teilnahme verwiesen (Selbstselektion) und im Unterkapitel 4.2 analysiert ob bei Nichtteilnehmenden Effekte in dieser Hinsicht zu erkennen sind („Non-Response“-Analyse). Die Analyse ergab keine Hinweise für eine Verzerrung solcher Art. Auch bei der vorliegenden Betrachtung der Evaluationsaktivität bietet sich ein Analyse an. Ein möglicher Hinweis auf eine starke Selbstselektion ist, wenn Mitarbeitende in Verwaltungseinheiten, die über eine hohe Evaluationsaktivität verfügen sich vermehrt an der Umfrage beteiligt haben. Die nachfolgende Tabelle 33 offenbart keine Anhaltspunkte für einen konsistenten Zusammenhang in diese Richtung. In der VOL Bern haben trotz kleinem Anteil von Mitarbeitenden mit Evaluationsaktivität und relativ (zur Anzahl Mitarbeitenden) kleiner Anzahl an Evaluationsberichten, überdurchschnittlich viele (72.3 %) Mitarbeitende an der Befragung teilgenommen. Des Weiteren hat ein hoher Anteil an Mitarbeitenden des BUWD Luzern und des GD Basel-Stadt, die über eine tiefe Evaluationsaktivität verfügen, an der Befragung teilgenommen.

Tabelle 33: Übersicht der Evaluationsaktivität

Department/Direktion		Anteile in %	Anzahl Berichte	Rücklauf in %	Mitarbeitende
Luzern	Gesundheits- und Sozialdepartement	37.5	24	38.2	170
	Bildungs- und Kulturdepartement	50.0	13	57.1	105
	Bau- Umwelt- und Wirtschaftsdepartement	39.4	10	60.8	125
Basel-Stadt	Gesundheitsdepartement	41.6	10	60.9	179
	Erziehungsdepartement	54.8	25	56.3	64
Bern	Gesundheits- und Fürsorgedirektion	29.8	34	53.4	234
	Erziehungsdirektion	42.5	41	25.3	289
	Bau- Verkehrs- und Energiedirektion	38.2	51	51.0	410
	Volkswirtschaftsdirektion	35.1	25	72.3	213
Total		38.3	226	50.7	907

Anmerkungen: Anteile = Prozentuale Anteile von Mitarbeitenden mit Evaluationsaktivität (Selbstangaben). Anzahl Berichte = Anzahl an Evaluationsberichten (Evaluationsrecherche ohne Schulevaluationen). Rücklauf = Prozentualer Anteil an Antwortenden. Mitarbeitende = Anzahl Mitarbeitende von zentralen Verwaltungseinheiten.

Quelle: Eigene Darstellung

Als nächstes werden die Anteile der Mitarbeitenden, die angeben, Evaluationen in ihre Arbeit einzubeziehen betrachtet. Diese Anteile sind in der Abbildung 25 aufgeführt

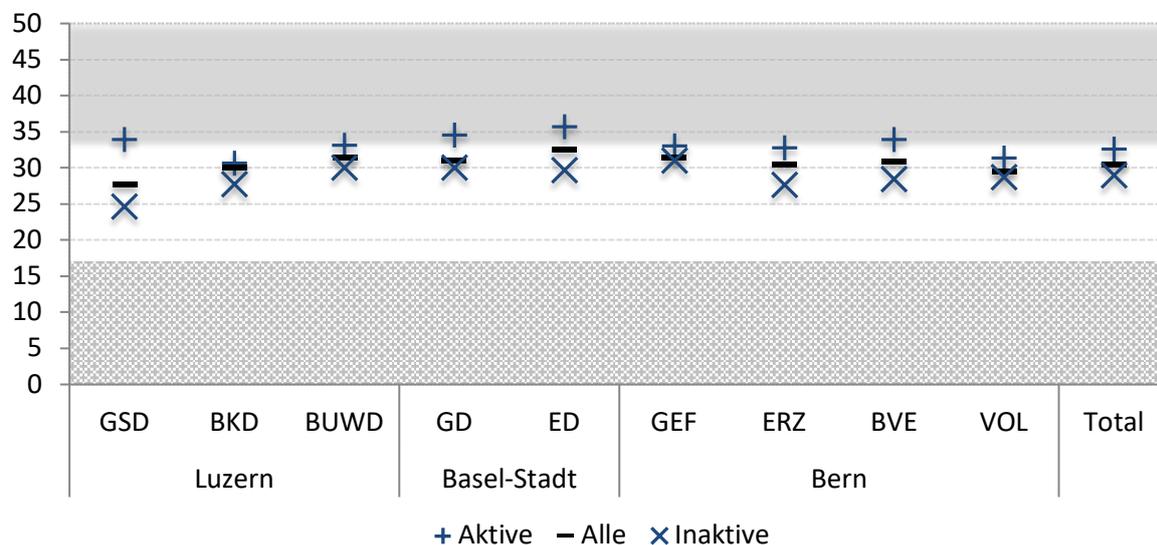
und zeigen diesbezüglich lediglich sehr geringe Unterschiede zwischen den Departementen und Direktionen auf. Rund 40 % der Befragten ziehen nie Evaluationen in ihre Arbeit mit ein. Die Anteile mit mindestens seltener Evaluationsverwendung reichen von 55.3 % für die VOL Bern bis zu 64.8 % für das GSD Luzern. Allerdings ist in diesem Zusammenhang zu erwähnen, dass in den vorherigen Unterkapiteln durchaus Unterschiede innerhalb der Departemente und Direktionen aufgezeigt werden konnten: Beispielsweise lässt sich die Feststellung machen, dass die Departementssekretariate und Generalsekretariate häufig über die tiefsten Anteile an Mitarbeitenden mit Evaluationsverwendung aufweisen, verglichen mit den anderen Verwaltungseinheiten.

Einstellung zur Evaluation

Sämtliche Einstellungswerte der Departemente und Direktionen liegen im neutralen oder positiven Bereich.¹⁰⁰ Verwaltungseinheiten, die der Evaluation gegenüber insgesamt negativ eingestellt sind (unterer grau schraffierter Bereich in Abbildung 26), gibt es nicht. Zudem sind die Einstellungswerte der Departemente und Direktionen generell sehr ähnlich: Berücksichtigen wir die Angaben der Mitarbeitenden mit Evaluationsaktivität („non-attitude“-bereinigte Stichprobe) reichen die Werte von 30.0 (BKD Luzern) bis zu dem Maximalwert von 35.7 (ED Basel-Stadt). Von den neun Verwaltungseinheiten verfügen vier über eine positive Einstellung, die anderen fünf verfügen über eine neutrale Einstellung zur Evaluation.

¹⁰⁰ Die Einstufung in die Wertebereiche in positive, negative und neutrale Einstellung zur Evaluation als auch die konkrete Operationalisierung und Bedeutung der Einstellung zur Evaluation wurde im vierten Kapitel vollzogen und kann dort nachvollzogen werden.

Abbildung 26: Einstellung zur Evaluation unterschieden nach Evaluationsaktivität



Anmerkungen: Die Einstellung zur Evaluation ergibt sich aus der Summierung von fünf Fragen und umfasst den Wertebereich 0 (minimale Einstellung zur Evaluation) bis 50 (maximale Einstellung zur Evaluation). Ein hoher Wert entspricht einer positiven Einstellung gegenüber der Evaluation. Untere graue Schattierung = Negative Einstellung zur Evaluation. Obere graue Schattierung = Positive Einstellung zur Evaluation. Keine Schattierung = Neutrale Einstellung zur Evaluation.

Quelle: Eigene Darstellung

Die positivste Einstellung zur Evaluation haben die Mitarbeitenden mit Evaluationsaktivität im ED Basel-Stadt. Ausserdem weisen die evaluationsaktiven Mitarbeitenden aus dem GD Basel-Stadt, dem GSD Luzern und der BVE Bern Werte im positiven Bereich auf. Die negativste Einstellung lässt sich bei den Mitarbeitenden des BKD Luzern und den Mitarbeitenden der VOL Bern verorten. Auffallend sind die beiden Departemente von Basel-Stadt, die über die positivsten Durchschnittswerte verfügen. Darüber hinaus lassen sich entlang der Kantone keine weiteren konsistenten Schlüsse ziehen. In Luzern haben die Mitarbeitenden des GSD Luzern eine positive Einstellung, das BKD Luzern und das BUWD Luzern haben eine neutrale Einstellung zur Evaluation. In Bern kann lediglich die BVE Bern in den positiven Einstellungsbereich eingestuft werden.

Zudem gibt es Divergenzen in der Einstellung zur Evaluation zwischen Mitarbeitenden unterschiedlicher Evaluationsaktivität: Mitarbeitende, die angeben Evaluationsaktivitäten auszuführen, haben in allen Verwaltungseinheiten eine positivere Einstellung zur Evaluation als Mitarbeitende ohne Evaluationsaktivität. Die Abbildung 26 offenbart,

dass die Mittelwerte zwischen evaluationsaktiven und -inaktiven Mitarbeitenden im GSD Luzern, im ED Basel-Stadt sowie in der BVE Bern am stärksten auseinandergehen. Statistisch nachweisbar sind die Unterschiede für das GSD Luzern, das GD Basel-Stadt und die BVE Bern. Ein Grund für die grosse Divergenz im GSD Luzern kann nicht eindeutig identifiziert werden. Wichtig ist die Bemerkung, dass diese Divergenzen besonders stark in der Dienststelle Soziales und Gesellschaft ausgeprägt ist.

Auf eine Verzerrung der Befragungsergebnisse aufgrund der Selbstselektion wurde bereits verwiesen, unter anderem auch im vorherigen Abschnitt der Evaluationsaktivität. Auch für die Einstellung zur Evaluation soll im Folgenden die Rücklaufquoten auf Hinsicht dieser Verzerrung betrachtet werden. Demgemäss sollten in Verwaltungseinheiten mit einer hohen Rücklaufquote auch generell positivere Einstellungswerte zu beobachten sein, da sich Mitarbeitende mit einer negativeren Einstellung weniger häufig beteiligen. Einen solchen Zusammenhang ist, bei der bivariaten Betrachtung, nicht konsistent beobachtbar. Zwar verfügt das ED Basel-Stadt bei einer hohen Rücklaufquote (56.3 %) über die positivste Einstellung. Demgegenüber steht aber die Situation in der VOL Bern, die trotz der höchsten Rücklaufquote (72.3 %) einen vergleichsweise tiefen Einstellungswert aufweist.

6 Wie lassen sich die Unterschiede in der Evaluationskultur erklären?

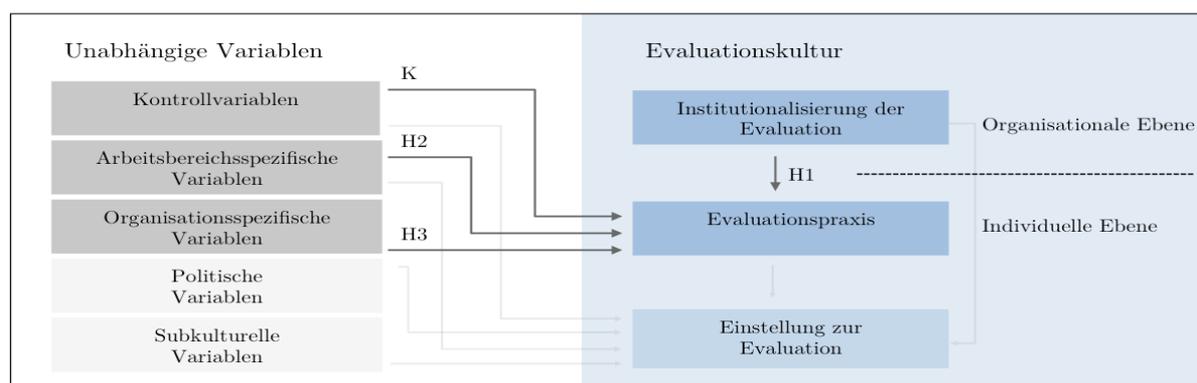
In diesem Kapitel werden die im dritten Kapitel erläuterten theoretischen Modelle empirisch untersucht. Zuerst wird die Evaluationspraxis analysiert, die sich aus Evaluationsaktivität und Evaluationsverwendung zusammensetzt. Danach wird die Einstellung zur Evaluation untersucht. Dementsprechend gliedert sich das vorliegende Kapitel in diese zwei Teile.

6.1 Untersuchung der Unterschiede in der Evaluationspraxis

Für die Untersuchung der Evaluationspraxis wird folgendes Vorgehen gewählt: Die formulierten Hypothesen werden zuerst im bivariaten und danach im multivariaten Modell getestet. Anhand der empirischen Erkenntnisse werden anschliessend die Hypothesen beantwortet.

Die zu untersuchenden Hypothesen wurden im dritten Kapitel hergeleitet und erläutert. In der nachfolgenden Abbildung 27 sind sie nochmals zusammenfassend dargestellt.

Abbildung 27: Modell für die Untersuchung der Evaluationspraxis



Anmerkungen: Die Pfeile geben die zu untersuchenden Einflussbeziehungen wieder.

Quelle: Eigene Darstellung

Die erste Hypothese adressiert den Einfluss der Institutionalisierung der Evaluation auf die Evaluationspraxis. Die Hypothesen zwei und drei berücksichtigen arbeitsbereichsspezifische und organisationsspezifische Variablen. Des Weiteren werden Kontrollvariablen bei der Analyse berücksichtigt.

6.1.1 Bivariate Analyse

In diesem Abschnitt werden die bivariaten Zusammenhänge zwischen den Variablen mithilfe von Korrelationskoeffizienten untersucht. Für einen korrekten Gebrauch der Korrelationskoeffizienten sind die verschiedenen Skalenniveaus zu berücksichtigen (siehe Benninghaus 2007: 66ff.; Bortz/Schuster 2010: 153ff.; De Vaus 2014: 241ff.; Gehring/Weins 2009: 141ff.). Entsprechend der Kombination der Skalenniveaus, werden verschiedene Korrelationskoeffizienten verwendet. Weitere Ausführungen und Informationen bezüglich der Charakteristiken der Korrelationskoeffizienten sind der Tabelle 46 im Anhang zu entnehmen.

Die nachfolgende Tabelle 34 zeigt die Korrelationskoeffizienten für die Evaluationsaktivität und für die Evaluationsverwendung.

Tabelle 34: Bivariate Zusammenhänge

	Aktivität (I)		Verwendung (O)	
	Korrelation	N	Korrelation	N
Institutionalisierung der Evaluation H1				
Evaluationsstelle (O)	0.22** (0.11**)	748	0.15* (0.07*)	818
Arbeitsbereichsspezifische Variablen H2				
Volatilität (I)	0.10**	716	0.14** (0.11**)	778
Komplexität (I)	0.13**	708	0.20** (0.15**)	770
Politikzyklus (No)				
Implementierung	0.03	423	0.09*	423
Formulierung	0.05	423	0.22**	455
Politikmassnahmen (No)				
Regulativ	-0.01	673	-0.04	730
Redistributiv	0.01	478	0.04	526
Organisationsspezifische Variablen H3				
Verwaltungsgrösse (I)	-0.01	752	-0.04	818
Kontrollvariablen K				
Alter (I)	0.08*	738	0.02 (0.02)	804
Weiblich (No)	-0.13**	705	-0.05	762
Hierarchieposition (O)	0.36** (0.22**)	747	0.27** (0.17**)	812
Bildung (O)	0.17** (0.10**)	748	0.21** (0.13**)	814
Kanton (No)				
Basel-Stadt	-0.01	818	-0.01	818
Bern	-0.06	818	-0.05	818
Luzern	0.09*	818	0.01	818

Anmerkungen: N=Anzahl Mitarbeitende. Die Skalenniveaus sind hinter den Variablen in Klammern aufgeführt: (No)=Nominales Skalenniveau. (O)=Ordinales Skalenniveau. (I)=Intervallskaliertes Skalenniveau. Die Korrelationskoeffizienten ergeben sich entsprechend der in der Tabelle 46 dargestellten Kombinationen der beiden untersuchten Variablen. In Klammern aufgeführt ist der Korrelationskoeffizient Kendalls-Tau-b. Die Stärke der Korrelation kann gemäss De Vaus (2014: 291) unterschieden werden zwischen kleiner (0.1), mittlerer (0.3) und grosser (0.5) Effektstärke. Der Korrelationskoeffizient Gamma, der bei ordinalen Merkmalen angewendet wird ergibt höhere Korrelationswerte (siehe De Vaus 2014: 286 ff.).

* = p < 0,05; ** = p < 0,01; 2-seitig.

Quelle: Eigene Darstellung

Zur Erinnerung: Die Aktivität bezieht sich auf die durchschnittliche Anzahl an Evaluationen, an denen die Befragten im Mittel pro Jahr mitarbeiten. Die Analyse zeigt, dass sämtliche sozioökonomische Variablen mit der Evaluationsaktivität korrelieren. Alter, Bildungsabschluss und Hierarchieposition hängen positiv mit der Evaluationsaktivität zusammen. Der negative Korrelationskoeffizient für das Geschlecht weist darauf hin, dass Frauen weniger häufig angeben, an der Erarbeitung von Evaluationen beteiligt zu sein, als Männer. Ausserdem sind die Mitarbeitenden in Luzern häufiger an der Erarbeitung von Evaluationen beteiligt, als in den beiden anderen Kantonen Basel-Stadt und Bern. Die Kontrolle der ersten Hypothese zeigt, dass die Institutionalisierung der Evaluation H1 tatsächlich positiv in die postulierte Richtung mit der Evaluationsaktivität zusammenhängt. Gleiches gilt für die Variablen Komplexität und Volatilität der zweiten Hypothese. Anders fällt die Beurteilung für die beiden weiteren Variablen dieser Hypothese aus - Phasen des Politikzyklus und Politikmassnahmen – hier bestehen keine nachweisbaren Zusammenhänge zur Evaluationsaktivität.¹⁰¹ Auch die Verwaltungsgrösse hängt nicht mit der Evaluationsaktivität zusammen, was mit der dritten Hypothese postuliert wurde.

Die zweite Spalte der Tabelle 34 illustriert die Korrelationen mit der Häufigkeit der Evaluationsverwendung. Auffallend ist bei dieser Spalte, dass kein Zusammenhang zum Alter und zum Geschlecht hergestellt werden kann. Auch gibt es keine Kantonsunterschiede in der Häufigkeit der Verwendung von Evaluationen. Bildung und Position hängen dagegen positiv mit der Verwendungshäufigkeit zusammen. Die erste Hypothese kann zudem auch für die Verwendungshäufigkeit bestätigt werden, nämlich, dass die Institutionalisierung der Evaluation positiv mit der Verwendungshäufigkeit zusammenhängt. Die Variablen der zweiten Hypothese zeigen Zusammenhänge für die

¹⁰¹ Es gilt zu beachten, dass bei der bivariaten Analyse von nominalen Variablen mit mehr als zwei Kategorien eine andere Referenzkategorie gilt, als bei der multivariaten Analyse. In der bivariaten Analyse werden die Unterschiede zwischen den Mitarbeitenden die der Kategorie entsprechen mit allen restlichen Mitarbeitenden verglichen. Beispielsweise der Vergleich der Mitarbeitenden die in der Implementierung tätig sind mit den Mitarbeitenden die nicht in der Implementierung (interne Dienstleistungen und Formulierung) tätig sind. Die multivariate Analyse dagegen, vergleicht mit der definierten Referenzkategorie. Beispielsweise der Vergleich der Mitarbeitenden die in der Implementierung tätig sind im Vergleich zu den Mitarbeitenden die interne Dienstleistungen (Referenzkategorie) übernehmen.

Komplexität und die Volatilität des Arbeitsbereiches auf. Ebenso die Implementierungsphase des Politikzyklus, in welchem die Mitarbeitenden beschäftigt sind, hängt mit der Verwendungshäufigkeit der Evaluationen zusammen. Einzig für die Variable Politikmassnahmen kann kein Zusammenhang zur Häufigkeit der Verwendung von Evaluationen hergestellt werden. Ebenso hängt die Verwaltungsgrösse, wie in der dritten Hypothese postuliert, nicht mit der Verwendungshäufigkeit zusammen.

Der Vollständigkeit halber wird die Korrelation zwischen den beiden Evaluationspraxisindikatoren Aktivität und Verwendung analysiert. Die beiden Variablen verfügen über einen Korrelationskoeffizienten von 0.67 (Gamma) respektive 0.48 (Kendalls-Tau-b).

6.1.2 Multivariate Analysen

In diesem Abschnitt werden die vorgestellten Einflussfaktoren auf die Evaluationsaktivität und die Evaluationsverwendung in multivariaten Regressionsmodellen untersucht.

Der Einfluss auf die beiden Indikatoren der Evaluationspraxis werden jeweils mittels dreier Modelle analysiert. Im ersten Modell wird der Einfluss der sozioökonomischen und geografischen Variablen auf die Evaluationspraxis geschätzt. Im zweiten Modell werden arbeitsbereichsspezifische Variablen in das Modell aufgenommen um schliesslich das dritte Modell mit allen Variablen zu berechnen. Dieses Vorgehen folgt der in Backhaus et al. (2011: 97) geschilderten „blockweisen Regressionsanalyse“, in der einzelne Blöcke in die Regressionsgleichung einbezogen werden. Dies ermöglicht die Interpretation der Veränderung und des Einflusses der einzelnen Blöcke.

Vor der konkreten Anwendung der Regressionsmodelle wird zunächst überprüft, ob die Voraussetzungen für die Anwendung eines solchen Verfahrens erfüllt sind. Regressionsanalysen liefern nur dann verlässliche Ergebnisse, wenn einige Voraussetzungen erfüllt sind. Die Untersuchung der Evaluationsaktivität basiert auf den Annahmen des generalisierten Regressionsmodells, die Daten zur Untersuchung der Evaluationspraxis haben den Annahmen der logistischen Regression zu genügen.

Regressionsanalysen beider Art basieren auf der Annahme, dass keine Multikollinearität vorliegt. Deshalb wird geprüft, inwiefern die unabhängigen Variablen lineare Abhängigkeiten untereinander aufweisen. Das Problem einer starken Korrelation von unabhängigen Variablen (Multikollinearität) ist bei der multivariaten Regressionsanalyse zu bedenken, da die Schätzungen unpräziser ausfallen (Backhaus et al. 2011: 93ff.; Urban/Mayerl 2011: 225ff.). Die paarweise Analyse der unabhängigen Variablen zeigt diesbezüglich keine problematisch hohen Korrelationskoeffizienten (siehe Tabelle 50 im Anhang).¹⁰² Der Varianz-Inflations-Faktor (VIF) ist eine weitere Masszahl, um die Daten auf Multikollinearität zu testen und weist Werte im dem Bereich auf, die keine starke Multikollinearität indizieren (siehe dazu Kopp/Lois 2012: 127; Urban/Mayerl 2011: 231ff.). Zusätzlich sind bei der Verwendung der logistischen Regression Ausreisser zu bedenken. Für eine Einschätzung diesbezüglich werden standardisierte Residuen berechnet (siehe Backhaus et al. 2011: 277ff.). Diese weisen keine Werte im kritischen Bereich auf.

Ein weiterer wichtiger zu klärender Punkt, bei der multivariaten Analyse von Daten mit hierarchischer Struktur, ist die Abklärung der Anwendung einer Mehrebenenanalyse. Es handelt sich hierbei insofern um einen substanziellen Punkt, als dass die Durchführung einer einfachen Regression die Unabhängigkeit der individuellen Beobachtungen voraussetzt (siehe Hox 2010: 4ff.). Gerade bei der hier zur Anwendung kommenden Datenstruktur¹⁰³ muss zwingend überprüft werden, ob die Evaluationspraxis der Verwaltungsmitarbeitenden von der Verwaltungseinheit abhängig ist. In der vorgestellten Literatur zur Organisationskultur wurde eine Angleichung der Kultur zwischen den Mitarbeitenden innerhalb der Organisationen aufgezeigt. Infolgedessen drängt sich auch theoretisch eine Mehrebenenanalyse auf. Für die Klärung des weiteren Vorgehens wird die Intraklassenkorrelation (englisch: intra class correlation, ICC)

¹⁰² Urban/Mayerl (2011: 230) erwähnen einen Grenzwert des Korrelationswertes in der Praxisanwendung von 0.8. In unserer Untersuchung wird dieser Wert bei den Korrelationskoeffizienten, die lineare Zusammenhänge berechnen, nicht erreicht.

¹⁰³ Der hier verwendete Datensatz ist hierarchisch strukturiert in mindestens drei Ebenen. Die Verwaltungsmitarbeitenden (Ebene 1) arbeiten in verschiedenen Ämtern und Bereichen (Ebene 2), diese sind wiederum zusammengeschlossen in Departementen und Direktionen (Ebene 3).

berechnet. Mit dieser Korrelation kann ermittelt werden, inwieweit die Direktionen und Departemente, respektive die Ämter und Bereiche, für die Varianz der Evaluationspraxis verantwortlich sind.

In der nachfolgenden Tabelle 35 sind die Resultate der sogenannten Nullmodelle (Modelle ohne unabhängige Variablen) aufgeführt. Anhand der Nullmodelle lassen sich die Anteile der Varianz auf Makro- oder Mikroebene zuordnen (Snijders/Bosker 2012: 17ff.).¹⁰⁴ Der „fixed“ Effekt berechnet die Schätzung zur abhängigen Variable ohne Einbezug von unabhängigen Variablen. Der „random“ Effekt zeigt die Varianz der Schätzungen zwischen den Departementen und Direktionen (respektive Ämtern und Bereichen) und die Varianz der Residuen auf der Ebene der Mitarbeitenden auf.

Tabelle 35 zeigt, dass der Intraklassenkorrelationskoeffizient der Evaluationsaktivität, berechnet für die Gruppeneffekte der Departemente und Direktionen, 0.005 beträgt. Das heisst, dass lediglich 0.5 Prozent der Varianz der Evaluationsaktivität auf die Departemente und Direktionen zurückgeführt werden kann. Der Intraklassenkorrelationskoeffizient der Evaluationsverwendung fällt sogar noch niedriger aus (nahezu null). Auch für die Ämter und Bereiche errechnen sich sehr geringe Anteile an der Varianz der Evaluationsaktivität und der Evaluationsverwendung.

Tabelle 35: Nullmodelle

	Erarbeitung	Verwendung
„Fixed“ Effekt		
Konstante	1.85*** (0.19)	0.42*** (0.09)
„Random“ Effekt		
Varianz auf Makro-Ebene Departemente und Direktionen	0.27 (0.29)	
Varianz auf Makro-Ebene Ämter und Bereiche	0.33 (0.37)	
Varianz auf Mikro-Ebene Mitarbeitende	4.03 (0.11)	
Intraklassenkorrelation (ICC) Departemente und Direktionen	0.005 (0.010)	0.000 (0.000)
Intraklassenkorrelation (ICC) Ämter und Bereiche	0.011 (0.015)	0.021 (0.022)
N (Anzahl Gruppen)	730 (9/49)	794 (9/50)

Anmerkungen: Nullmodelle mit drei Ebenen: Ebene 1: individuelle Mitarbeitende; Ebene 2: Ämter und Bereiche; Ebene 3: Departemente und Direktionen. Standardfehler in Klammern. Standardfehler in Klammern.

* = p < 0,05; ** = p < 0,01; *** = p < 0,001; 2-seitig.

Quelle: Eigene Darstellung

¹⁰⁴ In der vorliegenden Untersuchung bilden die Departemente und Direktionen sowie die Ämter und Bereiche die Makroebene. Die Mitarbeitenden bilden die Mikroebene (Individualebene).

Die Befunde zeigen somit, dass die beiden untersuchten Makro-Ebenen niedrige ICC-Werte aufweisen und 0 bis 2.1 Prozent der Varianz auf Unterschiede zwischen den Departementen und Direktionen oder zwischen den Ämtern und Bereichen zurückzuführen sind. Die Resultate verdeutlichen damit, dass die Evaluationspraxis primär von individuellen Charakteristiken abhängt.

Demzufolge ist aus statistischer Sicht die Anwendung von Mehrebenenanalysen nicht notwendig. Gemäss Snijders/Bosker (2012: 22 ff.) sind bei der hier vorliegenden Ausgangslage (geringer ICC) ebenso herkömmliche Regressionsmodelle für die Analyse denkbar. Gründe für die geringe Varianz auf Kontextebene können unter anderem im Forschungsdesign gefunden werden. Schwenkel (2016: 114) untersucht in seiner Dissertation das Vertrauen der Bevölkerung in die Gerichte und argumentiert, dass ein sehr ungleiches Verhältnis der Anzahl der Fälle, mit sehr vielen Beobachtungen (N) auf individueller Ebene und wenigen Fällen auf der Makro-Ebene, per se zu einer höheren Varianz auf individueller Ebene führt.¹⁰⁵ Bei der hier vorliegenden Untersuchung liegt mit 882 Mitarbeitenden in neun Departementen und Direktionen ebenfalls ein sehr ungleiches Verhältnis vor.

Trotzdem wird in dieser Untersuchung aufgrund der theoretischen Bedeutung der Makroebene die Mehrebenenanalyse zur Anwendung kommen.¹⁰⁶ Als ein weiteres Argument kann zudem der Beitrag von Steenbergen/Jones (2002) herangezogen werden. Dieser Artikel erwähnt, dass die Koeffizienten genauer geschätzt werden, wenn die Ebenen berücksichtigt werden. Zudem ist zu beachten, dass sowohl die neun Departemente und Direktionen als auch die 50 Ämter und Bereiche für die Ermittlung des ICC als Cluster berechnet werden und damit anders gruppiert sind, als die unabhängigen Makrovariablen dieser Untersuchung. Auch dies spricht für die Analyse mithilfe von

¹⁰⁵ Schwenkel (2016: 114) verweist auf die von Van Deth/Tausendpfund (2013: 447ff.) aufgeführten Erklärungsmöglichkeiten bezüglich der geringen Bedeutung von Kontexteinflüssen.

¹⁰⁶ Zum Vergleich werden auch die Einebenenmodelle berechnet. Diese Modelle sind im Anhang aufgeführt. Die Einebenenmodelle weisen keine wesentlichen Unterschiede zu den Mehrebenenmodellen auf.

Mehrebenenmodellen. Konkret werden in dieser Untersuchung sogenannte „Random-Intercept“-Modelle angewendet werden (siehe Snijders/Bosker 2012: 41ff.).

6.1.2.1 Evaluationsaktivität

Als Nächstes wird die Evaluationsaktivität im multivariaten Regressionsmodell untersucht. Die Evaluationsaktivität weist eine starke rechtsschiefe Verteilung (Schiefe 3.66) auf. Zudem ist eine starke Überdispersion¹⁰⁷ vorhanden und es handelt sich um Zähldaten. Bei multivariaten Analysen mit einer anhängigen Variable, die diese Charakteristiken aufweist, wird in der Literatur auf die Verwendung des Poisson Regressionsmodell oder der negativ binomialen Regression verwiesen (siehe Hilbe 2011: 2ff.). Der Vergleich zwischen den unterschiedlichen Modellen zeigt, dass die negative binomiale Regression die beste Anpassung an die Daten aufweist, weshalb sie für die Untersuchung der Evaluationsaktivität Verwendung finden wird.¹⁰⁸

In der nachfolgenden Tabelle 36 sind die negativ binomialen Regressionsmodelle für die Untersuchung der Evaluationsaktivität ersichtlich.

Neben den aufgeführten Regressionskoeffizienten wird als Interpretationshilfe hinsichtlich der Einflussstärke auch die IRR (englisch: incidence rate ratio) angegeben. Dieser gibt die Höhe der erwarteten Evaluationsaktivität bei Veränderung der entsprechenden unabhängigen Variable wieder. Ein konkretes Interpretationsbeispiel des IRR wird bei der Analyse des dritten Modells A3 vorgenommen.

¹⁰⁷ Die Varianz ist um 8.96 grösser als der Mittelwert.

¹⁰⁸ Das Akaike-Informationskriterium (AIC), das korrigierte Akaike-Informationskriterium (AICC) und das Bayes-Informationskriterium (BIC) zeigten die tiefsten Werte für das negativ binomiale Regressionsmodell auf im Vergleich zum linearen und zum Poisson Regressionsmodell.

Tabelle 36: Negativ binomiale Regressionsanalyse der Evaluationsaktivität

	Modell A1			Modell A2			Modell A3		
	Koeff.	SE	IRR	Koeff.	SE	IRR	Koeff.	SE	IRR
Konstante	0.05	0.44	1.04	-0.35	0.52	0.59	-0.80	0.65	0.45
Institutionalisierung der Evaluation H1									
<i>Evaluationsstelle (Ref.: Keine Evaluationsstelle)</i>									
Vollständige Evaluationsstelle							0.53+	0.27	1.69
Ansätze einer Evaluationsstelle							0.89*	0.42	2.43
Arbeitsbereichsspezifische Variablen H2									
Volatilität				-0.01	0.04	0.99	-0.01	0.04	0.99
Komplexität				0.13*	0.05	1.13	0.12*	0.05	1.11
<i>Politikzyklus (Ref.: Interne Dienstleistung)</i>									
Implementierung				0.02	0.21	1.02	-0.01	0.21	1.03
Formulierung				0.14	0.26	1.15	0.11	0.29	1.09
<i>Politikmassnahmen (Ref.: Distributiv)</i>									
Regulativ				-0.01	0.19	0.99	0.03	0.19	1.03
Redistributiv				-0.01	0.44	1.01	0.09	0.44	1.09
Organisationspezifische Variablen H3									
Verwaltungsgrösse							0.00	0.00	1.00
Kontrollvariablen K									
Alter	0.01	0.01	1.01	0.00	0.01	1.00	0.00	0.01	1.00
Weiblich	-0.46**	0.19	0.63	-0.40*	0.20	0.67	-0.42*	0.20	0.66
<i>Hierarchieposition (Ref.: Angestellte und Lernende)</i>									
Oberes Kader	0.24	0.38	1.88	0.14	0.38	1.15	0.11	0.38	1.11
Kader	0.63**	0.21	1.26	0.61**	0.21	1.84	0.64**	0.21	1.91
<i>Bildung (Ref.: Obligatorische Schule, Berufsbildung und Berufslehre)</i>									
Hochschulabschluss	-0.17	0.24	0.84	-0.42	0.26	0.66	-0.46+	0.25	0.63
Höhere Berufsbildung	0.48	0.29	1.61	0.17	0.30	1.19	0.19	0.29	1.21
Allgemeinbildung	-0.30	0.36	0.74	-0.34	0.39	0.71	-0.31	0.37	0.73
<i>Kanton (Ref.: Bern)</i>									
Basel-Stadt	0.29	0.28	1.33	0.35	0.31	1.42	0.09	0.36	1.09
Luzern	0.60**	0.25	1.82	0.64+	0.28	1.90	0.68*	0.29	1.97
N			672			621			621
AIC			2135			2016			2013
BIC			2189			2091			2097
Log Likelihood			-1055.34			-991.33			-987.50
Likelihood-Ratio-Test						7.00 (zu Modell 1)			7.66* (zu Modell 2)

Anmerkungen: Mehrebenenmodell („Random-Intercept“ Modell) mit drei Ebenen: Ebene 1: individueller Mitarbeitender; Ebene 2: Ämter und Bereiche; Ebene 3: Departemente und Direktionen. Koeff.= nicht standardisierter Regressionskoeffizient. SE=Standard Fehler. N=Anzahl Mitarbeitende. Ref.=Referenzkategorie zu den Dummyvariablen. AIC=Akaike-Informationskriterium. BIC=Bayessches-Informationskriterium. IRR=Incidence rate ratio.

+: p<0.1; *: p<0.05; **: p<0.01; ***: p<0.001; 2-seitig.

Quelle: Eigene Darstellung

Das erste Modell (*Modell A1*) berücksichtigt den Einfluss von sozioökonomischen Variablen und der beiden Kantonsvariablen auf die Evaluationsaktivität. Das Modell zeigt, dass Frauen auch bei Kontrolle der Hierarchieposition, der Bildung, des Alters und des Kantons, über eine signifikant geringere Evaluationsaktivität verfügen, als Männer. Verwaltungsmitarbeitende in einer Kaderposition weisen zudem über eine nachweisbar höhere Evaluationsaktivität auf als Angestellte, welche die Referenzkate-

gorie für diesen Vergleich bilden. Ebenso haben die Luzerner Verwaltungsmitarbeitenden eine höhere Evaluationsaktivität, als die Mitarbeitenden der Berner Verwaltungseinheiten.

Erweitert man die Analyse um die arbeitsbereichsspezifischen Variablen (*Modell A2*), bleiben die beschriebenen Einflüsse bestehen – mit der Ausnahme, dass die bisher signifikante Kantonsdifferenz für Luzern nicht mehr stark signifikant ist. Hingegen kann ein Einfluss der Komplexität auf die Evaluationsaktivität belegt werden. Für die drei anderen arbeitsbereichsspezifischen Variablen (Volatilität, Politikzyklus und Politikmassnahmen) kann kein signifikanter Einfluss nachgewiesen werden.

Im dritten Modell (*Modell A3*) werden alle Variablen in die Untersuchung einbezogen. Es lässt sich eine Wirkung der Institutionalisierung der Evaluation auf die Evaluationsaktivität belegen. Die Mitarbeitenden der Verwaltungseinheiten mit einer ansatzweisen Institutionalisierung verfügen im Vergleich zu Mitarbeitenden in Verwaltungseinheiten ohne Institutionalisierung über eine höhere Evaluationsaktivität. Für die Mitarbeitenden der Verwaltungseinheiten mit einer voll institutionalisierten Stelle ist der Einfluss zwar positiv, aber knapp nicht auf dem 95-Prozentsniveau signifikant. Die Komplexität übt auch im Gesamtmodell einen nachweisbar positiven Effekt auf die Aktivität aus. Damit ist die Komplexität die einzige arbeitsbereichsspezifische Variable mit einem Einfluss, der sich statistisch nachweisen lässt. Keinen Einfluss übt die Verwaltungsgrösse auf die Evaluationsaktivität aus. Des Weiteren zeigt die Analyse der Kontrollvariablen, dass das Geschlecht, die Hierarchieposition und auch die Kantonsvariable Luzern einen Einfluss auf die Evaluationsaktivität ausüben.

Die Stärke der Regressionskoeffizienten für die vorliegenden Modelle sind folgendermassen zu interpretieren: Bei einer Veränderung der unabhängigen Variable um eine Einheit wird sich die Differenz des Logarithmus für den Erwartungswert der abhängigen Variable um den Wert des Koeffizienten ändern. Diese nicht einfach zu interpretierenden Einflüsse werden mithilfe des IRR leichter nachvollziehbar. Die Einflüsse werden nachfolgend für das dritte Modell A3 erläutert. Der signifikante Einfluss (IRR 2.43) der Variable „Ansätze einer Evaluationsstelle“ bedeutet, dass Mitarbeitende, die

in einer Verwaltungseinheit mit einer ansatzweisen Institutionalisierung arbeiten, im Vergleich zu Mitarbeitenden, die in einer Verwaltungseinheit ohne Institutionalisierung arbeiten, im Jahr eine um 2.43 höhere erwartete Evaluationsaktivität haben. Die erwartete Anzahl Evaluationen pro Jahr steigt, auch bei der erhöhten Einstufung der Komplexität um eine Einheit, um den Faktor 1.11. Zur Erinnerung: die Befragten konnten ihren Arbeitsbereich als überhaupt nicht komplex (Wert null) bis sehr komplex (Wert zehn) einstufen. Des Weiteren zeigt das Modell, dass Frauen im Vergleich zu den Männern eine 0.66 tiefere erwartete Evaluationsaktivität im Jahr haben. Mitarbeitende in einer Kaderposition verfügen im Vergleich zu Angestellten über eine nahezu doppelt so hohe erwartete Evaluationsaktivität im Jahr. Ebenso verhält es sich mit Luzerner Kantonsangestellten verglichen mit den Berner Kantonsangestellten.

Als Nächstes werden drei Masse der Modellgüte zwecks Beurteilung und Vergleich der Erklärungskraft der gerechneten Modelle herangezogen. Der Likelihood-Ratio-Test eignet sich dazu, verschiedene Modelle zu vergleichen und gibt statistische Hilfestellung für den Entscheid, ob Variablen in das Modell aufgenommen werden. Hierbei werden die Werte des Log Likelihood der verschiedenen Modelle miteinander verglichen (siehe dazu Hilbe 2011: 67). Der Likelihood spiegelt die Wahrscheinlichkeit wieder, mit den Parameterschätzungen die empirisch erhobenen Beobachtungswerte zu erhalten und wird mit folgender Formel berechnet:

$$LR = -2\{L_{reduced} - L_{full}\}^{109}$$

Die Likelihood-Ratio-Tests in der Tabelle 36 beziehen sich auf den Vergleich der drei gerechneten Modelle. Die Werte weisen auf die statistisch nachweisbare Verbesserung der Modellgüte von Modell A2 auf Modell A3 hin. Eine zunehmende Erklärungskraft der unabhängigen Variablen von Modell A1 zu Modell A2 kann hingegen nicht nachgewiesen werden (LR-Test ist nicht signifikant). Auch die heute gebräuchlichsten Gü-

¹⁰⁹ Formel aus Hilbe (2011: 68).

temasse bauen auf dem Likelihood auf. Das Akaike-Informationskriterium (AIC) errechnet sich aus dem Likelihood (L) und zusätzlich der Anzahl der Merkmale (Parameter; k), die in dem Modell enthalten sind. Kleinere Werte des Akaike-Informationskriteriums bedeuten einen besseren Modellfit. Da die Anzahl Merkmale (Parameter; k) mit zwei multipliziert und addiert werden (+2k) berücksichtigt das AIC die Anzahl Parameter und belohnt kleinere und somit einfachere Modelle. Das Bayessches-Informationskriterium (BIC) bezieht zusätzlich zur Anzahl Merkmalen (Parametern; k) die Grösse der Stichprobe mit ein (n).¹¹⁰ Auch für das BIC bedeuten tiefere Werte einen besseren Modellfit.

$$AIC = -2L + 2k = -2(L - k)^{111}$$

$$BIC = -2L + k \ln(n)^{112}$$

Das Modell A1 weist unter den drei Modellen die höchsten AIC- und BIC-Werte auf und damit über die tiefste Modellgüte. Modell A2 und Modell A3 weisen beinahe identische Modellgütewerte auf (Modell A3 hat bessere AIC-Werte, Modell A2 hat bessere BIC-Werte).

Nachdem die Evaluationsaktivität untersucht wurde, können die sechs formulierten Hypothesen getestet werden. Diese sind in der nachfolgenden Tabelle 37 nochmals aufgeführt.

¹¹⁰ Es gibt ebenso eine Version des AIC, welche die Grösse der Stichprobe berücksichtigt. Für eine weiterführende Diskussion bezüglich der Informationskriterientests (siehe Hilbe 2011: 68 ff.).

¹¹¹ Formel aus Hilbe (2011: 69).

¹¹² Formel aus Hilbe (2011: 71).

Tabelle 37: Hypothesentest der Evaluationsaktivität

	Hypothese	Test
Institutionalisierung der Evaluation	H1: Wenn die Evaluation in der Verwaltungseinheit institutionalisiert ist, dann ist die Evaluationspraxis der Verwaltungsmitarbeitenden höher.	(✓)
Arbeitsbereichsspezifische Variablen	H2a: Je volatiler der Arbeitsbereich der Mitarbeitenden ist, desto höher ist die Evaluationspraxis, über die sie verfügen.	–
	H2b: Je komplexer der Arbeitsbereich der Mitarbeitenden ist, desto höher ist die Evaluationspraxis, über die sie verfügen.	✓
	H2c: Wenn die Verwaltungsmitarbeitenden in der Politikformulierung oder in der Politikimplementierung tätig sind, dann verfügen sie im Vergleich zu Verwaltungsmitarbeitenden, die interne Dienstleistungen, übernehmen über eine höhere Evaluationspraxis.	–
	H2d: Wenn die Verwaltungsmitarbeitenden Regulierungs- oder Redistributionsmassnahmen umsetzen, dann verfügen sie im Vergleich zu Verwaltungsmitarbeitenden mit distributiven Politiken über eine höhere Evaluationspraxis.	–
Organisationsspezifische Variable	H3: Je ressourcenstärker die Verwaltung ist, desto höher ist die Evaluationspraxis der Verwaltungsmitarbeitenden.	–

Anmerkungen: Angenommen werden die Hypothesen, wenn im multivariaten Modell mit allen Merkmalen (Modell V3) der Einfluss mindestens auf dem 95-Prozentsniveau signifikant ist. ✓=Hypothese wird angenommen. –=Hypothese wird verworfen. ()=Hypothese wird teilweise angenommen respektive teilweise verworfen.

Quelle: Eigene Darstellung

Einen positiven Einfluss der Institutionalisierung der Evaluation auf die Evaluationsaktivität lässt sich in Verwaltungseinheiten, in der Ansätze einer Evaluationsstelle vorhanden sind nachweisen. Für die Mitarbeitenden der Verwaltungseinheiten mit einer voll institutionalisierten Stelle ist der Einfluss zwar positiv, aber knapp nicht signifikant. Dementsprechend wird die Hypothese H1 nur teilweise angenommen.¹¹³ Bei den arbeitsbereichsspezifischen Variablen kann einzig die Hypothese H2b angenommen werden. Die Daten weisen darauf hin, dass mit zunehmender Komplexität des Arbeitsbereiches auch die Evaluationsaktivität steigt. Die restlichen Hypothesen, welche die arbeitsbereichsspezifischen Variablen betreffen (H2a, H2c und H2d), werden verworfen, da weder nachweisbare Einflüsse der Volatilität, des Politikzyklus, noch der Politikmassnahmen auf die Evaluationsaktivitäten nachweisbar sind. Ebenso wenig beeinflusst die Verwaltungsgrösse die Evaluationsaktivität. Aus diesem Grund wird die dritte Hypothese H3 ebenfalls verworfen.

¹¹³ In diesem Zusammenhang ist darauf zu verweisen, dass Hypothesen nach dem indirekten Prinzip des Falsifikationismus funktionieren. Genau genommen kann eine Hypothese nicht empirisch verifiziert werden sondern es kann lediglich die Falschheit nachgewiesen werden (siehe Urban/Mayerl 2011: 133ff.). Trotzdem wird in dieser Untersuchung die Begrifflichkeit von „angenommenen Hypothesen“ Verwendung finden.

6.1.2.2 Evaluationsverwendung

In diesem Unterabschnitt wird die Evaluationsverwendung im multivariaten Modell untersucht. Die nachfolgende Tabelle 38 zeigt die Ergebnisse der Untersuchung der Evaluationsverwendung im logistischen Regressionsmodell auf.

Tabelle 38: Logistische Regressionsanalyse der Evaluationsverwendung

	Modell V1			Modell V2			Modell V3		
	Koeff.	SE	OR	Koeff.	SE	OR	Koeff.	SE	OR
Konstante	-0.63	0.42	1.04	-1.60**	0.54	0.20	-1.96**	0.71	0.14
Institutionalisierung der Evaluation H1									
<i>Evaluationsstelle (Ref.: Keine Evaluationsstelle)</i>									
Vollständige Evaluationsstelle							-0.42	0.31	0.66
Ansätze einer Evaluationsstelle							1.06*	0.49	2.88
Arbeitsbereichsspezifische Variablen H2									
Volatilität				0.05	0.04	1.05	0.05	0.04	1.05
Komplexität				0.11*	0.05	1.11	0.11*	0.05	1.12
<i>Politikzyklus (Ref.: Interne Dienstleistung)</i>									
Implementierung				0.51*	0.20	1.67	0.48*	0.20	1.62
Formulierung				0.63**	0.28	2.50	0.92**	0.30	2.52
<i>Politikmassnahmen (Ref.: Distributiv)</i>									
Regulativ				-0.11	0.20	0.89	-0.11	0.20	0.89
Redistributiv				0.50	0.45	1.65	0.50	0.45	1.65
Organisationspezifische Variablen H3									
Verwaltungsgrösse				0.00	0.00	1.00	0.00	0.00	1.00
Kontrollvariablen K									
Alter	0.00	0.01	1.00	0.00	0.01	1.00	0.00	0.01	1.00
Weiblich	-0.07	0.18	0.94	0.05	0.19	1.06	0.10	0.19	1.10
<i>Hierarchieposition (Ref.: Angestellte und Lernende)</i>									
Oberes Kader	0.84*	0.37	2.32	0.61	0.39	1.84	0.71+	0.39	2.09
Kader	0.78***	0.21	2.19	0.79***	0.22	2.19	0.80***	0.22	2.03
<i>Bildung (Ref.: Obligatorische Schule, Berufsbildung und Berufslehre)</i>									
Hochschulabschluss	0.93***	0.22	2.53	0.56+	0.26	1.57	0.57+	0.25	1.60
Höhere Berufsbildung	0.68*	0.28	1.97	0.44	0.31	1.55	0.45	0.31	1.57
Allgemeinbildung	0.54+	0.32	1.72	0.51	0.35	1.66	0.50	0.35	1.65
<i>Kanton (Ref.: Bern)</i>									
Basel-Stadt	0.22	0.27	1.25	0.31	0.38	1.36	-0.21	0.39	0.81
Luzern	0.17	0.25	1.19	0.32	0.38	1.37	0.25	0.35	1.28
N			727			670			670
AIC			936			850			849
BIC			987			923			934
Log Likelihood			-457.30			-409.49			-405.65
Likelihood-Ratio-Test						19.72** (zu Modell 1)			7.69+ (zu Modell 2)

Anmerkungen: Mehrebenenmodell („Random-Intercept“ Modell) mit drei Ebenen: Ebene 1: individueller Mitarbeitender; Ebene 2: Ämter und Bereiche; Ebene 3: Departemente und Direktionen. Koeff.=Regressionskoeffizient. OR=Odds Ratios. SE=Standardfehler. Die Signifikanzen wurden mit dem Wald-Test geprüft. Ref.=Referenzkategorie zu den Dummyvariablen. AIC=Akaike-Informationskriterium. BIC=Bayessches-Informationskriterium.

+: p<0.1; *: p<0.05; **: p<0.01; ***: p<0.001; 2-seitig.

Quelle: Eigene Darstellung

Für die Analyse der Evaluationsverwendung wird zwischen Mitarbeitenden, die Evaluationen in ihre Arbeit miteinbeziehen, und Personen, die nie Evaluationen in ihre Arbeit miteinbeziehen, unterschieden.¹¹⁴

Das erste Modell (Modell V1), das lediglich sozioökonomische Variablen und die beiden Kantonsvariablen enthält, zeigt nachweisbare Einflüsse der Hierarchieposition und der Bildung auf die Evaluationsverwendung auf. Dagegen lassen sich keine signifikanten Differenzen in der Evaluationsverwendung zwischen den Geschlechtern und zwischen Personen unterschiedlichen Alters belegen.

Mit der Integration der arbeitsbereichsspezifischen Variablen in das Modell (*Modell V2*) werden die Einflüsse der Bildung und der Position auf die Evaluationsverwendung abgeschwächt. Der Unterschied zwischen beispielsweise dem oberen Kader und den Angestellten ist nicht mehr statistisch nachweisbar. Dagegen lassen sich Einflüsse sowohl für die Komplexität, als auch für den Politikzyklus belegen. Von den beiden anderen arbeitsbereichsspezifischen Variablen, Volatilität und Politikmassnahmen, gehen in diesem Modell keine signifikanten Wirkungen auf die Evaluationsverwendung aus.

Das dritte *Modell V3* berücksichtigt alle Variablen. Die Analyse der Institutionalisierung der Evaluation weist auf unterschiedliche Effekte auf die Evaluationsverwendung hin. Ein nachweisbarer Einfluss lässt sich bei Verwaltungsmitarbeitenden, die in einer Verwaltungseinheit arbeiten, in der sich die Evaluation ansatzweise institutionalisiert hat, verorten. Hingegen zeigen sich bei Mitarbeitenden, die in einer Verwaltungseinheit arbeiten, in der die Evaluation institutionalisiert ist, keine Einflüsse auf die Evaluationsverwendung. Die signifikante Differenz der Evaluationsverwendung zwischen Kader und Angestellten bleibt erhalten. Ebenso bleibt der Einfluss der beiden arbeitsbereichsspezifischen Variablen Politikzyklus und Komplexität statistisch nachweisbar. Keinen belegbaren Einfluss übt ausserdem die Verwaltungsgrösse auch in diesem Modell aus.

¹¹⁴ Die Operationalisierung dieser sowie der anderen verwendeten Variablen kann der Tabelle 48 im Anhang entnommen werden.

Die Stärke der Einflüsse lässt sich am einfachsten mittels der Odds Ratio (OR) interpretieren. Die OR bezieht sich auf die relative Wahrscheinlichkeit der Evaluationsverwendung. Die OR von 2.03 der Variable Kader und die OR von 2.88 der Variable „ansatzweise Institutionalisierung“ für das dritte Modell V3, können folgendermassen gelesen werden: Für Personen mit einer Kaderposition ist die Wahrscheinlichkeit der Evaluationsverwendung im Vergleich zu den Angestellten um 2.03 höher. Für Mitarbeitende in einer Verwaltungseinheit mit einer ansatzweisen Institutionalisierung ist die Wahrscheinlichkeit der Evaluationsverwendung sogar um 2.88 höher, als für Mitarbeitende, die in Verwaltungseinheiten ohne Institutionalisierung arbeiten.

Für die Beurteilung von logistischen Regressionsmodellen werden verschiedene Gütemasse berücksichtigt: Die Log Likelihood wurde bereits bei der Beurteilung der Evaluationsaktivitätsmodelle vorgestellt. Zur Erinnerung: dieser spiegelt die Wahrscheinlichkeit wieder, mit den Parameterschätzungen die empirisch erhobenen Beobachtungswerte zu erhalten. Der Likelihood-Ratio-Test vergleicht die Modelle miteinander. Es sind signifikante Verbesserungen der Modellgüte von Modell V1 auf Modell V2 belegbar. Die Verbesserungen von Modell V2 zu Modell V3 sind statistisch nur schwach nachweisbar.

Nachdem die Ergebnisse der Einflüsse auf die Evaluationsverwendung vorliegen, können die Hypothesen getestet werden. Diese sind in nachfolgender Tabelle 39 zur Übersicht nochmals aufgeführt.

Tabelle 39: Hypothesentest der Evaluationsverwendung

	Hypothese	Test
Institutionalisierung der Evaluation	H1: Wenn die Evaluation in der Verwaltungseinheit institutionalisiert ist, dann ist die Evaluationspraxis der Verwaltungsmitarbeitenden höher.	(✓)
Arbeitsbereichsspezifische Variablen	H2a: Je volatiler der Arbeitsbereich der Mitarbeitenden ist, desto höher ist die Evaluationspraxis, über die sie verfügen.	–
	H2b: Je komplexer der Arbeitsbereich der Mitarbeitenden ist, desto höher ist die Evaluationspraxis, über die sie verfügen.	✓
	H2c: Wenn die Verwaltungsmitarbeitenden in der Politikformulierung oder in der Politikimplementierung tätig sind, dann verfügen sie im Vergleich zu Verwaltungsmitarbeitenden, die interne Dienstleistungen, übernehmen über eine höhere Evaluationspraxis.	✓
	H2d: Wenn die Verwaltungsmitarbeitenden Regulierungs- oder Redistributionsmassnahmen umsetzen, dann verfügen sie im Vergleich zu Verwaltungsmitarbeitenden mit distributiven Politiken über eine höhere Evaluationspraxis	–
Organisationsspezifische Variable	H3: Je ressourcenstärker die Verwaltung ist, desto höher ist die Evaluationspraxis der Verwaltungsmitarbeitenden.	–

Anmerkungen: Angenommen werden die Hypothesen, wenn im multivariaten Modell mit allen Merkmalen (Modell V3) der Einfluss mindestens auf dem 95-Prozentsniveau signifikant ist. ✓=Hypothese wird angenommen. –=Hypothese wird verworfen. ()=Hypothese wird teilweise angenommen respektive teilweise verworfen.

Quelle: Eigene Darstellung

Die erste Hypothese H1 postuliert einen von der Institutionalisierung ausgehenden Einfluss auf die Evaluationsverwendung. Die Analyse hat gezeigt, dass die beiden Variablen, welche die Institutionalisierung beschreiben, unterschiedliche Wirkungen auf die Evaluationsverwendung haben. Zur Erinnerung: es wurde zwischen Verwaltungseinheiten mit einer ansatzweisen Institutionalisierung und solchen, in denen die Evaluation institutionalisiert ist, unterschieden. Für Mitarbeitende einer Verwaltungseinheit mit einer ansatzweisen Institutionalisierung, verglichen mit Mitarbeitenden, die in Verwaltungseinheiten arbeiten, in denen diese nicht institutionalisiert ist, ist die Evaluationsverwendung nachweisbar verbreiteter. Bei den Mitarbeitenden, die in Verwaltungen arbeiten, in denen die Evaluation institutionalisiert ist, lässt sich dieser Einfluss nicht belegen. Aus diesem Grund kann die Hypothese H1 nur teilweise angenommen werden. Gründe für diesen Unterschied lassen sich aus den Daten nicht direkt ableiten. Aus dem Vergleich mit den empirischen Ergebnissen der Evaluationsaktivität lässt sich aber schliessen, dass die Institutionalisierung der Evaluation nicht zwingend mit einer erhöhten Evaluationspraxis einhergehen muss, sondern dass sich die Evaluationsaktivität und die Evaluationsverwendung auch in einem Umfeld entwickeln konnte, in welchem die Evaluation nicht voll institutionalisiert ist. Einen nachweisbaren Ein-

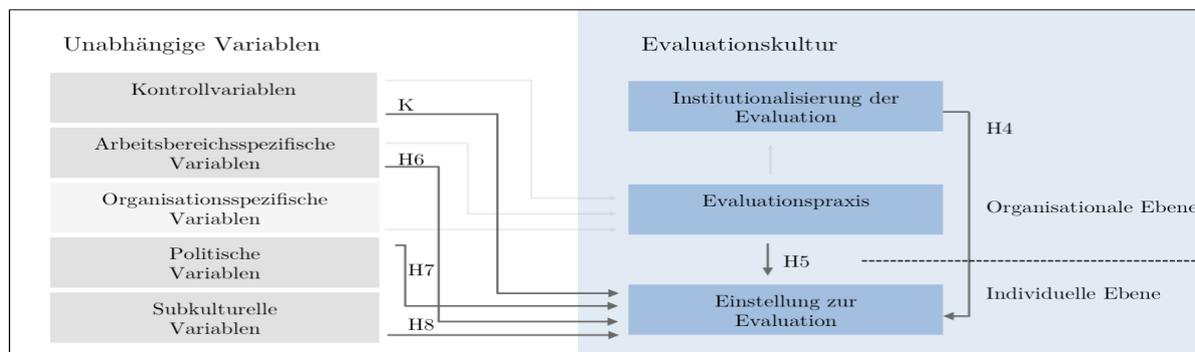
fluss üben zudem die arbeitsbereichsspezifischen Variablen aus. Die Komplexität des Arbeitsbereiches sowie die Phase im Politikzyklus üben auf die Evaluationsverwendung einen statistisch signifikanten Einfluss aus. Je komplexer der Arbeitsbereich einer Person ist, desto höher ist auch die Evaluationsverwendung. Personen, die in der Formulierung oder in der Implementierung von Politikmassnahmen tätig sind, verwenden auch häufiger Evaluationsberichte. Demzufolge werden die Hypothesen H2b und H2c angenommen. Für die beiden anderen Hypothesen bezüglich des Arbeitsbereichs (Volatilität H2a und Politikmassnahmen H2d) lässt sich keine empirische Evidenz nachweisen. Darüber hinaus übt die Departementsgrösse auch in diesem Modell keinen Einfluss auf die Evaluationsverwendung aus. Hypothese H3 wird somit ebenfalls verworfen.

6.2 Untersuchung der Unterschiede in der Einstellung zur Evaluation

Im theoretischen Teil wurde auf die Einstellung zur Evaluation als wichtige Dimension der Evaluationskultur hingewiesen. Erklärungsfaktoren für die Varianz in der Einstellung zur Evaluation werden nun in diesem Unterkapitel untersucht. Hierfür wird das gleiche Vorgehen gewählt, wie bei der Untersuchung der Evaluationspraxis. Die zu untersuchenden Hypothesen werden auf bivariate Zusammenhänge untersucht und anschliessend im multivariaten Modell getestet.

Die Hypothesen wurden bereits im dritten Kapitel hergeleitet und sind in der nachfolgenden Abbildung 28 ersichtlich. Die beiden Hypothesen vier und fünf erörtern Einflüsse der Evaluationspraxis und der Institutionalisierung der Evaluation auf die Einstellung. Die Hypothesen sechs bis acht berücksichtigen arbeitsbereichsspezifische, politische und subkulturelle Variablen. Zudem werden Kontrollvariablen für die Analyse berücksichtigt.

Abbildung 28: Modell für die Untersuchung der Einstellung zur Evaluation



Anmerkungen: Die Pfeile geben die zu untersuchenden Einflussbeziehungen wieder.

Quelle: Eigene Darstellung

6.2.1 Bivariate Analyse

In diesem Abschnitt werden die bivariaten Zusammenhänge zur Einstellung zur Evaluation analysiert.

Die nachfolgende Tabelle 40 führt die Korrelationskoeffizienten für die Einstellung zur Evaluation auf. Die Koeffizienten, welche die Hypothesen der Dimensionen der Evaluationskultur untersuchen (H4 und H5), finden keine Bestätigung für einen bivariaten Zusammenhang. Weder die Institutionalisierung der Evaluation, noch das Ausmass der Evaluationsaktivität und der Evaluationsverwendung der Verwaltungseinheit hängt mit der Einstellung zur Evaluation zusammen. Auch für die Hypothesen, die eine Subkultur von Personen der gleichen Position oder einem Studium technischer, exakter oder naturwissenschaftlicher Richtung postulieren, finden sich keine empirischen Nachweise. Demgegenüber zeigen die Hypothesen, die politische Variablen berücksichtigen (H7a und H7b), nachweisbare Zusammenhänge auf. Die traditionellere Selbsteinstufung hängt tatsächlich negativ mit der Einstellung zur Evaluation zusammen. Auch für die Links-Rechts-Einstufung der Mitarbeitenden kann ein Zusammenhang mit den Einstellungswerten statistisch nachgewiesen werden. Mitarbeitende, die sich stärker Links einstufen, sind positiver eingestellt. Des Weiteren unterscheiden sich Mitarbeitende, die in der Formulierungsphase des Politikzyklus tätig sind, signifikant in der Einstellung zur Evaluation. Sie sind, verglichen mit den anderen Mitarbeitenden, positiver eingestellt. Für Mitarbeitende, die in der Implementierungsphase tätig

sind, lassen sich dagegen signifikant negativere Einstellungen zur Evaluation beobachten.¹¹⁵

Tabelle 40: Bivariate Zusammenhänge

Einstellung zur Evaluation (I)	Einstellung zur Evaluation (I)	
	Korrelation	N
Institutionalisierung der Evaluation H4		
Evaluationsstelle (O)	0.04 (0.02)	352
Evaluationspraxis H5		
Evaluationsaktivität (I)	0.07	352
Evaluationsverwendung (I)	0.00	352
Arbeitsbereichsspezifische Variablen H6		
Politikzyklus (No)		
Implementierung	-0.11*	336
Formulierung	0.14*	336
Politische Variablen H7		
Links-Rechts (I)	-0.13**	319
Traditionalismus-Modernismus (I)	0.26**	305
Subkulturelle Variablen H8		
Hierarchieposition (O)	-0.09 (-0.07)	349
Technische-, Exakte- und Naturwissenschaften (No)	0.02	349
Kontrollvariablen K		
Alter (I)	-0.07	348
Weiblich (No)	-0.01	324
Kanton (No)		
Basel-Stadt	0.04	352
Bern	0.02	352
Luzern	-0.06	352

Anmerkungen: N=Anzahl Mitarbeitende. Die Skalenniveaus sind hinter den Variablen in Klammern aufgeführt: (No)=Nominales Skalenniveau. (O)=Ordinales Skalenniveau. (I)=Intervallskaliertes Skalenniveau. Die Korrelationskoeffizienten ergeben sich entsprechend der in der Tabelle 46 dargestellten Kombinationen der beiden untersuchten Variablen. In Klammern aufgeführt ist der Korrelationskoeffizient Kendalls-Tau-b. Die Stärke der Korrelation kann gemäss De Vaus (2014: 291) unterschieden werden zwischen kleiner (0.1), mittlerer (0.3) und grosser (0.5) Effektstärke.

* = $p < 0,05$; ** = $p < 0,01$; 2-seitig.

Quelle: Eigene Darstellung

Die Ergebnisse der bivariaten Analyse lassen sich somit einfach zusammenfassen: Politische und arbeitsbereichsspezifische Variablen hängen mit der Einstellung zur Evaluation zusammen, die anderen Variablen nicht. Überraschend ist, dass keine sozio-ökonomischen Zusammenhänge gefunden werden können und auch die beiden ande-

¹¹⁵ Es gilt zu beachten, dass bei der bivariaten Analyse von nominalen Variablen mit mehr als zwei Kategorien eine andere Referenzkategorie gilt, als bei der multivariaten Analyse. In der bivariaten Analyse werden die Unterschiede zwischen den Mitarbeitenden die der Kategorie entsprechen mit allen restlichen Mitarbeitenden verglichen. Beispielsweise der Vergleich der Mitarbeitenden die in der Implementierung tätig sind mit den Mitarbeitenden die nicht in der Implementierung (interne Dienstleistungen und Formulierung) tätig sind. Die multivariate Analyse dagegen, vergleicht mit der definierten Referenzkategorie. Beispielsweise der Vergleich der Mitarbeitenden die in der Implementierung tätig sind im Vergleich zu den Mitarbeitenden die interne Dienstleistungen (Referenzkategorie) übernehmen.

ren Dimensionen der Evaluationskultur nicht mit der Einstellung zur Evaluation zusammenhängen.

6.2.2 Multivariate Analyse

Als Nächstes werden die Einflüsse auf die Einstellung zur Evaluation im multivariaten Modell untersucht. Die Einflüsse auf die Einstellung werden, analog zum vorigen Unterkapitel, in drei Modellen analysiert. Beim ersten Modell E1 wird der Einfluss der Kontrollvariablen getestet. Im zweiten Modell E2 werden arbeitsbereichsspezifische, politische und subkulturelle Variablen in das Modell aufgenommen. Im letzten Modell E3 wird mit allen Variablen gerechnet, wobei die beiden Dimensionen der Evaluationskultur – Evaluationspraxis und Institutionalisierung der Evaluation – in das Modell integriert werden.

Die Untersuchung der Einstellung zur Evaluation erfolgt mit der linearen Regressionsanalyse. Bei diesem Verfahren ist auf die Einhaltung der sogenannten BLUE-Kriterien (englisch: best linear unbiased estimator) zu achten (Kopp/Lois 2012: 122 ff.; Urban/Mayerl 2011: 120 ff.). Eines der Kriterien ist die Multikollinearität. Auf das Problem einer starken Korrelation von unabhängigen Variablen (Multikollinearität) wurde bereits bei der Untersuchung der Evaluationspraxis hingewiesen. Die paarweise Analyse der unabhängigen Variablen (siehe Tabelle 51 im Anhang), als auch der Varianz-Inflations-Faktor (VIF) weisen auf keine Multikollinearität hin.¹¹⁶ Ein weiteres wichtiges Kriterium ist die Voraussetzung an die Fehlerterme, deren Varianz entlang der Werte der unabhängigen Variable konstant sein sollten (Homoskedastizität). Diesbezüglich weisen die Daten, wie auch die anderen BLUE-Kriterien, auf keine problematischen Ausprägungen hin.

Bevor die drei Modelle betrachtet werden, wird zuerst der Intraklassenkorrelationskoeffizient berechnet. Der ICC wurde bereits für die Untersuchung der Evaluationspraxis

¹¹⁶ Urban/Mayerl (2011: 232) erwähnen einen weichen Grenzwert des VIF-Wertes von 0,2, der nicht unterschritten werden sollte. In unserer Untersuchung wird dieser Wert deutlich nicht unterschritten.

verwendet und ausführlich beschrieben. Mit dem ICC wird ermittelt, inwieweit die Direktionen und Departemente, respektive die Ämter und Bereiche für die Varianz der Einstellung verantwortlich sind. Tabelle 41 bildet die Intraklassenkorrelationen (ICC) für diese beiden Makro-Ebenen ab und zeigt, dass sich keine Varianzen auf diesen Ebenen verorten lassen.

Tabelle 41: Nullmodelle

	Einstellung zur Evaluation
„Fixed“ Effekt	
Konstante	32.9*** (0.47)
„Random“ Effekt	
Varianz auf Makro-Ebene Departemente und Direktionen	0.00(0.00)
Varianz auf Makro-Ebene Ämter und Bereiche	0.00(0.00)
Varianz auf Mikro-Ebene Mitarbeitende	8.73(0.40)
Intraklassenkorrelation (ICC) Departemente und Direktionen	0.000(0.000)
Intraklassenkorrelation (ICC) Ämter und Bereiche	0.000(0.000)
N (Anzahl Gruppen)	350 (9/43)

Anmerkungen: Nullmodelle mit drei Ebenen: Ebene 1: individuelle Mitarbeitende; Ebene 2: Ämter und Bereiche; Ebene 3: Departemente und Direktionen. Standardfehler in Klammern.

* = $p < 0,05$; ** = $p < 0,01$; *** = $p < 0,001$; 2-seitig.

Quelle: Eigene Darstellung

Aufgrund der Intraklassenkorrelationskoeffizienten lässt sich aus statistischer Sicht keine Notwendigkeit für eine Mehrebenenanalyse ableiten. Trotzdem wird diese aufgrund der Datenstruktur und den Vorteilen von Mehrebenenanalysen für die Untersuchung der Einstellung zur Evaluation angewendet.¹¹⁷ Zum Vergleich werden die identischen Modelle ohne „random effects“ (Einebenenmodelle) berechnet. Diese Modelle sind im Anhang aufgeführt und weisen nur geringfügige Unterschiede zu den Mehrebenenmodellen auf.

Die nachfolgende Tabelle 42 zeigt die drei Mehrebenenmodelle. Das erste Modell berücksichtigt die Kontrollvariablen (*Modell E1*) und zeigt analog zur bivariaten Analyse keine signifikanten Ergebnisse. Alter, Geschlecht und Kanton nehmen keinen Einfluss auf die Einstellung zur Evaluation.

¹¹⁷ Eine ausführlichere Begründung, die für die Anwendung einer Mehrebenenanalyse spricht, ist im vorangehenden Unterkapitel bei der Untersuchung der Evaluationspraxis zu finden.

Tabelle 42: Lineare Regressionsanalyse der Einstellung zur Evaluation

	Modell E1		Modell E2		Modell E3	
	Koeff.	SE	Koeff.	SE	Koeff.	SE
Konstante	33.64***	2.55	24.79***	3.99	20.37***	5.26
Institutionalisierung der Evaluation H4						
<i>Evaluationsstelle (Ref.: Keine Evaluationsstelle)</i>						
Vollständige Evaluationsstelle					-0.95	1.57
Ansätze einer Evaluationsstelle					-1.64	2.51
Evaluationspraxis H5						
Evaluationsaktivität					0.15**	0.06
Evaluationsverwendung					-0.01	0.05
Arbeitsbereichsspezifische Variablen H6						
<i>Politikzyklus (Ref.: Interne Dienstleistungen)</i>						
Implementierung			-0.45	1.15	-0.70	1.17
Formulierung			1.52	1.44	1.70	1.48
Politische Variablen H7						
Links-Rechts			-0.32	0.28	-0.35	0.29
Traditionalismus-Modernismus			1.20***	0.31	1.18***	0.31
Subkulturelle Variablen H8						
<i>Hierarchieposition (Ref.: Angestellte und Lernende)</i>						
Oberes Kader			-3.69+	1.92	-3.00	2.05
Kader			-1.59	1.12	-1.17	1.14
<i>Studienrichtung (Ref.: Kein absolviertes Studium und andere Studienrichtungen)</i>						
Technische-, Exakte- und Naturwissenschaften			0.12	1.25	-0.61	1.26
Kontrollvariablen K						
Alter	-0.03	0.05	0.04	0.05	0.04	0.05
Weiblich	-0.09	1.03	-0.53	1.21	-0.40	1.26
<i>Kanton (Ref.: Bern)</i>						
Basel-Stadt	0.23	1.26	0.68	1.51	5.61*	2.55
Luzern	-0.67	1.25	-0.07	1.34	3.75*	1.95
N		315		258		247
AIC		2264		1824		1754
BIC		2289		1877		1820
Log Likelihood		-1124.82		-897.21		-858.10
Likelihood-Ratio-Test				24.34** (zu Modell1)		7.12+ (zu Modell 2)

Anmerkungen: Mehrebenenmodell („Random-Intercept“ Modell) mit drei Ebenen: Ebene 1: individueller Mitarbeitender; Ebene 2: Ämter und Bereiche; Ebene 3: Departemente und Direktionen. Koeff.= nicht standardisierter Regressionskoeffizient. SE=Standard Fehler. N=Anzahl Mitarbeitende. Ref.=Referenzkategorie zu den Dummyvariablen. AIC=Akaike-Informationskriterium. BIC=Bayessches-Informationskriterium.

+: p<0.1; *: p<0.05; **: p<0.01; ***: p<0.001; 2-seitig.

Quelle: Eigene Darstellung

Im zweiten Modell (*Modell E2*) werden arbeitsbereichsspezifische, politische und subkulturelle Variablen in die Analyse aufgenommen. Dieses Modell zeigt keine signifikanten Unterschiede in der Einstellung zur Evaluation zwischen Mitarbeitenden, die Formulierungstätigkeiten oder Implementierungstätigkeiten übernehmen und Mitarbeitenden, die interne Dienstleistungen ausführen. Auch die Links-Rechts-Selbsteinschätzung nimmt keinen Einfluss auf die Einstellung zur Evaluation. Im Gegensatz dazu lässt sich ein Einfluss der Traditionalismus-Modernismus-Einstufung

nachweisen. Demzufolge steigt der Einstellungswert um 1.20, bei der Veränderung in Richtung Modernismus um eine Einheit (in der Elferskala). Keine Anzeichen können dagegen in den Daten für die beiden Hypothesen gefunden werden, dass sich eine Subkultur der Personen der gleichen Studienrichtung und der gleichen Hierarchieposition ausmachen lässt. Bei den beiden subkulturellen Variablen zeigen sich lediglich Unterschiede zwischen Mitarbeitenden in einer oberen Kaderposition und den Angestellten. Das obere Kader verfügt über geschätzte 3.69 tiefere Evaluationseinstellungswerte, wobei dieser Unterschied knapp nicht auf dem 95-Prozentsniveau signifikant ist.

Beim dritten Modell (Modell E3) werden die beiden anderen Dimensionen der Evaluationskultur – Evaluationspraxis und die Institutionalisierung der Evaluation – miteinbezogen. Die Analyse zeigt, dass die Institutionalisierung der Evaluation keinen nachweisbaren Einfluss auf die Einstellung zur Evaluation nimmt. Der Einfluss der Evaluationspraxis wird mit den beiden Indikatoren Evaluationsaktivität und Evaluationsverwendung untersucht, was im multivariaten Modell zu unterschiedlichen Ergebnissen führt: die Höhe der Anteile der Mitarbeitenden mit Evaluationsverwendung nimmt keinen Einfluss auf die Einstellung der Mitarbeitenden. Im Gegensatz dazu lässt sich ein Einfluss der Höhe der Evaluationsaktivität in der Verwaltungseinheit auf die Einstellung zur Evaluation belegen. In Departementen und Direktionen, in denen mehr evaluiert wird, sind die Einstellungen zur Evaluation positiver. Des Weiteren sind die Unterschiede in der Einstellung zur Evaluation zwischen Mitarbeitenden, die in unterschiedlichen Phasen des Politikzyklus tätig sind, nicht signifikant. Im vorherigen Modell E2 konnte der Einfluss der Traditionalismus-Modernismus-Selbsteinstufung nachgewiesen werden. Dieser Einfluss bleibt auch bei der Kontrolle (Einbezug) der beiden anderen Dimensionen der Evaluationskultur – Evaluationspraxis und Institutionalisierung der Evaluation – im Modell E3 bestehen. Keinen Einfluss auf die Einstellung zur Evaluation geht von der Links-Rechts-Selbsteinstufung, von der Hierarchieposition und vom Studienhintergrund aus. Bemerkenswert sind die Ergebnisse für die Kontrollvariablen, die im dritten Modell E3 nachweisbare kantonale Unterschiede

in der Einstellung zur Evaluation offenbaren. Dieses Modell schätzt die Einstellungswerte der Basler Verwaltungsmitarbeitenden um 5.61 Punkte positiver ein, als die Werte der Berner Verwaltungsmitarbeitenden. Auch die Einstellungswerte der Luzerner Verwaltungsmitarbeitenden werden mit 3.75 Punkten wesentlich positiver geschätzt, als die Erwartungswerte der Berner Verwaltungsmitarbeitenden. Damit offenbart die multivariate Analyse, unter der Kontrolle der anderen unabhängigen Variablen, für Bern nachweisbare negativere Einstellungswerte, die in der bivariaten Analyse unentdeckt geblieben sind. Die Analyse zeigt, dass unter Berücksichtigung der beiden anderen Dimensionen der Evaluationskultur im Modell E3, die Einstellungswerte für Bern nachweisbar tiefer sind. Über die Gründe für die Unterschiede kann nur spekuliert werden. Naheliegend wäre die Grösse der Berner Verwaltungseinheiten, die für eine Verbreitung einer positiven Einstellung hinderlich sein könnte. Die Mitarbeiterstärke wurde im Modell berücksichtigt und kann als Erklärung für die Unterschiede ausgeschlossen werden.

Im nächsten Schritt wird die Modellgüte der drei vorgestellten Modelle beurteilt. Dazu wird der Likelihood-Ratio-Test, das Akaike-Informationskriterium (AIC) und das Bayessches-Informationskriterium (BIC) errechnet. Die Berechnungsformeln und eine ausführliche Beschreibung dieser Gütemasse wurden bereits bei den multivariaten Modellen zur Untersuchung der Evaluationsaktivität aufgeführt und können dort entnommen werden. Die Werte der Likelihood-Ratio-Tests in Tabelle 42 vergleichen die Modelle untereinander. Der Test zeigt eine statistisch nachweisbare Verbesserung der Modellgüte von Modell E1 auf Modell E2. Eine zunehmende Erklärungskraft der unabhängigen Variablen von Modell E2 zu Modell E3 ist hingegen nur schwach nachweisbar (LR-Test ist nicht auf dem 95-Prozentsniveau signifikant). Bei der Analyse der AIC- und BIC-Gütemasse, bei welcher tiefere Werte auf einen besseren Modellfit hinweisen, lässt sich für das Modell E3 die beste Modellgüte berechnen.

Nachdem die Einstellung zur Evaluation untersucht wurde, können die acht Hypothesen beurteilt werden. Diese sind in der nachfolgenden Tabelle 43 aufgeführt.

Tabelle 43: Hypothesentest der Einstellung zur Evaluation

	Hypothese	Test
Institutionalisierung der Evaluation	H4: Wenn die Evaluation in der Verwaltungseinheit institutionalisiert ist, dann ist die Einstellung zur Evaluation der Mitarbeitenden positiver.	-
Evaluationspraxis	H5a: Je höher die Evaluationsaktivität in der Verwaltungseinheit ist, desto positiver ist die Einstellung zur Evaluation.	✓
	H5b: Je höher die Anteile der Evaluationsverwendung in der Verwaltungseinheit sind, desto positiver ist die Einstellung zur Evaluation.	-
Arbeitsbereichsspezifische Variablen	H6: Wenn die Verwaltungsmitarbeitenden in der Politikformulierung oder in der Politikimplementierung tätig sind, dann verfügen sie im Vergleich zu Mitarbeitenden, die interne Dienstleistungen tätigen über eine positivere Einstellung zur Evaluation.	-
Politische Variablen	H7a: Je weiter links sich die Mitarbeitenden in der Links-Rechts-Skala einstufen, desto positiver ist die Einstellung zur Evaluation.	-
	H7b: Je traditionalistischer sich die Mitarbeitenden in der Modernismus-Traditionalismus-Skala einstufen, desto negativer ist die Einstellung zur Evaluation.	✓
Subkulturelle Variablen	H8a: Je höher die Position der Mitarbeitenden in der Verwaltungshierarchie ist, desto positiver ist die Einstellung zur Evaluation.	-
	H8b: Mitarbeitende, die ein Studium technischer, exakter oder naturwissenschaftlicher Richtung absolviert haben, verfügen im Vergleich zu den anderen Mitarbeitenden über eine positivere Einstellung zur Evaluation.	-

Anmerkungen: Angenommen werden die Hypothesen, wenn im multivariaten Modell mit allen Merkmalen (Modell E3) der Einfluss mindestens auf dem 95-Prozentsniveau signifikant ist. ✓=Hypothese wird angenommen. -=Hypothese wird verworfen. ()=Hypothese wird teilweise angenommen respektive teilweise verworfen.

Quelle: Eigene Darstellung

Mit den Hypothesen H4 und H5 wird für einen Einfluss der beiden Dimensionen der Evaluationskultur – Evaluationspraxis und Institutionalisierung der Evaluation – auf die Einstellung zur Evaluation argumentiert. Die multivariate Analyse zeigt keinen nachweisbaren Einfluss der Institutionalisierung der Evaluation auf die Einstellung zur Evaluation auf. Aus diesem Grund kann die Hypothese H4 nicht angenommen werden. Auch die Höhe der Anteile an Mitarbeitenden in den Verwaltungseinheiten, die Evaluationen verwenden, nimmt keinen Einfluss auf die Einstellung der Mitarbeitenden. Demzufolge ist auch die Hypothese H5b zu verwerfen. Im Gegensatz dazu hängt die Höhe der Evaluationsaktivität in den Departementen und Direktionen mit der Einstellung ihrer Mitarbeitenden in positiver Weise zusammen. Folglich wird die Hypothese H5a angenommen. Die sechste Hypothese adressiert einen Einfluss der Formulierungs- oder Implementierungstätigkeit auf die Einstellung. Ein solcher lässt sich empirisch nicht nachweisen und diese Hypothese wird verworfen. Demgegenüber weisen die Daten darauf hin, dass mit zunehmender Modernismus-Einstufung die Einstellung zur Evaluation positiver wird. Dieser Einfluss entspricht der in der Hypothese H7b

formulierten Richtung, die demzufolge angenommen werden kann. Die restlichen Hypothesen (H7a, H8a und H8b), werden verworfen, da keine nachweisbaren Einflüsse der Links-Rechts-Selbsteinschätzung, der Hierarchieposition und für den Studienhintergrund nachweisbar sind.

C. Synthese und Schlussfolgerungen

Im nachfolgenden Synthesekapitel werden die in der Einleitung formulierten Forschungsfragen nochmals aufgegriffen und zusammenfassend beantwortet. Zusätzlich werden die gewonnenen Befunde diskutiert. Das letzte Kapitel wird schliesslich auf der Basis der Forschungserkenntnisse zwei wichtige Schlussfolgerungen ziehen.

7 Synthese

Die vorliegende Dissertation orientiert sich an folgenden drei Forschungsfragen:

1. *Was ist die organisationale Evaluationskultur?*
2. *Welche Unterschiede in der organisationalen Evaluationskultur gibt es zwischen den Departementen und Direktionen?*
3. *Wie lassen sich Unterschiede in der organisationalen Evaluationskultur erklären?*

In den folgenden drei Unterkapiteln werden diese Forschungsfragen zusammenfassend beantwortet.

7.1 Was ist die organisationale Evaluationskultur?

Erstmals wurde in dieser Untersuchung der Begriff der organisationalen Evaluationskultur vertieft theoretisch aufgearbeitet und diskutiert. Dazu wurden Forschungsbeiträge sowohl aus der Forschung über Evaluation als auch aus verwandten Forschungsdisziplinen zur Konzeptualisierung und für die Definition herangezogen. Diese Vorgehensweise orientiert sich am von De Vaus (2014: 41ff.) erläuterten Vorgang für die Konkretisierung von Konzepten.

Die Forschungsbeiträge zur politischen Kultur zeigen, dass solche Untersuchungen hauptsächlich auf Werte und Einstellungen gegenüber der Politik und dem politischen System fokussieren. Darüber hinaus kann der Literatur zur politischen Kulturforschung auch die Feststellung entnommen werden, dass Kultur zwischen Mikro- und

Makroebene anzusiedeln ist. Aus den Beiträgen der Organisationskulturforschung sind die Debatten über die unterschiedlichen Herangehensweisen an die Kultur – mittels eines eher qualitativen oder eines eher quantitativen Ansatzes für die Konzeptualisierung der Evaluationskultur – von Nutzen.

Des Weiteren wurden im zweiten Kapitel Beiträge erfasst, welche die Evaluationskultur oder ähnliche Konzepte untersuchen. Dabei offenbarte sich eine grosse Vielfalt im Hinblick sowohl auf die Definition der Evaluationskultur als auch auf deren Untersuchungsgegenstände. Zwecks Strukturierung der Bandbreite wird in Anlehnung an Barbier (2012) die Evaluationskultur in drei Bestandteile aufgeteilt: Die erste Kategorie von Bestandteilen fokussiert auf die Analyse der Strukturen und Prozesse der Organisation bezüglich Evaluation. Die zweite Kategorie adressiert die Evaluationspraxis der Mitarbeitenden und gibt wieder, inwieweit die Evaluation Einzug in deren Arbeitsalltag gehalten hat. Die dritte Kategorie von Eigenschaften umfasst Einschätzungen von Personen hinsichtlich der Evaluation und evaluative Wertvorstellungen. Abgeleitet aus diesen drei Kategorien wird die organisationale Evaluationskultur folgendermassen definiert:

Die organisationale Evaluationskultur umfasst die gemeinsamen *Einstellungen* und Wertvorstellungen der Mitarbeitenden zur Evaluation und findet Ausdruck im Verhalten (in der *Evaluationspraxis*) und in der Struktur (in der *Institutionalisierung* der Evaluation) der Organisation.

Folglich ergibt sich für die organisationale Evaluationskultur eine multidimensionale Konzeptualisierung, bestehend aus drei konstitutiven Dimensionen, die in der Definition kursiv hervorgehoben werden. Im zweiten Kapitel wurde aufgezeigt, dass durch die Berücksichtigung dieser drei Dimensionen eine umfassende Betrachtung der Evaluationskultur in Organisationen gewährleistet werden kann.

Das vierte Kapitel thematisierte die konkrete Messung der organisationalen Evaluationskultur. Dafür wurden für die drei Dimensionen unterschiedliche Indikatoren erläu-

tert und hinsichtlich ihrer Anwendung diskutiert. Die nachfolgende Auflistung gibt die Indikatoren für die drei Dimensionen wieder, die in der vorliegenden Untersuchung zur Anwendung kommen:

- Die Untersuchung berücksichtigt zwei Indikatoren für die Messung der **Institutionalisierung der Evaluation** (erste Dimension): das Vorhandensein sowohl einer *Evaluationsstelle* als auch *spezifischer Unterlagen*, welche die Evaluation betreffen.
- Die **Evaluationspraxis** bildet die zweite Dimension der Evaluationskultur und soll aufzeigen, inwieweit die Evaluation in den Arbeitsalltag der Mitarbeitenden Einzug gehalten hat. Konkret wird die Evaluationspraxis mit den beiden Indikatoren *Evaluationsaktivität* und *Evaluationsverwendung* gemessen.
- Die **Einstellung zur Evaluation** (dritte Dimension) wird mithilfe von fünf Fragen ermittelt und zu einer Skala addiert. Die Skala besteht aus den Werten null bis 50 und gibt die Grundstimmung gegenüber Evaluationen wieder. Ein hoher Wert steht unter anderem für die Zuschreibung einer grossen Nützlichkeit von Evaluationen und für eine positive Haltung gegenüber der Evaluation.

In diesem Zusammenhang ist darauf hinzuweisen, dass die aufgeführten Dimensionen und Indikatoren zwar die wesentlichen Elemente der Evaluationskultur aufweisen, es aber zweckmässig ist, die zu untersuchenden Bestandteile der Evaluationskultur spezifisch auf die entsprechende Organisation und Fragestellung der Untersuchung anzupassen.

7.2 Welche Unterschiede bezüglich der Evaluationskultur lassen sich in den neun Departementen und Direktionen feststellen?

Eine weitere Zielsetzung der Dissertation ist es, einen ersten Einblick in die Ausgestaltung der Evaluationskultur in den Kantonen zu geben. Im fünften Kapitel wurde die organisationale Evaluationskultur in den neun Verwaltungseinheiten beschrieben. Dabei wurde bewusst auf einen systematischen Vergleich der Evaluationskultur innerhalb unterschiedlicher Politikfelder respektive der drei Kantone verzichtet. Ein Vergleich dieser Art findet sich wie erwähnt bereits in dem Beitrag von Dolder et al. (2017), der teilweise auf Daten der vorliegenden Untersuchung basiert.¹¹⁸ Im Folgenden werden die gefundenen Unterschiede der organisationalen Evaluationskultur in vier Punkten zusammenfassend dargelegt.

Erstens gibt es grosse Unterschiede bei der Institutionalisierung der Evaluation. In den untersuchten neun Verwaltungseinheiten findet sich die ganze Bandbreite verschiedener Ausprägungen der Institutionalisierung:

- Vollständig institutionalisiert ist die Evaluation in der gesamten Erziehungsdirektion (ERZ) des Kantons Bern, in welchem auf Departementsstufe eine Evaluationsstelle sowie spezifische Grundlegendokumente betreffend Evaluation vorhanden sind.
- Von einer teilweisen Institutionalisierung der Evaluation kann beim Bildungs- und Kulturdepartement (BKD) des Kantons Luzern gesprochen werden, weil in einer Dienststelle beide Indikatoren der Institutionalisierung zu finden sind.

¹¹⁸ Der Vergleich berücksichtigt die gleichen drei Kantone Basel-Stadt, Luzern und Bern wie die vorliegende Untersuchung und berücksichtigt die beiden Politikfelder Gesundheit und Bildung. Anhand der Auswertung der Entwicklung der Periode 2000 bis 2013 kann aufgezeigt werden, dass es einen politikfeldspezifischen Entwicklungspfad der Evaluationskultur gibt. Eine kantonsspezifische Entwicklung lässt sich für die sechs Verwaltungseinheiten hingegen nicht ausmachen.

- Erste Ansätze einer Institutionalisierung sind in je einer Dienststelle des Gesundheitsdepartementes (GD) des Kantons Basel-Stadt und des Erziehungsdepartementes (ED) des Kantons Basel-Stadt zu finden.
- Nicht institutionalisiert ist die Evaluation im Gesundheits- und Sozialdepartement (GSD) des Kantons Luzern und im Bau-, Umwelt- und Wirtschaftsdepartement (BUWD) des Kantons Luzern. Im Kanton Bern ist die Evaluation weder in der Volkswirtschaftsdirektion (VOL) noch in der Bau-, Verkehrs- und Energiedirektion (BVE) oder der Gesundheits- und Fürsorgedirektion (GEF) institutionalisiert.

Ferner zeigt die Untersuchung, dass Verwaltungseinheiten mit einer Evaluationsstelle gleichzeitig auch über Grundlagendokumente betreffend Evaluation verfügen. Verwaltungseinheiten, bei denen erste Ansätze einer Evaluationsstelle vorhanden sind – beispielsweise ein kleines Team, das für Evaluationsfragen zuständig ist – verfügen über keine vergleichbaren Grundlagendokumente.

Zweitens lassen sich kleinere Unterschiede in der Evaluationsaktivität erkennen. Die Evaluationsaktivität wird mittels zweier Quellen eruiert: einerseits mithilfe der Online-Befragung, welche die individuelle Evaluationsaktivität der Mitarbeitenden erfasst, andererseits mit der Evaluationsrecherche, welche die gesamte Anzahl der in Auftrag gegebenen oder durchgeführten Evaluationen pro Verwaltungseinheit erhebt. Beide Quellen zeigen, dass in jeder Verwaltungseinheit ein Mindestmass an Evaluationsaktivität vorhanden ist, denn eine gewisse Anzahl an Evaluationsberichten und ein gewisser Anteil an Mitarbeitenden mit Evaluationsaktivität sind in jedem Departement und jeder Direktion zu finden. Wie die Beschreibung der Departemente und Direktionen zeigt, gibt es innerhalb der Verwaltungseinheiten durchaus Differenzen. Gemessen an der Anzahl der Evaluationsberichte (Evaluationsrecherche für den Zeitraum 2005 bis 2013) verfügen die Berner Verwaltungseinheiten über die höchste Evaluationsaktivität.

- Am meisten wird in der BVE Bern evaluiert, gefolgt von der ERZ Bern und der GEF Bern.

- Die kleinste Evaluationsaktivität haben das BKD Luzern, das BUWD Luzern und das GD Basel-Stadt.
- Dazwischen sind die Evaluationsaktivitätshäufigkeitswerte der VOL Bern, des ED Basel-Stadt und des GSD Luzern.

Diese Ergebnisse deuten damit in die gleiche Richtung wie die Untersuchung von Balthasar/Rieder (2009), die Bern (für den Zeitraum 2003 bis 2008) im Kantonsvergleich eine sehr hohe Evaluationsaktivität zuweist und für Basel-Stadt und Luzern eine ähnliche, und im Vergleich zu Bern, wesentlich geringere Evaluationsaktivität verzeichnet. In diesem Zusammenhang ist nochmals darauf hinzuweisen, dass ein Kantonsvergleich in dieser Untersuchung nur bedingt gemacht werden kann, da die Departemente und Direktionen sehr unterschiedlich zusammengesetzt sind. So ist beispielsweise die ERZ Bern neben Bildung auch für Kultur zuständig, während im ED von Basel-Stadt Kultur nicht betreut wird, die Themenbereiche Jugend und Sport dagegen schon.

Drittens sind die Anteile der Mitarbeitenden, die in ihrer Arbeit Evaluationen verwenden, in den neun Verwaltungseinheiten nahezu identisch: Bei allen untersuchten Departementen und Direktionen bewegt sich der Anteil derjenigen, die nie Evaluationen in ihre Arbeit miteinbeziehen, bei an die 40 %. Allerdings zeigt die Analyse innerhalb der Departemente und Direktionen, dass die Departements- und Generalsekretariate häufig die geringsten Anteile von Mitarbeitenden mit Evaluationsverwendung aufweisen. Ein Grund besteht sicherlich in der Eigenschaft der Sekretariate als Stabstelle.

Viertens sind nur geringe Unterschiede in der Einstellung zur Evaluation zwischen den Departementen und Direktionen nachweisbar: Die Mitarbeitenden sind insgesamt positiv oder zumindest neutral gegenüber der Evaluation gestimmt. Eine starke Ablehnung oder gar ein Widerstand gegen die Evaluation kann in keiner Verwaltungseinheit festgestellt werden.

- Die positivste Einstellung zur Evaluation haben die Mitarbeitenden im ED Basel-Stadt. Ausserdem weisen die Mitarbeitenden aus dem GD Basel-Stadt, dem GSD Luzern und der BVE Bern Werte im positiven Bereich auf.
- Die restlichen fünf Verwaltungseinheiten verfügen über eine neutrale Einstellung zur Evaluation.

7.3 Wie lassen sich die Unterschiede in der Evaluationskultur erklären?

Die dritte Forschungsfrage adressiert die Untersuchung der Evaluationskultur selbst. Die zu untersuchenden Einflussbeziehungen wurden im dritten Kapitel in zwei theoretischen Modellen ausführlich beschrieben. Im ersten Modell werden Einflüsse auf die Evaluationspraxis untersucht, während im zweiten Modell unterschiedliche Einflussfaktoren auf die Einstellung zur Evaluation analysiert werden. Beide Untersuchungen basieren auf dem Verfahren der multivariaten Regressionsanalyse. Die Ergebnisse dieser Analysen werden in den nächsten beiden Abschnitten zusammenfassend dargestellt.

7.3.1 Untersuchung der Einflüsse auf die Evaluationspraxis

Die Evaluationspraxis wird mithilfe zweier Indikatoren untersucht. Entsprechend werden die Hypothesen für die Untersuchung der Evaluationspraxis mit zwei abhängigen Variablen – Evaluationsaktivität und Evaluationsverwendung – überprüft. Tabelle 44 zeigt die Ergebnisse für die beiden Indikatoren auf. Die Ergebnisse der Hypothesentests wurden bereits im vorangegangenen Kapitel ausführlich erläutert und sind ausserdem in untenstehender Tabelle 44 zusammenfassend aufgeführt.¹¹⁹

¹¹⁹ Es wird darauf verzichtet die einzelnen Hypothesen nochmals explizit anzunehmen beziehungsweise zu verwerfen. Der Hypothesentest wurde im fünften Kapitel beschrieben und vollzogen.

Tabelle 44: Gesamtübersicht der Hypothesen bezüglich der Evaluationspraxis

	Variable		Aktivität	Verwendung
Institutionalisierung der Evaluation	Evaluationsstelle	H1	(✓)	(✓)
Arbeitsbereichsspezifische Variablen	Volatilität	H2a	-	-
	Komplexität	H2b	✓	✓
	Politikzyklus	H2c	-	✓
	Politikmassnahmen	H2d	-	-
Organisationsspezifische Variable	Verwaltungsgrösse	H3	-	-

Anmerkungen: Angenommen werden die Hypothesen, wenn in den multivariaten Modellen mit allen Merkmalen (Modell A3 und V3) der Einfluss mindestens auf dem 95-Prozentsniveau signifikant ist. ✓=Hypothese wird angenommen. -=Hypothese wird verworfen. ()=Hypothese wird teilweise angenommen respektive teilweise verworfen.

Quelle: Eigene Darstellung

H1: Die erste Hypothese postuliert einen Einfluss der Institutionalisierung der Evaluation auf die Evaluationspraxis. Die Analyse zeigt einen nachweisbaren positiven Zusammenhang für Mitarbeitende einer Verwaltungseinheit mit einer ansatzweisen Institutionalisierung der Evaluation mit der Evaluationsaktivität und mit der Evaluationsverwendung. Bei den Mitarbeitenden, die in Verwaltungen arbeiten, in denen die Evaluation institutionalisiert ist, lässt sich dieser Einfluss nicht belegen. Dementsprechend wird die Hypothese H1 nur teilweise angenommen.

H2a: Die Volatilität des Arbeitsbereichs nimmt entgegen der in der Hypothese formulierten Erwartung keinen Einfluss auf die Evaluationspraxis.

H2b: Im Gegensatz dazu zeigt sich, dass die Komplexität des Arbeitsbereichs positiv mit der Evaluationspraxis zusammenhängt. Deckungsgleich mit der formulierten Hypothese kann dieser Einfluss sowohl für die Evaluationsaktivität als auch für die Evaluationsverwendung bestätigt werden. Je komplexer der Arbeitsbereich der Mitarbeitenden ist, desto höher ist die Evaluationsaktivität und desto höher ist auch die Wahrscheinlichkeit der Evaluationsverwendung.

H2c: Die multivariate Analyse zeigt, dass die Einstufung der Mitarbeitenden in die unterschiedlichen Phasen des Politikzyklus anders auf die Evaluationsaktivität wirkt als auf die Evaluationsverwendung: Einerseits ist die Wahrscheinlichkeit der Evaluationsverwendung für Mitarbeitende, die in der Politikformulierung oder in der Politikimplementierung tätig sind, im Vergleich zu Mitarbeitenden, die interne Dienstleistungen übernehmen, nachweislich höher. Dies entspricht der Richtung der formulierten

Hypothese. Andererseits lässt sich kein Einfluss dieser Unterscheidung auf die Evaluationsaktivität belegen.

H2d: Diese Hypothese postuliert einen Einfluss der unterschiedlichen Politikmassnahmen auf die Evaluationspraxis. Ein solcher kann aufgrund der multivariaten Analyse nicht nachgewiesen werden: Verwaltungsmitarbeitende, die Regulierungs- oder Redistributionsmassnahmen umsetzen, verfügen im Vergleich zu Verwaltungsmitarbeitenden mit distributiven Politiken über keine signifikant höhere Evaluationspraxis.

H3: Die Ressourcenstärke der Verwaltung wurde mit der Verwaltungsgrösse operationalisiert und sollte aufgrund der formulierten Hypothese die Evaluationspraxis positiv beeinflussen. Dieser Einfluss lässt sich empirisch nicht belegen.

Des Weiteren zeigt die multivariate Analyse, dass Frauen im Vergleich zu Männern eine nachweisbar geringere Evaluationsaktivität angeben. Zudem verfügen Mitarbeitende in einer Kaderposition im Vergleich zu Angestellten über eine höhere Evaluationsaktivität, und es lässt sich für die Luzerner Kantonsangestellten eine höhere Evaluationsaktivität nachweisen als für die Berner Kantonsangestellten. Bei der Untersuchung der Evaluationsverwendung dagegen zeigen sich keine Kantonsdifferenzen, es lassen sich lediglich signifikante Differenzen zwischen Kader und Angestellten belegen.

7.3.2 Untersuchung der Einflüsse auf die Einstellung zur Evaluation

In diesem Abschnitt werden die Ergebnisse zur Untersuchung der Einflüsse auf die Einstellung zur Evaluation zusammenfassend präsentiert. Die nachfolgende Tabelle 45 zeigt die untersuchten Hypothesen.¹²⁰

¹²⁰ Wie im vorherigen Abschnitt wird auch hier darauf verzichtet, die einzelnen Hypothesen nochmals explizit anzunehmen beziehungsweise abzulehnen.

Tabelle 45: Hypothesentest der Einstellung zur Evaluation

	Variable	Hypothese	Test
Institutionalisierung der Evaluation	Evaluationsstelle	H4	-
Evaluationspraxis	Evaluationsaktivität	H5a	✓
	Evaluationsverwendung	H5b	-
Arbeitsbereichsspezifische Variablen	Politikzyklus	H6	-
Politische Variablen	Links-Rechts	H7a	-
	Traditionalismus-Modernismus	H7b	✓
Subkulturelle Variablen	Hierarchieposition	H8a	-
	Studienrichtung	H8b	-

Anmerkungen: Angenommen werden die Hypothesen, wenn im multivariaten Modell mit allen Merkmalen (Modell E3) der Einfluss mindestens auf dem 95-Prozentsniveau signifikant ist. ✓=Hypothese wird angenommen. -=Hypothese wird verworfen. ()=Hypothese wird teilweise angenommen respektive teilweise verworfen.

Quelle: Eigene Darstellung

H4: Die vierte Hypothese postuliert einen Einfluss der Institutionalisierung der Evaluation auf die Einstellung zur Evaluation. Für diesen Einfluss lässt sich empirisch keinen Hinweis finden. Mitarbeitende, die in Verwaltungseinheiten arbeiten, wo die Evaluation institutionalisiert ist, weisen im Vergleich zu den anderen im multivariaten Modell keine signifikant positiveren Einstellungswerte auf.

H5a: Im Gegensatz dazu zeigt sich ein Zusammengehen mit der Evaluationsaktivität. Die Höhe der Evaluationsaktivität in den Departementen und Direktionen (Anzahl Evaluationsberichte) hängt mit der Einstellung ihrer Mitarbeitenden in positiver Weise zusammen.

H5b: Die Evaluationsverwendung hingegen nimmt keinen nachweisbaren Einfluss auf die Einstellung zur Evaluation. Die Anzahl der Mitarbeitenden, die in der Verwaltungseinheit Evaluationen verwenden, beeinflusst demzufolge die Einstellung nicht signifikant.

H6: In der sechsten Hypothese wird die Vermutung geäußert, dass Verwaltungsmitarbeitende, die in der Politikformulierung oder in der Politikimplementierung tätig sind, im Vergleich zu Mitarbeitenden, die interne Dienstleistungen übernehmen, über eine positivere Einstellung zur Evaluation verfügen. Dieser Zusammenhang lässt sich empirisch nicht belegen.

H7a: In dieser Hypothese wird postuliert, dass die Links-Rechts-Selbsteinstufung der Mitarbeitenden für die Erklärung der Einstellung zur Evaluation von Bedeutung ist.

Dieser Einfluss kann aufgrund der multivariaten Analyse nicht nachgewiesen werden. Es findet sich keine empirische Evidenz, dass die Links-Rechts-Selbsteinstufung mit der Einstellung zur Evaluation zusammenhängt.

H7b: Im Gegensatz dazu weisen die Daten darauf hin, dass mit zunehmender Modernismus-Einstufung die Einstellung zur Evaluation positiver wird. Dieser Einfluss entspricht der in Hypothese H7b formulierten Richtung und zeigt, dass die Einstellung zur Evaluation mit traditionalistischen respektive modernistischen Werten zusammenhängt.

H8a und H8b: Keine empirischen Indizien lassen sich im multivariaten Modell für die Bildung einer Subkultur erkennen. Es wurde vermutet, dass Mitarbeitende der gleichen Verwaltungshierarchie oder auch Mitarbeitende mit einem Studium technischer, exakter oder naturwissenschaftlicher Richtung eine Subkultur mit gleicher Einstellung bilden. Allerdings lassen sich weder entlang der Verwaltungshierarchie noch entlang des Studienhintergrundes nachweisbare Unterschiede in der Einstellung zur Evaluation feststellen.

Des Weiteren offenbart die multivariate Analyse signifikant positivere Einstellungen zur Evaluation der Mitarbeitenden in Basel-Stadt und Luzern im Vergleich zu Bern.

7.4 Diskussion

Im Folgenden wird der Forschungsbeitrag dieser Untersuchung aufgezeigt und es werden ausgewählte methodische Fragen diskutiert.

Die vorliegende Studie leistet einen wichtigen konzeptionellen Beitrag, indem die organisationale Evaluationskultur auf der Basis einer vertieften theoretischen Auseinandersetzung in drei konstitutive Dimensionen strukturiert wird. Darüber hinaus zeigt die Untersuchung auf, wie diese Dimensionen konkret gemessen werden können. Es wird argumentiert, dass die Berücksichtigung der drei Dimensionen – Einstellung, Praxis und Institutionalisierung – eine umfassende Einschätzung der Evaluationskultur in Organisationen gewährleistet. Angesichts der Ergebnisse des konkreten Anwendungsparadigmas in den neun kantonalen Verwaltungseinheiten kann davon ausgegangen wer-

den, dass es mit dieser Konzeptualisierung gelingt, die wesentlichen Elemente der Evaluationskultur zu erfassen und zu beschreiben. So mussten weder nachträglich neue Indikatoren hinzugenommen werden, noch gab es deren überflüssige. Im Unterschied zu bisherigen empirischen Untersuchungen hat die vorliegende Analyse die Einstellung zur Evaluation stärker in den Fokus gerückt. Die Bedeutung der Einstellungen und Werte im Hinblick auf die Kultur haben bereits die Organisationskulturforschung (beispielsweise Schein 2010: 23ff.) und die Forschung der politischen Kultur (Greiffenhagen 2009; Pickel/Pickel 2006: 49ff.; Westle 2009) deutlich gemacht. In einigen Beiträgen wird denn auch auf Einstellungen zur Evaluation im Zusammenhang mit der Evaluationskultur hingewiesen (vorwiegend jene basierend auf Barbier 2012), empirisch ist dies aber bislang kaum untersucht worden. Dies mag sicherlich forschungsökonomische Gründe haben; die Untersuchung von evaluativen Wertvorstellungen und der Einstellungen zur Evaluation bedingt die Befragung einer gewissen Mindestanzahl von Personen, was mit erheblichem Aufwand verbunden ist. Zum anderen sind weder Einstellung noch Werte „hard facts“, die eindeutig gemessen werden können. Dies kann ein weiterer wichtiger Grund für das bisherige Ausserachtlassen sein.

Daraus leitet sich ein weiterer Forschungsbeitrag der vorliegenden Untersuchung ab: Der Einbezug individueller und organisationaler Daten erlaubt es, empirische Analysen durchzuführen und wichtige Erkenntnisse der organisationalen Evaluationskultur zu gewinnen. Dies bietet die Möglichkeit, die Gründe für die oft erwähnten Unterschiede zwischen den Politikfeldern hinsichtlich ihrer Evaluationspraxis (siehe Balthasar 2007; Balthasar/Rieder 2009; Furubo/Sandahl 2002; Hense et al. 2013; Jacob et al. 2015; Pattyn 2014, 2015; Widmer 2008; Widmer/Beywl 2009), die bisher vorwiegend theoretisch diskutiert wurden, empirisch zu analysieren. Die multivariate Analyse zeigt, dass dieses Vorgehen sich bewährt hat, indem aufgezeigt wird, dass die Evaluationsaktivität und die Evaluationsverwendung mit der Komplexität des Arbeitsbereiches zusammenhängen. Zudem nimmt mit dem Politikzyklus eine weitere arbeitsbereichsspezifische Eigenschaft Einfluss auf die Evaluationsverwendung.

Des Weiteren ist es gelungen, neue Erkenntnisse über die Einstellung zur Evaluation zu gewinnen, insbesondere weil dieses Vorgehen es erstmals erlaubt, geäußerte Vermutungen empirisch zu testen. So lässt sich etwa der von Schwandt/Dahler-Larsen (2006) beobachtete Widerstand gegenüber der Evaluation, der mit traditionellen Werten einhergeht, in dieser Untersuchung statistisch nachweisen. Auch die Tatsache, dass die Evaluationsaktivität der Departemente und Direktionen die Einstellung zur Evaluation positiv beeinflusst, die Hierarchieposition hingegen nicht mit der Einstellung zusammenhängt, kann dank der angewendeten quantitativen Analysen aufgezeigt werden.

Die eben beschriebenen Forschungsergebnisse zeigen folglich auf, dass die Untersuchung der Evaluationskultur mittels quantitativer Analyseverfahren wichtige Erkenntnisse in Bezug auf die organisationale Evaluationskultur mit sich bringt. Insofern hat sich dieses Vorgehen bewährt. Ein weiterer Vorteil der multivariaten Regressionsmethode besteht darin, dass sich der Einfluss einer bestimmten unabhängigen Variable, unter Kontrolle aller anderen unabhängigen Variablen ermitteln lässt (Kopp/Lois 2012: 116ff.). So war es möglich zu zeigen, dass Unterschiede zwischen den Kantonen sowohl hinsichtlich der Evaluationspraxis als auch bezüglich der Einstellung zur Evaluation bestehen.¹²¹

Die Verwendung der Mehrebenenanalyse für die vorliegende Untersuchung wurde bereits ausführlich diskutiert. Aufgrund der hierarchischen Datenstruktur und wegen ihrer Vorteile (siehe Steenbergen/Jones 2002) kam sie entsprechend zum Einsatz. Zur Kontrolle wurden auch die entsprechenden Einebenenmodelle berechnet. Die Differenzen sowohl mit Blick auf die Koeffizienten als auch bei den Standardfehlern sind – wie der Vergleich aller untersuchten Modelle zeigt – gering. Eine Verzerrung aufgrund der vorhandenen hierarchischen Datenstruktur kann somit ausgeschlossen werden oder ist zumindest nur sehr schwach ausgeprägt. In diese Richtung weist auch der berechne-

¹²¹ Die Unterschiede zwischen den Kantonen wurden erst in der multivariaten Analyse ersichtlich.

te Intraklassenkorrelationskoeffizient (englisch: intra class correlation, ICC).¹²² Eine genauere Beschreibung und Analyse des ICC sowie eine ausführliche Diskussion hinsichtlich der Anwendung der Mehrebenenanalyse ist im sechsten Kapitel zu finden.

Was die beschränkte Generalisierbarkeit der Ergebnisse betrifft, so wurde bereits im vierten Kapitel auf die Limitierung der Aussagekraft der vorliegenden Befragungsdaten hingewiesen. Problematische Punkte sind die Selbstselektion (englisch: self-selection-bias) und die Möglichkeit von Falschangaben (englisch: misreporting), die zu verzerrten Ergebnissen führen können (für eine Auflistung siehe Bethlehem 2010). Dass Verzerrungen dieser Art in der vorliegenden Untersuchung vorhanden sind, kann nicht ausgeschlossen werden. Allerdings deuten weder die im vierten Kapitel durchgeführte „Non-Response“-Analyse noch die Untersuchung der Rücklaufquote im fünften Kapitel auf solche Verzerrungen hin. Zudem ist darauf hinzuweisen, dass in den ausgewählten Verwaltungseinheiten eine Totalerhebung durchgeführt wurde. Häder (2010: 139) verweist in diesem Zusammenhang darauf, dass bei einer Totalerhebung „bestimmte Fehler, die bei einer Stichprobenziehung auftreten können, ausgeschlossen sind und die gewonnenen Befunde unter Umständen belastbarer und genauer sein dürften“. Des Weiteren liegt mit einer Rücklaufquote von 50,7 % bei der Online-Mitarbeitendenbefragung eine hohe Befragungsteilnahme vor.¹²³

Zu betonen ist jedoch, dass die Aussagekraft der Ergebnisse, die mit dem Verfahren der Regressionsanalyse gewonnen worden sind, sich auf die ausgewählten Verwaltungseinheiten beschränkt. Inwieweit sie auch über die untersuchten Fälle hinaus zutreffen, ist schwierig abzuschätzen und dürfte für jeden nachgewiesenen Zusammenhang unterschiedlich sein. Beispielsweise wurde die Untersuchung des Zusammenhangs zwischen der Traditionalismus-Modernismus-Selbsteinstufung und der Einstellung zur Evaluation von Schwandt/Dahler-Larsen (2006) übernommen, die diesen Zu-

¹²² Der ICC berechnet das Verhältnis zwischen den Anteilen der verschiedenen Ebenen anhand der Varianz und dient als Entscheidungsgrundlage für die Verwendung von Mehr- und Einebenenmodellen.

¹²³ Allerdings schließt eine hohe Rücklaufquote Selbstselektionsverzerrungen nicht aus. Eine hohe Rücklaufquote ist aber in jeder Untersuchung wünschenswert und gilt als wichtiges Kriterium zur Beurteilung der Befragung (siehe De Vaus 2014: 125f.).

sammenhang in dänischen Schulen beobachten konnten. Dass er sich auch in der vorliegenden Untersuchung im schweizerischen Kontext nachweisen lässt, spricht für eine Generalisierbarkeit auf andere Begleitumstände und Kantone. Allerdings muss hier angesichts des schweizerischen Föderalismus, der sich neben wirtschaftlichen, kulturellen und politischen Eigenheiten der Kantone (Freitag 2005; Vatter 2002: 17f.) auch in der unterschiedlichen Ausgestaltung der kantonalen Verwaltungen (siehe Koller 2013) manifestiert, zurückhaltend vorgegangen werden. Insofern ist bei Aussagen, die auf einer Generalisierung unter Ausserachtlassung dieser Aspekte beruhen, grosse Vorsicht geboten.

8 Schlussfolgerungen

Aus den empirischen Ergebnissen lassen sich zwei Hauptschlussfolgerungen ziehen: Einerseits hat die vorliegende Untersuchung gezeigt, dass sich empirische Hinweise für einen Einfluss der Komplexität des Arbeitsbereichs auf die Evaluationspraxis finden lassen. Andererseits wird deutlich, dass die Einstellung zur Evaluation mit modernistischen, veränderungsfreundlichen Wertvorstellungen zusammenhängt. Beide Schlussfolgerungen werden im Folgenden näher dargelegt.

Die vorliegende Untersuchung kann einen Hinweis auf die festgestellten und bisher vorwiegend theoretisch diskutierten Unterschiede der Evaluationspraxis in diversen Politikfeldern geben (siehe Balthasar 2007; Balthasar/Rieder 2009; Furubo/Sandahl 2002; Hense et al. 2013; Jacob et al. 2015; Pattyn 2014, 2015; Widmer 2008; Widmer/Beywl 2009). Allerdings zeigt die Analyse, dass sich auch innerhalb der einzelnen Politikfelder selbst erhebliche Unterschiede in der Evaluationspraxis finden lassen. Beispielsweise können, wie im dritten Kapitel dargelegt, dem Politikfeld Bildung zwar sehr viele Evaluationsberichte zugeordnet werden, innerhalb dieses Feldes gibt es aber auch Bereiche, in denen sehr wenig evaluiert wird. Aus diesem Grund wurden arbeitsbereichsspezifische Variablen, welche die spezifischen Eigenschaften des Arbeitsbereiches der Mitarbeitenden berücksichtigen, auf ihren Einfluss auf die Evaluationspraxis untersucht. Mit dieser Vorgehensweise geht die vorliegende Untersuchung der in den aufgeführten Beiträgen zugrundeliegenden Frage nach den Gründen für die Unterschiede in der Evaluationspraxis nach. Die erhobenen Daten zeigen, dass die Komplexität des Arbeitsbereiches hier eine wichtige Rolle spielt. Ein komplexer Arbeitsbereich zeichnet sich dadurch aus, dass die Massnahmen und/oder die Zusammenhänge unklar sind, also viel Technik- und oder Fachwissen vorausgesetzt wird. Ein weiteres Merkmal der Komplexität ist der Einbezug von empirischen Daten. Die multivariaten Modelle dieser Untersuchung offenbaren, dass die Komplexität positiv mit der Evaluationspraxis der Mitarbeitenden zusammenhängt. Damit stützt der empirische Befund die im vierten Kapitel erläuterten Effekte, wie sie Frey (2012) mit Blick auf die Verwendung von forschungsbasierten Wirkungsinformationen erwähnt hat.

Auf Grundlage der Thesen von Radaelli (1999) und von Boswell (2009) betont Frey (2012: 70f.), dass die Unsicherheit und die Komplexität von Problemen und deren Lösungen den Bedarf und die Verwendung von Forschungskenntnissen moderieren. Dieser Effekt lässt sich in der vorliegenden Untersuchung für die Evaluation beobachten. Die Komplexität des Arbeitsbereiches schlägt sich dabei in beiden Indikatoren der Evaluationspraxis nieder: Sie hängt sowohl mit der Evaluationsaktivität als auch mit der Evaluationsverwendung zusammen und ist ausserdem in allen berechneten Modellen dieser Untersuchung konsistent nachweisbar.¹²⁴ Je komplexer der Arbeitsbereich der Mitarbeitenden ist, desto höher ist die Evaluationsaktivität und desto höher ist auch die Wahrscheinlichkeit der Evaluationsverwendung. Konkret steigt bei der erhöhten Einstufung der Komplexität um eine Einheit – die Befragten konnten ihren Arbeitsbereich als überhaupt nicht komplex (Wert null) bis sehr komplex (Wert zehn) einstufen – die erwartete Anzahl an Evaluationen pro Jahr um den Faktor 1.11, die Wahrscheinlichkeit der Evaluationsverwendung um den Faktor 1.12. Damit stützen die empirischen Ergebnisse den Befund, dass die Komplexität ein wichtiger Faktor für die Evaluationspraxis ist: Sie generiert eine höhere Nachfrage nach Wissen in Form von Evaluationen, um die ungewissen Wirkungszusammenhänge und Situationen zu verstehen. Zweitens zeigt die empirische Analyse, dass die Einstellung zur Evaluation mit modernistischen und traditionalistischen Wertvorstellungen zusammenhängt. Die Ergebnisse der vorliegenden empirischen Untersuchung machen deutlich, dass die Traditionalismus-Modernismus-Selbsteinstufung mit der Einstellung zur Evaluation zusammenhängt, während sich für die Links-Rechts-Selbsteinstufung in den multivariaten Modellen hier keine Unterschiede erkennen lassen. Dies ist insofern bemerkenswert, als in diesem Fall nicht die in der Politikwissenschaft oft verwendete Links-Rechts-Dimension von Bedeutung ist, sondern vielmehr die weniger gebräuchliche Traditio-

¹²⁴ Mittels dem blockweisen Vorgehen wurden verschiedene Regressionsmodelle berechnet. Ausserdem wurde zum Vergleich neben den Mehrebenenmodellen auch Einebenenmodelle berechnet. In allen Modellen zeigt sich, dass die Komplexität positiv mit der Evaluationsaktivität und mit der Evaluationsverwendung zusammenhängt (siehe dazu das sechste Kapitel) und dieser Zusammenhang damit konsistent nachweisbar ist.

nalismus-Modernismus-Dimension einen Erklärungsansatz bietet. Die Auswertung der Daten lässt darauf schliessen, dass diese Selbsteinstufung ein wichtiger und signifikanter Prädiktor für die Einstellung zur Evaluation ist. Der errechnete Wert des Regressionskoeffizienten zeigt, dass der geschätzte Einstellungswert um 1.18 (mit einer Skala von null bis 50) bei der Veränderung in Richtung Modernismus um eine Einheit (in der Elferskala) steigt. Stuft sich beispielsweise jemand in der Traditionalismus-Modernismus-Skala um fünf Punkte höher ein, erhöht sich sein vorhergesagter Einstellungswert um 6.35. Damit stützen die Befunde der vorliegenden Untersuchung die im Artikel von Schwandt/Dahler-Larsen (2006) geäusserte Beobachtung, dass Personen mit modernistischen Wertvorstellungen einen anderen Umgang mit evidenzbasiertem Wissen haben. Besonders bemerkenswert ist, dass diese Beobachtung in dänischen Schulen gemacht wurde und der Zusammenhang nun auch im schweizerischen Kontext nachgewiesen werden kann.

Aufgrund des geschilderten Zusammenhanges lässt sich argumentieren, dass die Traditionalismus-Modernismus- ähnlich wie die Links-Rechts-Selbsteinstufung ein Glaubenssystem mit spezifischen Wertvorstellungen darstellt, die mit der Einstellung zur Evaluation einhergehen.¹²⁵ Allerdings ist es bei der Traditionalismus-Modernismus-Unterscheidung schwierig zu erfassen, welche Themen und Aspekte diese Unterscheidung konkret beinhaltet. Epple (2013) verweist in Anlehnung an Stolz (2001) auf vier Fragen zur Messung des Traditionalismus-Modernismus-Index. Die Fragen beinhalten die Themen Chancengleichheit für Ausländerinnen und Ausländer, Verteidigung von Tradition, Stärkung der Armee sowie Ruhe und Ordnung. Diese Fragen werden in der vorliegenden Untersuchung nicht angewendet, da im Pretest, der im Rahmen dieser Untersuchung durchgeführt wurde, die ausgewählten Verwaltungsmitarbeitenden auf diese vier Fragen keine konsistenten Antworten gaben.¹²⁶ Aus diesem Grund wurde die

¹²⁵ In der Literatur wird darauf hingewiesen, dass die politische Ideologie über den Prädiktor des Wahlentscheids hinaus ein Glaubenssystem darstellt (siehe Gauchat 2015; Martin/Desmond 2010; Perrin et al. 2014).

¹²⁶ Siehe dazu die Erläuterungen im dritten Kapitel. Der Traditionalismus-Modernismus wurde, wie die Links-Rechts-Unterscheidung, mittels einer Frage erhoben (siehe Tabelle 49 im Anhang).

Traditionalismus-Modernismus-Unterscheidung mit einer einzelnen Frage ermittelt, sodass sich die Bedeutung der zugrundeliegenden Themen oder Werte nicht weiter empirisch untersuchen lässt. Einen Anhaltspunkt auf entsprechende Elemente gibt dagegen der Beitrag von Stolz (2001: 87). Er verweist auf den stark veränderungsfeindlichen Charakter des Konzeptes des Traditionalismus, das er mithilfe von sechs unterschiedlichen Variablen bildet: „Wir gelangten so zur Einsicht, dass in allen sechs Variablen ein gemeinsames, Veränderung ablehnendes Element enthalten ist, und dass diese daher eher als Teildimensionen eines grösseren, veränderungsfeindlichen Phänomens angesehen werden können. Dieses Phänomen nennen wir Traditionalismus [...]“. Auch der Artikel von Schwandt/Dahler-Larsen (2006) betont, dass Personen mit modernistischen Wertvorstellungen den Glauben haben, „Dinge verändern und steuern zu können“. Diese Beiträge legen die Vermutung nahe, dass die Haltung gegenüber Veränderungen im abstrakten Konzept des Traditionalismus ein wesentliches Element darstellt, und wie die vorliegenden Ergebnisse zeigen, ist sie wahrscheinlich auch für die Einstellung zur Evaluation entscheidend.

In der Einleitung wurde darauf hingewiesen, dass beträchtliche Ressourcen in Evaluationen investiert werden. Dies ist wahrscheinlich mit ein Grund, weshalb der Nutzen der Evaluation in der bisherigen Forschung im Zentrum stand – sowohl im schweizerischen (Widmer et al. 2016) als auch im internationalen Kontext (Hense et al. 2013). Die vorliegende Studie hat unter anderem die Einstellungen von Verwaltungsmitarbeitenden zur Evaluation untersucht und dabei zeigen können, mit welchen Faktoren sie zusammenhängen. Auf der Basis psychologischer Untersuchungen lässt sich argumentieren, dass im Zusammenhang mit der Nutzung der Evaluation auch die Einstellung zur Evaluation verstärkt erforscht werden sollte. Eine Reihe von Forschungsbeiträgen indizieren einen Einfluss der persönlichen Einstellung auf das Verhalten (siehe dazu Ajzen 2001). Insofern ist anzunehmen, dass die Einstellung zur Evaluation mehr als nur die Bewertung der Evaluation darstellt. Vielmehr dürfte sie darüber hinaus auch den Umgang mit der Evaluation im Arbeitsalltag beeinflussen. Es ist somit ein Zusammenhang zwischen der Einstellung zur Evaluation und der Nutzung der Evaluation

zu vermuten. In der empirischen Prüfung dieses Zusammenhangs besteht eine wichtige Forschungslücke, die es zu füllen gilt. Damit könnte die Bedeutsamkeit der Einstellung noch differenzierter beurteilt werden. Im Gegensatz zu dieser Frage, die offengelassen werden muss, kann die vorliegende Studie der Forschungsgemeinde folgende Erkenntnis weitergeben: Zukünftige Forschungsvorhaben, welche die Einstellung zur Evaluation oder die Einstellung gegenüber Evidenzen im Allgemeinen untersuchen, sollten auch modernistische und traditionalistische Wertvorstellungen als Erklärungsfaktoren berücksichtigen.

Anhang I Literatur- und Quellenverzeichnis

- Abdul Rashid, Zabid/Sambasivan, Murali/Abdul Rahman, Azmawani (2004): The Influence of Organizational Culture on Attitudes Toward Organizational Change. In: *Leadership & organization development Journal*, 25 (2), 161–179.
- Ajzen, Icek (2001): Nature and Operation of Attitudes. In: *Annual Review of Psychology*, 52 (1), 27–58.
- Allport, Gordon Willard (1954): *The Nature of Prejudice*. 5. Auflage. Massachusetts: Addison-Wesley.
- Almond, Gabriel A./Verba, Sidney (1963): *The Civic Culture: Political Attitudes and Democracy in Five Nations*. 3. Auflage. Princeton: Princeton University Press.
- Alvesson, Mats (2013): *Understanding Organizational Culture*. 2. Auflage. London: Sage Publications.
- Anderfuhren-Biget, Simon/Varone, Frédéric/Giauque, David (2013): Policy Environment and Public Service Motivation. In: *Public Administration*, 92 (4), 807–825.
- ARGEV, Bern (2016): Externe Schulevaluation in Bern.
http://www.argev.ch/page.php?language=de&pages_id=24 (Zugriff am 2.4.2016).
- Arnold, Mary E. (2006): Developing Evaluation Capacity in Extension 4-H Field Faculty: A Framework for Success. In: *American Journal of Evaluation*, 27 (2), 257–269.
- Ashkanasy, Neal M/Wilderom, Celeste/Peterson, Mark F (2011): *The Handbook of Organizational Culture and Climate*. Thousand Oaks: Sage Publications.
- AVD, Bern (2015): Volksschule der Erziehungsdirektion des Kanton Bern.
http://www.erz.be.ch/erz/de/index/direktion/organisation/amt_fuer_kindergarten Volksschuleundberatung/ueber_uns/volksschule_d.html (Zugriff am 19.10.2015).
- Backhaus, Klaus/Erichson, Bern/Plinke, Wulff/Weiber, Rolf (2011): *Multivariate Analysemethoden eine anwendungsorientierte Einführung*. 13. Auflage. Berlin: Springer.
- Balthasar, Andreas (2007): *Institutionelle Verankerung und Verwendung von Evaluationen*. Zürich/ Chur: Rüegger Verlag.
- Balthasar, Andreas (2015): L'utilisation de l'évaluation par l'administration fédérale. In: Horber-Papazian, Katia (Hrsg.), *Regards croisés sur l'évaluation en Suisse*. Lausanne: Presses polytechniques et universitaires romandes, 115–132.

- Balthasar, Andreas/Müller, Franziska (2014): Die Verbreitung evidenzbasierter und gleichstellungssensitiver Informationen in den Entscheidungsprozessen kantonaler Steuer- und Sozialtransferpolitik: Eine quantitative Analyse. In: *Swiss Political Science Review*, 20 (1), 70–95.
- Balthasar, Andreas/Rieder, Stefan (2009): Wo ist evidenzbasierte Politik möglich?: Die Verbreitung von Evaluationen auf kantonaler Ebene. In: Vatter, Adrian/ Varone, Frédéric/Sager, Fritz (Hrsg.), *Demokratie als Leidenschaft*. Bern: Haupt, 403–429.
- Balthasar, Andreas/Strotz, Chantal (2017): Verbreitung und Verankerung von Evaluation in der Bundesverwaltung. In: Sager, Fritz/Widmer, Thomas/Balthasar, Andreas (Hrsg.), *Evaluation im politischen System der Schweiz: Entwicklung, Bedeutung und Wechselwirkungen*. Zürich: Neue Zürcher Zeitung NZZ Libro, 89–118.
- Bandelow, Nils C. (2015): Advocacy Coalition Framework. In: Wenzelburger, Georg/Zohnhöfer, Reimut (Hrsg.), *Handbuch Policy-Forschung*. Wiesbaden: Springer Fachmedien Wiesbaden, 305–324.
- Barbier, Jean-Claude (2012): Introduction. In: Barbier, Jean-Claude/Hawkins, Penny (Hrsg.), *Evaluation Cultures. Sense Making in Complex Times*. New Brunswick: Transaction Publishers, 1–18.
- Barnes, Marian/Matka, Elizabeth/Sullivan, Helen (2003): Evidence, Understanding and Complexity: Evaluation in Non-Linear Systems. In: *Evaluation*, 9 (3), 265–284.
- Behrens, Fritz/Heinze, Rolf G./Hilbert, Josef/Stöbe-Blossey, Sybille (Hrsg.) (2005): *Ausblicke auf den aktivierenden Staat. Von der Idee zur Strategie*. Berlin: Edition Sigma.
- Benninghaus, Hans (2007): *Deskriptive Statistik: eine Einführung für Sozialwissenschaftler*. 11. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Best, Samuel J./McDermott, Monika L. (2007): Measuring Opinions vs. Non-Opinions - The Case of the USA Patriot Act. In: *The Forum*, 5 (2), Article 7.
- Bethlehem, Jelke (2010): Selection Bias in Web Surveys: Selection Bias in Web Surveys. In: *International Statistical Review*, 78 (2), 161–188.
- Beywl, Wolfgang/Widmer, Thomas (2009): Evaluation in Expansion: Ausgangslage für den intersektoralen Dreiländer-Vergleich. In: Widmer, Thomas/Beywl, Wolfgang/Fabian, Carlo (Hrsg.), *Evaluation*. Wiesbaden: VS Verlag für Sozialwissenschaften, 13–23.

- BFS (2014): Bevölkerungsstand und -struktur.
http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/02/blank/key/raeumliche_verteilung/kantone__gemeinden.html (Zugriff am 20.1.2015).
- BKD, Luzern (2014): Organigramm des Bildungs- und Kulturdepartementes des Kantons Luzern. http://www.lu.ch/verwaltung/BKD/bkd_dienststellen (Zugriff am 6.11.2014).
- BKD, Luzern (2016): Bildungs- und Kulturdepartement des Kantons Luzern.
http://www.lu.ch/verwaltung/BKD/bkd_unser_departement (Zugriff am 8.3.2016).
- BKD VB Evaluation, Luzern (2016): Externe Schulevaluation der Volksschulbildung des Kantons Luzern.
https://volksschulbildung.lu.ch/aufsicht_evaluation/ae_externe_sev (Zugriff am 5.4.2016).
- BKD VB, Luzern (2016): Dienststelle Volksschulbildung des Kantons Luzern.
<https://volksschulbildung.lu.ch/> (Zugriff am 30.3.2016).
- Blank, Joshua M./Shaw, Daron (2015): Does Partisanship Shape Attitudes toward Science and Public Policy? The Case for Ideology and Religion. In: *The ANNALS of the American Academy of Political and Social Science*, 658 (1), 18–35.
- Blatter, Joachim/Bombach, Clara/Wiprächtiger, Roman (2015): Enhancing Gender Equity Through Evidence-Based Policymaking? Theorizing and Tracing the Use of Systematic Knowledge in Family and Tax Policy Reforms. In: *European Policy Analysis*, 1 (1), 3–34.
- Blatter, Joachim/Haverland, Markus (2014): *Designing Case Studies: Explanatory Approaches in Small-N Research*. Basingstoke: Palgrave Macmillan.
- Bornstein, Robert F. (1989): Exposure and Affect: Overview and Meta-Analysis of Research 1968-1987. In: *Psychological Bulletin*, 106, 265–89.
- Bortz, Jürgen/Döring, Nicola (2006): *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler*. 4. Auflage. Heidelberg: Springer.
- Bortz, Jürgen/Schuster, Christof (2010): *Statistik für Human- und Sozialwissenschaftler*. 7. Auflage. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.
- Boswell, Christina (2009): *The Political Uses of Expert Knowledge: Immigration Policy and Social Research*. Cambridge: Cambridge University Press.
- Botcheva, Luba/White, Catherine Roller/Huffman, Lynne C. (2002): Learning Culture and Outcomes Measurement Practices in Community Agencies. In: *American Journal of Evaluation*, 23 (4), 421–434.

- Boyle, Richard/McNamara, Gerry/O'Hara, Joe (2012): Riding the Celtic Tiger: Forces Shaping Evaluation Culture in Ireland in Good Times and Bad. In: Barbier, Jean-Claude/Hawkins, Penny (Hrsg.), *Evaluation Cultures. Sense Making in Complex Times*. New Brunswick: Transaction Publishers, 45–64.
- Brandon, Paul R./Higa, Terry Ann F. (2004): An Empirical Study of Building the Evaluation Capacity of K-12 Site-Managed Project Personnel. In: *Canadian Journal of Program Evaluation*, 19 (1), 125–142.
- Bundi, Pirmin (2016): What Do We Know About the Demand for Evaluation? Insights From the Parliamentary Arena. In: *American Journal of Evaluation*, 37 (4), 522–541.
- Bundi, Pirmin/Varone, Frédéric/Gava, Roy/Widmer, Thomas (2016): Self-Selection and Misreporting in Legislative Surveys. In: *Political Science Research and Methods*, 1-19.
- Buri, Christof/Schneider, Gerald (1993): Gründe und Scheingründe für das Schweizer Abstimmungsverhalten. In: *Schweizerische Zeitschrift für Soziologie*, 19 (2), 389–417.
- Bussmann, Werner (1997): *Einführung in die Politikevaluation*. Basel, Frankfurt am Main: Helbing und Lichtenhahn.
- Bussmann, Werner (2007): *Institutionalisierung der Evaluation in der Schweiz. Verfassungsauftrag, Konkretisierungsspielräume und Umsetzungsstrategie*. In: Schäfer, Heinz (Hrsg.), *Evaluierung der Gesetze/Gesetzesabfolgenabschätzung (II)*. Wien: Manz, 1–20.
- Bustelo, María (2006): The Potential Role of Standards and Guidelines in the Development of an Evaluation Culture in Spain. In: *Evaluation*, 12 (4), 437–453.
- BUWD LAWÄ, Luzern (2016): Dienststelle Landwirtschaft und Wald des Bau-, Umwelt- und Wirtschaftsdepartementes des Kantons Luzern. https://www.lawa.lu.ch/ueber_uns (Zugriff am 10.3.2016).
- BUWD, Luzern (2014): Organigramm des Bau-, Umwelt- und Wirtschaftsdepartementes des Kantons Luzern. http://www.lu.ch/verwaltung/BUWD/buwd_dienststellen (Zugriff am 7.3.2014).
- BUWD, Luzern (2016): Bau-, Umwelt- und Wirtschaftsdepartement des Kantons Luzern. http://www.lu.ch/verwaltung/BUWD/buwd_unser_department (Zugriff am 8.3.2016).

- BUWD VIF, Luzern (2016): Dienststelle Verkehr und Infrastruktur des Bau-, Umwelt- und Wirtschaftsdepartementes des Kantons Luzern.
https://www.vif.lu.ch/ueber_uns (Zugriff am 10.3.2016).
- BVE, Bern (2014): Organigramm der Bau-, Verkehrs- und Energiedirektion des Kantons Bern.
<http://www.bve.be.ch/bve/de/index/direktion/organisation/organigramm.html> (Zugriff am 29.8.2014).
- BVE, Bern (2016): Bau-, Verkehrs- und Energiedirektion des Kantons Bern.
<http://www.bve.be.ch/bve/de/index/direktion/ueber-die-direktion.html> (Zugriff am 29.3.2016).
- Camilleri, Emanuel (2007): Antecedents Affecting Public Service Motivation. In: *Personnel Review*, 36 (3), 356–377.
- Castaing, Sébastien (2006): The Effects of Psychological Contract Fulfilment and Public Service Motivation on Organizational Commitment in the French Civil Service. In: *Public Policy and Administration*, 21 (1), 84–98.
- Cobb, Michael D./Nance, Mark T. (2013): The Consequences of Measuring Non-Attitudes about Foreign Trade Preferences. In: *Survey Practice*, 4 (6).
- Compton, Donald W./Glover-Kudon, Rebecca/Smith, Iris E./Eden Avery, Mary (2002): Ongoing Capacity Building in the American Cancer Society (ACS) 1995–2001. In: *New Directions for Evaluation*, 2002 (93), 47–62.
- Cousins, J. Bradley/Earl Lorna M. (1995): *Participatory Evaluation in Education: Studies in Evaluation use and Organizational Learning*. London: Falmer Press.
- Cousins, J. Bradley/Elliott, Catherine/Amo, Courtney/Bourgeois, Isabelle/Chouinard, Jill/Goh, Swee C./Lahey, Robert (2008): Organizational Capacity to Do and Use Evaluation: Results of a Pan-Canadian Survey of Evaluators. In: *Canadian Journal of Program Evaluation*, 23 (3), 1–35.
- Cousins, J. Bradley/Goh, Swee C./Clark, Shannon/Lee, Linda L. (2004): Integrating Evaluative Inquiry Into the Organizational Culture: A Review and Synthesis of the Knowledge Base. In: *Canadian Journal of Program Evaluation*, 19 (2), 99–141.
- Cousins, J. Bradley/Goh, Swee C./Elliott, Catherine J./Bourgeois, Isabelle (2014): Framing the Capacity to Do and Use Evaluation. In: *New Directions for Evaluation*, 2014 (141), 7–23.
- Dahler-Larsen, Peter/Schwandt, Thomas A. (2012): Political Culture as Context for Evaluation. In: *New Directions for Evaluation*, 2012 (135), 75–87.

- De Peuter, Bart/Pattyn, Valérie (2009): Evaluation Capacity: Enabler or Exponent of Evaluation Culture? In: Fouquet, Annie/Méasson, Ludovic (Hrsg.), *L'évaluation des politiques publiques en Europe: Cultures et futurs*. Paris: L'Harmattan, 133–142.
- De Vaus, David (2014): *Surveys in Social Research*. 6. Auflage. London: Routledge.
- De Vries, Catherine E./Hakhverdian, Armen/Lancee, Bram (2013): The Dynamics of Voters' Left/Right Identification: The Role of Economic and Cultural Attitudes. In: *Political Science Research and Methods*, 1 (2), 223–238.
- Derlien, Hans-Ulrich (1990): Genesis and Structure of Evaluation Efforts in Comparative Perspective. In: Rist, Ray C. (Hrsg.), *Program Evaluation and the Management of Government: Patterns and Prospects Across Eight Nations*. New Brunswick: Transaction Publishers, 147–176.
- Dolder, Olivier (2018): *Evaluation und NPM: Gegenseitiger Einfluss von Evaluationskultur und New Public Management*. Dissertation (in Vorbereitung). Luzern.
- Dolder, Olivier/Rohrbach, Walter/Varone, Frédéric (2017): Evaluationskultur auf kantonaler Ebene: politikfeld- oder kantonsspezifische Entwicklungspfade? In: Sager, Fritz/Widmer, Thomas/Balthasar, Andreas (Hrsg.), *Evaluation im politischen System der Schweiz: Entwicklung, Bedeutung und Wechselwirkungen*. Zürich: NZZ Libro, 119–153.
- Duscha, Markus/Klemisch, Herbert/Meyer, Wolfgang (2009): Umweltevaluation in Deutschland – Entwicklungstrends mit Fokus auf dem Energiesektor. In: Widmer, Thomas/Beywl, Wolfgang/Fabian, Carlo (Hrsg.), *Evaluation*. VS Verlag für Sozialwissenschaften, 203–212.
- Easton, David (1965a): *A Framework for Political Analysis*. Englewood Cliffs N.J.: Prentice-Hall.
- Easton, David (1965b): *A Systems Analysis of Political Life*. New York: Wiley.
- Eberli, Daniela/Bundi, Pirmin (2017): Parlament und Evaluation: Guts Meets Brain. In: Sager, Fritz/Widmer, Thomas/Balthasar, Andreas (Hrsg.), *Evaluation im politischen System der Schweiz: Entwicklung, Bedeutung und Wechselwirkungen*. Zürich: Neue Zürcher Zeitung NZZ Libro, 243–278.
- ED, Basel-Stadt (2014): Organigramm der Erziehungsdirektion des Kantons Basel-Stadt. <http://www.ed.bs.ch/ueber-das-departement/bereiche.html> (Zugriff am 2.8.2014).
- ED, Basel-Stadt (2016): Erziehungsdepartement des Kantons Basel-Stadt. <http://www.ed.bs.ch/ueber-das-departement/bereiche.html> (Zugriff am 2.3.2016).

- ED HS, Basel-Stadt (2016): Bereich Hochschule des Erziehungsdepartementes des Kantons Basel-Stadt. <http://www.hochschulen.bs.ch/ueber-uns/organisation.html> (Zugriff am 7.4.2016).
- ED JFF, Basel-Stadt (2016): Abteilung Jugend, und Familienförderung des Erziehungsdepartementes des Kantons Basel-Stadt. <http://www.jfs.bs.ch/ueber-uns/aufgaben-organisation/jugend-und-familienfoerderung/planung.html> (Zugriff am 31.3.2016).
- ED SPD, Basel-Stadt (2016): Schulpsychologischer Dienst SPD des Erziehungsdepartementes des Kantons Basel-Stadt. <https://www.edubs.ch/dienste/spd> (Zugriff am 7.4.2016).
- Epple, Ruedi (2013): Die Berechnung des Traditionalismus-Index. Working Paper. https://lettres.unifr.ch/fileadmin/Documentation/Departements/Sciences_sociales/Soziologie__Sozialpolitik_und_Sozialarbeit/Forschung/NFP60/Trad_Index.pdf (Zugriff am 2.6.2017).
- ERZ B&E, Bern (2016): Abteilung Bildungsplanung und Evaluation der Erziehungsdirektion des Kantons Bern. http://www.erz.be.ch/erz/de/index/direktion/organisation/generalsekretariat/bildungsplanung_und-evaluation1.html.
- ERZ, Bern (2014): Organigramm der Erziehungsdirektion des Kantons Bern. <http://www.erz.be.ch/erz/de/index/direktion/organisation/organigramm.html> (Zugriff am 29.8.2014).
- ERZ, Bern (2016): Erziehungsdirektion des Kantons Bern. <http://www.erz.be.ch/erz/de/index/direktion/ueber-die-direktion.html> (Zugriff am 29.3.2016).
- ERZB, Bern (2015): Erziehungsberatung der Erziehungsdirektion des Kantons Bern. <http://www.erz.be.ch/erz/de/index/erziehungsberatung/erziehungsberatung/ueber-uns.html> (Zugriff am 19.10.2015).
- Forss, Kim (2012): Four Organizations-Four Evaluation Cultures. In: Barbier, Jean-Claude/Hawkins, Penny (Hrsg.), *Evaluation Cultures. Sense Making in Complex Times*. New Brunswick: Transaction Publishers, 177–204.
- Forss, Kim/Bandstein, Sara (2008): Evidence-Based Evaluation of Development Cooperation: Possible? Feasible? Desirable? In: *Nonie Working Papers Nr. 8.*, 39 (1), 82–89.
- Freire, André (2006): Bringing Social Identities Back. In: *The Social Anchors of Left-Right Orientation in Western Europe*. In: *International Political Science Review*, 27 (4), 359–378.

- Freire, André/Kivistik, Kats (2013): Western and Non-Western Meanings of the Left-Right Divide Across Four Continents. In: *Journal of Political Ideologies*, 18 (2), 171–199.
- Freitag, Markus (2005): Labor Schweiz: Vergleichende Wahlbeteiligungsforschung bei kantonalen Parlamentswahlen. In: *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 57 (4), 667–690.
- Frey, Kathrin (2012): Evidenzbasierte Politikformulierung in der Schweiz: Gesetzesrevisionen im Vergleich. Baden-Baden: Nomos.
- Frey, Kathrin/Di Giorgi, Vanessa/Widmer, Thomas (2017): Die Evaluationstätigkeit in der Schweiz: ausgewählte Entwicklungen im Vergleich. In: Sager, Fritz/Widmer, Thomas/Balthasar, Andreas (Hrsg.), *Evaluation im politischen System der Schweiz: Entwicklung, Bedeutung und Wechselwirkungen*. Zürich: Neue Zürcher Zeitung NZZ Libro, 67–87.
- Frey, Kathrin/Ledermann, Simone (2017): Nutzung von Evaluationen in Regierung und Verwaltung. In: Sager, Fritz/Widmer, Thomas/Balthasar, Andreas (Hrsg.), *Evaluation im politischen System der Schweiz: Entwicklung, Bedeutung und Wechselwirkungen*. Zürich: Neue Zürcher Zeitung NZZ Libro, 211–242.
- Furubo, Jan-Eric/Rist, Ray C./Sandahl, Rolf (2002): *International Atlas of Evaluation*. Brunswick: Transaction Publishers.
- Furubo, Jan-Eric/Sandahl, Rolf (2002): A Diffusion Perspective on Global Developments in Evaluation. In: Furubo, Jan-Eric/Rist, Ray C./Sandahl, Rolf (Hrsg.), *International Atlas of Evaluation*. New Brunswick: Transaction Publishers.
- Gauchat, Gordon (2015): The Political Context of Science in the United States: Public Acceptance of Evidence-Based Policy and Science Funding. In: *Social Forces*, 94 (2), 723–748.
- GD, Basel-Stadt (2014): Organigramm des Gesundheitsdepartementes des Kantons Basel-Stadt. <http://www.gd.bs.ch/ueber-das-departement/bereiche-abteilungen.html> (Zugriff am 29.8.2014).
- GD, Basel-Stadt (2016): Gesundheitsdepartement des Kantons Basel-Stadt. <http://www.gd.bs.ch/ueber-das-departement.html> (Zugriff am 29.3.2016).
- GEF, Bern (2014): Organigramm der Gesundheits- und Fürsorgedirektion des Kantons Bern. <http://www.gef.be.ch/gef/de/index/direktion/ueber-die-direktion/organigramm.html> (Zugriff am 28.8.2014).

- GEF, Bern (2016): Gesundheits- und Fürsorgedirektion des Kantons Bern.
<http://www.gef.be.ch/gef/de/index/direktion/ueber-die-direktion.html> (Zugriff am 29.3.2016).
- GEF GS, Bern (2016): Generalsekretariat der Gesundheits- und Fürsorgedirektion des Kantons Bern.
http://www.gef.be.ch/gef/de/index/direktion/organisation/gs/ueber_uns.html (Zugriff am 5.4.2016).
- Gehring, Uwe W./Weins, Cornelia (2009): Grundkurs Statistik für Politologen und Soziologen. 5. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Giauque, David/Ritz, Adrian/Varone, Frédéric/Anderfuhren-Biget, Simon (2010): Public Service Motivation. First Empirical Evidence in Swiss Municipalities. In: Yearbook of Swiss Administrative Sciences. Zürich: Schweizerische Gesellschaft für Verwaltungswissenschaften, 159–170.
- Giauque, David/Ritz, Adrian/Varone, Frédéric/Anderfuhren-Biget, Simon/Waldner, Christian (2011): Putting Public Service Motivation into Context: A Balance between Universalism and Particularism. In: International Review of Administrative Sciences, 77 (2), 227–253.
- Greiffenhagen, Sylvia (2009): Theorie(n) der Politischen Kultur. In: Salzborn, Samuel (Hrsg.), Politische Kultur: Forschungsstand und Forschungsperspektiven. Frankfurt am Main: P. Lang, 11–29.
- GSD DS, Luzern (2016): Departementssekretariat des Gesundheits- und Sozialdepartement des Kantons Luzern.
http://www.lu.ch/verwaltung/GSD/gsd_dienststellen/gsd_departementssekretariat (Zugriff am 14.9.2016).
- GSD, Luzern (2014): Organigramm des Gesundheits- und Sozialdepartementes des Kantons Luzern. <http://www.lu.ch/verwaltung/GSD> (Zugriff am 11.6.2014).
- GSD, Luzern (2016a): Gesundheits- und Sozialdepartement des Kantons Luzern.
http://www.lu.ch/verwaltung/GSD/gsd_unser_department (Zugriff am 22.2.2016).
- GSD, Luzern (2016b): Kennzahlen des Gesundheits- und Sozialdepartementes des Kantons Luzern.
http://www.lu.ch/verwaltung/GSD/gsd_unser_department/kennzahlen (Zugriff am 22.2.2016).
- GSD WIRA DLZ, Luzern (2016): Dienstleistungszentrum für Arbeitsmarktliche Angebote der Dienststelle Wirtschaft und Arbeit des Kantons Luzern.

- https://wira.lu.ch/abteilungen/arbeitsmarkt/am_evaluation_amm (Zugriff am 19.9.2016).
- Guba, Egon G./Lincoln, Yvonna (1987): The Countenance of Fourth-Generation Evaluation: Description, Judgement and Negotiation. In: Palumbo, Dennis James (Hrsg.), *The Politics of Program Evaluation*. Newbury Park: Sage Publications, 202–234.
- Häder, Michael (2010): *Empirische Sozialforschung: Eine Einführung*. 2. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Heinelt, Hubert (2006): Do Policies Determine Politics? In: Fischer, Frank/Miller, Gerald J./Sidney, Mara S. (Hrsg.), *Handbook of Public Policy Analysis*. Boca Raton: CRC Press, 109–119.
- Henry, Gary T./Mark, Melvin M. (2003): Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. In: *American Journal of Evaluation*, 24 (3), 293–314.
- Hense, Jan/Rädiker, Stefan/Böttcher, Wolfgang/Widmer, Thomas (Hrsg.) (2013): *Forschung über Evaluation: Bedingungen, Prozesse und Wirkungen*. Münster: Waxmann.
- Hense, Jan/Widmer, Thomas (2013): Ein Überblick zum internationalen Stand der Forschung über Evaluation. In: Hense, Jan/Rädiker, Stefan/Böttcher, Wolfgang/Widmer, Thomas (Hrsg.), *Forschung über Evaluation: Bedingungen, Prozesse und Wirkungen*. Münster: Waxmann, 251–278.
- Hilbe, Joseph M. (2011): *Negative Binomial Regression*. 2. Auflage. Cambridge: Cambridge University Press.
- Hindmarsh, Richard (2013): Nuclear Disaster at Fukushima Daiichi. Introducing the Terrain. In: Hindmarsh, Richard (Hrsg.), *Nuclear Disaster at Fukushima Daiichi: Social, Political and Environmental Issues*. London: Routledge, 1–21.
- Hofstede, Geert H. (1980): *Culture's Consequences: International Differences in Work-Related Values*. Beverly Hills, California: Sage Publications.
- Horber-Papazian, Katja (2006): La place de l'évaluation des politiques publiques en Suisse. In: Chappellet, Jean-Loup (Hrsg.), *Contributions à l'action publique*. Lausanne: Presses polytechniques et universitaires romandes, 131–144.
- Horber-Papazian, Katja/Jacot-Descombes, Caroline (2012): Is Evaluation Culture Shaped by the Swiss Political System and Multiculturalism? In: Barbier, Jean-Claude/Hawkins, Penny (Hrsg.), *Evaluation Cultures. Sense Making in Complex Times*. New Brunswick: Transaction Publishers, 89–104.

- Horber-Papazian, Katja/Thévoz, Laurent (1990): Switzerland: Moving towards Evaluation. In: Rist, Ray C. (Hrsg.), *Program Evaluation and the Management of Government: Patterns and Prospects across Eight Nations*. New Brunswick: Transaction Publishers, 59–73.
- Howlett, Michael (2009): Policy Analytical Capacity and Evidence-Based Policy-Making: Lessons from Canada. In: *Canadian Public Administration*, 52 (2), 153–175.
- Hox, Joop (2010): *Multilevel Analysis: Techniques and Applications*. 2. Auflage. New York: Routledge.
- Hupe, Peter/Bufat, Aurélien (2014): A Public Service Gap: Capturing Contexts in a Comparative Approach of Street-Level Bureaucracy. In: *Public Management Review*, 16 (4), 548–569.
- Jacob, Steve (2005): *Institutionnaliser l'évaluation des politiques publiques: étude comparée des dispositifs institutionnels en Belgique, en France, en Suisse et aux Pays-Bas*. Bruxelles : New York: P.I.E.-P. Lang.
- Jacob, Steve/Boisvert, Yves (2010): To Be or Not to Be a Profession: Pros, Cons and Challenges for Evaluation. In: *Evaluation*, 16 (4), 349–369.
- Jacob, Steve/Speer, Sandra/Furubo, Jan-Eric (2015): The Institutionalization of Evaluation Matters: Updating the International Atlas of Evaluation 10 Years Later. In: *Evaluation*, 21 (1), 6–31.
- Jacob, Steve/Varone, Frédéric (2003): *Evaluer l'action publique: Etat des lieux et perspectives en Belgique*. Gent: Academia Press.
- Jacob, Steve/Varone, Frédéric (2004): Cheminement institutionnel de l'évaluation des politiques publiques en France, en Suisse et aux Pays-Bas (1970-2003). In: *Politiques et Management Public*, 22 (2), 135–152.
- Jahn, Detlef (2005): Fälle, Fallstricke und die komparative Methode in der vergleichenden Politikwissenschaft. In: Kropp, Sabine/Minkenbergh, Michael (Hrsg.), *Vergleich in der Politikwissenschaft*. Wiesbaden: VS Verlag für Sozialwissenschaften, 55-75.
- Jahn, Detlef (2013): *Einführung in die vergleichende Politikwissenschaft*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Jann, Werner/Wegrich, Kai (2006): Theories of the Policy Cycle. In: Fischer, Frank/Miller, Gerald J./Sidney, Mara S. (Hrsg.), *Handbook of Public Policy Analysis*. Boca Raton: CRC Press, 43–62.

- Johnson, Kelli/Greenseid, Lija O./Toal, Stacie A./King, Jean A./Lawrenz, Frances/Volkov, Boris (2009): Research on Evaluation Use: A Review of the Empirical Literature from 1986 to 2005. In: *American Journal of Evaluation*, 30 (3), 377–410.
- Jost, John T./Federico, Christopher M./Napier, Jaime L. (2009): Political Ideology: Its Structure, Functions, and Elective Affinities. In: *Annual Review of Psychology*, 60 (1), 307–337.
- Jost, John T./Nosek, Brian A./Gosling, Samuel D. (2008): Ideology: Its Resurgence in Social, Personality, and Political Psychology. In: *Perspectives on Psychological Science*, 3 (2), 126–136.
- Julnes, Patria De Lancer (2006): Performance Measurement an Effective Tool for Government Accountability? The Debate Goes On. In: *Evaluation*, 12 (2), 219–235.
- Kahan, Dan M./Peters, Ellen/Wittlin, Maggie/Slovic, Paul/Ouellette, Lisa Larrimore/Braman, Donald/Mandel, Gregory (2012): The Polarizing Impact of Science Literacy and Numeracy on Perceived Climate Change Risks. In: *Nature Climate Change*, 2 (10), 732–735.
- Kallemeyn, Leanne M./Hall, Jori/Friche, Nanna/McReynolds, Clifton (2015): Cross-Continental Reflections on Evaluation Practice: Methods, Use, and Valuing. In: *American Journal of Evaluation*, 36 (3), 339–357.
- Katz, Steven/Sutherland, Stephanie/Earl, Lorna (2002): Developing an Evaluation Habit of Mind. In: *Canadian Journal of Program Evaluation*, 17 (2), 103–119.
- Kaufman, Roger A./Guerra-López, Ingrid/Platt, William A. (2006): *Practical Evaluation for Educators: Finding What Works and What Doesn't*. Thousand Oaks, California: Corwin Press.
- King, Jean A./Stevahn, Laurie/Ghere, Gail/Minnema, Jane (2001): Toward a Taxonomy of Essential Evaluator Competencies. In: *American Journal of Evaluation*, 22 (2), 229–247.
- Kirkhart, Karen E. (2000): Reconceptualizing Evaluation Use: An Integrated Theory of Influence. In: *New Directions for Evaluation*, 2000 (88), 5–23.
- Kissling-Näf, Ingrid/Knoepfel, Peter (1994): Politikorientierte Lernprozesse: Konzeptionelle Überlegungen. In: Bussmann, Werner (Hrsg.), *Lernen in Verwaltungen und Policy-Netzwerken*. 1. Auflage. Chur: Rüegger, 99–129.

- Klöti, Ulrich/Schneider, Gerald (1989): Die Informationsbeschaffung des Gesetzgebers: Eine vergleichende Evaluationsstudie zur Bedeutung von Wirkungsanalysen im Schweizer Legiferierungsprozess. Grösch: Rüegger.
- Knutsen, Oddbjörn (1995): Value Orientations, Political Conflicts and Left-Right Identification: A Comparative Study. In: *European Journal of Political Research*, 28 (1), 63–93.
- Koller, Christophe (2013a): Die kantonalen Verwaltungen. In: Ladner, Andreas/Chappelet, Jean-Loup/Emery, Yves/Knoepfel, Peter/Mader, Luzius/Soguel, Nils/Varone, Frédéric (Hrsg.), *Handbuch der öffentlichen Verwaltung in der Schweiz*. Zürich: Verlag Neue Zürcher Zeitung, 127–148.
- Koller, Christophe (2013b): Profil des Personals der öffentlichen Hand. In: Ladner, Andreas/Chappelet, Jean-Loup/Emery, Yves/Knoepfel, Peter/Mader, Luzius/Soguel, Nils/Varone, Frédéric (Hrsg.), *Handbuch der öffentlichen Verwaltung in der Schweiz*. Zürich: Verlag Neue Zürcher Zeitung, 499–520.
- Kopp, Johannes/Lois, Daniel (2012): *Sozialwissenschaftliche Datenanalyse: Eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kübler, Daniel/Widmer, Thomas (2007): Subnationale Unterstützung eines nationalen Drogenprogramms. Eine Panelanalyse mit Ereignisdaten in den Schweizer Kantonen. In: Holzinger, Katharina/Jörgens, Helge/Knill, Christoph (Hrsg.), *Transfer, Diffusion und Konvergenz von Politiken*, 200–220.
- Kuckartz, Udo/Rädiker, Stefan/Ebert, Thomas/Schehl, Kulia (2010): *Statistik - Eine verständliche Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Labin, Susan N./Duffy, Jennifer L./Meyers, Duncan C./Wandersman, Abraham/Lesesne, Catherine A. (2012): A Research Synthesis of the Evaluation Capacity Building Literature. In: *American Journal of Evaluation*, 33 (3), 307–338.
- Läubli Loud, Marlène (2014): Institutionalization and Evaluation Culture—Interplay Between the One and the Other: Lessons From the Swiss Federal Office of Public Health (FOPH). In: Läubli Loud, Marlène/Mayne, John (Hrsg.), *Enhancing Evaluation Use*. Thousand Oaks: Sage Publications, 55–82.
- Lauth, Hans-Joachim/Pickel, Gert/Pickel, Susanne (2015): *Methoden der vergleichenden Politikwissenschaft*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Ledermann, Simone (2014): Evidenz und Expertise im vorparlamentarischen Gesetzgebungsprozess: Die Rolle von Verwaltung und externen Experten. In: *Swiss Political Science Review*, 20 (3), 453–485.

- Lee, Robert/Lawrence, Peter (2013): *Organizational Behaviour: Politics at Work*. London: Routledge.
- Leeuw, Frans L. (2000): Evaluation in Europe. In: Stockmann, Reinhard (Hrsg.), *Evaluationsforschung: Grundlagen und ausgewählte Forschungsfelder*. Opladen: Leske + Budrich, 57–76.
- Leeuw, Frans L./Rozendal, Piet J. (1994): Policy Evaluation and the Netherlands' Government: Scope, Utilization and Organizational Learning. In: Leeuw, Frans L./Rist, Ray C./Sonnichsen, Richard C. (Hrsg.), *Can Governments Learn? Comparative Perspectives on Evaluation and Organizational Learning*. New Brunswick: Transaction Publishers, 67–89.
- Linder, Wolf (2012): *Schweizerische Demokratie. Institutionen - Prozesse - Perspektiven*. 3. Auflage. Bern: Haupt Verlag.
- Lipset, Seymour M. (1959): Some Social Requisites of Democracy: Economic Development and Political Legitimacy. In: *American Political Science Review*, 53 (01), 69–105.
- Lipset, Seymour M. (1981): *Political Man: The Social Bases of Politics*. Baltimore: John Hopkins.
- Lipsky, Michael (2010): *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services (30th Anniversary Expanded Edition)*. New York: Russell Sage Foundation.
- Lowi, Theodore J. (1972): Four Systems of Policy, Politics, and Choice. In: *Public Administration Review*, 32 (4), 298–310.
- Lutz, Georg (2016): *Eidgenössische Wahlen 2015. Wahlteilnahme und Wahlentscheid*. Lausanne: FORS. <http://forscenter.ch/wp-content/uploads/2013/10/Selects-2015-Brochure-DE1.pdf> (Zugriff am 25.5.2017).
- Mackay, Keith (2002): The World Bank's ECB Experience. In: *New Directions for Evaluation*, 2002 (93), 81–100.
- Mader, Luzius (1985): *L'évaluation législative : pour une analyse empirique des effets de la législation*. Lausanne: Payot.
- Mader, Luzius (2009): Die institutionelle Einbettung der Evaluationsfunktion in der Schweiz. In: Widmer, Thomas/Beywl, Wolfgang/Fabian, Carlo (Hrsg.), *Evaluation*. Wiesbaden: VS Verlag für Sozialwissenschaften, 52–63.
- Martin, Christian W. (2009): Theoretische Modellbildung. In: Pickel, Susanne/Pickel, Gert/Lauth, Hans-Joachim/Jahn, Detlef (Hrsg.), *Methoden der vergleichenden Po-*

- litik- und Sozialwissenschaft. Wiesbaden: VS Verlag für Sozialwissenschaften, 37–52.
- Martin, Joanne (2002): *Organizational Culture: Mapping the Terrain*. Thousand Oaks: Sage Publications.
- Martin, John Levi/Desmond, Matthew (2010): Political Position and Social Knowledge. In: *Sociological Forum*, 25 (1), 1–26.
- Mayne, John (2011): Building an Evaluative Culture: The Key to Effective Evaluation and Results Management. In: *Canadian Journal of Program Evaluation*, 24 (2), 1–30.
- Mayne, John/Rist, Ray C. (2006): Studies Are Not Enough: The Necessary Transformation of Evaluation. In: *Canadian Journal of Program Evaluation*, 21 (3), 93–120.
- McCright, Aaron M./Dunlap, Riley E. (2011): The Politicization of Climate Change and Polarization in the American Public's Views of Global Warming, 2001–2010. In: *The Sociological Quarterly*, 52 (2), 155–194.
- McNamara, Gerry/O'Hara, Joe/Boyle, Richard/Sullivan, Conor (2009): Developing a Culture of Evaluation in the Irish Public Sector The Case of Education. In: *Evaluation*, 15 (1), 101–112.
- Meyer, Wolfgang/Rech, Jörg (2013): Determinanten der Evaluationskultur- Das Beispiel Entwicklungszusammenarbeit. In: Hense, Jan/Rädiker, Stefan/Böttcher, Wolfgang/Widmer, Thomas (Hrsg.), *Forschung über Evaluation*. Münster: Waxmann, 189–210.
- Mihalache, Roxana (2010): A Developing Evaluation Culture in Romania: Myths, Gaps and Triggers. In: *Evaluation*, 16 (3), 323–332.
- Milic, Thomas (2008): Links, Mitte, Rechts: Worauf gründet die ideologische Selbstidentifikation der Schweizer und Schweizerinnen? In: *Swiss Political Science Review*, 14 (2), 245–285.
- Montoya, R. Matthew/Horton, Robert S./Vevea, Jack L./Citkowicz, Martyna/Lauber, Elissa A. (2017): A Re-Examination of the Mere Exposure Effect: The Influence of Repeated Exposure on Recognition, Familiarity, and Liking. In: *Psychological Bulletin*, 143 (5), 459–498.
- Moynihan, Donald P./Pandey, Sanjay K. (2007): The Role of Organizations in Fostering Public Service Motivation. In: *Public Administration Review*, 67 (1), 40–53.

- Mummendey, Hans Dieter (1988): Verhalten und Einstellung: Untersuchung der Einstellungs- und Selbstkonzeptänderung nach Änderung des alltäglichen Verhaltens. Berlin: Springer.
- Nielsen, Steffen Bohni/Ejler, Nicolaj (2008): Improving Performance? Exploring the Complementarities between Evaluation and Performance Management. In: *Evaluation*, 14 (2), 171–192.
- Nielsen, Steffen Bohni/Lemire, Sebastian/Skov, Majbritt (2011): Measuring Evaluation Capacity—Results and Implications of a Danish Study. In: *American Journal of Evaluation*, 32 (3), 324–344.
- Norpoth, Helmut/Lodge, Milton (1985): The Difference between Attitudes and Nonattitudes in the Mass Public: Just Measurements? In: *American Journal of Political Science*, 29 (2), 291–307.
- OECD (1997): Promouvoir l'utilisation de l'évaluation de programme. Paris: OECD.
- OECD (2010): Evaluation in Development Agencies.
<http://www.oecd.org/development/evaluation/dcdndep/49592538.pdf>.
- Opp, Karl-Dieter (2010): Kausalität als Gegenstand der Sozialwissenschaften und der multivariaten Statistik. In: Wolf, Christof/Best, Henning (Hrsg.), Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS Verlag für Sozialwissenschaften, 9–38.
- Owen, John M. (2003): Evaluation Culture: a Definition and Analysis of its Development within Organisations. In: *Evaluation Journal of Australasia*, 3 (1), 43–47.
- Owen, John M./McDonald, Diane E. (1999): Creating an Evaluation Culture in International Development Cooperation Agencies. In: *Journal of International Cooperation in Education*, 2 (2), 41–53.
- Pandey, Sanjay K./Stazyk, Edmund C. (2008): Antecedents and Correlates of Public Service Motivation. In: Perry, James L./Hondegheim, Annie (Hrsg.), *Motivation in Public Management: The Call of Public Service*. Reprinted. New York: Oxford University Press, 101–117.
- Patton, Michael Quinn (1998): Discovering Process Use. In: *Evaluation*, 4 (2), 225–233.
- Pattyn, Valérie (2014): Why Organizations (Do Not) Evaluate? Explaining Evaluation Activity through the Lens of Configurational Comparative Methods. In: *Evaluation*, 20 (3), 348–367.

- Pattyn, Valérie (2015): Explaining Variance in Policy Evaluation Regularity. The Case of the Flemish Public Sector. In: *Public Management Review*, 17 (10), 1475–1495.
- Perrin, Andrew J./Micah Roos, J./Gauchat, Gordon W. (2014): From Coalition to Constraint: Modes of Thought in Contemporary American Conservatism. In: *Sociological Forum*, 29 (2), 285–300.
- Personalamt, Bern (2014): Personalkennzahlen der Verwaltungsangestellten des Kantons Bern.
<http://www.fin.be.ch/fin/de/index/personal/anstellungsbedingungen/statistik.html>
 (Zugriff am 14.10.2015).
- Pettigrew, Thomas F./Tropp, Linda R. (2006): A Meta-Analytic Test of Intergroup Contact Theory. In: *Journal of Personality and Social Psychology*, 90 (5), 751–783.
- Pickel, Susanne/Pickel, Gert (2006): Politische Kultur- und Demokratieforschung: Grundbegriffe, Theorien, Methoden: eine Einführung. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Porst, Rolf (2011): Fragebogen: Ein Arbeitsbuch. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Preskill, Hallie/Boyle, Shanelle (2008): A Multidisciplinary Model of Evaluation Capacity Building. In: *American Journal of Evaluation*, 29 (4), 443–459.
- Preskill, Hallie/Torres, Rosalie T. (1999): Building Capacity for Organizational Learning Through Evaluative Inquiry. In: *Evaluation*, 5 (1), 42–60.
- Radaelli, Claudio M. (1999): The Public Policy of the European Union: Whither Politics of Expertise? In: *Journal of European Public Policy*, 6 (5), 757–774.
- RAV (2015): Arbeitsvermittlung (Arbeit) Volkswirtschaftsdirektion - Kanton Bern.
<http://www.vol.be.ch/vol/de/index/arbeit/arbeitsvermittlung.html> (Zugriff am 19.10.2015).
- Rhyn, Heinz (2009): Evaluation im Bildungsbereich in der Schweiz. In: Widmer, Thomas/Beywl, Wolfgang/Fabian, Carlo (Hrsg.), *Evaluation*. Wiesbaden: VS Verlag für Sozialwissenschaften, 182–192.
- Rieder, Stefan/Widmer, Thomas (2007): Kantone im Wandel. Reformaktivitäten der Schweizer Kantone zwischen 1990 und 1999: Ursachen, Ausgestaltung und Konsequenzen. Bern: Haupt Verlag.
- Rist (2006): The „E“ in Monitoring and Evaluation: Using Evaluative Knowledge to Support a Resultsbased Management System. In: Rist, Ray C./Stame, Nicoletta

- (Hrsg.), *From Studies to Streams: Managing Evaluative Systems*. New Brunswick: Transaction Publishers, 3–22.
- Rodríguez-Bilella, Pablo/Monterde-Díaz, Rafael (2010): Evaluation, Valuation, Negotiation: Some Reflections Towards a Culture of Evaluation. In: *The Canadian Journal of Program Evaluation*, 25 (3), 1–10.
- Rohe, Karl (1996): Politische Kultur: Zum Verständnis eines theoretischen Konzeptes. In: Niedermayer, Oskar/Beyme, Klaus (Hrsg.), *Politische Kultur in Ost- und Westdeutschland*. Opladen: Leske + Budrich, 1–21.
- Sabatier, Paul A. (1987): Knowledge, Policy-Oriented Learning, and Policy Change; An Advocacy Coalition Framework. In: *Knowledge: Creation, Diffusion, Utilization*, 8, 649–692.
- Sabatier, Paul A./Jenkins-Smith, Hank C. (1993): *Policy Change and Learning. An Advocacy Coalition Approach*. Boulder: Westview Press.
- Sabatier, Paul A./Jenkins-Smith, Hank C. (1999): The Advocacy Coalition Framework: An Assessment. In: Sabatier, Paul A. (Hrsg.), *Theories of the Policy Process*. Boulder: Westview Press.
- Salzborn, Samuel (2009): Einführung. In: Salzborn, Samuel (Hrsg.), *Politische Kultur: Forschungsstand und Forschungsperspektiven*. Frankfurt am Main: P. Lang, 7–10.
- Sanders, James R. (2002): Presidential Address: On Mainstreaming Evaluation. In: *American Journal of Evaluation*, 23 (3), 253–259.
- Sanders, James R. (2003): Mainstreaming Evaluation. In: *New Directions for Evaluation*, 2003 (99), 3–6.
- Schäfer, Thomas (2010): *Statistik I: Deskriptive und Explorative Datenanalyse*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Scheerens, Jaap (2004): The Evaluation Culture. In: *Studies in Educational Evaluation*, 30 (2), 105–124.
- Schein, Edgar H. (2010): *Organizational Culture and Leadership*. 4. Auflage. San Francisco: Jossey-Bass.
- Schrefler, Lorna (2010): The Usage of Scientific Knowledge by Independent Regulatory Agencies. In: *Governance*, 23 (2), 309–330.
- Schwandt, Thomas/Dahler-Larsen, Peter (2006): When Evaluation Meets the ‘Rough Ground’ in Communities. In: *Evaluation*, 12 (4), 496–505.

- Schwenkel, Christof (2016): Der Einfluss kantonaler Justizsysteme auf das Vertrauen der Bevölkerung in die Gerichte. Bern: Stämpfli Verlag.
- Shadish, William R./Cook, Thomas D./Leviton, Laura C. (1991): Foundations of Program Evaluation: Theories and Practice. Newbury Park: Sage Publications.
- Smircich, Linda (1983): Concepts of Culture and Organizational Analysis. In: *Administrative Science Quarterly*, 28 (3), 339–358.
- Snijders, Tom A. B./Bosker, Roel J. (2012): Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. 2. Auflage. Los Angeles: Sage Publications.
- Spector, Paul E. (2006): Method Variance in Organizational Research: Truth or Urban Legend? In: *Organizational Research Methods*, 9 (2), 221–232.
- Speer, Sandra (2012): Sectoral Evaluation Cultures: A Comparison of the Education and Labour Market Sectors in Germany. In: Barbier, Jean-Claude/Hawkins, Penny (Hrsg.), Evaluation Cultures. Sense Making in Complex Times. New Brunswick: Transaction Publishers, 65–88.
- Spinatsch, Markus (2002): Evaluation in Switzerland: Moving toward a Decentralized System. In: Furubo, Jan-Eric/Rist, Ray C./Sandahl, Rolf (Hrsg.), International Atlas of Evaluation. New Brunswick: Transaction Publishers, 375–391.
- Staatskanzlei, Bern (2016): Staatskalender des Kantons Bern.
http://www.sta.be.ch/sta/de/index/staatskanzlei/staatskanzlei/downloads_publicationen/staatskalender.html#originRequestUrl=www.be.ch/staatskalender (Zugriff am 11.11.2016).
- Stame, Nicoletta (2012): The Culture of Evaluation in Italy. In: Barbier, Jean-Claude/Hawkins, Penny (Hrsg.), Evaluation Cultures. Sense Making in Complex Times. New Brunswick: Transaction Publishers, 19–43.
- Stamm, Margrit (2003): Evaluation und ihre Folgen für die Bildung: Eine unterschätzte pädagogische Herausforderung. 1. Auflage. Münster: Waxmann.
- Statistisches Amt, Basel-Stadt (2014): Personalkennzahlen der Verwaltungsangestellten des Kantons Basel-Stadt. <http://www.statistik.bs.ch/zahlen/tabellen/18-oeffentliche-finanzen/staatspersonal.html> (Zugriff am 2.3.2015).
- Steenbergen, Marco R./Jones, Bradford S. (2002): Modeling Multilevel Data Structures. In: *American Journal of Political Science*, 46 (1), 218–237.
- Stern, Elliot (2006): Contextual Challenges for Evaluation Practice. In: Shaw, Ian F./Greene, Jennifer C./Mark, Melvin M. (Hrsg.), The Sage Handbook of Evalua-

- tion: Policies, Programs and Practices. 1. Auflage. Kalifornien: Thousand Oaks, 292–314.
- Stockdill, Stacey Hueftle/Baizerman, Michael/Compton, Donald W. (2002): Toward a Definition of the ECB Process: A Conversation with the ECB Literature. In: *New Directions for Evaluation*, 2002 (93), 7–26.
- Stockmann, Reinhard (2000): Evaluation in Deutschland. In: Stockmann, Reinhard (Hrsg.), *Evaluationsforschung: Grundlagen und ausgewählte Forschungsfelder*. Opladen: Leske + Budrich, 11–40.
- Stolz, Jörg (2001): Traditionalismus und das Fremde: Einstellungen zu Ausländern und Ausländerinnen. In: Hoffmann-Nowotny, Hans-Joachim (Hrsg.), *Das Fremde in der Schweiz: Ergebnisse soziologischer Forschung*. Zürich: Seismo-Verlag, 81–105.
- Stufflebeam, Daniel L. (2001): Evaluation Checklists: Practical Tools for Guiding and Judging Evaluations. In: *American Journal of Evaluation*, 22 (1), 71–79.
- Stufflebeam, Daniel L. (2002): Institutionalizing Evaluation Checklist. <http://www.wmich.edu/evalctr/checklists> (Zugriff am 11.2.2014).
- Sutherland, Stephanie (2004): Creating a Culture of Data Use for Continuous Improvement: A Case Study of an Edison Project School. In: *American Journal of Evaluation*, 25 (3), 277–293.
- Taut, Sandy (2007): Studying Self-Evaluation Capacity Building in a Large International Development Organization. In: *American Journal of Evaluation*, 28 (1), 45–59.
- Taylor-Ritzler, Tina/Suarez-Balcazar, Yolanda/Garcia-Iriarte, Edurne/Henry, David B./Balcazar, Fabricio E. (2013): Understanding and Measuring Evaluation Capacity A Model and Instrument Validation Study. In: *American Journal of Evaluation*, 34 (2), 190–206.
- Torres, Rosalie T./Preskill, Hallie (2001): Evaluation and Organizational Learning: Past, Present, and Future. In: *American Journal of Evaluation*, 22 (3), 387–395.
- Toulemonde, Jacques (2000): Evaluation Culture(s) in Europe: Differences and Convergence between National Practices. In: *Vierteljahrshefte zur Wirtschaftsforschung*, 69 (3), 350–357.
- Trochim, William M. K. (2006): An Evaluation Culture. Research Methods Knowledge Base. <http://www.socialresearchmethods.net/kb/evalcult.php> (Zugriff am 29.3.2016).

- Uchtenhagen, Ambros (2009): Evaluation im Gesundheitsbereich: Schweiz. In: Widmer, Thomas/Beywl, Wolfgang/Fabian, Carlo (Hrsg.), Evaluation. Wiesbaden: VS Verlag für Sozialwissenschaften, 341–352.
- Urban, Dieter/Mayerl, Jochen (2011): Regressionsanalyse: Theorie, Technik und Anwendung. 4. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ridde, Valéry /Shakir, Sahibullah (2005): Evaluation Capacity Building and Humanitarian Organization. In: *Journal of MultiDisciplinary Evaluation*, 2 (3), 78–112.
- Van der Knaap, Peter (1995): Policy Evaluation and Learning Feedback, Enlightenment or Argumentation? In: *Evaluation*, 1 (2), 189–216.
- Van Deth, Jan W./Tausendpfund, Markus (2013): Schlussbetrachtung: Warum ist nicht alle Politik lokale Politik? In: Van Deth, Jan W./Tausendpfund, Markus (Hrsg.), Politik im Kontext: Ist alle Politik lokale Politik? Wiesbaden: Springer Fachmedien Wiesbaden, 441–455.
- Varone, Frédéric (2007): Développer les capacités d'évaluation. In: Genard, Jean-Louis/Jacob, Steve/Varone, Frédéric (Hrsg.), L'évaluation des politiques au niveau régional. Brüssel: P.I.E. Peter Lang, 157–177.
- Varone, Frédéric (2015): Capacités évaluatives et recherche académique: quelle dynamique en Suisse? In: Horber-Papazian, Katia (Hrsg.), Regards croisés sur l'évaluation en Suisse. Lausanne: Presses polytechniques et universitaires romandes, 247–265.
- Varone, Frédéric/Jacob, Steve/De Winter, Lieven (2005): Polity, Politics and Policy Evaluation in Belgium. In: *Evaluation*, 11 (3), 253–273.
- Vatter, Adrian (2002): Kantonale Demokratien im Vergleich: Entstehungsgründe, Interaktionen und Wirkungen politischer Institutionen in den Schweizer Kantonen. Opladen: Leske + Budrich.
- Vatter, Adrian (2014): Das politische System der Schweiz. Baden-Baden: Nomos.
- Vedung, Evert (1999): Evaluation im öffentlichen Sektor. Wien: Böhlau.
- VOL beco SF, Bern (2016): Standortförderung der Volkswirtschaftsdirektion des Kantons Bern.
<https://www.berninvest.be.ch/berninvest/de/index/dienstleistungen/dienstleistungen/evaluationsbesuche.html> (Zugriff am 29.3.2016).
- VOL, Bern (2014): Organigramm der Volkswirtschaftsdirektion des Kantons Bern.
<http://www.vol.be.ch/vol/de/index/direktion/ueber-die-direktion/organigramm.html> (Zugriff am 29.3.2014).

- VOL, Bern (2016): Volkswirtschaftsdirektion des Kantons Bern.
<http://www.vol.be.ch/vol/de/index/direktion/ueber-die-direktion.html> (Zugriff am 29.3.2016).
- VOL GS, Bern (2016): Generalsekretariat der Volkswirtschaftsdirektion des Kantons Bern. http://www.vol.be.ch/vol/de/index/direktion/organisation/gs/ueber_uns.html (Zugriff am 28.10.2016).
- Volkov, Boris B./King, Jean A. (2007): A Checklist for Building Organizational Evaluation Capacity. <http://vermontaccess.net/wp-content/uploads/2013/04/NPOs-Capacity-Building-Criteria.pdf> (Zugriff am 3.12.2013).
- Wagner, Ulrich/Christ, Oliver/Pettigrew, Thomas F./Stellmacher, Jost/Wolf, Carina (2016): Prejudice And Minority Proportion: Contact Instead Of Threat Effects. In: *Social Psychology Quarterly*, 69 (4), 380-390.
- Weible, Christopher M. (2008): Expert-Based Information and Policy Subsystems: A Review and Synthesis. In: *Policy Studies Journal*, 36 (4), 615–635.
- Weible, Christopher M./Sabatier, Paul A./Jenkins-Smith, Hank C./Nohrstedt, Daniel/Henry, Adam Douglas/deLeon, Peter (2011): A Quarter Century of the Advocacy Coalition Framework: An Introduction to the Special Issue. In: *Policy Studies Journal*, 39 (3), 349–360.
- Weible, Christopher M./Sabatier, Paul A./McQueen, Kelly (2009): Themes and Variations: Taking Stock of the Advocacy Coalition Framework. In: *Policy Studies Journal*, 37 (1), 121–140.
- Weins, Cornelia (2010): Uni-und bivariate deskriptive Statistik. In: Wolf, Christof/Best, Henning (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS Verlag für Sozialwissenschaften, 65–89.
- Westle, Bettina (2009): *Politische Kultur: Eine Einführung*. 1. Auflage. Baden-Baden: Nomos.
- Widmer, Thomas (1996): *Meta-Evaluation Kriterien zur Bewertung von Evaluationen*. Bern: Haupt.
- Widmer, Thomas (2008): Evaluation in der Aussenpolitik: Gründe für eine Evaluationslücke. In: *Zeitschrift für Internationale Beziehungen*, 15 (1), 125–136.
- Widmer, Thomas (2017): Geschichte der Evaluation im schweizerischen politischen System. In: Sager, Fritz/Widmer, Thomas/Balthasar, Andreas (Hrsg.), *Evaluation im politischen System der Schweiz: Entwicklung, Bedeutung und Wechselwirkungen*. Zürich: Neue Zürcher Zeitung NZZ Libro, 51–66.

- Widmer, Thomas/Beywl, Wolfgang (2009): Divergente Entwicklungspfade der Evaluation. In: Widmer, Thomas/Beywl, Wolfgang/Fabian, Carlo (Hrsg.), *Evaluation*. Wiesbaden: VS Verlag für Sozialwissenschaften, 509–527.
- Widmer, Thomas/De Rocchi, Thomas (2012): *Evaluation: Grundlagen, Ansätze und Anwendungen*. Zürich/Chur: Rüegger.
- Widmer, Thomas/Eberli, Daniela/Ackermann, Günter/Balzer, Lars/Bundi, Pirmin/Hirschi, Christian/Kasper, Tanya/Neuenschwander, Peter/Neuhaus, Björn/Rohrbach, Walter/et al. (2016): Forschung über Evaluation in der Schweiz: Stand und Aussichten. In: *LeGes: Gesetzgebung & Evaluation*, 27 (3), 459–483.
- Widmer, Thomas/Neuenschwander, Peter (2004): Embedding Evaluation in the Swiss Federal Administration Purpose, Institutional Design and Utilization. In: *Evaluation*, 10 (4), 388–409.
- Widmer, Thomas/Rüegg, Erwin/Neuenschwander, Peter (2001): Stand und Aussichten der Evaluation beim Bund (EvalBund). Schlussbericht zuhanden der Schweizerischen Bundeskanzlei vom 9. März 2001. Zürich: Institut für Politikwissenschaft der Universität Zürich.
- Windhoff-Héritier, Adrienne (1987): *Policy-Analyse: Eine Einführung*. Frankfurt, New York: Campus Verlag.
- Wood, Wendy (2000): Attitude Change: Persuasion and Social Influence. In: *Annual Review of Psychology*, 51 (1), 539–570.
- Zajonc, Robert B. (1968): Attitudinal Effects of Mere Exposure. In: *Journal of Personality and Social Psychology Monographs*, 9, 1–27.
- Zimmermann, Willi/Kolly, Doris/Mahon, Pascal (1989): *Die wissenschaftliche Evaluation von Bundespolitik. Eine bibliographische und methodische Exploration*. Bern: EJPD, Arbeitsgruppe Gesetzesevaluation.

Anhang II Abkürzungsverzeichnis

ACF	Advocacy Coalition Framework
AGEVAL	Arbeitsgruppe Gesetzesevaluation
AIC	Akaike-Informationskriterium
BAG	Bundesamt für Gesundheit
BAK	Bundesamt für Kultur
BFS	Bundesamt für Statistik
BKD	Bildungs- und Kulturdepartement des Kantons Luzern
BIC	Bayes-Informationskriterium
BLUE	Best Linear Unbiased Estimator
BUWAL	Bundesamt für Umwelt, Wald und Landwirtschaft
BUWD	Bau-, Umwelt- und Wirtschaftsdepartement des Kantons Luzern
BVE	Bau-, Verkehrs- und Energiedirektion des Kantons Bern
DEZA	Direktion für Entwicklung und Zusammenarbeit
ECB	Evaluationskapazitätsbildung
ED	Erziehungsdepartement des Kantons Basel-Stadt
EJPD	Eidgenössisches Justiz- und Polizeidepartement
ERKOS	Erfolgskontrolle von Staatsbeiträgen im Kanton Bern
ERZ	Erziehungsdirektion des Kantons Bern
ERZB	Erziehungsberatung der Erziehungsdirektion des Kantons Bern
et al.	Et alii (und andere)
fedpol	Bundesamt für Polizei
GD	Gesundheitsdepartement des Kantons Basel-Stadt
GEF	Gesundheits- und Fürsorgedirektion des Kantons Bern
GPK	Geschäftsprüfungskommission
GSD	Gesundheits- und Sozialdepartement des Kantons Luzern
f.	folgende
ff.	fortfolgende

ICC	Intraklassenkorrelation
IDHEAP	Institut Des Hautes Etudes en Administrations Publiques
IFES	Institut für Externe Schulevaluation auf der Sekundarstufe II
IRR	Incidence Rate Ratio
LR	Likelihood-Ratio-Tests
NFP	Nationales Forschungsprogramm
NPM	New Public Management
OECD	Organisation for Economic Co-operation and Development
OR	Odds Ratios
PRE	Proportional Reduction of Error
PSM	Public Service Motivation
RAV	Regionales Arbeitsvermittlungszentrum
SE	Standard-Fehler
SECO	Staatssekretariat für Wirtschaft
SEVAL	Schweizerische Evaluationsgesellschaft
SNF	Schweizerischer Nationalfonds
VIF	Varianz-Inflations-Faktor
VOL	Volkswirtschaftsdirektion des Kantons Bern

Anhang III Anmerkungen für die bivariate Analyse

Die nachfolgende Tabelle 46 zeigt die verwendeten Korrelationskoeffizienten dieser Untersuchung auf.

Tabelle 46: Übersicht der verwendeten Korrelationskoeffizienten

Skalenniveau Variable y	Skalenniveau Variable x		
	Intervallskala	Dichotome Variable	Ordinalskala
Intervallskala	Produkt-Moment-Korrelation	Punktbiseriale Korrelation	Gamma (γ) (Kendalls-Tau-b)
Dichotome Variable		Phi-Koeffizient*	Gamma (γ) (Kendalls-Tau-b)
Ordinalskala			Gamma (γ) (Kendalls-Tau-b)

Anmerkungen: Die Auswahl der Koeffizienten wurde in Anlehnung an Bortz/Schuster (2010), Benninghaus (2007) und De Vaus (2014) getroffen.

* Der Phi-Koeffizient wird nicht angewendet, da diese Kombination der Variablen in dieser Untersuchung nicht vorkommt. Er wird lediglich der Vollständigkeit halber aufgeführt.

Quelle: Eigene Darstellung in Anlehnung an Bortz/Schuster (2010: 171) und De Vaus (2014: 291)

Der Produkt-Moment-Korrelationskoeffizient wird angewendet, wenn beide Variablen intervallskaliert sind (auch Bravais-Pearson-Korrelation siehe Bortz/Schuster (2010: 156)). Für die Berechnung der Korrelation zwischen zwei ordinalen Variablen wird ein Rangkorrelationskoeffizient verwendet. Hierfür bietet sich die Rangkorrelation von Spearman an. Allerdings ist dieser problematisch, wenn viele doppelt besetzte Ränge (verbundene Rangplätze) vorliegen. Bortz/Schuster (2010: 179) erwähnen einen Anteil an verbundenen Rangplätzen von maximal 20 Prozent bei der Verwendung des Koeffizienten von Spearman. Dieser Wert wird im verwendeten Datensatz in den ordinalen Merkmalen deutlich überschritten. Aus diesem Grund wird für die Berechnung zwischen ordinalen Variablen das Zusammenhangsmass Gamma (γ) und Kendalls-Tau-b verwendet. Ebenso wird der Zusammenhang zwischen dichotomen und ordinalen Variablen mit diesen beiden Koeffizienten dargestellt (siehe De Vaus 2014: 291ff.). Der Zusammenhang zwischen einem dichotomen und einem intervallskalierten Merkmal wird mit einer punktbiserialen Korrelation, der Zusammenhang zwischen einem ordinalen und einem intervallskalierten Variable mit dem Korrelationskoeffizienten Gamma und Kendalls-Tau-b, dargestellt (Bortz/Schuster 2010; De Vaus 2014: 291f.).

Die Korrelationskoeffizienten sind zudem unterschiedlich zu interpretieren. Die nachfolgende Tabelle 47 zeigt die wichtigsten Charakteristiken der zur Anwendung kommenden Korrelationskoeffizienten auf.

Tabelle 47: Charakteristiken der verwendeten Korrelationskoeffizienten

Korrelationskoeffizient	Wertebereich	PRE-Interpretation	Richtung	Nur Linear?
Phi-Koeffizient	0-1	Nein	Nein	Nein
Gamma (γ)	0-1	Ja	Ja	Ja
Kendalls-Tau-b	0-1	Ja	Ja	Ja
Produkt-Moment-Korrelation	0-1	Ja, wenn quadriert (r^2)	Ja	Ja
Punktbiseriale Korrelation	0-1	Ja, wenn quadriert	Nein	Nein

Anmerkungen: Der Gamma Korrelationskoeffizient gibt höhere Werte im Vergleich zu anderen ordinalen Koeffizienten (Kendalls-Tau-b).
Quelle: Eigene Darstellung in Anlehnung an De Vaus (2014: 288f.)

Bei den Korrelationskoeffizienten gibt es zwei Typen: einerseits die Chi-Quadrat basierten und andererseits die PRE (proportional reduction of error) basierten Korrelationskoeffizienten. Die Zusammenhangsmasse, aufbauend auf Chi-Quadrat, untersuchen die Differenz zwischen beobachteter und erwarteter Häufigkeit. Die PRE-basierten Koeffizienten drücken die Prognostizierbarkeit aus. Beispielsweise kann bei einem Korrelationskoeffizienten von 0.7 der Wert der anderen Variable um 70% exakter vorhergesagt werden. Die bei dieser Untersuchung angewendeten Korrelationskoeffizienten sind diesem Korrelationstypen zuzuordnen.

Anhang IV Operationalisierung, bivariate Zusammenhänge und Einebenenmodelle

Tabelle 48: Operationalisierung und Deskription der Variablen für die Analyse der Evaluationspraxis

Variablen	Operationalisierung	Kategorien	%	N	Min	Max	Mittel	SA	Q
Evaluationsaktivität	Bitte geben Sie die ungefähre Anzahl an Evaluationen an, an deren Erarbeitung Sie während einem Jahr durchschnittlich beteiligt sind.		100.0	752	0	30	1.84	4.05	S
Evaluationsverwendung	Wie häufig ziehen Sie Evaluationen in Ihre Arbeit mit ein? Dummy-Variable: Wert 0 für „Nie“, Wert 1 für mindestens „Seltener“.	Nie Seltener Mehrere Male im Monat Einmal in der Woche Mehrere Male in der Woche Täglich Total	39.7 36.4 15.4 4.5 2.8 1.1 100.0	325 298 126 37 23 9 818					S
Evaluationsstelle	Vorhandensein einer Evaluationsstelle. Bivariate Analyse: Wert 2 für „Vollständige“, Wert 1 für „Ansatzweise“, Wert 0 für „Keine“. Multivariate Analyse: zwei Dummy-Variablen. - Vollständige Evaluationsstelle - Ansatzweise Evaluationsstelle - Keine Evaluationsstelle (Referenzkategorie)	Vollständige Ansatzweise Keine Total	12.1 7.0 80.8 100.0	107 62 713 882					R
Volatilität	Wie würden Sie den Bereich, in dem Sie arbeiten, bezüglich Veränderungen beurteilen? Likert-Skala: Werte 0 für „Überhaupt keine Veränderung“ bis 10 für „Sehr starke Veränderung“.		100.0	787	0	10	5.66	2.45	S
Komplexität	Wie würden Sie den Bereich, in dem Sie arbeiten, bezüglich Komplexität beurteilen? Likert-Skala: Werte 0 für „Überhaupt nicht komplex“ bis 10 für „Sehr stark komplex“.		100.0	779	0	10	6.92	2.14	S
Politikzyklus	Welcher Kategorie würden Sie Ihre Tätigkeit am ehesten zuschreiben? Zwei Dummy-Variablen: - Formulierung und Planung - Implementierung	Formulierung und Planung Implementierung Interne Dienstleistungen Total	14.1 40.9 45.0 100.0	110 318 350 778					S

	- Interne Dienstleistungen (Referenzkategorie)								
Politikmassnahmen	Sind die Politiken/Massnahmen an denen Sie mitarbeiten mehrheitlich Redistributiv, Regulativ oder Distributiv?	Mehrheitlich Redistributiv	4.7	36					S
	Zwei Dummy-Variablen:	Mehrheitlich Regulativ	31.3	242					
		Mehrheitlich Distributiv	18.5	143					
		- Regulativ	Keine Zuordnung möglich	45.6	353				
		- Redistributiv	Total	100.0	774				
	- Distributiv sowie keine Zuordnung möglich (Referenzkategorie)								
Verwaltungsgrösse	Gesamtzahl angefragter Mitarbeitenden der Direktionen und Departemente.		100.0	882	64	410	238.04	106.2	R
Geschlecht	Bitte geben Sie ihr Geschlecht an. Wert 1 für „Weiblich“, Wert 0 für „Männlich“.	Männlich	48.2	395					S
		Weiblich	51.8	425					
		Total	100.0	820					
Alter	Bitte geben Sie ihr Alter an.		100.0	859	15	66	44.4	11.5	S
Hierarchieposition	In welcher Funktion arbeiten Sie?	Lernende/Praktikant/-in	4.6	40					S
	Bivariate Analyse: Wert 1 für „Lernende“ bis Wert 4 für „Oberes Kader“.	Angestellte/-r	61.4	532					
		Kader	26.8	232					
		Oberes Kader	7.3	63					
		Multivariate Analyse: zwei Dummy-Variablen.	Total	100.0	867				
	- Oberes Kader								
	- Kader								
	- Angestellte und Lernende (Referenzkategorie)								
Bildung	Bitte geben Sie Ihre höchste abgeschlossene Bildung an.	Obligatorische Schule	1.8	16					S
	Bivariate Analyse: Wert 1 für „obligatorische Schule [...]“ bis Wert 5 für „Hochschule“.	Berufsbildung, Berufslehre	18.6	162					
		Fachmittelschulen, Maturitätsschulen	9.1	79					
		Höhere Berufsbildung	14.2	124					
		Multivariate Analyse: drei Dummy-Variablen.	Hochschulen	56.3	490				
		- Hochschulen	Total	100.0	896				
	- Höhere Berufsbildung								
	- Fachmittelschulen, Maturitätsschulen								
	- Obligatorische Schule, Berufsbildung und Berufslehre (Referenzkategorie)								

Anmerkungen: N=Anzahl Mitarbeitende. %=Anteile in den Kategorien in Prozent. Min=Minimaler Wert. Max=Maximaler Wert. Mittel=Mittelwert (auch arithmetisches Mittel). SA=Standardabweichung. Q=Quelle (S=Online Mitarbeitendenbefragung; R=eigene Recherchen).

Quelle: Eigene Darstellung

Tabelle 49: Operationalisierung und Deskription der Variablen für die Analyse der Einstellung zur Evaluation

Variablen	Operationalisierung	Kategorien	%	N	Min	Max	Mittel	SA	Q
Einstellung zur Evaluation	Skala die aus fünf Fragen zur Einstellung zur Evaluation gebildet wird.		100.0	369	0	50	30.6	9.25	S
Evaluationsstelle	Vorhandensein einer Evaluationsstelle.	Ja	76.6	58					R
	Bivariate Analyse: Wert 2 für „Vollständige“, Wert 1 für „Ansatzweise“, Wert 0 für „Keine“.	Ansatzweise	9.6	50					
		Keine	13.9	320					
	Multivariate Analyse: zwei Dummy-Variablen. - Vollständige Evaluationsstelle - Ansatzweise Evaluationsstelle - Keine Evaluationsstelle (Referenzkategorie)	Total	100.0	418					
Evaluationsaktivität	Anzahl an recherchierten Evaluationsberichten für den Zeitraum 2005 bis 2013 auf Stufe der Departemente und Direktionen.		100.0	418	10	51	29.13	15.13	R
Evaluationsverwendung	Prozentanteile von Mitarbeitenden mit mindestens seltener Verwendung von Evaluationen in die Arbeit auf Stufe der Ämter und Bereiche.		100.0	335	33.3	87.5	64.73	11.20	S
Politikzyklus	Welcher Kategorie würden Sie Ihre Tätigkeit am ehesten zuschreiben?	Formulierung und Planung	19.4	67					S
		Implementierung	46.4	160					
	Zwei Dummy-Variablen: - Formulierung und Planung - Implementierung - Interne Dienstleistungen (Referenzkategorie)	Interne Dienstleistungen	34.2	118					
		Total	100.0	345					
Traditionalismus-Modernismus	In der Politik wird manchmal von „Traditionalismus“ und von „Modernismus“ gesprochen. Wo würden Sie sich selber auf dieser Skala einordnen? Die Werte der Einstufung in der Likert-Skala wurden übernommen. Wert 0 für „Traditionalistisch“ bis Wert 10 für „Modernistisch“.		100.0	315	1	10	7.22	1.65	S
Links-Rechts	In der Politik wird manchmal von „Links“ und „Rechts“ gesprochen. Wo würden Sie sich selbst auf dieser Skala einordnen? Likert-Skala wurden übernommen. Wert 0 für „Links“ bis Wert 10 für „Rechts“.		100.0	329	0	10	3.66	1.97	S

Hierarchieposition	In welcher Funktion arbeiten Sie?	Lernende/Praktikant/-in	3.2	13					S
	Bivariate Analyse: Wert 1 für „Lernende“ bis Wert 4 für „Oberes Kader“.	Angestellte/-r	50.6	205					
		Kader	35.8	145					
	Multivariate Analyse: zwei Dummy-Variablen:	Oberes Kader	10.4	42					
		- Oberes Kader	Total	100.0	408				
- Kader									
Studienrichtung	Welches war der Schwerpunkt Ihrer Studienfachrichtung? Eine Dummy-Variable: - Technische, exakte- und Naturwissenschaften - Kein absolviertes Studium und die anderen Studienfachrichtungen (Referenzkategorie)	Geistes- und Sozialwissenschaften	25.6	104					S
		Wirtschaftswissenschaften	11.3	46					
		Recht	3.9	16					
		Exakte- und Naturwissenschaften	8.9	36					
		Medizin und Pharmazie	3.7	15					
		Technische Wissenschaften	15.5	63					
		Interdisziplinäre und andere Studienrichtungen	9.1	37					
		Kein absolviertes Studium	22.1	90					
	Total	100.0	407						
Alter	Bitte geben Sie ihr Alter an.		100.0	405	15	65	45.7	11.0	S
Geschlecht	Bitte geben Sie ihr Geschlecht an. Wert 1 für „Weiblich“, Wert 0 für „Männlich“.	Männlich	55.5	212					S
		Weiblich	44.5	170					
		Total	100.0	382					
Kanton	Zwei Dummy-Variablen: - Basel-Stadt - Luzern - Bern (Referenzkategorie)	Basel-Stadt	18.9	79					S
		Bern	58.4	244					
		Luzern	22.7	95					
		Total	100.0	418					

Anmerkungen: N=Anzahl Mitarbeitende. %=Anteile in den Kategorien in Prozent. Min=Minimaler Wert. Max=Maximaler Wert. Mittel=Mittelwert (auch arithmetisches Mittel). SA=Standardabweichung. Q=Quelle (S=Online Mitarbeitendenbefragung; R=eigene Recherchen). Für die Untersuchung der Einstellung zur Evaluation werden lediglich die Angaben der Evaluationsaktiven verwendet. Aus diesem Grund beruht diese Analyse auf einer kleineren Anzahl an Mitarbeitenden im Vergleich zu der Untersuchung der Evaluationspraxis.

Quelle: Eigene Darstellung

Tabelle 50: Bivariate Zusammenhänge der unabhängigen Variablen für die Untersuchung der Evaluationspraxis

	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
1. Alter (I)	-0.19**	0.41**	-0.09**	0.07*	0.09**	0.07+	-0.04	0.01	-0.05	0.08	0.02	-0.02	0.13	0.01
		0.30**	-0.07**							0.04				
2. Geschlecht (N)		-0.46**	-0.18**	-0.12**	-0.26**	-0.06+	-0.08*	-0.14**	0.07*	0.32**	-0.17**	0.12**	-0.09*	-0.00
		-0.25**	-0.10**							0.13**				
3. Hierarchieposition (O)			0.55**	0.25**	0.30**	0.18**	0.34**	0.08	-0.09	0.16*	0.03	-0.04	0.05	-0.03
			0.32**	0.17**	0.20**	0.09**	0.14**	(0.04)	-0.02	0.07*	0.02	-0.01	0.02	-0.01
4. Bildung (O)				0.12**	0.37**	0.20**	0.70**	0.10	0.01	0.20**	0.03	-0.02	0.06	-0.07
				0.09**	0.27**	0.11**	0.23**	0.05	0.00	0.09**	0.05	-0.01	0.03	-0.03
5. Volatilität (I)					0.44**	0.11**	0.04	0.04	-0.02	0.01	-0.11	0.03	-0.08*	0.07*
										0.00				
6. Komplexität (I)						0.13**	0.16**	0.12**	-0.06	0.00	-0.02	0.02	-0.04	0.04
										0.00				
7. Implementierung (N)								0.23**	0.07+	0.22**	-0.05	-0.06+	-0.05	0.11**
										0.11**				
8. Formulierung (N)								0.03	-0.02	0.04	-0.02	0.07*	-0.02	-0.04
										0.01				
9. Regulativ (N)									-0.15**	-0.53**	0.03	-0.13**	0.03	0.08*
										-0.18**				
10. Redistributiv (N)										0.13	-0.09**	0.09**	-0.10**	0.03
11. Evaluationsstelle (O)											-0.16**	0.43**	-0.34**	-0.02
											-0.10**	0.20**	-0.15**	-0.01
12. Verwaltungsgrösse (I)												-0.37**	0.73**	-0.53**
13. Basel-Stadt (N)													-0.59**	-0.24**
14. Bern (N)														-0.67**
15. Luzern (N)														

Anmerkungen: Die Skalenniveaus sind hinter den Variablen in Klammern aufgeführt: (N) nominales Skalenniveau; (O) ordinales Skalenniveau; (I) intervallskaliertes Skalenniveau. Die Korrelationskoeffizienten ergeben sich entsprechend der in der Tabelle 46 dargestellten Kombinationen der beiden untersuchten Merkmale. Die Korrelation von Variablen mit ordinalem Skalenniveau sind mit zwei Korrelationskoeffizienten beschrieben: Gamma und Kendalls-Tau-b (zweite Zeile).

* = $p < 0,05$; ** = $p < 0,01$; 2-seitig.

Quelle: Eigene Darstellung

Tabelle 51: Bivariate Zusammenhänge der unabhängigen Variablen für die Untersuchung der Einstellung zur Evaluation

	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. Alter (I)	-0.24**	0.38**	0.08	-0.04	0.09	0.02	-0.05	0.01	0.04	0.13*	0.03	0.00	0.07
		0.30**									0.02		
2. Geschlecht (N)		-0.50**	0.02	-0.06	-0.01	-0.27	0.17**	-0.12*	-0.03	-0.29**	0.42**	-0.14	0.05
		-0.28**									0.21**		
3. Hierarchieposition (O)			0.09	0.29**	0.19**	0.05	-0.08	0.03	0.03	-0.67	0.10	0.04	0.00
			0.05	0.14**	0.13**	0.04	-0.04	0.02	0.02	-0.29**	0.05	0.03	0.00
4. Implementierung (N)				-0.46**	-0.13*	0.09	-0.03	-0.05	0.08	0.07	-0.01	-0.01	0.08
											0.01		
5. Formulierung (N)					0.20**	-0.12*	0.05	-0.01	-0.03	0.02	0.03	0.01	0.12*
											-0.01		
6. Traditionalismus-Modernismus (I)						-0.22**	-0.01	-0.00	0.01	-0.02	0.23**	0.01	0.02
											0.13**		
7. Links-Rechts (I)							-0.14*	0.06	0.05	0.01	-0.35**	0.00	-0.20**
											-0.22**		
8. Basel-Stadt (N)								-0.57**	-0.26**	-0.24**	0.38**	-0.47**	0.20**
											0.21**		
9. Bern (N)									-0.64**	0.21**	-0.36	0.79**	-0.30**
											-0.18**		
10. Luzern (N)										-0.02	0.04	-0.50**	0.16**
											0.01		
11. Technische-, Exakte- und Naturwissenschaften (N)											-0.79**	0.27**	-0.11
											-0.26**		
12. Evaluationsstelle (O)												-0.14*	0.03
												-0.09*	0.02
13. Evaluationsaktivität (I)													-0.07
14. Evaluationsverwendung (I)													

Anmerkungen: Die Skalenniveaus sind hinter den Variablen in Klammern aufgeführt: (N) nominales Skalenniveau; (O) ordinales Skalenniveau; (I) intervallskaliertes Skalenniveau. Die Korrelationskoeffizienten ergeben sich entsprechend der in der Tabelle 46 dargestellten Kombinationen der beiden untersuchten Merkmale. Die Korrelation von Variablen mit ordinalem Skalenniveau sind mit zwei Korrelationskoeffizienten beschrieben: Gamma und Kendalls-Tau-b (zweite Zeile). Für die Untersuchung der Einstellung zur Evaluation werden lediglich die Angaben der Evaluationsaktiven verwendet. Aus diesem Grund basiert diese Untersuchung auf einer kleineren Anzahl an Mitarbeitenden und darin liegt auch der Grund für die Unterschiede der Korrelationskoeffizienten im Vergleich zu der vorigen Tabelle.

* = $p < 0,05$; ** = $p < 0,01$; 2-seitig.

Quelle: Eigene Darstellung

Tabelle 52: Negativ binomiale Regressionsanalyse der Evaluationsaktivität (Einebenenmodell)

	Modell A1		Modell A2		Modell A3	
	Koeff.	SE	Koeff.	SE	Koeff.	SE
Konstante	0.04	0.43	-0.34	0.53	-0.87	0.65
Institutionalisierung der Evaluation H1						
Evaluationsstelle (Ref.: Keine Evaluationsstelle)						
Vollständige Evaluationsstelle					0.51+	0.27
Ansätze einer Evaluationsstelle					0.89*	0.42
Arbeitsbereichsspezifische Variablen H2						
Volatilität			-0.01	0.04	-0.01	0.04
Komplexität			0.12**	0.05	0.12*	0.05
Politikzyklus (Ref.: Interne Dienstleistungen)						
Implementierung			0.04	0.21	0.01	0.21
Formulierung			0.14	0.25	0.16	0.24
Politikmassnahmen (Ref.: Distributiv)						
Regulativ			-0.08	0.19	0.00	0.19
Redistributiv			-0.02	0.44	0.08	0.44
Organisationspezifische Variablen H3						
Verwaltungsgrösse					0.00	0.00
Kontrollvariablen K						
Alter	0.01	0.01	0.00	0.01	0.00	0.01
Weiblich	-0.41**	0.18	-0.36+	0.19	-0.40*	0.19
Hierarchieposition (Ref.: Angestellte und Lernende)						
Oberes Kader	0.14	0.38	0.10	0.37	0.09	0.37
Kader	0.55**	0.21	0.56**	0.21	0.64**	0.21
Bildung (Ref.: Obligatorische Schule, Berufsbildung und Berufslehre)						
Hochschulabschluss	-0.09	0.23	0.39	0.25	-0.47+	0.25
Höhere Berufsbildung	0.55+	0.29	0.24	0.30	0.23	0.29
Allgemeinbildung	-0.23	0.36	-0.23	0.38	-0.31	0.37
Kanton (Ref.: Bern)						
Basel-Stadt	0.29	0.23	0.24	0.24	0.09	0.36
Luzern	0.58**	0.21	0.46**	0.22	0.68*	0.29
N		689		628		626
AIC		2178		2034		2028
BIC		2228		2109		2117
Pseudo R ²		0.017		0.020		0.025
Log Likelihood		-1080.95		-1000.18		-994.30
Likelihood-Ratio-Test				6.53 (zu Modell 1)		9.69*(zu Modell 2)

Anmerkungen: Koeff.= nicht standardisierter Regressionskoeffizient. SE=Standard Fehler. N=Anzahl Mitarbeitende. Ref.=Referenzkategorie zu den Dummyvariablen. AIC=Akaike-Informationskriterium. BIC= Bayessches-Informationskriterium. Pseudo R²=Pseudo Bestimmtheitsmass.

+: p<0.1; *: p<0.05; **: p<0.01; ***: p<0.001; 2-seitig.

Quelle: Eigene Darstellung

Tabelle 53: Logistische Regressionsanalyse der Evaluationsverwendung (Einebenenmodell)

	Modell V1			Modell V2			Modell V3		
	Koeff.	SE	OR	Koeff.	SE	OR	Koeff.	SE	OR
Konstante	-0.57	0.01	0.56	-1.45**	0.51	0.23	-1.78**	0.65	0.17
Institutionalisierung der Evaluation H1									
Evaluationsstelle (Ref.: Keine Evaluationsstelle)									
Vollständige Evaluationsstelle							-0.41	0.27	0.67
Ansätze einer Evaluationsstelle							1.02*	0.43	2.78
Arbeitsbereichsspezifische Variablen H2									
Volatilität				0.04	0.04	1.04	0.04	0.04	1.04
Komplexität				0.10*	0.05	1.12	0.11*	0.05	1.12
Politikzyklus (Ref.: Interne Dienstleistung)									
Implementierung				0.53**	0.19	1.70	0.46*	0.05	1.59
Formulierung				0.96**	0.29	2.60	0.92**	0.20	2.52
Politikmassnahmen (Ref.: Distributiv)									
Regulativ				-0.11	0.19	0.89	-0.10	0.20	0.91
Redistributiv				0.55	0.44	1.73	0.52	0.45	1.68
Organisationspezifische Variablen H3									
Verwaltungsgrösse				0.00	0.00	1.00	0.00	0.00	1.00
Kontrollvariablen K									
Alter	0.00	0.01	1.00	0.00	0.01	1.00	0.00	0.01	1.00
Weiblich	-0.04	0.17	0.96	0.07	0.19	1.07	0.15	0.19	1.16
Hierarchieposition (Ref.: Angestellte und Lernende)									
Oberes Kader	0.88*	0.36	2.42	0.68+	0.37	1.98	0.80*	0.38	2.22
Kader	0.82***	0.20	2.27	0.83***	0.22	2.29	0.85***	0.22	2.33
<i>Bildung (Ref.: Obligatorische Schule, Berufsbildung und Berufslehre)</i>									
Hochschulabschluss	0.81***	0.21	2.24	0.31+	0.24	1.36	0.39	0.25	1.48
Höhere Berufsbildung	0.58*	0.27	1.78	0.34	0.30	1.41	0.39	0.30	1.48
Allgemeinbildung	0.43	0.32	1.53	0.43	0.34	1.54	0.44	0.34	1.56
Kanton (Ref.: Bern)									
Basel-Stadt	0.30	0.22	1.35	0.24	0.29	1.27	-0.28	0.34	0.75
Luzern	0.16	0.20	1.18	0.14	0.29	1.15	0.18	0.29	1.19
N			746			678			676
Log Likelihood			-470.89			-415.64			-406.17
Likelihood-Ratio-Test				20.70***(zu Modell 1)			9.61*(zu Modell 2)		

Anmerkungen: Koeff.=Regressionskoeffizient. OR=Odds Ratios. SE=Standardfehler. Die Signifikanzen wurden mit dem Wald-Test geprüft. Ref.=Referenzkategorie zu den Dummyvariablen.

+: p<0.1; *: p<0.05; **: p<0.01; ***: p<0.001; 2-seitig.

Quelle: Eigene Darstellung

Tabelle 54: Lineare Regressionsanalyse der Einstellung zur Evaluation (Einebenenmodell)

	Modell E1		Modell E2		Modell E3	
	Koeff.	SE	Koeff.	SE	Koeff.	SE
Konstante	35.42***	2.42	25.33***	4.16	20.37***	5.44
Institutionalisierung der Evaluation H4						
Evaluationsstelle (Ref.: Keine Evaluationsstelle)						
Vollständige Evaluationsstelle					-1.64	1.62
Ansätze einer Evaluationsstelle					-0.95	2.59
Evaluationspraxis H5						
Evaluationsaktivität					0.15*	0.06
Evaluationsverwendung					-0.01	0.05
Arbeitsbereichsspezifische Variablen H6						
Politikzyklus (Ref.: Interne Dienstleistung)						
Implementierung			-0.55	1.18	-0.70	1.21
Formulierung			1.53	1.47	1.70	1.53
Politische Variablen H7						
Traditionalismus-Modernismus			1.23***	0.32	1.18***	0.32
Links-Rechts			-0.35	0.27	-0.35	0.30
Subkulturelle Variablen H8						
Hierarchieposition (Ref.: Angestellte und Lernende)						
Oberes Kader			-3.87+	1.97	-3.00	2.12
Kader			-1.79	1.14	-1.17	1.18
Studienrichtung (Ref.: Kein absolviertes Studium und andere Studienrichtungen)						
Technische-, Exakte- und Naturwissenschaften			0.31	1.24	-0.61	1.31
Kontrollvariablen K						
Alter	-0.04	0.05	0.04	0.05	0.04	0.05
Weiblich	-0.38	1.02	-0.79	1.23	-0.40	1.30
Kanton (Ref.: Bern)						
Basel-Stadt	0.33	1.25	0.83	1.40	5.61*	2.63
Luzern	-0.56	1.21	-0.00	1.26	3.75+	2.02
N		321		259		247
R2		0.004		0.100		0.123
Angepasstes-R2		-0.008		0.060		0.069
Log Likelihood		-1146.6		-904.74		-858.10
Likelihood-Ratio-Test			26.74*** (zu Modell1)		7.47 (zu Modell 2)	

Anmerkungen: Koeff.=nicht standardisierter Regressionskoeffizient. SE=Standard Fehler. N=Anzahl Mitarbeitende. Ref.=Referenzkategorie zu den Dummyvariablen. R=Bestimmtheitsmass. Angepasstes R'=Angepasstes Bestimmtheitsmass.

+: p<0.1; *: p<0.05; **: p<0.01; ***: p<0.001; 2-seitig.

Quelle: Eigene Darstellung

Anhang V Online-Fragebogen

Anmerkung: Der Fragebogen wurde in Deutsch und auch in Französisch verfasst. Im Folgenden ist der deutschsprachige Fragebogen aufgeführt.

Frage 1

Bitte geben Sie Ihr Geschlecht an.

Männlich

Weiblich

Frage 2

Bitte geben Sie ihr Alter an.

Frage 3

Bitte geben Sie Ihre höchste abgeschlossene Bildung an.

Obligatorische Schule nicht abgeschlossen, Obligatorische Schule abgeschlossen, 10. Schuljahr, allgemeine Berufsvorbereitung, Handelsschuljahr, Haushaltslehrjahr, Sprachaufenthalt

Berufsbildung, Berufslehre, Vollzeitberufsschule

Allgemeinbildung, Fachmittelschule, Maturitätsschule, Berufsmaturität, Lehrerseminar

Höhere Berufsbildung (Eidg. Diplom, Eidg. Fachausweis, Techniker- oder Fachschule, HTL, HWV)

Hochschulen (Universität, Fachhochschule, Pädagogische Hochschule, ETH)

Frage 4

Folgefrage von Frage 3

Welches war der Schwerpunkt Ihrer Studienfachrichtung?

Geistes- und Sozialwissenschaften	<input type="checkbox"/>
Wirtschaftswissenschaften	<input type="checkbox"/>
Recht	<input type="checkbox"/>
Exakte- und Naturwissenschaften	<input type="checkbox"/>
Medizin und Pharmazie	<input type="checkbox"/>
Technische Wissenschaften	<input type="checkbox"/>
Interdisziplinäre und andere Studienrichtungen	<input type="checkbox"/>

Frage 5

In welcher Funktion arbeiten Sie?

Lernende/Praktikant/-in	<input type="checkbox"/>
Hilfskraft	<input type="checkbox"/>
Angestellte/-r	<input type="checkbox"/>
Kader	<input type="checkbox"/>
Oberes Kader	<input type="checkbox"/>
Andere:	<input type="checkbox"/>

Frage 6

Wie ist Ihre grundsätzliche Haltung gegenüber Evaluationen?

<input type="checkbox"/> 00	<input type="checkbox"/> 01	<input type="checkbox"/> 02	<input type="checkbox"/> 03	<input type="checkbox"/> 04	<input type="checkbox"/> 05	<input type="checkbox"/> 06	<input type="checkbox"/> 07	<input type="checkbox"/> 08	<input type="checkbox"/> 09	<input type="checkbox"/> 10
Sehr negativ					Sehr positiv					

Frage 7

Wie stark stimmen Sie der folgenden Aussage zu?

Die Evaluationsaktivität der Verwaltung sollte ausgeweitet werden.

<input type="checkbox"/> 00	<input type="checkbox"/> 01	<input type="checkbox"/> 02	<input type="checkbox"/> 03	<input type="checkbox"/> 04	<input type="checkbox"/> 05	<input type="checkbox"/> 06	<input type="checkbox"/> 07	<input type="checkbox"/> 08	<input type="checkbox"/> 09	<input type="checkbox"/> 10
Stimme	überhaupt							Stimme stark		
nicht zu								zu		

Frage 8

Wie stark stimmen Sie der folgenden Aussage zu?

Die Ressourcen (beispielsweise Geld und Zeit) die in Evaluationen investiert werden, sind es wert.

<input type="checkbox"/> 00	<input type="checkbox"/> 01	<input type="checkbox"/> 02	<input type="checkbox"/> 03	<input type="checkbox"/> 04	<input type="checkbox"/> 05	<input type="checkbox"/> 06	<input type="checkbox"/> 07	<input type="checkbox"/> 08	<input type="checkbox"/> 09	<input type="checkbox"/> 10
Stimme	überhaupt							Stimme stark		
nicht zu								zu		

Frage 9

Als wie nützlich bewerten Sie Evaluationen (oder wären nützlich) in der Verwaltung um bessere Produkte, Programme, Prozesse oder Dienstleistungen zu entwickeln?

<input type="checkbox"/> 00	<input type="checkbox"/> 01	<input type="checkbox"/> 02	<input type="checkbox"/> 03	<input type="checkbox"/> 04	<input type="checkbox"/> 05	<input type="checkbox"/> 06	<input type="checkbox"/> 07	<input type="checkbox"/> 08	<input type="checkbox"/> 09	<input type="checkbox"/> 10
Überhaupt nicht nützlich								Sehr nützlich		

Frage 10

Als wie sinnvoll bewerten Sie (oder würden Sie bewerten) den Einbezug von Evaluationen in Ihre Arbeit?

00 01 02 03 04 05 06 07 08 09 10

Überhaupt nicht sinn-
voll

Sehr sinnvoll

Frage 11

Sind Sie bei Ihrer Arbeit an der Erarbeitung von Evaluationen beteiligt?

Ja

Nein

Frage 12

Folgefrage von Frage 11

Bitte geben Sie die ungefähre Anzahl an Evaluationen an, an deren Erarbeitung Sie während einem Jahr durchschnittlich beteiligt sind:

Frage 13

Ziehen Sie in Ihrer Arbeit Evaluationen mit ein? Beispielsweise um Entscheide zu fällen oder Informationen zu erhalten.

Ja

Nein

Frage 14

Folgefrage von Frage 13

Wie häufig ziehen Sie Evaluationen in Ihre Arbeit mit ein?

Beispielsweise um Entscheide zu fällen oder wichtige Informationen zu erhalten.

Mehrmals in der Woche	<input type="checkbox"/>
Einmal in der Woche	<input type="checkbox"/>
Mehrmals im Monat	<input type="checkbox"/>
Seltener	<input type="checkbox"/>
Nie	<input type="checkbox"/>

Frage 15

Gibt es in Ihrer Verwaltungseinheit eine eigene Evaluationsstelle?

Ja	<input type="checkbox"/>
Nein	<input type="checkbox"/>
Weiss nicht	<input type="checkbox"/>

Frage 16

Bitte geben Sie die Verwaltungseinheit an, in der Sie arbeiten:

Frage 17

Sind die Politiken/Massnahmen an denen Sie mitarbeiten mehrheitlich redistributiv, regulativ oder distributiv?

Mehrheitlich redistributiv. Gemeint sind Umverteilungsmassnahmen.

Mehrheitlich regulativ. Gemeint sind Regulierungen.

Mehrheitlich distributiv. Gemeint sind Verteilungsmassnahmen.

Meine Aufgaben/Massnahmen, an denen ich mitarbeite, können keiner dieser Kategorien zugeordnet werden.

Frage 18

Welcher Kategorie würden Sie Ihre Tätigkeit am ehesten zuschreiben?

Formulierung und Planung von öffentlichen Politiken/Massnahmen.

Implementierung (Umsetzung) von öffentlichen Politiken/Massnahmen.

Erbringung von internen Dienstleistungen.

Frage 19

Wie würden Sie den Bereich, in dem Sie arbeiten, bezüglich Veränderungen beurteilen? Gemeint ist damit für diese Untersuchung wie stark der Tätigkeitsbereich, in dem Sie arbeiten von Veränderungen betroffen ist. Ein stark veränderliches Tätigkeitsfeld ist eines, welches häufig von Reformen (Gesetzes- und Verordnungsänderungen) betroffen ist

00 01 02 03 04 05 06 07 08 09 10

Überhaupt keine Veränderungen

Sehr starke Veränderungen

Frage 20

Wie würden Sie den Bereich, in dem Sie arbeiten, bezüglich Komplexität beurteilen?

Ein komplexer Tätigkeitsbereich zeichnet sich dadurch aus, dass die Massnahmen und/oder die Zusammenhänge unklar sind, viel Technik- und/oder Fachwissen verlangt (beispielsweise ökonomisches Know-how). Ebenso der Einbezug von empirischen Daten und Technisierung kennzeichnen die Komplexität.

<input type="checkbox"/> 00	<input type="checkbox"/> 01	<input type="checkbox"/> 02	<input type="checkbox"/> 03	<input type="checkbox"/> 04	<input type="checkbox"/> 05	<input type="checkbox"/> 06	<input type="checkbox"/> 07	<input type="checkbox"/> 08	<input type="checkbox"/> 09	<input type="checkbox"/> 10
Überhaupt		nicht								Sehr starke
komplex										komplex

Frage 21

In der Politik wird manchmal von „Traditionalismus“ und von „Modernismus“ gesprochen.

Wo würden Sie sich selber auf dieser Skala einordnen?

<input type="checkbox"/> 00	<input type="checkbox"/> 01	<input type="checkbox"/> 02	<input type="checkbox"/> 03	<input type="checkbox"/> 04	<input type="checkbox"/> 05	<input type="checkbox"/> 06	<input type="checkbox"/> 07	<input type="checkbox"/> 08	<input type="checkbox"/> 09	<input type="checkbox"/> 10	<input type="checkbox"/>
Traditionalistisch									Modernistisch		Diese Frage
											möchte ich
											nicht be-
											antworten

Frage 22

In der Politik wird manchmal von „Links“ und „Rechts“ gesprochen. Wo würden Sie sich selber auf dieser Skala einordnen?

<input type="checkbox"/> 00	<input type="checkbox"/> 01	<input type="checkbox"/> 02	<input type="checkbox"/> 03	<input type="checkbox"/> 04	<input type="checkbox"/> 05	<input type="checkbox"/> 06	<input type="checkbox"/> 07	<input type="checkbox"/> 08	<input type="checkbox"/> 09	<input type="checkbox"/> 10	<input type="checkbox"/>	
Links											Rechts	Diese Frage möchte ich nicht beantworten
