A sequential explanatory mixed methods investigation of e-government outcomes evaluation practice in developing countries: South Africa as context

By

Charles Boamah-Abu

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

In the Department of Information Systems

University of Cape Town

Supervisor: Prof Michael Kyobe

May, 2021

Plagiarism declaration

I, the undersigned, do hereby declare that:

- This thesis is my handiwork
- Any contribution to this research by (or work of) others has been acknowledged and cited accordingly.

C Boamah-Abu

Acknowledgement

My life is a journey; there is a story— of people, things, and events— to be told. This thesis is dedicated to all those who contributed to my "becoming". So this is to:

- My siblings: Ante, for being the "deputy" mother that you are; Bro Kofi who instilled the culture of reading and shaped my life; Akos and Maame Yaa.
- Sharon, Bertha, Obaa Yaa, and the "Abu Asuafo", home and abroad. There have been days I felt like giving up but could not: you're the reason. Let this be your inspiration.
- Childhood friends: we all made it; mine was last, and of course, *the most enjoyable*
- The memories of my parents, Paapa and Awo, who taught me from childhood, that hard work pays off; and siblings—Felicia, Kate, and Aboo, who are not here to see the dream manifest and bask in the glory of their handiwork; Maame, who serenaded my childhood with stories, and waited every night to take a book away from my embrace; Nana Nyamekye: for a conversation with my father on that fateful dawn, which set my destiny; maternal grandfathers, who showed us that contentment comes from, not how much you had, but how much you shared (to them, the extended family was also their family); Wofa, whose library became my treasure and sanctuary; Wofa Ampofo, for the pearls of wisdom; teachers at Kotokuom Methodist Primary and L. A. Middle Schools who ignited the flame of academic excellence; Bro Abora, for the tutelage after school and during holidays; Abass, my dearest cousin and friend; and Abrantie— a dog is, indeed, man's best friend.

Also, I am sincerely grateful to:

- SB for being a lifetime companion and Tano: a trusted friend.
- Peter Nuako, for your interest in my personal development.
- Prof T Schatzki, for your advice in the early stages, when all was blurry, the resources you offered, and the compliment "you write so well"— from you whose writing I admire.
- Prof AL Browne, for the resources which opened my eyes to quantitative and mixed methods in practice research
- Finally, to my supervisor, Prof Michael Kyobe for your patience and support, especially, when I fell ill on commencement of this course and almost gave up.

This is part of the story to be told.

“Dwuma pa biara a yedi,
efi Nyame no mmoam;
Enti, yɛbɛto Ebenezer aky’rɛ se
nea Nyame aboa y’abedu nen.
Ebenezer nea Nyame aboa y’abedu nen
Ebenezer yiw Nyame N’adom ara kwa”
In Presbyterian Church of Ghana Hymn Book (p. xx)

Abstract

E-government has become *de rigueur* in public service delivery. Attendant to the growing adoption is the need to establish the extent to which investments are yielding desired outcomes and how the services can be improved. E-government evaluation is, therefore, an important research problem— more so in developing countries, where resources are scarce, and the opportunity cost is higher. This, notwithstanding, there is a paucity of research.

This research explored the "lay of the land", to describe e-government outcomes evaluation practice in developing countries. South Africa was selected as context; the sample consisted of public sector employees. Practice theory was adopted to frame e-government evaluation as what people do. A two-phased sequential explanatory mixed methods strategy, supported by post-positivism, was employed. This allowed examination of the different aspects of practice— the tangible (e.g., activities) and the intangibles (e.g., affective issues).

In the quantitative phase, questionnaire data from 106 public employees (i.e., managers and non-managers from IT, Evaluation, and Other backgrounds) was analysed with factor and cluster analyses techniques to identify the important elements which described e-government outcomes evaluation practice and the patterns of performance. This was followed by a qualitative phase in which interview transcripts of 12 participants drawn from the sample of the preceding phase, four from each background (i.e., two from management and non-management respectively), were analysed with content analysis to help understand the results of the quantitative phase.

The findings show an inadequate organizational capacity to evaluate e-government outcomes and use the results. The primary purpose of evaluation is compliance, and organizational outcomes are mostly measured. While there may be some degree of use (e.g., learning, i.e., conceptual) recommendations may not be implemented. Furthermore, there were significant differences among participants; evaluation and IT managers with the requisite expertise were likely to have a positive perception of e-government outcomes evaluation practice.

This research adds value to e-government outcomes evaluation research and practice in developing countries, and also to the methodological development of practice theory in Information Systems (IS). Consistent with research on practice, recommendations are drawn to help e-government evaluation stakeholders improve future practice and, thereby, ameliorate the high e-government failure rate. The findings shed light on current practice, e.g., what is

done well and challenges, and add to the otherwise sparse body of knowledge in the domain under investigation. A deterrent to the adoption of practice theory is the paucity of theoretical and conceptual frameworks. This research illustrates how Schatzki's theory can be applied to investigate an IS problem and helps remedy the dearth of empirical research on practice. Furthermore, it contributes to the current effort to develop mixed methods in IS— and practice research in general. The agenda for future research can help advance the, hitherto, under-researched domain of e-government evaluation in developing countries.

Table of content

Plagiarism declaration i
Acknowledgement ii
Abstract iv
Table of content vi
List of tables xi
List of figures xiv
Chapter 1 : Introduction 1
1.1 Problem statement 2
1.2 Theoretical and philosophical perspective 3
1.3 Research strategy 4
1.4 Research context and sample 6
1.5 Summary of results 6
1.6 Delimitation 6
1.7 Relevance and contribution 7
1.8 Outline of thesis 8
Chapter 2 : Review of literature 9
2.1 Literature review strategy 9
2.2 E-government and public sector performance 10
2.2.1 E-government 10
2.2.2 The public sector 11
2.2.3 Public sector reforms: an overview 11
2.2.4 Public value: voices in the field 13
2.2.5 Public service values 15
2.2.6 Public value and evaluation 16
2.2.7 Public value research: current issues 16
2.2.8 IT in the public sector 18
2.3 E-government in developing countries 19
2.3.1 Development assistance and public sector reforms 19
2.3.2 E-government and development 20
2.4 Evaluation practice 20
2.4.1 Practice 21
2.4.2 Evaluation 23
2.4.3 Evaluation as social practice 29
2.4.4 Evaluation quality: factors of influence 30

2.4.5 Research on evaluation practice 31
2.4.6 Evaluation practice in developing countries 32
2.5 IT outcomes evaluation practice 34
2.5.1 Types of IT evaluation 35
2.5.2 IT outcomes evaluation: challenges 35
2.5.3 Research on IT evaluation practice 36
2.5.4 Context 37
2.5.5 What is measured 37
2.5.6 Resources 37
2.5.7 Activities 38
2.5.8 Outcomes 38
2.5.9 Theory 39
2.6 E-government outcomes evaluation practice 41
2.6.1 E-government outcomes evaluation: challenges 41
2.6.2 Context 42
2.6.3 What is measured 43
2.6.4 Resources 43
2.6.5 Activities 44
2.6.6 Outcomes 45
2.6.7 Theory 45
2.6.8 E-government evaluation practice in developing countries 46
2.7 South Africa as context 48
2.7.1. Public sector reform 48
2.7.2 Evaluation practice 49
2.7.3 E-government and evaluation practice 51
Summary 52
Chapter 3 : Theoretical and conceptual frameworks 55
3.1 The role of theory in this research 55
3.2 Practice theory 57
3.2.1 Why practice theory? 57
3.2.2 What is practice theory? 57
3.2.3 The methodological debate 61
3.2.4 Schatzki's site ontology 62
3.2.5 Implications for this research 65
3.3 Practice theory and IS practice-based research 66
3.4 Conceptual model 67
Summary 71
Chapter 4 : Research design 72
4.1 Elements of research design 73

4.2 Types of research by purpose 74

4.3 Philosophical assumptions in research 75

4.4 Research paradigms 78

4.4.1 Positivism 79
4.4.2 Constructivism 79
4.4.3 Post-positivism 79
4.4.4 Implications for design 80

4.5 Research strategies 80

4.5.1 Case study 81
4.5.2 Survey 81
4.5.3 Implications for design 81

4.6 Research methods 82

4.6.1 Logic of inquiry 82
4.6.2 Sampling 83
4.6.3 Data collection 86
4.6.4 Data analysis 87
4.6.5 Validation 88
4.6.6 Quantitative, qualitative and mixed methods research 89
4.6.7 Features of mixed methods research 91

4.7 Time horizon 94

4.8 Adopted strategy: sequential explanatory mixed methods (SEMM) survey 95

4.8.1 Phase 1: Quantitative research 96
4.8.2 Phase 2: Qualitative research 97
4.8.3 Organization and presentation of research 97

4.9 Ethics 97

Summary 99

Chapter 5 : Phase 1 (Quantitative data collection and analysis) 100

5.1 Sampling 101

5.2 Data collection 102

5.2.1 Questionnaire construction 102
5.2.2 Questionnaire administration 108
5.2.3 Response rate 109

5.3 Data preparation 110

5.4 Analytical methods 110

5.4.1 Descriptive analysis 111
5.4.2 Correlation 111

5.4.3 Comparison 111
5.4.4 Exploratory factor analysis (EFA) 112
5.4.5 Cluster analysis 116

5.5 Data analysis and preliminary results 119

5.5.1 Descriptive analysis 119
5.5.2 Exploratory factor analysis 126
5.5.3 Further analysis of elements: comparing by group 136
5.5.4 Cluster analysis 139

Summary 150

Chapter 6 : Phase 2 (Qualitative data collection and analysis) 151

6.1 Ethics and quality issues 152

6.2 Sampling 154

6.3 Data collection 155

6.3.1 Interview protocol 155
6.3.2 Interview administration 157

6.4 Analytic technique 158

6.4.1 Content analysis 161
6.4.2 Analysis strategy 162

6.5 Data analysis and preliminary results 165

6.5.1 Characteristics of participants 166
6.5.2 Organizational capacity 166
6.5.3 Policies and values 168
6.5.4 Participation 170
6.5.5 Roles and responsibilities 171
6.5.6 Use of results 172
6.5.7 Effect of participation 174

Summary 176

Chapter 7 : Discussions and conclusion 177

7.1 Integrated results and discussion 177

7.1.1 Framework for integration 178
7.1.2 The elements of e-government outcomes evaluation 181
7.1.3 The patterns of performance 183
7.1.4 Summary of findings 183

7.2 Recommendations for improving practice 184

7.3 Validation 189

7.3.1 Validation of Phase 1 190
7.3.2 Validation of Phase 2 191

7.3.3 Validation of mixed methods 192
7.3.4 Limitations 193
7.4 Contribution 193
7.5 Further research 195
7.6 Summary 196
References 198
Appendices 218
Appendix A: Cover letter 218
Appendix B: Consent form 219
Appendix C: Questionnaire 220
Appendix D: Interview protocol and initial coding frame 222
Appendix D1: Interview protocol 222
Appendix D2: Initial coding scheme 223
Appendix E: Additional quantitative results 224
E1: Descriptive statistics 224
E2: Explanatory factor analysis 225
E3: Propositions and decisions 226
Appendix F: Ethics approval 228

List of tables

Table 1.1: Mixed methods research checklist. (Creamer, 2018; Fetters & Freshwater, 2015; Jalongo & Saracho, 2016) 5
Table 2.1: Overview of literature review strategy 9
Table 2.2: Public sector reforms 12
Table 2.3: Summary of common public value, public values, and public service values frameworks 16
Table 2.4: Public value research: relevant issues 17
Table 2.5: Summary of evaluation practice literature 34
Table 2.6: Relevant extant research on IT evaluation 40
Table 2.7: Summary of IT outcomes evaluation literature review 41
Table 2.8: Dimensions of e-government outcomes 44
Table 3.1: Practice theory: summary of key ontological issues 60
Table 3.2: Schatzki's ontology: relevant concepts 65
Table 3.3: E-government outcomes evaluation practice as Assessment and Use bundles 68
Table 4.1: Types of research by purpose (Collis & Hussey, 2014; Sekaran & Bougie, 2016; Sue & Ritter, 2012) 75
Table 4.2: Basic philosophical assumptions (Bunge, 2006; Kovalainen & Anderson, 2014; Niiniluoto, 2014) 77
Table 4.3:Research paradigm (Cecez-Kecmanovic & Kennan, 2018; Lincoln et al., 2011; Mertens, 2015) 78
Table 4.4: Summary of design decisions 96
Table 5.1: Research strategy for Phase 1 100
Table 5.2: Operationalization of items 104
Table 5.3: Perceptions of e-government outcomes evaluation 105
Table 5.4: Distribution of participants by background 120
Table 5.5: Distribution of participants by background, experience (years and knowledge) and number of evaluation 121
Table 5.6: Evaluation strategies 122
Table 5.7: E-government outcomes evaluated 123
Table 5.8: The purpose of evaluation 123
Table 5.9: Descriptive Statistics: Likert items (N = 106) 124

Table 5.10: Descriptive statistics-collapsed scale (N = 106) 125
Table 5.11: Total Variance Explained 128
Table 5.12: KMO and Bartlett's Test 129
Table 5.13: Goodness-of-fit Test 129
Table 5.14: Pattern matrix (Principal Axis Factor, Promax rotated) 130
Table 5.15: Structure Matrix 131
Table 5.16: Factor Correlation Matrix 131
Table 5.17: Excluded items 131
Table 5.18: Descriptive and reliability statistics of factors (N = 106) 132
Table 5.19: Reliability Statistics 132
Table 5.20: Inter-Item Correlation Matrix 132
Table 5.21: Mann-Whitney test (Differences by management position and experience) 138
Table 5.22: Kruskal-Wallis test (Differences by background and knowledge, df=2) 138
Table 5.23: Descriptive statistics of elements 140
Table 5.24: Pearson Correlation among factors 140
Table 5.25: Auto-Clustering 141
Table 5.26: Centroids of cluster variables (on a 5-point Strongly disagree to Strongly agree Likert scale) 144
Table 5.27: ANOVA of cluster means (df=2) 145
Table 5.28: Cluster composition by management position; background, experience and knowledge 148
Table 5.29: Summary of quantitative results (N=106) 149
Table 6.1 Research strategy for Phase 2 151
Table 6.2: Techniques to improve trustworthiness (Bazeley, 2013; Bryman, 2016; Elo et al., 2014; Noble & Smith, 2015; Sousa, 2014) 153
Table 6.3:Qualitaive research strategy 154
Table 6.4: Distribution of participants (N=6) 155
Table 6.5: Aspects of EOEP examined in Phase 2 156
Table 6.6: Qualitative analysis- important factors 161
Table 6.7: Characteristics of participants 166
Table 6.8: Organizational capacity 168
Table 6.9: Policies and values (Batho Pele and NEPF) 170
Table 6.10: Participation 171

Table 6.11: Understanding of roles....172
Table 6.12: Use of reports....174
Table 6.13: Effect of participation....175
Table 6.14: Summary of preliminary results....176
Table 7.1: Integration framework for analysis: combines results of the quantitative and qualitative phases (Tables 5.29 and 6.14)....179
Table 7.2: Summary of results....184
Table 7.3: Positive and negative or mixed findings....185
Table 7.4: Summary of recommendations:....188
Table 7.5: Validation of Phase 1 (Quantitative)....190
Table 7.6: Validation of Phase 2 (Qualitative)....191
Table 7.7: Validation: mixed methods research....192

List of figures

Figure 1.1: Research process 4
Figure 2.1: Structure of chapter 9
Figure 2.2: Program model, adapted from Chen (2015) 23
Figure 2.3: Program logic model adapted from (McLaughlin & Jordan, 2015) 24
Figure 2.4: Theories used in evaluation 27
Figure 2.5: Evaluation practice model adapted from (King & Alkin, 2018; Mark & Henry, 2004) 29
Figure 3.1: Structure of this chapter 55
Figure 3.2: E-government outcomes evaluation practice as a site 69
Figure 3.3: E-government evaluation practice in the public sector as "site of sites" 69
Figure 4.1: Structure of chapter 73
Figure 4.2: The quantitative-qualitative methods continuum 90
Figure 4.3: Strategy for this research 95
Figure 5.1: Structure of chapter 101
Figure 5.2: Distribution of participants by background 120
Figure 5.3: Distribution of participants by background, experience (years and knowledge) and number of evaluation 121
Figure 5.4: Responsibility for evaluation: initiation and composition of teams 122
Figure 5.5: Scree plot 129
Figure 5.6: Agglomeration schedule coefficients (Centroid method) 141
Figure 5.7: Cluster quality 143
Figure 5.8: Relative sizes of clusters 144
Figure 5.9: Centroids of clusters 145
Figure 5.10: Components of clusters 146
Figure 5.11: Characteristics of participants and variants of practice 148
Figure 6.1: Structure of chapter 152
Figure 6.2: Data analysis strategy adapted from (Cho & Lee, 2014; Mayring, 2014; Rose et al., 2015) 162
Figure 7.1: Structure of chapter 177

Chapter 1 : Introduction

Evaluation of Information Technology (IT) outcomes (also called value, benefits, etc.) is as old as its introduction into the public sector. Today, as was the case many decades ago, researchers and practitioners are preoccupied with the extent to which IT outcomes are being realized. IT managers continue to rank this as an important issue (Kappelman et al., 2016). Evaluation is an important IT management tool: the results can be used to improve and, thereby, achieve the desired outcomes. Thus, it is important to understand how the evaluation of outcomes is done and the results used.

Electronic government (e-government, hereafter), *tout court*, refers to the use of IT in the public sector (Bolívar et al., 2016; Sundberg & Larsson, 2017). E-government outcomes evaluation practice (EOEP), i.e., how the evaluation of outcomes is done and results used, is the overarching remit of this research.

The literature is replete with potential benefits. For example, public sector efficiency, effective delivery of public services (e.g., health and education), engagement with the public, and inclusive growth and sustainable development (Sabani et al., 2018; UN, 2016). Well-managed e-government, therefore, cannot be over-emphasised. According to UN (2016, 2018, 2020), member states of the United Nations have invested in some form of e-government. The optimism, however, may be tempered with caution. Brown (2015) posit an "e-government paradox", i.e., the benefits are incommensurate with investments. Hiziroglu et al. (2017) contend the benefits are not supported by evidence from evaluation.

Attendant to the growing e-government investments are demands on governments to demonstrate outcomes thereof. This is especially so in developing countries, where calls for good governance by development agencies, e.g., the World Bank, have heightened the need for transparency and accountability (Nag, 2018). E-government outcomes evaluation is, therefore, important (Pandey & Gupta, 2018).

Evaluation of outcomes and use of results foster an evaluation culture, e.g., positive benefits can motivate stakeholders to achieve more benefits (Estevez & Montoya, 2015). According to Ceric (2015), evidence suggests that organizations which evaluate outcomes have a higher perceived IT value. However, the literature on e-government outcomes evaluation is scant (Sterrenberg, 2017).

Evaluation contexts in developed and developing countries are different. The latter is characterized by, e.g., a lack of resources, poor management, and donor dependence, thus, e-government evaluation practices from developed countries may not be applicable (Deng et al., 2018). This, notwithstanding, the literature on developing countries is under-represented (Hiziroglu et al., 2017).

E-government is characterized by a higher failure rate, attributable to poor implementation and management (Hiziroglu et al., 2017; Sterrenberg, 2017). Most projects are either abandoned or completed behind schedule, over budget, and fail to deliver expected outcomes. In developing countries, e-government competes with social services (e.g., education and health) for the same resources, therefore, the opportunity cost is higher. Recent surveys (e.g., UN, 2016, 2018, 2020) show developing countries ranked lower in their capacity to use e-government to deliver public services.

There is a need to improve how e-government achieves outcomes in developing countries. Evaluation can be an important management tool (Hiziroglu et al., 2017). However, according to OECD (2013, p. 122), the lack of understanding of e-government evaluation practice "hampers… the capacity to improve e-government development and implementation based on lessons learned". Misuraca et al. (2013) noted that how e-government evaluation is done is relatively neglected, and called for research in this regard. Pekkola and Päivärinta (2019) observed scant IS evaluation practice research in the public sector.

1.1 Problem statement

The preceding section draws attention to two under-developed areas of research, which need attention: e-government evaluation in developing countries and how e-government outcomes evaluation is done. This research addressed EOEP in developing countries. Not much is known about this, therefore, the research purpose was descriptive— to explore the current state of practice. How evaluation is done was conceived as practice and examined with practice theory. This allowed an investigation of practice as entity and performance, i.e., what people do and how this is done (Higginson et al., 2014; Welch, 2016). Thus, the main research objectives were to identify and describe the important elements which constituted EOEP and how these were performed by different evaluation stakeholders. An ancillary objective, inherent to practice research, was to improve future practice.

According to Walliman (2018), a research problem is best expressed as an abstract question and sub-questions to represent sub-problems. Consequently, to achieve the aforementioned objectives, the following questions were formulated for investigation.

Main question: *How is e-government outcomes evaluation done in developing countries*?

Sub-questions:

RQ1: What are the factors which describe how e-government outcomes evaluation is done?
RQ2: What are the patterns of performance among stakeholder groups?
RQ3: How can e-government evaluation be improved?

RQ2 is about how different groups perform their practice, and may be reframed as: "*How do different groups perform e-government outcomes evaluation*"?

The unit of analysis defines what may be included in the research. The focus was "how e-government outcomes evaluation is done". Therefore, practice (i.e., EOEP) constituted the unit of analysis. The research problem, purpose, objectives, and questions were consistent with research underpinned by practice theory (Browne et al., 2014; Nicolini & Monteiro, 2017). Theories served various purposes in this research.

1.2 Theoretical and philosophical perspective

E-government was conceived as a program and evaluation thereof as practice. General assumptions about program evaluation were therefore relevant. Doing evaluation is a social endeavour (Rog, 2015) and, thus, social theory was an appropriate lens. Also, public value is widely used in e-government evaluation (e.g., Sabani et al., 2019). Schatzki's version of practice theory (Schatzki, 2012, 2019) was applied to frame evaluation (and, therefore, e-government evaluation) as practice.

The ontological construct of practice has implications for what can be investigated, and how. Practices are considered "real" (Schatzki, 2017b). Some parts (e.g., activities, people, and tools) are tangible while others (e.g., affective issues) are intangible (Schatzki, 2012). Thus, the post-positivism paradigm—which supports mixed methods (i.e., quantitative and qualitative)— was adopted to examine the different perspectives.

A conceptual model was constructed to integrate the relevant concepts from practice theory, evaluation, and public value, to examine how evaluation of e-government outcomes is done. This model, in conjunction with post-positivism, informed the design and conduct of this research.

1.3 Research strategy

To account for the different perspectives of EOEP, this research adopted a descriptive, sequential explanatory mixed methods (SEMM) survey, i.e., quantitative research followed by qualitative. Figure 1 illustrates the overall research strategy.

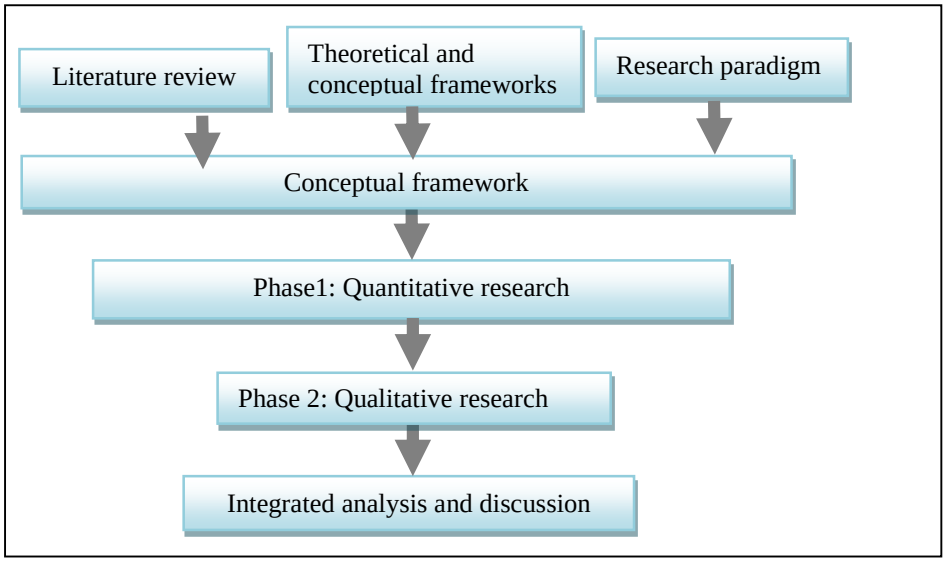


Figure 1.1: Research process

In the quantitative phase, purposive sampling was adopted to select participants with knowledge and experience of EOEP. Questionnaire-based data was collected and submitted to exploratory factor analysis to identify the factors which described participants' perceptions. Subsequently, cluster analysis was conducted to explore the patterns of performance among the different groups of participants. The objective of the qualitative phase was to understand the quantitative results. Semi-structured interviews were conducted to collect data from a subset of the sample in Phase 1; this was analysed with content analysis. Both results were integrated to draw conclusions. This strategy supported the research questions and objectives set out in section 1.1.

The structure of mixed methods research departs considerably from conventional research. Thus, checklists (e.g., Creamer, 2018; Jalongo & Saracho, 2016) are recommended for methodological transparency and also to facilitate understanding of the research audience. Table 1.1 outlines the strategy and tactics implemented throughout the research process.

Table 1.1: Mixed methods research checklist. (Creamer, 2018; Fetters & Freshwater, 2015; Jalongo & Saracho, 2016)

	Aspect of research	Guideline	Tactic
1	Title	Mention the topic and type of mixed methods	Both mentioned in title
2	Abstract	Outline the research topic, type of mixed methods, data sources, data integration, and what was learnt	All discussed
3	Introduction	Establish the need for research	Formulated in problem statement (section 1.1)
		Establish overall rationale for mixed methods	Stated: to examine the different aspects of practice; i.e., to gain insight into EOEP
		Where necessary, formulate both quantitative and qualitative questions to collects quantitative and qualitative data	RQ1: quantitative, about "what" (i.e., to identify elements) RQ2: qualitative, about how (i.e., to understand patterns of performance— how something is done)
		Outline research strategy	Explicated in this section
4	Literature review	Review of mixed methods literature for audience	Reviewed the literature; outlined basic features, main typologies, etc. (section 4.6.7)
		Establish deficit of mixed methods in the area of study	Chapter 2 (Literature review) established paucity in IS, public value and practice research
5	Research design	Overall rationale for mixed methods	Explanatory (stated in Abstract, Introduction, and section 4.8
		Detailed description of applicable typology	Overview of SEMM (section 4.8)
		Visual illustration of quantitative and qualitative methods in the order of phases	Illustrated by diagrams (Figure 4.3)
		Detailed description of methods according to phases	Overview of quantitative (section 4.8) Purpose: to describe practice Sampling: purposive, snowball Data collection: questionnaire Data analysis: descriptive, exploratory factor and cluster analyses Overview of qualitative (section 4.8) Purpose: to understand results of Phase 1 Sampling: purposive; sampling frame: Phase 2 sample Data collection: semi-structured interview Data analysis—content analysis
		Mixing	Sampling; results
6	Implementation	Phase 1: Quantitative (Chapter 5) Phase 2: Qualitative; follow-up on Phase 1 (Chapter 6)	
7	Discussion and conclusion	Integrate results of phases and draw meta-inferences (Chapter 7)	
8	Contribution	Domain, i.e., e-government outcomes evaluation (Chapter 7) Methodology, i.e. practice research and mixed methods (Chapter 7)	
9	Reporting	Present implementations in respective phases and integrate results; write for audience: use headings, visual display, etc.	

1.4 Research context and sample

South Africa was selected as the context for the investigation. This was considered an appropriate case because the country has undergone public sector reform underpinned by e-government (DPSA, 2001) and evaluation (Podems et al., 2014). E-government is well-developed (Bwalya & Mutula, 2014) and some programs (e.g., e-filing and e-Justice) are considered successful and have been studied (Naidoo et al., 2011). Also, the country's e-government ranking is comparatively higher among developing countries (UN, 2014, 2016, 2018, 2020). Thus, it is likely that the phenomenon of interest (i.e., e-government evaluation practice) existed and could be explored.

Practitioners can give accounts of their practice. Public sector employees are likely to be involved in the use of e-government evaluation to improve service delivery and, therefore, constituted relevant data sources. Also, it was easier to identify public employees with knowledge of e-government evaluation practice, compared to external stakeholders. Participants were selected from IT, Evaluation and Other backgrounds.

1.5 Summary of results

Participants' perceptions of EOEP may be described by six elements (i.e., Evaluation framework; Use of results, Organizational capacity; Shared understanding; Stakeholder participation, and Affective issues) and three distinct patterns of performance linked to the demography of participants (i.e., IT and Evaluation managers with extensive experience; IT and Evaluation non-managers with moderate experience and Other non-managers with little experience).

Overall, the research showed an inadequate internal organizational capacity to evaluate and use results; the main effect was individual and organizational learning; lack of implementation of recommendations for improvement; predominance of organizational (compared to democratic and development) e-government outcomes; and significant differences among participants (e.g., managers with Evaluation backgrounds and experience were likely to have a positive perception of their practice). Several challenges to current EOEP were identified and recommendations were drawn for improvement.

1.6 Delimitation

Upfront specification of scope furthers the understanding of research stakeholders. E-government evaluation is generally classified as pre-, during, and post-implementation. This

research focused on post-implementation and, specifically, how public organizations in developing countries evaluate e-government outcomes. This elided evaluation of specific e-government to establish outcomes *per se* or development of frameworks for such a purpose. The research context was South Africa; the sample consisted of only public sector employees.

The research purpose was descriptive, which excluded testing theory or establishing causality. Consistent with practice theory, the primary objective was to explore the elements of practice and patterns of performance. The following elements of EOEP were examined: context, resources, activities, theory (i.e., underlying assumptions), and outcomes. Notwithstanding its value, as with all research, this has limitations.

1.7 Relevance and contribution

This research addressed the clarion call for IS research to generate value for both researchers and practitioners. Although there are no standard evaluation criteria, rigour (i.e., validity of the research process and, thereby, outcomes thereof) and relevance (i.e., the extent to which the research addresses an enduring, pervasive or current problem) are commonly used (Hassan, 2014; Lyytinen & Rowe, 2017; Toffel, 2016). The value to researchers is measured by a relevant problem, rigorous process which contributes to a domain's body of knowledge, innovative use of theories and concepts, etc. To practitioners, this is about a relevant problem and the extent to which the outcomes are suitable, and can be implemented. This research pursued both rigour and relevance.

Section 1.1 established the following: lack of evaluation as one of the primary causes of e-government failure, the need for e-government evaluation in developing countries, and paucity of research on how e-government evaluation is done (i.e., EOEP). Simonofski et al. (2017) aver evaluation research from practitioners' viewpoint is much elided. How e-government evaluation is done in developing countries is, therefore, relevant to both practitioners and researchers.

This research contributes to both e-government (and, in general, IS) evaluation theory and practice. The empirical evidence illuminates current practice (e.g., what is done well, challenges, etc.) and adds to the literature. Recommendations for action are drawn to improve both practice and research. Furthermore, the research design contributes to the development of theoretical, methodological, and empirical approaches to both practice and mixed methods research in IS. The following is an overview of the thesis.

1.8 Outline of thesis

This thesis bears a unique structure, due to the mixed method strategy adopted: Phase 2 depended on the results of Phase 1. This chapter provided an overview of this research—identified gaps in the literature, justified the relevance of the research problem, and specified the objectives, scope, and expected contributions.

Chapter 2 reviewed the literature of the disparate domains to situate EOEP in developing countries in previous research and identify the relevant concepts.

Chapter 3 built on the preceding chapter to explicate the theoretical framework (i.e., Schatzki's ontology), draw implications, and construct a conceptual model to inform this research.

Chapter 4 discussed the research design, a plan to assure consistency among the research elements (i.e., paradigm, strategy and applicable methods, etc., with the research purpose). Post-positivism was adopted to inform the SEMM research.

Chapter 5 focused on the quantitative component (Phase 1) and described the assumptions, implementations, and results. The objective was to describe the main elements of e-government outcomes evaluation and the patterns of performance.

Chapter 6 described the implementation and results of the qualitative component (Phase 2). The main objective was to follow up on the results of Chapter 5 to understand how e-government outcomes evaluation is done.

Chapter 7 integrated the results of the respective phases to draw conclusions, develop recommendations for improvement on current EOEP, and identify opportunities for future research.

Chapter 2 : Review of literature

The preceding chapter formulated the research problem as how e-government outcomes evaluation is done in developing countries. This review located the problem in the context of public sector performance evaluation and social practice. Broad themes from the relevant domains were identified and integrated to facilitate understanding. Figure 2.1 illustrates the structure of this chapter.

Figure 2.1: Structure of chapter

2.1 Literature review strategy

Several typologies have been proposed; important factors to consider include the state of the problem domain and the purpose, breadth, and structure of the review. Table 2.1 shows the review strategy adopted.

Table 2.1: Overview of literature review strategy

Aspect of review	*Description*
Purpose	To understand the research problem
Scope	Adequate coverage; relevant literature from various sources
Process	Iterative
Inclusion criterion	Relevance

A review for understanding helps to conceptualize a, hitherto, under-researched problem. It involves interpretation of literature (and integration of resultant knowledge) from disparate domains and, therefore, is usually narrative and detailed (Rowe, 2014). This was appropriate: a reasonable collection of relevant literature was reviewed to understand the problem. The sources included books, journals, conference proceedings, and documents.

The main criterion for inclusion was relevance: each source was read to decide whether or not to include. Acquisition involved database search and expert recommendation. A scheme of keywords was developed and various database search techniques were employed. The following broad areas were covered: e-government, public management (reforms, IT and public value), practice (particularly evaluation practice, with emphasis on IT and e-government), and their relevance to developing country context. Furthermore, experts, e.g., Theodore Schatzki and Alison Browne offered their recent works and also pointed to other sources.

The selected sources were analyzed to: understand the problem and the underlying theoretical and philosophical assumptions; specify gaps; identify important concepts, strategies, and methods; justify design decisions and formulate implementation plans. An iterative process ensured improved understanding, which could be applied to refining the constituent stages (e.g., further search, acquisition, analysis, and interpretation).

The outcomes of the review are presented in three related parts. This chapter concerns e-government outcomes evaluation in developing countries; Chapter 3: theoretical and conceptual framework; and Chapter 4: research design.

2.2 E-government and public sector performance

This section situated e-government outcomes evaluation within public sector performance evaluation and outlined: the place of IT in public sector reforms and service delivery, how the value contributed is assessed, and research thereof.

2.2.1 E-government

There is no universally accepted definition of e-government. One way is to define by the stakeholders and corresponding outcomes (Rose et al., 2018). Internal stakeholders are public organizations and employees while external encompass the public in general —citizens, businesses, etc. For this research, e-government refers to the use of IT in the public sector to achieve internal efficiency; effective service delivery and interaction between public sector

organizations and citizens, businesses; and improved accountability and transparency (Deng et al., 2018; Rose et al., 2018). Where necessary, the umbrella definition, "the use of IT in the public sector", is used as a placeholder for the foregoing definition. An understanding of the public sector as a context of e-government, and evaluation thereof, is important.

2.2.2 The public sector

The public sector is characterized by public service and a wide range of stakeholders. It consists of two main parts: government, which sets policy and public administration, responsible for policy implementation and management (Shafritz et al., 2015). However, the two are often used interchangeably (Cloete, 2012). This research focuses on the public administration component.

There is a need to measure the performance of public organizations to establish whether or not the mandate (i.e., public service) is being achieved. However, the nature of the public sector makes performance measurement difficult. Outcomes encompass both financial and non-financial; are difficult to track, measure from the perspectives of diverse stakeholders, and attribute to specific programs or organizations (Hartley et al., 2017; Hughes, 2012). Consequently, surrogates (e.g., activities and outputs) are widely measured, instead of outcomes (Hartley et al., 2017).

There are differences among public organizations, subsectors (health, defence, education, etc.), and levels of government (e.g., national, provincial, and local), therefore, differences in practices (Hodgkinson et al., 2017). Although important for decision-making, performance information is often not used (Hammerschmid et al., 2013).

The foregoing describes the public sector as the context of e-government and its evaluation. The public sector has undergone transformations over the years, with implications for delivery systems. This has spawned public service management practices, theories and frameworks, and research.

2.2.3 Public sector reforms: an overview

Various authors (e.g., Bryson et al., 2014; Radnor et al., 2016) have chronicled public sector reforms. Table 2.2 summarized the characteristics of each era. The main features of public administration (the period before the 1980s) were efficiency and legitimacy— assured by hierarchical structures; objective (i.e., apolitical) professional public servants that implemented government policy; etc. Citizens' involvement in the policy process amounted to voters,

constituency, etc. Public administration failed because the sector was steeped in politics and officials served their interests.

Table 2.2: Public sector reforms

Era	Reform	Characteristics
Up to 1980s	Public administration	Efficiency and legitimacy; politics-administration dichotomy; professionalism
1980s to 2000s	New public management (NPM)	Efficiency and effectiveness; IT, decentralized organizations; private sector management practices; citizen as customer
Post-NPM	Called Public value, Digital governance, etc.)	Efficiency; effectiveness; social and democratic values; co-production of service by public organizations and external stakeholders; performance management

The New Public Management (NPM) emerged in the I980s and 1990s. The factors which occasioned this included bureaucratic and inefficient public sector, IT and the redesign of processes, and globalization (Hughes, 2012). The basic tenet was running public organizations like a business. Private sector techniques (e.g., decentralization, performance measurement to decision-making, and users as customers) were introduced into the public sector to improve organizational efficiency and effectiveness in service delivery. The term public management usurped public administration. The private sector values were at variance with public service and failed to address exigencies of collaborative government, social and democratic objectives (social cohesion; social inequalities; justice, etc.), and the environment (Bryson et al., 2014; Moore, 2013). NPM showed private sector practices are not well-suited to the public sector (Hughes, 2012).

The shortcomings of NPM have engendered an alternative paradigm of public management (Radnor et al., 2016). Notwithstanding the disagreement on nomenclature (i.e., there are various labels), there is consensus that this encompasses market as well as social and democratic values. The underlying principle is collaboration (among public organizations and with the public) in service delivery. Public employees are not apolitical, but catalysts to this end. Accountability extends beyond internal structures to politicians, citizens, the law, etc. A key feature is IT— to facilitate greater access to information and decision-making processes. For this research, public value was adopted to characterize the management practices and attendant theories and research of the era.

There is growing recognition that public value provides an alternative perspective on performance measurement (Hartley et al., 2017). Developments in the public value field converge from three principal streams, *viz*., managerial, social, and psychological (Bryson et al., 2014).

2.2.4 Public value: voices in the field

The central themes in the literature are public value and public values (Bryson et al., 2014). Although related, these concepts are distinguishable and provide different lenses for examining phenomena in the public sector (Witesman, 2016). Value, values, and public are polysemic; there are no universally acceptable definitions in literature. The result is a conceptual muddle (Nabatchi, 2012), which permeates both research and practice.

In general, value concerns the worth of an entity; the public refers to a collective social entity, e.g., a group of individuals; organization; community; state; etc. (Meynhardt, 2015). Public value may be defined as what the public considers worthy— created and sustained by public organizations in conjunction with the public. Public values, as a concept, prescribes an appropriate relationship between the public sector and the public. The following highlights the key characteristics of the three main streams in the literature. The purpose is not to sustain the raging debates in the field but to draw relevant concepts to inform this research.

2.2.4.1 Moore's public value framework

Originally, Moore (1995) developed public value to help public managers to create value— analogous to how private sector managers create private value. It has since been extended to include other service delivery stakeholders (e.g., other public employees, politicians, businesses, and citizens). Public value is both a process and an outcome.

As a process, public value is an organizing framework. The strategic triangle consists of the public value to create; legitimacy and support (e.g., authority and resources from stakeholders) for action; and operational capability (e.g., resources and appropriate processes) to achieve such value. The components are interrelated, e.g., stakeholder involvement can improve legitimacy (or trust) in public organizations and, thereby, improve public value— which may lead to the allocation of more resources to create more public value.

As an outcome (i.e., what is created or added), public value is a multifaceted construct which encompasses efficiency, effectiveness, accountability, engagement, satisfaction, justice and fairness, etc. Moore (2013) broadly divides into the cost of using public assets and achievement of public outcomes. Performance measurement focuses on each stage of the value chain (to obviate what destroys, while improving what adds, value), and the public is best placed to determine what constitutes value (Benington & Moore, 2011).

The public value framework has been adopted by various disciplines, governments and public organizations worldwide, including sub-Saharan Africa (Benington & Moore, 2011; Bryson et

al., 2014). The criticism includes the managerial focus, which can encumber public value creation (Bryson et al., 2014; Vandenabeele et al., 2013). Table 2.3 shows the main dimension of public value as outcomes.

2.2.4.2 Bozeman's public values framework

While public value concerns the creation of public value, public values concerns the relationship between actors in the public sector and social entities *apropos* public service (Witesman, 2016). The focus is on the social outcomes of policy (Nabatchi, 2012). The assumption is that public values can be systematically identified from relevant sources (constitution, policies, mission statements, etc.) using literature review, case study, survey, etc., and analyzed to build criteria to assess the extent to which public values have been achieved (Welch et al., 2015). By comparing the expected value against achieved, it can be determined whether specific public value failure, or otherwise, has occurred.

Nabatchi (2012) categorize public values into five main "frames". Organizational values, e.g., administrative efficiency, objectivity, and professionalism. Market values, e.g., cost-efficiency, cost-effectiveness, and productivity. Political values (linked to democracy), e.g., participation, representation, responsiveness, liberty, and equality. Legal values, e.g., rights of the individual, fairness, and equity. Itinerant values, common to the other frames, e.g., accountability; legitimacy, and trust. Table 2.3 shows this and other categorizations of public values.

The public values framework has been adopted in various disciplines (e.g., policy evaluation research) and by various governments and public organizations to promote governance (van der Wal et al., 2015; Welch et al., 2015). The main criticism is the lack of a predefined set of mutually exclusive values— the elements may be incompatible and overlapping.

2.2.4.3 Meynhardt's public value theory

Meynhardt (2009, 2015) locates public value in valuing, i.e. evaluation. The framework is based on human-psychological needs. Value is the result of evaluating an object (e.g., e-government service) by a subject (e.g., the public), and turns on emotional-motivational characteristics, e.g., satisfaction. This means evaluation, as a process, is subjective— grounded in peoples' values (i.e., beliefs, perceptions, etc.).

The framework combines both public value and public values (Witesman, 2016). The process of "valuing" is rooted in the context of shared values. Public value can, therefore, be viewed as a shared value about the quality of relationships with public organizations (Meynhardt, 2015).

Table 2.3 shows there are four dimensions. Moral-ethical: about actions which ensure equity and justice; utilitarian-instrumental: efficient processes and effective creation and use of service; political-social: groups and their interests; and hedonistic-aesthetic: satisfaction with experience (Meynhardt, 2015).

The differences among the streams permeate disciplines and practical uses. However, there are striking similarities, e.g., in the categories by Bozeman, Meynhardt and Nabatchi. A common thread is the need to establish value, i.e., valuing or evaluation (Bryson et al., 2014; Rutgers, 2015). Some authors (e.g., Benington & Moore, 2011; Witesman, 2016) contend there has been a cross-pollination of the literature— public value and public values can be combined, both in theory and practice.

According to Vandenabeele et al. (2013), a combination of the streams can further an understanding of public sector performance. Public values can be managed to define the public value to create, gain legitimacy and support, and build capacity to achieve outcomes. As stated above, public values may guide employee behaviour; however, specific values may be derived for the public sector.

2.2.5 Public service values

Public values and public service values are distinguishable— albeit used incorrectly as substitutes. According to Turner (2015), the latter is a subset of social, professional, ethical, and other values which directly inform people's roles and actions in public service— to promote consistent practices and, thereby, gain citizens' trust. Thus, public service values are important for creating public value. For example, organizational values (e.g., professionalism and adherence to hierarchical authority) are imperative to value creation in the military. de Graaf et al. (2016) identified ten categories of values (see Table 2.3).

The relationship between public value, public values, and public service value is evident. In many countries, public service values are distilled from public values *writ large* and coded into legislation, rules, policies, etc., to guide employees. In South Africa, public service values are enshrined in the constitution and translated into public service principles (i.e., *Batho Pele Principles,* section 2.7). Public value, i.e., valuing, is an appropriate lens to examine public sector performance (Bryson et al., 2015; Vandenabeele et al., 2013).

Table 2.3: Summary of common public value, public values, and public service values frameworks

Author	Focus	Value dimensions
Moore (1995, 2013)	Public value as process and outcomes of public value creation	Efficiency; effectiveness; accountability; engagement; satisfaction; justice and fairness. Broadly divided into cost of using public assets and achievement of public outcomes
Jørgensen and Bozeman (2007)	Public values (prescribes behaviour)	The public sector's contribution to society; the relationship between public administration and politicians; the relationship between public administration and its environment; internal function and organization; public employees
Meynhardt (2009, 2009, 2015)	Psychological; valuing (combines public value and public values)	Moral-ethical, utilitarian-instrumental, political-social, and hedonistic-aesthetic. Emphasizes affective issues
Kelly, Mulgan, and Muers, 2002)	Public value	Service (e.g., fair access and use); Outcomes (e.g., high employment as an outcome of education as service) Trust/confidence/legitimacy in the relationship with the public sector
Nabatchi (2012)	Public values	Organizational, market, political, legal and itinerant
de Graaf et al. (2016)	Public service values	Openness; participation; accountability; legitimacy; effectiveness; efficiency; integrity; lawfulness; professionalism; treating equal cases equally

2.2.6 Public value and evaluation

Evaluation is important, to determine whether public value is created or not (Bryson et al., 2015). The brief exposition above shows public value may be viewed either, and both, as process and outcome. It is important to assess, from the perspective of the public (i.e., group of individuals, organizations, etc.) whether public value (i.e., as outcome) has been created (or destroyed). Public values (i.e., what the public values) and public sector values (which prescribe desired behaviour of employees) form the necessary backdrop. Furthermore, as pointed out in Meynhardt's framework, affective issues, e.g., satisfaction, are important.

Although not well-defined, the dimensions of public value include efficiency, effectiveness, democratic values, and accrue internally to the public sector and externally to the public. It accommodates both financial and non-financial outcomes as well as unique contexts, as public values play a role. However, according to Bryson et al. (2014), evaluation of democratic values is often ignored. Public participation can bring different perspectives to shape values and processes. Notwithstanding the opportunities, there are also several challenges for public value research.

2.2.7 Public value research: current issues

The profusion of frameworks present challenges to research. According to Marcon (2014), different frameworks yield different research outcomes. Hartley et al. (2017) attribute the paucity of empirical research to the complex nature of public value and the conceptual muddle.

According to Bryson et al. (2015) and Fukumoto and Bozeman (2019), how public value evaluation is done, by whom, and what is measured, need attention. Hartley et al. (2017) contend research on public value outcomes is relatively ignored, compared to activities and outputs. There are differences among public organizations (Hodgkinson et al., 2017); thus, there may be differences in public value and assessment thereof. Hartley et al. (2017) call for comparative research in this direction.

Empirical research may be conducted from several perspectives (Hartley et al., 2017). Co-production of public values means the involvement of diverse stakeholders; however, traditionally, the focus is on managers— which fail to paint a true picture. Also, there is tension between whether to investigate public value from the point of view of individuals or collective (e.g., group of citizens, organization, etc.).

From an ontological and methodological standpoint, Hartley et al. (2017) noted that parts of public value are real while others are constructed. Therefore, several strategies (case study, survey, etc.) and methods (i.e., quantitative, qualitative, mixed) may apply to a variety of purposes (exploratory, descriptive, explanatory, etc.); however, case studies are predominant.

Table 2.4 is a summary of the literature on public value, as outlined above. This research contributes to addressing some of the issues.

Table 2.4: Public value research: relevant issues

Aspect	*Description*
Public value, as outlined above	Adequate lens for performance management research in the public sector Main concepts: public value, public values, public sector values Focuses on both public value as outcome and process As outcome: the components are efficiency, effectiveness and democratic values As process: about co-production of public value, which emphasizes stakeholder involvement As valuing: underpinned by public values and affective issues Stakeholders include public organizations, public employees, politicians, businesses, citizens, etc.
Research problems: gaps	Scant public value research attributable to complexity and ambiguity How public value is evaluated needs attention Democratic values under-researched compared to efficiency and effectiveness Performance mainly examined as activities and output; outcomes under-researched Differences among organizations, therefore, a need to examine similarities and differences
Ontology	Some aspects of public value are real, others constructed
Research perspectives	Multi-stakeholder perspective; managerial is predominant Assumptions: managerial, social, psychological Individualism-collectivism; individualism predominant
Methodology	Important to use different methodologies to examine the different perspectives of public value Quantitative research to describe organizational performance; qualitative and mixed methods research for understanding Case study is predominant; a need for methodological diversity

The importance of public value as a lens for investigating public sector performance and the relevant research issues have been established. As stated above, IT is a significant cog in public service delivery. Various authors (e.g., Bannister & Grönlund, 2017; Moon et al., 2014) used the co-evolution of public sector reforms (as outlined above) and IT to chronicle the management and research implications.

2.2.8 IT in the public sector

Two main epochs may be identified in the IT in the public sector literature. Early uses of IT concerned organizational operations and control, and management of resources, e.g., human, information, and infrastructure (Bannister & Grönlund, 2017; Moon et al., 2014). Thus, this era was defined by application IT to achieve internal organizational performance, i.e., efficiency. The locus of control was the IT unit, and evaluation concerned the impact on organizations and the work of employees.

The advent of the Internet (especially the Web) enabled seamless integration of processes, collaboration among different public organizations, and engagement with the public in political and service delivery processes. This engendered changes in implementation, use, and management to include both internal and external stakeholders. Thus, the locus of control extended beyond the IT unit.

The concept of e-government emerged during the Internet era and coincided with NPM, hence, the parallels with e-commerce in the private sector (Cordella & Bonina, 2012). The underlying assumptions emphasised market-related outcomes, e.g., efficiency, cost reduction, etc. The focus on service delivery explains the undue emphasis on technology as a medium and the narrow definitions of e-government, e.g., as the Internet for public service delivery (Moon et al., 2014). Evaluation, therefore, mainly concerned internal, economic outcomes.

E-government is a tool to support the public sector to maximize public value; the *raison d'être* is public value creation (Bannister & Connolly, 2014). Thus, there have been calls (Pang et al., 2014; Scott et al., 2016) to establish not only organizational efficiency and effective service delivery but also outcomes that accrue to the public at large. Furthermore, public value, as described above, is considered an appropriate lens for examining e-government outcomes (Al Rawahi et al., 2016; Deng et al., 2018; Scott et al., 2016).

Differences in evaluation contexts, e.g., stakeholders, values, resources, processes, etc., may result in different outcomes. This research concerned developing countries, therefore, an understanding thereof is germane. The "e-government divide" is often used to distinguish

between developing and developed countries, e.g., the former is characterized by a lack of resources; contrasting against developed countries provides a framework for systematic inquiry. Studies (e.g., Deng et al., 2018; UN, 2020) have taken advantage of such distinction.

2.3 E-government in developing countries

Developing countries are distinguishable from developed countries by unique historical, social, political, and economic factors (Turner et al., 2015). Most inherited public sector characterized by, *inter alia*, lack of financial and human resources— which hamper policy development and implementation, and service delivery (Richmond, 2015). The consequence is a general failure to support economic and social development (Zafarullah & Sarker, 2016). Most countries, therefore, resort to development assistance.

2.3.1 Development assistance and public sector reforms

According to Turner et al. (2015), development as a concept broadly encapsulates activities to reduce poverty and improve people's lives, while ensuring sustainable use of resources. It involves assistance from developed countries to recipients, mostly developing countries. Thus, development is a common feature of developing countries. Over the years, development agencies, e.g., the World Bank and the International Monetary Fund (IMF), have targeted the public sector of developing countries through conditionality to drive development. Two such conditionalities are public sector reforms and governance reforms (Turner et al., 2015; Zafarullah & Sarker, 2016).

Many developing countries adopted NPM during the 1980s, supported by e-government. NPM did not achieve the expected goal of improved performance because the strategies failed to address the unique contexts of developing countries (e.g., lack of resources, low capacity to develop and implement policies, etc.) Development agencies introduced governance reforms as a condition for assistance to developing countries in the 1990s.

Governance has economic, social, democratic, and environmental focus, and involves the public and private sectors as well as civil society (Zafarullah & Sarker, 2016). The characteristic features are participation, rule of law, equity and inclusiveness, transparency, responsiveness, consensus and legitimacy, effectiveness and efficiency, and accountability (Nag, 2018). As Zafarullah and Sarker observed, governance is consonant with the concept of public value in public management.

The centrepiece of development assistance is IT (Lal et al., 2018). Recent initiatives (e.g., the Sustainable Development Goals) also reinforce the role of e-government. The significant role of IT in development has engendered a cognate field, *viz*., IT for development (ICT4D). Development issues (e.g., poverty reduction and good governance) pervade e-government research in developing countries (Nam, 2018; Wahid, 2013).

2.3.2 E-government and development

Contextual factors mediate between e-government investment and outcomes. Thus, e-government models for developed countries may not be applicable in developing countries. Developing countries are characterized by a lack of IT infrastructure, management expertise; skilled employees; digitally skilled public; low income; high cost of access; weak institutions, etc. (Al Rawahi et al., 2016; Keramati et al., 2018; Ramli, 2017; UN, 2016).

The implication for e-government is the higher failure rate, due to poor implementation, management, and usage. Most projects are either abandoned or completed late, above budget, and do not meet user needs (Idoughi & Adbelhakim, 2018). Developing countries have consistently ranked lower in the United Nations' surveys (e.g., UN, 2016, 2018, 2020), which compare member states (according to IT infrastructure, human capacity, and online services) on their capacity to use e-government for service delivery.

The higher failure rate also means a higher opportunity cost. In developing countries, e-government competes with basic services (water, education, health, etc.) for limited resources. Thus, it is important to determine the contribution of e-government in developing countries. Furthermore, most e-government projects are funded as development assistance (Lal et al., 2018). There is a growing demand by development agencies for feedback on investments (Pandey & Gupta, 2018).

Goldkuhl and Lagsten (2012) encourage IS evaluation researchers to draw on program evaluation as a reference discipline. For this research, e-government was viewed as a program. Various authors (e.g., Misuraca et al., 2013 and Savoldelli et al., 2013) have similarly conceptualized e-government.

2.4 Evaluation practice

There is growing interest in the study of practices (i.e., what people do) to understand social phenomena (Hui et al., 2017). This way, the interconnected parts, e.g., people, activities, and material things (i.e., resources, physical context, etc.) can be examined at once. Rog (2015),

Goldkuhl and Lagsten (2012) and Misuraca et al. (2013) respectively conceptualized evaluation, IS evaluation, and e-government evaluation as practice. An understanding of what is practice, what is evaluation, and, thereby, their interconnection is essential.

2.4.1 Practice

Practice is polysemantic (Hui et al., 2017). Some common uses include: a learning method, e.g., practise a song; occupation or field of activity, e.g., evaluation and medicine; the way something is done, e.g., driving, evaluating, and teaching; and culture, e.g., a way of eating with a specific implement (Hui et al., 2017; Reckwitz, 2002; Welch, 2016). In organizations, practices represent how work is done (Erden et al., 2014). Work involves interactions among people, the physical environment, and its objects (Gherardi, 2012). More importantly, it is goal-oriented, i.e., to create or achieve an outcome.

There are two ways to view practices: entities and performances (Welch, 2016). As the former, practices are considered as real, composed of elements. For example, evaluating, marriage, etc. may be viewed as things which exist. However, such entities may only come into existence through action or performance. That is, evaluating, marriage, etc., only exist when performed by people. A focus on both perspectives (i.e., entity and performance) enables an investigation— at once– of what is done and how it is done.

There is no consensus on the constituent elements or content of practice (Higginson et al., 2015; Spaargaren et al., 2016). According to Schatzki (2012), practice consists of interconnected human activities organized by understandings; rules; and teleoaffective elements (e.g., outcomes and affect). Activities (e.g., knowing, thinking, saying, and doing) are directed towards an outcome (e.g., to acquire new knowledge, create an artefact, etc.) and involve not only physical, but also cognitive and affective human characteristics.

There is a growing recognition that affect influences human cognition and behaviour (Zhang, 2013). Affect is polysemic (Gherardi, 2017), involves emotions, moods, feelings, attitudes, satisfaction, etc., and is directed at a target. For example, an employee may have a strong dislike for a specific activity. Affective issues have gained currency in social research (Welch, 2016). Cognition concerns thinking and involves the knowledge to understand and evaluate circumstances, which include the physical environment; people; the outcome to achieve; appropriate activity to perform; tools; etc.

Schatzki's description is consistent with the characterization by Reckwitz (2002) that practice consists of interconnected elements: bodily activities; mental activities; things and their use;

background knowledge of understanding, know-how; emotions; and motivation. According to Welch (2016), others have categorized the elements as understanding (e.g., know-how and practical understanding); procedures (e.g., rules, principles, and instructions), and engagement (i.e., a set of ends, affective and normative perspectives, etc.); and material (i.e., objects, tools, and infrastructure) for performance.

The elements of practice may be arranged or combined differently under different circumstances, thus, there can be different patterns of practice. For example, based on different skills and experience, physical context, etc., different people may perform activities differently (Browne et al., 2014; Schatzki, 2005).

Although undertaken by individuals, practice is a collective endeavour— "people's activities" (Schatzki, 2012, p. 13); Feldman and Worline (2016, p. 307) averred "practices are enacted through the actions of individuals but are never simply the actions of an individual". The rules, sense-making, intended outcomes, resources, acceptable actions, etc., may not be exclusive to the individual but rooted in, and guided by, general understanding and interaction with other actors in the work practices. Thus, there is a shift in emphasis from the individual to practice as the unit of analysis (Feldman & Worline, 2016; Nicolini, 2017).

An examination of work practices may range from a simple description of what people do (i.e., (i.e., the constituent elements of practice as an entity) to an understanding of, e.g., what actions are performed and what knowledge and external objects are used. Further to the unit of analysis, practice provides a theoretical perspective and terminology (Nicolini, 2017).

The study of work as social practices offers, among other things, an opportunity to improve current practices. For example, according to Welch (2016), practice as entity focuses on the elements (e.g., actors and their interactions) which inform practice as performance (i.e., how work is done— the observed behaviour). Therefore, the former can be targeted to achieve changes in observed behaviour.

For this research, practice is defined as how work is done in organizations and viewed as a real object. This supports the research objectives (i.e., to identify and describe the elements of practice, patterns of performance), and suggests ways of improving current practice. However, an evaluation may be characterized variously. Thus, the next section examines and situates evaluation in social practice.

2.4.2 Evaluation

A program is an open system and consists of subsystems, which interact with each other and with the environment (Chen, 2015). Figure 2.2 shows a program model. Processes are activities which transform input (i.e., resources) into outputs (e.g., products) within a context. Feedback is information about (and which may be used to adjust, e.g., improve) the program. The system model is widely used in IS research (Burton-Jones et al., 2015).

The conceptualization of e-government as a program is appropriate. Resources (e. g., human; financial; technological; etc.) are transformed by processes (e.g., IT implementation and management; workshops; etc.) into output (i.e., services). The feedback from evaluation may be applied to improve parts of the system (e.g., improve processes and, thereby, services).

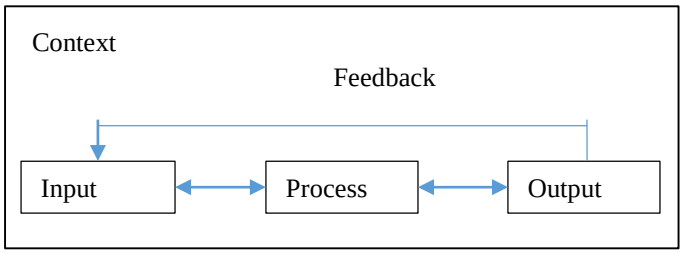


Figure 2.2: Program model, adapted from Chen (2015)

2.4.2.1 Evaluation defined

Many terms, including evaluation itself, lack universal definition (Patton, 2018; Schwandt, 2018). This may be attributed to the multidisciplinary nature of the field (Chouinard et al., 2017). Common definitions include: making a judgement about the merits and worth of an evaluand (Patton, 2018); how organizations assess their performance, programs, and services (Rogers & Gullickson, 2018). Furthermore, an evaluation may be classified variously; e.g., according to program stage, i.e., pre-, during, and post-implementation (also referred to as planned, ongoing, and completed). This research concerned how public organizations assess outcomes of completed e-government programs.

There is a growing demand for outcomes evaluation in the public sector (Wholey, 2015). This is attributable to the corresponding demand for transparency and accountability by stakeholders for governments and public organizations to demonstrate the "value of money" of investments. Evaluation of outcomes establishes how and why such outcomes are being achieved (Chen, 2015; Connelly & Vanderhoven, 2018), and provides an "understanding of strengths and weakness" of programs (Phillips & de Wet, 2017, p. 103). Two important decisions in evaluation are what to evaluate and how.

A program logic model embodies assumptions about program implementation and management thereof and serves as a useful framework to guide evaluation practice and research (McLaughlin & Jordan, 2015; Reynolds et al., 2016). For example, Mark and Henry (2004) use a logic model to represent the effect of evaluation.

2.4.2.2 Basic features of evaluation

There are several versions of logic models— different in content, i.e., the number of elements (e.g., in some models, context subsume resources) and form (e.g., diagram, narrative and tabular). In general, the models are depicted as simple and linear, rather than complex and dynamic (Chen, 2015; Reynolds et al., 2016). Figure 2.3 illustrates how evaluation is done. The arrows explicitly show interaction among the elements.

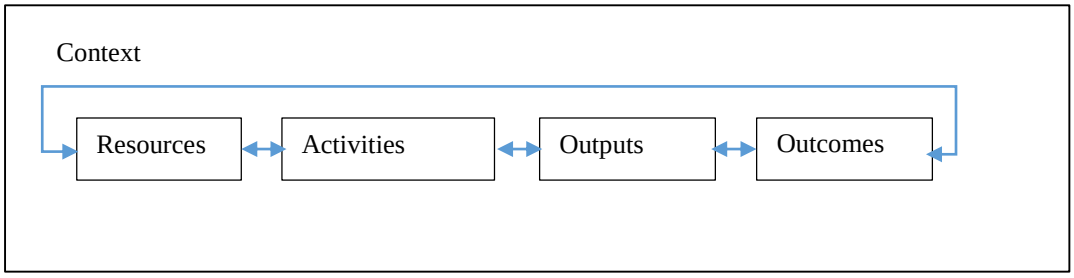


Figure 2.3: Program logic model adapted from (McLaughlin & Jordan, 2015)

Resources include the requisite human, financial, physical, and other input. Evaluation activities represent appropriate actions e.g., planning, implementing, and disseminating results. Planning is the overarching design to ensure results are both valid and relevant. This involves specifying stakeholders and their interests, roles, etc.; relevant data; strategies; data sources; appropriate data collection and analyses methods, and dissemination of results to stakeholders. Implementation executes the evaluation plan.

Although not clear cut (and not separated in some models), outputs, and outcomes are distinguishable. Output is the direct result or product of doing an evaluation (e.g., a written report which consists of findings and recommendations) while outcome is the effect (i.e., change, influence, consequence, etc.) of an evaluation. The change may be in awareness, attitudes, knowledge, skill, behaviour, etc., attributable to the evaluation (Schwandt, 2015). Outcome is widely used and, therefore, preferred to value, benefits, performance, contribution, impact (i.e., long-term outcome), etc., in the e-government evaluation literature.

Evaluation context foregrounds the spatial and temporal features, hence the characterization as situated (Coldwell, 2019). Contextual factors may be organized into organizational (internal and external); evaluation stakeholder; and program. Internal organizational factors include

leadership; structures; processes; politics; and evaluation culture (e.g., shared values about evaluation, preference for paradigms and methods, capacity building, commitment to regular evaluation and use of results); etc. External factors include legislation, economy, technology, professional issues, relationships with other organizations, etc. Program characteristics include what is to be evaluated (i.e., evaluand), extant knowledge about the program and evaluation beneficiaries.

Context may be conceptualized variously (Chouinard & Miley, 2016; Coldwell, 2019), oversimplified (e.g., as organizational context to examine drivers and barriers of evaluation) or complex (e.g., as social context to examine the dynamic interactions as people do evaluation).

2.4.2.3 What is evaluated

Two aspects of a program, which are often measured, are efficiency and effectiveness (Bourgeois & Naré, 2015; Kroll, 2015). Efficiency is the ratio between input and output while effectiveness is that between output and outcome. Thus, efficiency concerns the program (i.e., how well the program working) and common variables include economic (e.g., reduced cost), and technical (productivity, cost per unit of service delivered, etc.). On the other hand, effectiveness is about outcomes (i.e., how well the desired outcomes are being achieved) and the variables include cost effectiveness and social and other values. This is consistent with the current paradigm of public sector management, based on public value (section 2.2.4).

2.4.2.4 Evaluation use

Information use is central to evaluation (King & Alkin, 2018). Evaluation information helps to understand how and why outcomes are being achieved and, therefore, fosters learning for improvement. Important dimensions include the source and purpose of the information; stakeholders; and what part of the evaluand to apply to (Alkin & King, 2017). The primary sources are participation in the evaluation process and use of the product (e.g., results). Thus, evaluation use (i.e., effect, outcomes, etc.) may be defined by the impacts of product use and process use (Bourgeois & Naré, 2015; Fleischer & Christie, 2009; Gagnon et al., 2018). The effect may be measured on individual, group, and organizational levels (Alkin & King, 2017; Mark & Henry, 2004). For example, as a consequence of involvement, people or organizations may develop technical knowledge and a better understanding of evaluation.

Information may be used for accountability, learning, and continuous improvement (King & Alkin, 2018). There is no consensus on the types of evaluation use. However, the following are commonly used. Symbolic: i.e., to comply with demands, e.g., accountability to funders;

instrumental: i.e., to use results as is, to support decisions about a program, e.g., to modify; and conceptual, i.e., to improve knowledge about a program (Alkin & King, 2016).

There is disagreement on the extent of evaluation use in the public sector. For example, according to Wholey (2015), evaluation information is often not used. Alkin and King (2016), however, point to a noticeable improvement due to increased research and the heightened demand to establish the "value for money" of programs. This demand is more so in developing countries: most programs are development assistance, and development agencies need evidence of improvement in people's lives (Norgbey, 2016; Phillips & de Wet, 2017).

The growing emphasis on evaluation use to understand context implicates post-positivism, constructivism, and system theory (Vo & Christie, 2015). According to Phillips and de Wet (2017), positivism cannot account for context and is, therefore, inappropriate. Stakeholder involvement can also promote use (Chouinard & Miley, 2016; King & Alkin, 2018).

2.4.2.5 Stakeholder involvement

Stakeholders are those with interest in an evaluation, as participants in the process, users of findings, or both. These include policymakers; program managers, employees, evaluators; development agencies and members of the public.

Stakeholder involvement is characterized by interaction, collaboration, and shared understanding among people (Daigneault, 2014; Fleischer & Christie, 2009). Expertise is necessary for meaningful participation (Fielding, 2013). Stakeholder-centred approaches, therefore, emphasize capacity building (Preskill & Boyle, 2008a; Schwandt, 2015). Some factors to consider are demography, roles, and responsibilities.

Stakeholder involvement presents benefits (Chouinard & Milley, 2018; Daigneault, 2014). Different stakeholder groups can provide diverse perspectives and, thereby, help paint a more complete picture. Furthermore, stakeholders are likely to accept and commit to use information from an evaluation in which they participated. However, there are challenges; e.g., there is no agreement on this concept (Chouinard & Milley, 2018). Stakeholder involvement is central to evaluation (Daigneault, 2014), especially in international development (Chouinard & Miley, 2016).

In the complex landscape of evaluation, people rely on assumptions to make judgements (Rog, 2015). The role of theory is evident— to guide how evaluation is actually done (i.e., practice). There is growing interest in the link between evaluation theory and practice (Chouinard et al., 2017; Leviton, 2015; Rog, 2015)

2.4.2.6 "Theory" in evaluation

In the evaluation literature, the terms theory, approach, model, and framework are ill-defined and often used interchangeably (Rog, 2015). Arbour (2020) describes an evaluation framework as a structure which organizes assumptions, shared values and rules, outcomes, indicators, stakeholders, and how to assess an evaluand. It is prescriptive and context-dependent, and can take many forms (enforceable policy, logic models, explicit (e.g., written), formal, etc.). According to Schwandt (2015), evaluation theory is not a scientific theory but rather a set of ideas, which informs how evaluation is done—"a vague map to facilitate and guide action" (Miller, 2010, p.397) and focuses on how to practise evaluation (Rog, 2015). Figure 2.4 illustrates the three types of theory. Each fulfils a different role (Leeuw & Donaldson, 2015).

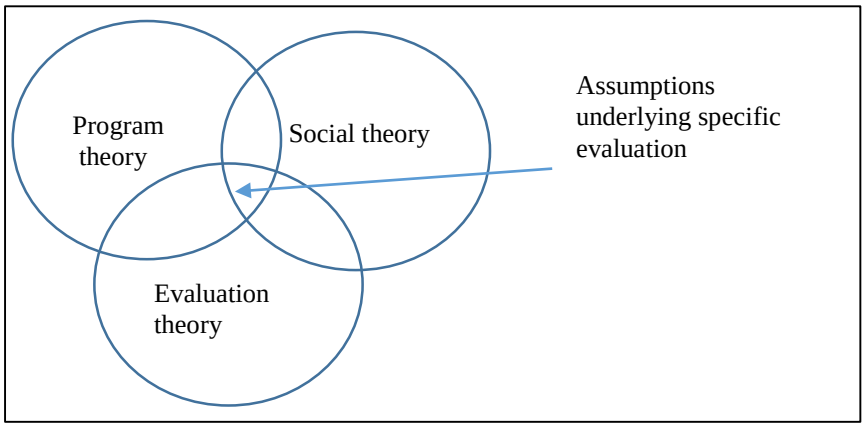


Figure 2.4: Theories used in evaluation

Program theory is a set of assumptions about the program. Social theory explains social phenomena and may be used to design how evaluation should be done. For example, concepts such as decision-making, motivation, etc., may help explain the actions of evaluation stakeholders. Evaluation theory is specific to the field, and prescribes "how to do evaluation in an adequate and efficient way" (Leeuw & Donaldson, 2015, p. 469). This includes decisions about the paradigm, methodology, strategy, and methods.

All three influence how people do evaluation (Rog, 2015). The confluence, as shown in Figure 2.4, is the specific perspective, which informs how to conduct evaluation (Leeuw & Donaldson, 2015). This has been developed into several approaches with idiosyncratic assumptions and practices, three of which are discussed below.

2.4.2.7 Participatory, theory-based and development evaluation

The participatory (also called collaborative, stakeholder-based, etc.) approach is a collection of approaches (e.g., utilization and emancipatory). The common characteristic is stakeholder involvement. Important dimensions include control of decision-making; stakeholder diversity; and level of participation (Chouinard & Milley, 2018; Cousins, 2013).

As explicated above, evaluation of outcomes makes an explanation imperative. Theory-driven evaluation (also called theory-based approach) integrates theories into the evaluation process (Coryn et al., 2011; Dahler-Larsen, 2018) and supports the evaluation of outcomes (Coldwell, 2019; Norgbey, 2016). This approach is well-developed; and is widely used in many continents (Dahler-Larsen, 2018), in other approaches (Mertens & Wilson, 2018), and by development agencies, e.g., World Bank (Coryn et al., 2011).

Development evaluation has its roots in the evaluation of international development assistance in developing countries (Carden & Alkin, 2012; Chouinard & Cousins, 2015). Key features include collaboration between the funders and recipients; change (i.e., improvement in people's lives); and capacity building in the recipient context (Chouinard & Miley, 2016; Norgbey, 2016). Participatory evaluation originated from development evaluation (Chouinard & Cousins, 2015). The demand by development agencies for evidence of development assistance places attention on both evaluation of outcomes as accountability and learning for improvement (Phillips & de Wet, 2017). Stakeholder involvement facilitates knowledge of local context, learning, and use.

The common thread through development, theory-driven, and participatory approaches is the need to understand how and why outcomes are being achieved— for learning and continuous improvement. While the tensions may not be resolved, this research draws on their common features. E-government is considered a development program, therefore, development evaluation is applicable. Furthermore, stakeholder involvement is central to the evaluation of public value (section 2.2.6). Theory-driven approach allows explicit integration of theories (e.g., public value and practice theory) to understand how evaluation is done.

The evaluation logic model above (Figure 2.3) may be extended to explicitly depict the integration of theory, as illustrated in Figure 2.5. Hansen et al. (2013) also adopted such a perspective. For this research, Figure 2.5 forms the basis of how evaluation of e-government outcomes is done. Moreover, the figure shows evaluation involves the assessment and use to effect change.

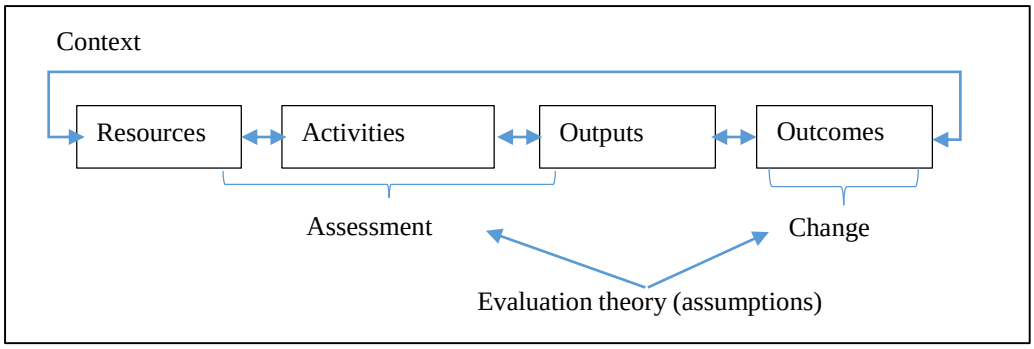


Figure 2.5: Evaluation practice model adapted from (King & Alkin, 2018; Mark & Henry, 2004)

From Figure 2.5, the relationship between evaluation theory and practice is evident. Although distinguishable, the two are intertwined. Chouinard et al. (2017) contend practice is the application of principles from theory. Chelimsky (2013) asserts evaluation practice is based on theory and performed in the real world by people. Smith and Brandon (2008; p. 135) averred theory explains the *why* of evaluation while practice is "the actual doing of evaluation".

As stated in section 2.4.1, practice is polysemic. This pervades evaluation; the term evaluation practice has several meanings. It is, therefore, important to connect evaluation and practice and to discuss the meaning adopted for this research as well as implications thereof.

2.4.3 Evaluation as social practice

There are various conceptualizations of evaluation; examples are professional practice and social practice. The distinguishing factor is what constitutes valid knowledge (Kallermeyn et al., 2015; Patton, 2018). As a profession (i.e., a speciality, discipline, etc.), evaluation signifies a branch of science, i.e., the science of valuing (Patton, 2018). Evaluation is viewed as a systematic endeavour: a rigorous application of scientific methods and tools to provide objective information. Thus, positivism is suitable (Kallermeyn et al., 2015).

By contrast, social practice emphasizes human activity. As Breidahl et al. (2016) observed, evaluation is not entirely rational; e.g., stakeholders, their interests, and values are important. Thus, this perspective lends itself to how people do their daily work. It emphasises the social context and facilitates the use of evaluation information for change (Saunders, 2012; Stern & Saunders, 2015) and is being adopted for evaluation of outcomes (Coldwell, 2019; Ivaldi et al., 2015; Saunders, 2012). Evidently, positivism is inappropriate as it cannot provide adequate explanation to facilitate use (Connelly & Vanderhoven, 2018). Furthermore, the popular strategies (e.g., experiments) may be unethical and impractical in cases where people are involved (Fielding, 2013). Non-experimental strategies (e.g., surveys, case study, focus

groups) are being employed. It is acknowledgement that mixed methods can support understanding, use and learning (Mertens, 2018).

The evaluation literature supports the social practice view. Kallermeyn et al. (2015) defined evaluation practice as the everyday work of doing an evaluation and using the information for change. This is consistent with the meaning of practice as how people do their work (see section 2.4.1) adopted for this research. It foregrounds the "social", situated, and improvement (i.e., change) about practice.

Developments in evaluation practice highlight the significance of psychology, affect, and knowledge. Mark and Henry (2004) identified cognitive, affective, motivational, and behavioural outcomes among individuals and groups as important factors. Christie (2012) argued that evaluation entails issues of human behaviour, and encouraged research to draw on psychology. According to Brandon and Fukunaga (2014), evaluation involves interaction among stakeholders, which underscores affective issues.

People apply knowledge to decisions and actions in specific situations when doing an evaluation. This includes education, experience, values and preferences (e.g., methodology) *apropos* evaluation, domain (e.g., public sector), program (e.g., e-government), stakeholders and their interests, context (e.g., organizational and political), etc. (Christie, 2012).

The foregoing shows that evaluation as a social practice is consistent with practice as conceptualized in section 2.4.1— and further explained in practice theory (Chapter 3). For example, the emphasis on people and what they do, context, resources, knowledge, assumptions, outcomes, and affective issues. As with practice, aspects of evaluation practice (e.g., people and observable activities) are tangible while others (e.g., affect) are not. A study of evaluation practice, therefore, needs to encompass both aspects. Others (e.g., Ryan, 2015; Schwandt, 2018) have similarly conceptualized evaluation as a social practice.

There is an ongoing debate on whether rigour or utility is the appropriate yardstick to measure the quality of evaluation. According to Chen (2015), both are important.

2.4.4 Evaluation quality: factors of influence

The quality of evaluation may be measured by the extent to which the process is rigorous and credible, and the findings can be used (Alkin & King, 2017). The literature is replete with factors which can assure quality. For example, independent evaluation structure; transparency; legitimacy; stakeholder participation; evaluation use; capacity building, and management

support (Chen, 2015; Chouinard & Cousins, 2015; Connelly & Vanderhoven, 2018; de Laat & Williams, 2014; Mayne, 2014).

Independence denotes the absence of interference and bias, and may be achieved through independent structures and individuals. Transparency is about the process, i.e., the formulation of a shared view of the evaluation (e.g., object, purpose, and use); selection of stakeholders with requisite expertise; adoption of appropriate strategies, methods, and tools; etc. Legitimacy, i.e., the credibility of the process and the extent to which results are used, is important in the social practice perspective, where shared understanding defines what is acceptable. Stakeholder participation can increase transparency and legitimacy, but also introduce other issues (e.g., degree of participation, locus of authority in decision-making, relationships, etc.). Furthermore, how outcomes are used can enhance or diminish commitment to future evaluation. Follow-up (to implement recommendations) can promote use and learning.

The factors above are interrelated, e.g., an independent structure may improve transparency, active stakeholder involvement and, thereby, enhance legitimacy and commitment to the use of results. By contrast, misuse of outcomes (e.g., for political witch-hunt by management) can erode legitimacy and stakeholder involvement.

The following may also facilitate evaluation and use (Gagnon et al., 2018; Preskill & Boyle, 2008a, 2008b; Rogers & Gullickson, 2018). Organizational structures, e.g., for communication and teamwork (to support dissemination and access to evaluation reports, and collaborative learning respectively). Recognition and reward systems to encourage stakeholder participation, competitions to create awareness, etc. Evaluation capacity building (frequent evaluation, professional development, evaluators as mentors, etc.) to develop resources (e.g., expertise, processes, and tools). Management support (e.g., evaluation champion) to facilitate resource acquisition and create a conducive environment for evaluation.

As stated above, evaluation in the real world is a complex endeavour. There is a longstanding call for empirical evidence to improve practice.

2.4.5 Research on evaluation practice

Research on evaluation (ROE) is the systematic examination of some aspects of evaluation to gather evidence to guide future evaluation practices (e.g., to help prevent bad practices) and theory development (Christie, 2012; Coryn et al., 2015; Smith, 1993). Seminal efforts to demarcate the ROE field include the work by Henry and Mark (2003) and Mark (2008).

Mark (2008) provides a framework to classify ROE by: what phenomenon (i.e., the aspects of evaluation, e.g., stakeholders, activities, and theories) to investigate and how (i.e., methods, e.g., classification, description, and causality). Although the classification may not be clear-cut, e.g., a study may examine the link between stakeholder involvement and evaluation use (e.g., Daigneault, 2014), the framework provides a structure for ROE and has been widely adopted.

Henry and Mark (2003) and Mark (2008) both identify evaluation practice (i.e., how an evaluation is actually conducted) as an important agenda in ROE. According to Christie (2012), research on evaluation practice (ROEP) investigates behaviour in the real world, the unit of analysis is an aspect of evaluation practice (i.e., how evaluation is done) Thus, the stakeholders are the knowers and an appropriate source of data.

Research on evaluation practice may be conducted in a specific domain (e.g., e-government, non-government organizations (NGOs), etc.). The scope may be wide-ranging. For example, strategies (experiments and non-experiments) and data collection methods employed (Christie & Fleischer, 2010); use of findings (Bourgeois & Naré, 2015; Coryn et al., 2015); evaluation capacity (Gagnon et al., 2018); and models for research on evaluation (Henry & Mark, 2003; Mark, 2008; Mark & Henry, 2004).

Kang et al. (2012) focused on how NGOs conducted evaluation and examined evaluation characteristics, evaluation purpose, dissemination of results, and challenges. To understand how public sector reform evaluation is done, Breidahl et al. (2016) compared two evaluations, according to purpose, planning, results and use. Fierro and Christie (2017) compared perceptions of managers and evaluation contact points on evaluation capacity building and evaluation practice. In general, survey is the strategy of choice in ROEP (Coryn et al., 2015).

This research concerned developing countries and, therefore, a review of the unique evaluation context and practice thereof is relevant.

2.4.6 Evaluation practice in developing countries

Section 2.3 described e-government contexts in developing countries. There is also a difference between the public sector, as an evaluation context, in developing and developed countries. The former is characterized by unique factors, e.g., resource poverty; diverse stakeholder groups, weak institutions vulnerable to political power, and development issues (Ofir & Kumar, 2013).

Most evaluations concern the impact of development on human lives and are undertaken by development agencies, in conjunction with the recipient developing country (Chouinard & Miley, 2016; Chouinard & Milley, 2018). The asymmetric power relationship means donors may impose decisions (e.g., on stakeholders, methods, etc.) on the recipient. Therefore, development evaluation primarily concerns accountability, which privileges quantitative approaches and methodological rigour but disregards the local context (Dighe & Sarode, 2019). This inhibits use, learning, and improvement (Chouinard & Cousins, 2015).

The shortcomings have spawned interest in the theory-driven approach, stakeholder involvement, and qualitative and mixed methods (Armytage, 2011; Chouinard & Cousins, 2015; Dighe & Sarode, 2019; Norgbey, 2016). These support, e.g., collaboration among diverse groups, transparency, access to context and local knowledge, use, and capacity building (e.g., skills developed through training and participation). Evidently, development evaluation may be characterized as a social practice. The shift from positivist, quantitative approaches is appropriate.

The preceding sections have described e-government; public sector reforms and IT use and management; evaluation as social practice; research on evaluation; and evaluation practice in developing countries, in particular. Table 2.5 summarizes the literature. This sets the scene for a discussion of IT and —more specifically— e-government evaluation practice.

Consistent with the foregoing and the research purpose, IT and e-government evaluation may be conceptualized as social practice (i.e., how people do their work). The following is an analysis, according to Figure 2.5: context, resources, activities, output, outcomes, and underlying assumptions.

Table 2.5: Summary of evaluation practice literature

Concept	Summary
Program	An open system; consists of resources, activities, outcomes and feedback
Evaluation	Polysemic; multidisciplinary *For this research:* assessment of an evaluand
Evaluation practice	*For this research*: how evaluation is done and results used For this research: the elements context, resources, activities, outputs, outcomes and theory
What is evaluated	Generally, efficiency and effectiveness; others (social and democratic) neglected
Evaluation context	Environment in which evaluation takes place; factors include organizational, program, stakeholder, political, etc.
Evaluation resources	Human and non-human resources for evaluation
Evaluation activities	The actions performed, e.g., planning, implementation, management response, dissemination of report
Evaluation output	Products of evaluation, e.g., report
Evaluation outcome	The effect, influence, etc. of evaluation process and product Main types include symbolic, instrumental and conceptual
Evaluation theory	Assumptions from the program, social and evaluation theories which inform how to do evaluation
Evaluation stakeholders	Participants, users of product, or both
Evaluation structure	Internal or external structure responsible for evaluation
Evaluation approach	Framework, model, etc.; specific assumptions which prescribe how evaluation *should* be done Participatory evaluation: characterized by stakeholder involvement Theory-driven evaluation: explicitly or implicitly, integrate theory for explanation Development evaluation: concerns development assistance; integrates theory and stakeholder involvement *This research* draws on concepts from the three approaches
Evaluation as practice	Professional practice: rational and objective; predominantly positivism and experimental strategies Social practice: situated in social context, not rational; predominantly non-positivism and non-experimental strategies *This research* views evaluation as social practice
Evaluation quality	About rigour and utility Improved by: Independent structures, transparent processes, stakeholder involvement, evaluation use, capacity building, management support
Research on evaluation practice	Systematic examination of aspects of evaluation, e.g., stakeholders, activities, etc. Paucity of research Survey is widely used
Evaluation practice in developing countries	Characterized by lack of resources, diverse stakeholders, development assistance, etc. A shift from positivist approaches to accommodate theory (for explanation), stakeholder involvement, use, learning and capacity building

2.5 IT outcomes evaluation practice

According to Saunders and Brynjolfsson (2016), IT constitutes about one-third of the total investments in organizations. Thus, it is important to determine the contribution to organizational performance and also support continuous improvement. IT evaluation remains a key IT management issue (Kappelman et al., 2016) and research theme (Palvia et al., 2015, 2017).

How IT is evaluated has evolved over the years, in synchrony with its use and outcomes in organizations. Thus, an understanding is important to draw implications for future practice.

Others (e.g., Chaysin et al., 2016; DeLone & McLean, 2016; Pather, 2017) trace such developments. Early applications involved routine operations (e.g., payroll, etc.), progressed through production and analysis of information for planning and decision-making to enabling the transformation of business processes.

The Internet-enabled the interconnection of processes in different parts of organizations and with external stakeholders. Consequently, IT outcomes extend beyond organizations (e.g., benefits to employees, improved management and processes, etc.) to include improved service and user satisfaction. The developments in use, locus of control and outcomes have changed the role of IT management, with implications for IT outcomes evaluation. As discussed in section 2.4.2.1, evaluation may be classified according to stages of implementation. This pervades IT evaluation.

2.5.1 Types of IT evaluation

Pre-, during, and post-implementation evaluation may be identified (Pather, 2017). The purpose and, therefore, what can be evaluated may differ according to stage. Pre-implementation evaluation assesses investment decisions on a proposed system. During implementation is about an on-going system, to improve the processes. Post-implementation may assess any of the following: output, use, and outcomes. Examples of output evaluation include IT application (e.g., evaluation of usability) and services delivered (e.g., evaluation of service maturity). In general, adoption (or acceptance) evaluation focuses on the actual usage of services.

IT outcomes evaluation is about IT in use. According to Coombs (2015), there is a dearth of literature on post-implementation compared to pre-implementation. Pekkola and Päivärinta (2019) also noted the paucity of IT outcomes evaluation. This research concerned post-implementation assessment; was limited to how evaluation of outcomes is performed; and elided development of measurement frameworks, evaluation of product (e.g., websites) and services as well as their use *per se*.

2.5.2 IT outcomes evaluation: challenges

Organizations often fail to evaluate IT outcomes at all; others use inappropriate methods and measurements (Berghout & Powell, 2010; Pather, 2017; Pekkola & Päivärinta, 2019; Schryen, 2013). The literature is replete with challenges. The outcomes are multifaceted; difficult to identify; and, thus, cannot be completely measured by available methods (Berghout & Powell, 2010; Chaysin et al., 2016). Evaluation of non-financial outcomes remains a perennial

challenge (Chaysin et al., 2016; Santarek & Obluska, 2012). Moreover, methods become obsolete and inappropriate over time because IT (and outcomes thereof) is ever-changing (Chaysin et al., 2016; Nguyen et al., 2015). Measurements are often ad-hoc (Frisk et al., 2015) and little is known about the extent to which IT evaluation is used (Song & Letch, 2014). These challenges perpetuate the IT paradox (Coombs, 2015; Guerreiro, 2016; Hajli & Sims, 2015). However, how IT evaluation is done has received scant attention (Pekkola & Päivärinta, 2019).

2.5.3 Research on IT evaluation practice

As discussed in section 2.4.1, research on practice in organizations investigates how people perform their work. Research on evaluation, the various aspects (including evaluation practice), and benefits have also been discussed and, thus, need no further exposition. While there is thriving research on practices in other fields, this is not the case in IT (Pekkola & Päivärinta, 2019). The following is an analysis of some relevant literature.

Goldkuhl and Lagsten (2012) conceptualized IT evaluation as a practice, research thereof as the study of how evaluation is done, and the unit of analysis as an aspect of IT evaluation—and list evaluation strategies and methods, activities, stakeholders, outcomes as examples. The authors identified two main research contributions, *viz*., theoretical (e.g., richer insight into how IT evaluation is performed) and practical (e.g., development of guidelines, frameworks, etc., to improve practice).

According to Song and Letch (2012, p. 279), research on IT evaluation practice examines "actual evaluation practices and report important finding(s)". This is consistent with research on evaluation practice, as elaborated herein. However, it is acknowledged that rigour is over-emphasized in IS research while practice issues have received scant attention (Pekkola & Päivärinta, 2019; Song & Letch, 2012).

There are several ways to examine IT evaluation practices. Nguyen et al. (2015) examined the theoretical perspective, stakeholder perspective, the unit of analysis, strategies, and methods. The Content, Context and Process (CCP) framework (see section 2.5.9) analyses the what, why, how, when, and who of evaluation (Song & Letch, 2012, 2014). This research applies the practice theory to examine evaluation context, resources, activities, output and outcomes, and theory (see Figure 2.5). The similarity among the three approaches is evident. As the analysis below shows, this research covers the elements of CCP (i.e., what is evaluated, purpose, activities, and stakeholders, in post-implementation evaluation)— and more (e.g., affective issues).

2.5.4 Context

From section 2.4.2.2, program, organization (internal and external) and stakeholder factors are important. IT is generally viewed as a socio-technical artefact, i.e., as technology embedded in technical, social, and other interconnected systems (McCoy, 2017). The socio-technical view of IT evaluation has long been recognized. For example, Stockdale et al. (2008) located IT evaluation in a social, technical and political context; according to Irani and Love (2008), IT evaluation is influenced by behaviour and biases and, therefore, cannot be viewed as completely rational but as part of social activity. This view is consistent with the social practice perspective of evaluation (see section 2.4.3). The socio-technical perspective to the study of technology is also established in practice theory. Evidently, context affects (and is affected by) what to measure, the purpose (i.e., why), resources, processes, effect, and underlying assumptions.

2.5.5 What is measured

The contribution of IT to organizational performance is referred to as IT value, contribution, performance, benefit, net benefit, effectiveness, outcome, impact, etc. (DeLone & McLean, 2016; Nguyen et al., 2015; Pekkola & Päivärinta, 2019; Zhang & Gable, 2014). The differences in nomenclature is indicative of the dissensus on the definition and operationalization of the concept (Schryen, 2013). This is attributable to the dynamic nature and complex uses of IT. As explained in section 2.4.2.2, consistent with the evaluation literature, outcome is preferable.

Concepts such as "value for money" exemplify the multi-dimensional nature of IT outcomes. Various constructs have been specified and studied, e.g., productivity, efficiency, and effectiveness, and user satisfaction (Santarek & Obluska, 2012); improved decision-making, increased productivity, reduced cost, increased profit, etc. (DeLone & McLean, 2016). According to DeLone and McLean, the constructs depend on— and may differ according to— the program, organizational context, etc.

2.5.6 Resources

Resources include people, budget, tools, management support, etc. Complementary factors can leverage IT resources to improve outcomes generated therefrom. For example, IT management, *ceteris paribus*, can improve IT outcomes (Dahlberg & Kivijärvi, 2018; Pather, 2017). IT evaluation may be used as a tool to foster organizational learning to improve skills, use of resources, processes, etc., and, consequently, outcomes. Different stakeholder groups (e.g., funders, developers, and users) have different expertise as well as perspectives of IT outcomes,

which can be leveraged to improve how evaluation is done and results used (Guerreiro, 2016; Nguyen et al., 2015). Outsourcing may be an important strategy (Guerreiro, 2016). Appropriate evaluation activities depend on the purpose, what to evaluate and how results may be used.

2.5.7 Activities

From section 2.4.2.2, this concern processes; thus, what is done (i.e., planning, implementation, and dissemination of results), by whom and how, are important. Stakeholder inclusion, benefits, and challenges have been discussed (section 2.4.2.5). An evaluation may be descriptive, explanatory, or predictive (Zhang & Gable, 2014). There may be several strategies, e.g., experiment, survey, case study, each with its implications for data collection and analysis methods; units of analysis, e.g., individuals, groups, organizations, and governments; and participants, e.g., management, employees, users (Nguyen et al., 2015; Song & Letch, 2012).

A plethora of measurement frameworks are available; what is appropriate depends on the outcomes to evaluate (Chaysin et al., 2016; DeLone & McLean, 2016; Nguyen et al., 2015). For example, objective measurements may suit tangible financial outcomes (e.g., operational cost efficiency) while subjective, qualitative measurements may be appropriate for intangible outcomes (e.g., changes to people's opinions and attitudes, user satisfaction, etc.). The former is prevalent because measurement instruments are readily available and easier to use (Berghout & Powell, 2010; Chaysin et al., 2016). Evaluation of non-financial outcomes is relatively ignored (Chaysin et al., 2016; Santarek & Obluska, 2012).

Song and Letch (2012) found: top management and IS and finance departments often participated; effectiveness constructs, financial and economic measurements, and survey were predominant; and scant literature on dissemination, although access to output (e.g., evaluation reports) is important to use.

Section 2.4.2 showed theory, non-experimental strategies, and mixed methods support tangible and non-tangible outcomes, understanding, learning, and use. There is growing adoption of IT evaluation approaches which emphasize context (Song & Letch, 2012).

2.5.8 Outcomes

This denotes use and effect thereof, is linked to evaluation purpose and may be classified into: symbolic, i.e., to comply with demands, e.g., accountability by funders; instrumental, i.e., to use results as is, to support decisions; and conceptual, i.e., to improve knowledge (Song & Letch, 2014). This is consistent with section 2.4.2.4.

Section 2.5.7 showed that purpose can suggest an applicable evaluation strategy and corresponding methods, e.g., objective strategies may suit accountability and decision-making while subjective may support understanding and learning.

Frameworks and research on use are scant (Song & Letch, 2014). Coombs (2015) observed results of post-implementation evaluation are not used. In a rare example, Song and Letch (2014) viewed IT evaluation use as an outcome, extended the CCP framework to investigate four IS projects and found evidence of symbolic use. According to Nguyen et al. (2015), theory is important in IT evaluation research.

2.5.9 Theory

Theory can provide explanations and facilitate improvement (Guerreiro, 2016) and, therefore, is fundamental to IT evaluation. This is consistent with section 2.4.2.6. However, there is no universally applicable IT evaluation theory (Guerreiro, 2016; Schryen, 2013); developments parallel uses and outcomes.

Early attempts examined productivity, e.g., cost-benefit analysis (Coombs, 2015; Frisk et al., 2015; Guerreiro, 2016). Other theories (organization, e.g., agency; socio-politico-technical, e.g., embeddedness, stakeholder, and resource-based view (RBV)) have emerged to accommodate intangible outcomes (Frisk et al., 2015; Guerreiro, 2016).

The CCP framework (based on the socio-technical view) is commonly used (e.g., Song & Letch, 2012, 2014; Stockdale et al., 2008; Stockdale & Standing, 2006). The dimensions are content, context, and process. Content concerns *what* to evaluate and encompasses tangible and intangible outcomes. Context addresses the *why* (i.e., purpose— linked to the type of use) and *who* (i.e., stakeholders). Process is about *how* (e.g., methods) and *when* (i.e., the stage of the program: pre-, during, and post- implementation). The components are interrelated, e.g., what to evaluate may determine why, who and how. Complementary factors can improve IT outcomes (section 2.5.6) and, therefore, are important to IT evaluation theory. Complementarity forms part of socio-technical perspective (Töhönen et al., 2015).

Variance, process, and system models are also used in IT outcomes evaluation (Burton-Jones et al., 2015; DeLone & McLean, 2016; Guerreiro, 2016). A variance model identifies factors and relationships, and has spawned studies which examine success factors (or otherwise) that enable (or inhibit) IT outcomes. All three models may be combined, e.g., the IS success model (DeLone & McLean, 2016).

IT outcomes evaluation is often ad-hoc, not planned for use (Song & Letch, 2012) and measurements are not based on theory (Frisk et al., 2015). Table 2.6 summarizes some relevant research.

Table 2.6: Relevant extant research on IT evaluation

Source	Research focus	Description of study
Goldkuhl and Lagsten (2012)	Classification of IT evaluation	Reviewed the literature to classify IT evaluation; identified IT evaluation practice as a key research theme
Chaysin et al. (2016)	Financial methods of evaluation	Reviewed the literature on evaluation of financial outcomes in IT and identified some of the challenges
Pather (2017)	Classification of IT evaluation	Reviewed the literature to classify IT evaluation
Song and Letch (2012)	IT evaluation practice	Reviewed how IT evaluation is done based on the CCP framework, according to why, what, how, when and who
Nguyen et al. (2015)	IT evaluation practices (i.e., how IT evaluation is done in practice)	Reviewed IS success literature, according to methodology, perspective (i.e., evaluation stakeholder); unit of analysis and data collection and analysis
Schryen (2013)	Analysis of IT evaluation literature	Reviewed the literature to propose an IT outcomes evaluation model; shows IT and non-IT resources, context and lag time influence processes which, in turn, influence organizational outcomes
Töhönen et al (2015)	Analysis of IT evaluation from a system perspective	Reviewed the literature to identify main features of IT outcomes and a system to conceptualize IT evaluation
DeLone and McLean (2016)	Framework for IT evaluation	Extended the dimensions of the DeLone and McLean IS success framework; net benefits (i.e., internal and external outcomes) as a key dimension
Dahlberg and Kivijärvi (2018)	Complementarity	Examined the influence of IT management on IT outcomes
Song and Letch (2014)	IT evaluation use	Extended the CCP framework to investigate IT evaluation use in organizations; considered outcomes process and product; found symbolic use

The preceding sections have analysed IT outcomes evaluation practice by context, resources, activities, output, outcomes (i.e., effect), and theory. Table 2.7 is a summary of the literature.

The difference between the public and private sectors (highlighted in section 2.2.2) permeates the literature. For example, the reference to IT outcomes in the private and public sectors as IT business value and IT public value lends credence. Sections 2.4 and 2.5 also outlined evaluation practice and IT outcomes evaluation practice. These have implications for how e-government outcomes evaluation is done.

Table 2.7: Summary of IT outcomes evaluation literature review

Concept	Summary
IT outcome	Value, benefit, etc.; the contribution of IT to organizational performance
	Paucity of research on post-implementation, particularly IT outcomes
IT outcomes evaluation practice	How IT outcomes evaluation is done
	Important and enduring research theme
	CCP (i.e., content, context and process) is the dominant framework for analysis
	For this research, analyzed by context, resources, activities, output outcomes and theory
Context	Socio-technical
	IT outcomes evaluation practice is situated
	Different contexts (e.g., public-private sectors, developed-developing countries, etc.) means different practices
What is measured	Ill-defined and multi-dimensional (i.e., quantitative and qualitative, etc.)
	Tangible often measured, compared to intangible
	Correspond to epochs in IT uses; each era adds to existing values e.g., pre-Internet era: organizational efficiency; post-Internet era adds customer satisfaction
	Relevant constructs depend on the type of use, context, etc.
Resources	Human and non-human
	Complementary resources important
	Outsourcing importance
Activities	Processes (e.g., design, implementation and dissemination of output)
	Stakeholders and their roles are important
	Unit of analysis: individual, group, organizational, government (e.g., local, national), etc.
	Evaluation strategies include experiments, surveys, case studies, etc.
	Suitable strategy and methods depend on outcomes to measure; context; type of use, etc.
	Objective measurement and survey are predominant
	Shift towards qualitative, mixed methods to support use, learning and improvement
	Not much is known about the dissemination of output
Outcomes	Effect, influence, use, etc. of IT evaluation on stakeholders
	Classified as symbolic, instrumental, and conceptual uses
	Type of use determines how assessment is done, e.g., subjective strategies support learning
	Little is known about the extent to which IT evaluation is used
Theory	Theories have evolved from economic to address social, technical and political perspectives
	A shift from rational towards the social practice view (see 2.4.3)
	CCP framework widely used
	Resource-based view emphasizes complementarity
	Variance, process and system models also used
	Variance dominant; identifies success factors (i.e., enablers) and inhibitors
	Lack of universal IS evaluation theory;
	Ad-hoc evaluation, no theoretical basis and little plan for use

2.6 E-government outcomes evaluation practice

As aforementioned, the following working definitions apply. Practice refers to how work is done. Evaluation is the assessment of an evaluand to determine its value. Evaluation outcome denotes the effect of use. E-government is the use of IT in the public sector. Therefore, e-government outcomes evaluation practice refers to how the assessment of outcomes of IT in the public sector is done.

2.6.1 E-government outcomes evaluation: challenges

According to the UN (2016, 2018, 2020), all member states have implemented e-government. However, there is a higher failure rate— about 70 to 80 percent worldwide (Sterrenberg, 2017). According to Bannister and Grönlund (2017), the failure is under-researched. Although e-

government outcomes evaluation can support learning for improvement, this is often ignored—and where undertaken, it is often ad-hoc, focuses on private sector values and the results are hardly used (Assar & Boughzala, 2016; Cha & Park, 2018; Kasimin, Aman, & Noor, 2013). It is, therefore, important to understand how e-government outcomes evaluation is done (Sterrenberg, 2017).

E-government outcomes evaluation is an important research theme (Pekkola & Päivärinta, 2019; Sterrenberg, 2017). However, this has received scant attention because the focus is mainly on IT evaluation in the private sector (Frisk et al., 2015). According to Pang et al. (2014), applicable theories are sparse; Weerakkody et al. (2015) observed conceptual and descriptive research are predominant.

Section 2.5 showed IT outcomes evaluation can be examined in several ways; this also applies to how e-government outcomes evaluation is done. Assar and Boughzala (2016) analyzed by what is evaluated and how. This research applied context, resources, activities, outcomes, and theory (see Figure 2.5).

2.6.2 Context

Section 2.4.2.2 discussed the contextual factors which influence evaluation (i.e., internal, external, and program). E-government evaluation occurs in a complex, dynamic social, technological and political environment (Sterrenberg, 2017). The public sector has been discussed in detail in section 2.2.2 and needs no further elaboration. Maximization of public value (see section 2.2.4 and 2.2.5) and diverse stakeholders and their interests bear on evaluation practice. There are many systems, of which evaluation of e-government outcomes is only a part. Context may be examined as local (i.e., municipal, state, county, provincial, etc.) or national government (Zahran et al., 2015).

Various aspects of context have been studied; e.g., the link between decision-making and outcomes based on public value (Sundberg & Larsson, 2017); public sector values, stakeholder groups, their relationships and perspectives of outcomes in national and local public organizations (Rose et al., 2018); benefits, enablers, and inhibitors in local government (Coombs, 2015); comparison of the capacity of national governments to use IT for service delivery (UN, 2016, 2018) and developing countries (Deng et al., 2018). What to evaluate—and how— depends on context.

2.6.3 What is measured

Section 2.2.8 showed IT outcomes correspond to epochs of IT use in the public sector. Section 2.2.4 outlined the shift to public value. The conceptual muddle permeates e-government outcomes evaluation. There is little consensus on what to evaluate (Frost & Lal, 2019; Scott et al., 2016).

Seminal endeavours to measure public value include Kearns (2004) and eGep (2006). Kearns explicated value from service delivery (e.g., improved quality and access, and reduced cost); outcomes (e.g., improved health and education, and poverty reduction), and trust (in public organizations through legitimate processes, professionalism, etc.). eGep identified the following values: organizational (i.e., efficiency, e.g., reduced cost and improved internal operation which accrue to public organizations), user (i.e., effective service delivery to, and access by, the public) and political (i.e., democratic, e.g., participation by citizens in democratic processes; accountability and transparency of public organizations, etc.).

Table 2.8 shows some of the first-order constructs for measuring public value in the e-government evaluation literature. Organizational efficiency, effective service delivery, and democratic values are common. This is consistent with the outcomes of e-government, as exemplified by the working definition adopted herein (section 2.2.1), namely, "the use of IT in the public sector to achieve internal efficiency; effective service delivery and interaction between public sector organizations and citizens, businesses; and improved accountability and transparency". Empirical studies (e.g., Al Rawahi et al., 2016; Scott et al., 2016; Sundberg & Larsson, 2017) adopted such constructs. In practice, evaluation involves resources.

2.6.4 Resources

Sections 2.4.2.2 and 2.5.6 show that resources include adequate expertise, budget, tools, time, and complementary resources. Various governments also developed policies, frameworks, and tools to guide practice (Deng et al., 2018), e.g., French Merava; German WiBe, and Danish e-Government Signpost. Research on e-government evaluation often draws on such frameworks (eGep, 2006; Stanimirovic & Vintar, 2013; Zahran et al., 2015). Public value emphasizes adequate resources (see the strategic triangle, section 2.2.4.1); top management plays an important role in the acquisition of the necessary resources to perform an evaluation (Pang et al., 2014).

Table 2.8: Dimensions of e-government outcomes

Source	Public value dimensions
Kearns (2004)	Service delivery
	Outcomes
	Trust
eGep (2006)	Efficiency
	Effectiveness
	Democratic
Jansen (2012)	Efficiency
	Effectiveness
	Democracy
	Innovation
OECD (2014)	Public service legitimacy and innovation
	Justice and fairness
	Fairness and efficiency in distribution
	Public institutions which reflect public value
	Efficiency and effectiveness in production
Rose et al. (2018)	Professionalism of the public sector
	Operational efficiency
	Service delivery to and access by the public
	Engagement
Pang et al. (2014)	Efficiency
	Effectiveness
	Improved democracy
	Legal
Scott et al. (2016)	Efficiency
	Effectiveness
	Improved democracy
Deng et al. (2018)	Service quality
	Effectiveness
	Social outcomes

2.6.5 Activities

From sections 2.4.2 and 2.5.5, important factors include what is done (i.e., design, implementation, and dissemination), by whom (i.e., stakeholders, their roles and extent of involvement), and how (i.e., strategies and methods). Assar and Boughzala (2016) analyzed by strategy (e.g., survey) and methods (i.e., quantitative, qualitative, or mixed).

There are several national, regional, and international e-government evaluation models— sets of theories, strategies, instruments, and methods to guide practice (Stanimirovic & Vintar, 2013; Zahran et al., 2015).

Section 2.6.3 showed a shift to public value. However, efficiency and effectiveness are commonly measured (Cha & Park, 2018; Sterrenberg, 2017). Hébert (2013) observed that positivism is dominant, although it cannot explain the e-government evaluation context. This is consistent with evaluation practice and IT evaluation practice (see sections 2.4.3 and 2.5.7). Thus, non-experimental strategies (surveys, case studies, user panels, etc.) and mixed methods may be appropriate. Suitable data sources include government statistics and documents (e.g.,

policies) and internal and external e-government stakeholders. Hébert also observed that self-reporting may not accurately represent experiences.

Dissemination of reports to stakeholders improves accountability and transparency; and enables access and evaluation use. However, there is a dearth of literature on how this is done. Some national policies and legislation (e.g., South African evaluation framework (DPME, 2011)) stipulate what to report, how and to whom. Heeks (2008) outlined strategies for preparing and disseminating e-government evaluation reports. UN (2020) found many countries publish results of citizen satisfaction evaluation online. Participation begs the question: what is the effect of evaluation on the stakeholder?

2.6.6 Outcomes

Sections 2.4.2.4 and section 2.5.8 discussed the effect of evaluation on individuals, groups, and organizations (i.e., symbolic, instrumental, and conceptual), and also showed the rate of use in the public sector is mixed while frameworks and research on IT evaluation use are scant. Not much is known about e-government outcomes evaluation use. According to Kasimin et al. (2013), e-government outcomes evaluation is ad-hoc and rarely used. Heeks (2008) noted governments seldom use e-government benchmarking results.

However, efforts (e.g., eGep, 2006; OECD, 2005, 2009, 2013, 2016) bear testament that OECD member countries regularly review e-government and use the results to improve practices. Also, many OECD countries have e-government evaluation policies and tools (see section 2.6.4). Kasimin et al. (2013) studied the use of results for organizational learning. E-government outcomes evaluation involves complex decisions and can benefit from theory (i.e., general assumptions) to guide practice.

2.6.7 Theory

As with evaluation (section 2.4.2.6), e-government evaluation literature refers to theory as a model, approach, framework, etc. Developments in theory correspond to public sector reforms and IT uses and outcomes. The theories have evolved from economics (e.g., cost reduction and productivity) to accommodate management uses (e.g., quality of information for decision-making) and transformation of business processes (e.g., integration of activities of organizations). The advent of e-government has spawned assumptions about external outcomes (Cha & Park, 2018; Sterrenberg, 2017; Sundberg & Larsson, 2017).

Stakeholder theory has been applied to evaluate stakeholder interactions and outcomes (e.g., Rose et al., 2018). The relationships are government to government (G2G); government to

public employees (G2E), government to citizens (G2C), and government to businesses (G2B). The focus of G2G and G2E is internal while G2C and G2B is external.

Public value, good governance, scorecards, etc., are also in use. Section 2.2.6 showed public value is appropriate and prevalent; and has been applied to e-government outcomes evaluation. The central concepts (i.e., public value, public values, public service values, and combinations thereof (sections 2.2.4 to 2.2.7) have been adopted for various studies (e.g., Al Rawahi et al., 2016; Bannister & Connolly, 2014; Deng et al., 2018; Pang et al., 2014; Rose et al., 2018; Savoldelli et al., 2014; Sterrenberg, 2017; Sundberg & Larsson, 2017).

The public value lens, however, is not without challenges: conceptual disarray, lack of practical guidelines; the paucity of empirical research, etc., have been highlighted in section 2.2.7. Some studies have combined public value with other theories; e.g., with stakeholder theory (Rose et al., 2018), service ecosystem (Sterrenberg, 2017), and decision theory (Sundberg & Larsson, 2017).

Section 2.3 discussed good governance. Nag (2018) identified the following dimensions: participation, rule of law, equity and inclusiveness, transparency, responsiveness, consensus and legitimacy, effectiveness and efficiency, and accountability. Kettani and Moulin (2014) operationalized as transparency, efficiency and effectiveness, responsiveness, equity, rule of law, and accountability. The similarities with the dimensions of public value (see section 2.6.3) are evident— both are underpinned by public values (Bannister & Connolly, 2014).

Variance, process and system models, and combinations thereof (see section 2.5.9) may be used to: identify success (or failure) factors, understand how IT achieves outcomes, and examine interactions in a context respectively (Frost & Lal, 2019; Pang et al., 2014).

Notwithstanding the importance of theory, there is no e-government evaluation theory (Sterrenberg, 2017). According to Pang et al. (2014), theory on e-government outcomes evaluation is underdeveloped.

As explained above, the unique context means evaluation practice in developing countries may be different from developing countries. This section analyzed EOEP and the next connects to the unique context of developing countries.

2.6.8 E-government evaluation practice in developing countries

Sections 2.3 and 2.4.6 provided an overview of the public sector in developing countries as the context for e-government and evaluation practice respectively. The unique characteristics

include lack of resources, diverse stakeholders, development assistance, and a lower capacity to use e-government to deliver service— a generally higher failure rate and opportunity cost. It is, therefore, important to understand how e-government outcomes evaluation in developing countries is done— to improve theory and practice.

E-government outcomes evaluation in developing countries is an important research theme (Hiziroglu et al., 2017; Wahid, 2013). However, the literature is underdeveloped. Joseph (2013) observed scant research on e-government in general while Otieno and Omwenga (2015) noted a paucity of research on e-government outcomes. Nonetheless, some studies relevant to this research are available.

Aman and Kasimin (2011) investigated evaluation use and found conceptual use, i.e., for learning. Kasimin et al. (2013) examined evaluation practice and developed a framework to support learning. Malaysia was the context of both studies. OECD (2013) examined practices in Egypt, the outcomes measured included improved decision-making, responsiveness, engagement and participation, and internal processes. Córdoba (2014) developed a conceptual framework to investigate stakeholder involvement in Columbia. Deng et al. (2018) developed and validated a framework for e-government outcomes evaluation in Sri Lanka. Other studies (e.g., Kamau & Wausi, 2015; Otieno & Omwenga, 2015) also developed frameworks. Hiziroglu et al. (2017) examined outcomes from the perspectives of external stakeholders.

From section 2.3, development (and therefore, development evaluation) is central to the e-government outcomes evaluation literature. Concepts, e.g., improving human lives, stakeholder participation, learning for improvement, capacity building, etc., are relevant. Bhatia et al. (2009) operationalized poverty reduction in a survey, Kettani and Moulin (2014) examined the effect of e-government on good governance, and Aladwani (2016) and Nam (2018) investigated corruption and e-government.

Sections 2.4.2.6, 2.5.9, and 2.6.7 showed theory is important to IT and e-government evaluation. Wahid (2013) encouraged the use of theory to increase understanding of how e-government evaluation is done in developing countries. However, much of the research is ad-hoc, without reference to theory or context, and identifies potential causes of success and failure, not measure outcomes (Palvia & Baqir, 2015).

Some of the theories which have been adopted are public value (Deng et al., 2018; Hiziroglu et al., 2017), governance (Kettani & Moulin, 2014; Suhardi et al., 2015), and design-gap (Palvia & Baqir, 2015). Sections 2.3, 2.6.3, and 2.6.7 showed the similarity between public value and

good governance and their applicability to the development context. Palvia and Baqir (2015) adjured the adoption of a widely used framework, in the absence of a grand theory.

Section 2.2.4 showed public value is used worldwide and considered an adequate lens for investigating public sector performance. Furthermore, there on-going effort to develop the literature. Consequently, public value was adopted for this research.

Public value has found scant application in e-government outcomes evaluation in developing countries. The literature mainly focuses on developed countries (Kamau & Wausi, 2015) and available frameworks are not appropriate for developing countries (Deng et al., 2018). This has spawned efforts to develop the literature on public value for e-government outcomes evaluation in developing countries (Hiziroglu et al., 2017)— and as evidenced by the foregoing studies. Nevertheless, how e-government outcomes evaluation is done is underdeveloped. It is this noticeable void that this research addressed by examining the practice in South Africa. Other studies (e.g., Deng et al., 2018) have also examined specific cases to contribute to the literature.

2.7 South Africa as context

South Africa has undergone reforms underpinned by IT, is among the first countries to promulgate e-government policy, ranks higher among developing countries in recent international comparative surveys on e-government and some programs (e.g., e-filing; e-Natis and e-Justice) are considered successful. Furthermore, evaluation constitutes an important component of public management. The following describes the context for this research.

2.7.1. Public sector reform

South Africa embarked on a democratic transformation in 1994. The Constitution (DOJ, 1996) mandates, *inter alia*, performance evaluation and relevant institutions and specifies the roles and responsibilities of stakeholders. Demands for accountability and transparency by donors, in addition to the constitutional imperatives, have heightened the importance of evaluation (Abrahams, 2014; DPSA, 1997; Podems et al., 2014)

In 1994, the new government inherited an apartheid bureaucracy, which needed transformation to serve all citizens. Public sector reform, underpinned by NPM, was introduced to improve efficiency (Cameron, 2015). New structures and systems were established, e.g., a three-tier national, provincial and local government. The Department of Public Service and Administration (DPSA) is responsible for national and provincial government while local

government is autonomous, under the Department of Cooperative Governance and Traditional Affairs.

Chapter 10, section 195(1) of the Constitution (DOJ, 1996), promotes a development-oriented public service and stipulates guiding values and principles. For example, professional ethics; efficient, economic and effective use of resources; impartial, fair, and equitable provision of services; responsiveness to people's needs; public participation in policy-making; accountable public administration; transparency; and human resource management. These exemplify public value tenets (Cameron, 2015; Naidoo, 2015). John Benington, a leading authority in the field, was an advisor at the time (Benington & Moore, 2010).

The Department of Public Service and Administration (DPSA) developed a policy framework to guide the reform (DPSA, 1997). At the core is service delivery, achievable through internal management and service delivery reforms. Aptly entitled *Batho Pele* (i.e., People First), the framework is based on eight principles, namely: consultation with the public about service delivery; service standards by which the public can assess services; equal access to services; courtesy to the public; provision of complete, accurate, and up-to-date information on services; openness and transparency, including reporting performance; redressing public dissatisfaction; and value for money in service delivery. The public sector is enjoined to apply the principles to craft policies to guide service delivery.

2.7.2 Evaluation practice

Reviews (Abrahams, 2014; Phillips et al., 2014; Podems et al., 2014) showed that evaluation in the public sector commenced about 1995. In addition to public sector reform, development initiatives (especially, the Millennium Development Goals (MDG) and, subsequently Sustainable Development Goals (SDG) have increased evaluation efforts.

In 2005, Government-wide Monitoring and Evaluation (GWME) was adopted. This plan emphasised evaluation across the public sector and autonomous evaluation units in public organizations. IT was, therefore, considered a key component of the system— to support information management. The challenges introduced by GWME include disparate reporting systems (Phillips et al., 2014).

The Department of Performance, Monitoring and Evaluation, now Department of Planning, Monitoring and Evaluation) was established in 2010 to oversee evaluation in the public sector and across all tiers of government (Goldman et al., 2015; Phillips et al., 2014). DPME

spearheaded efforts to develop a national evaluation system, including a framework and capacity.

2.7.2.1 National evaluation strategy

The National Evaluation Policy Framework (DPME, 2011) prescribes how evaluation of programs, should be performed and the results used to support national development. The underlying principles include ethics, rigour, utilization, transparency and accountability, participation, and learning. The strategy addresses issues such as evaluation quality, capacity development, purpose, stakeholders and their roles and responsibilities, processes, and use. It also underscores outcomes evaluation, public access to reports, and evaluation of the evaluation process itself, i.e., how evaluation is performed.

The strategy supports use for decision making, planning and resource allocation, and public accountability. The evaluation purposes outlined include: learning to improve performance, accountability to justify investments, decision-making about programs, and development of knowledge to benefit other programs. These are consistent with the types of evaluation use in section 2.4.2.4. Efficiency and effectiveness are generally assessed; there is little emphasis on democratic outcomes.

Rigour of process and use of results are also emphasized, consistent with section 2.4.5. Contributory factors outlined include stakeholder involvement, relevance, timeliness, legitimacy (e.g., inclusive process), and independence. An evaluation may be initiated by a public organization or demanded by another and undertaken by internal, external, or mixed evaluation teams.

The framework prescribes an evaluation process: design, implementation, development of reports and management response; dissemination of reports (including online publication) and follow-up (i.e., to implement improvement plan). The policy states: DPME "will ensure that evaluations are carried out to measure the impact of evaluation itself" (DPME, 2011)— a clear indication of research on evaluation. The framework has been adopted to evaluate various initiatives (Goldman et al., 2015).

2.7.2.2 Evaluation practice, opportunities and challenges

The DPME compiles the National Evaluation Plan, i.e., national priority programs to be evaluated. This is updated annually to show completed, lessons learnt, plans for improvement, etc. DPME (2011), thus far, has conducted evaluations of all types, including outcomes.

Some of the evaluation challenges are lack of expertise, limited evaluation capacity, poor information management, conflicts between the various evaluation structures (e.g., national, provincial, and local), focus on activities rather than outcomes, lack of data and indicators to evaluate outcomes, and evaluation for compliance, rather than learning and improvement (Abrahams, 2014; Goldman et al., 2015; Phillips et al., 2014). Madzivhandila et al. (2010) observed a paucity of empirical research on evaluation practices.

Notwithstanding the challenges, some aspects of evaluation are developed, e.g., requisite structures (including evaluation units in public organizations), a "whole of government approach" (which links resources, activities, output, and outcomes), political support, leadership at DPME, demand for evaluation and evidence of use (Abrahams, 2014; Phillips et al., 2014). Several national and provincial departments are conducting evaluations (Phillips et al., 2014). To develop capacity and competency, DPME has, e.g., built partnerships with international agencies and countries of similar context (DPME, 2011). As discussed above, e-government underpins the *Batho Pele* inspired public sector reform.

2.7.3 E-government and evaluation practice

E-government was adopted in 1995 to reform the public sector— to enable equal access to public service and inclusive economic participation. The Information Society was adopted as part of an overarching strategy to promote a development agenda to reduce poverty and redress inequality (DTPS, 2016). This emphasizes the role of e-government for development. DTPS (2017) provides a synopsis of e-government.

2.7.3.1 E-government strategy

South Africa's e-government policy (DPSA, 2001) was among the first in the world (United Kingdom's e-Government Framework was in 2000[1]). Based on the Batho Pele principles, it focused on creating IT value (through increased productivity; cost-effectiveness; and improved service delivery for stakeholders) and specified the roles and responsibilities of public organizations (e.g., departments, agencies, and structures). The Department of Public Service and Administration (DPSA) is responsible for e-government. In 2004, the Batho Pele Gateway project was implemented; this saw various departments and tiers of government provide online services.

[1] www.oecd.org/internet/ieconomy/1952944.pdf

Other national policies have been introduced. DTPS (2016, 2017) outlines the role of e-government in the Information Society (e.g., user-centred services, cost-effective service delivery, and enhanced engagement with citizens) and also successes and challenges.

2.7.3.2 Some successes

Many national, provincial, and local governments, and public organizations have implemented e-government. Some have been labelled successful (i.e., have achieved the desired outcomes), and are well-cited in the literature. For example, South African Revenue Service e-filing for tax management, Independent Electoral Commission election system, Department of Home Affairs e-Hanis and Smart Identification card for citizens, National Transport Information System e-Natis for driver and vehicle licence registration, Western Cape and Gauteng provinces and Johannesburg, Cape Town and Ethekweni municipalities (Bwalya & Mutula, 2014; DTPS, 2017; Naidoo et al., 2011).

In recent United Nations e-government surveys which assessed the capacity of member states to use IT to deliver service, South Africa ranked 76, 68, and 78 out of 193 member states; and three, two and three in Africa—in 2016, 2018, and 2020 respectively (UN, 2016, 2018, 2020). However, some challenges have been reported.

2.7.3.3 E-government challenges

DTPS (2017) acknowledges that e-government has not made progress. One of the primary reasons is the lack of coordination and collaboration. While some countries (e.g., United Kingdom, (UN, 2020)) have implemented an integrated strategy and created a central structure to oversee e-government, there is no overarching strategy in South Africa. To the contrary, disparate policies arrogate responsibilities to various departments and structures. Roles and responsibilities are not clearly defined, and the result is fragmented initiatives, duplication and incompatibilities, and lack of accountability. Unlike other countries (e.g., France and Germany), there is no national e-government evaluation strategy (DTPS, 2017).

The following sums up the review and draws conclusions to inform the remainder of this research.

Summary

Research context (i.e., problem, purpose, etc.) determines the purpose, scope, structure, procedures, etc. of literature review. This research concerned how e-government outcomes evaluation is done in developing countries. The problem is under-researched and spans many

domains. Thus, an iterative review to understand and consolidate relevant knowledge was appropriate. Table 2.1 shows the review strategy.

E-government, practice, IT and public sector reform, public value, and evaluation practice (i.e., program, IT, and e-government) were considered from a developing country perspective. Tables 2.3 to 2.8 summarize the review. Some of these concepts are polysemic, thus, their applications in this research were defined.

E-government refers to the use of IT in the public sector. Practice, in organizational research, denotes how work is done. This involved people, resources and their interaction, requisite knowledge, shared understanding, affect and actions, directed towards outcomes. Research on practice— to provide evidence to improve both theory and future practice— has gained currency.

The review showed public value provides an adequate lens to examine performance evaluation in the public sector. The relevant concepts include public value, public values, and public service values. Commonly measured outcomes are efficiency, effectiveness, and democratic values. Stakeholder involvement is central.

Evaluation may be considered a social practice, i.e., how people do evaluation. This implicates non-experimental strategies (e.g., survey, case study, and mixed methods), to understand both tangible and intangible aspects. This is consistent with the research problem. There are several evaluation approaches (also called frameworks), which prescribe how evaluation is done e.g., participatory, development, and theory-driven evaluation. This research draws on the common factors, e.g., stakeholder involvement, learning for improvement, etc.

IT use, outcomes, and applicable evaluation theories and practices have evolved in correspondence with public sector reforms. Public value has, therefore, emerged as an appropriate lens for e-government outcomes evaluation research.

The public sector in developing countries is characterized by lack of resources, higher e-government failure and opportunity cost, lower evaluation capacity, development issues (e-government is a development program), etc. The review showed e-government outcomes evaluation needs attention. Furthermore, development evaluation is applicable.

The review identified several gaps in the literature, which justified the need for this research. E-government outcomes evaluation remains an important research theme, however, there is a paucity of empirical research. While public value may be an appropriate lens for investigation, exemplar practical applications in research, especially quantitative, are sparse.

Notwithstanding the turn to research on practices in organizations, there is low adoption in IS, particularly IS evaluation. Although there is a need for research on e-government outcomes evaluation in developing countries, the focus is mainly on developed countries. This research addressed these gaps.

South Africa was considered an appropriate case. The public sector has undergone reform underpinned by e-government and evaluation. The country ranks higher among developing countries on the capacity to use e-government to support service delivery and some initiatives have achieved the desired outcomes. There are structures, support, and demand for evaluation as well as evidence of use.

The review also showed several ways to examine how IS evaluation and, therefore, e-government outcomes evaluation; the CCP framework is widely used. This research framed evaluation as a practice to explicitly foreground oft-ignored aspects, e.g., affect issues and change (i.e., efforts to improve future practice). Figure 2.5 showed the elements of e-government outcomes evaluation practice to consider, i.e., context, resources, activities, output, outcomes and theory.

Theories, frameworks, and models facilitate an understanding and investigation of phenomena, and are central to research. Conceptualization of e-government outcomes evaluation as practice affords the use of practice theory as a theoretical framework and it is to this, which the next chapter turns.

Chapter 3 : Theoretical and conceptual frameworks

As with all problem-solving endeavours, research is informed by assumptions. Consequently, an effective research design needs to explicate the role of theory. Assumptions about the research problem, e.g., purpose and appropriate questions, were outlined in Chapter 1. In Chapter 2, a broad, pragmatic literature review was conducted to understand the problem domain. The results showed that e-government outcomes evaluation can be framed as practice, i.e., how people performed their work.

Theories serve several purposes, e.g., social theory (i.e., a lens to study a social phenomenon), philosophical assumptions, etc. The main objective of this chapter is threefold, to: outline the role of theory, introduce practice theory (to justify its suitability and draw on Schatzki's version to frame e-government evaluation as how work is done), and construct a conceptual model to inform this research. Figure 3.1 illustrates the structure of this chapter.

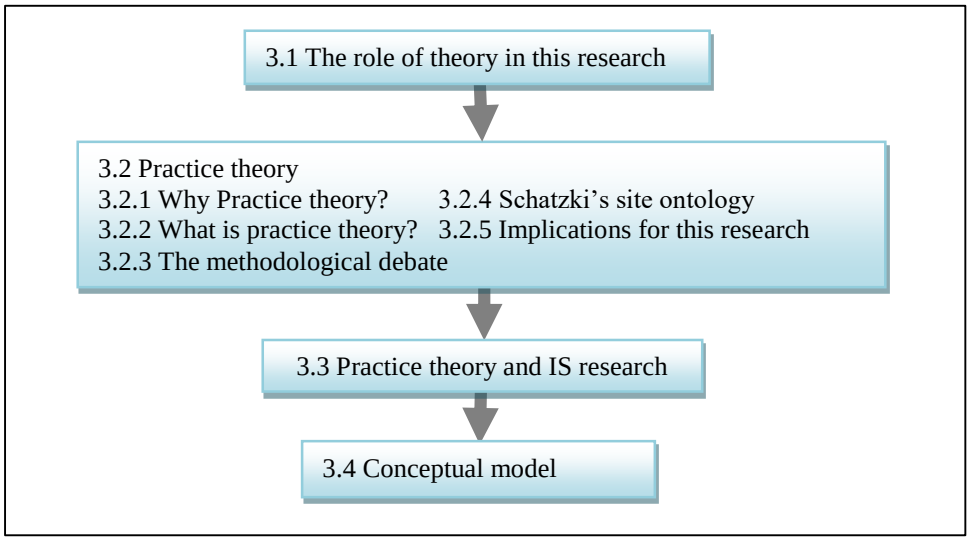


Figure 3.1: Structure of this chapter

3.1 The role of theory in this research

The primary purpose of research is to contribute to existing knowledge. Literature review, theoretical and conceptual frameworks, and other elements (e.g., research objectives, questions, etc.) facilitate an understanding of what is to be known and how and, therefore, are fundamental to research design.

This notwithstanding, the research landscape is littered with ambiguities. The terms are not clearly defined and often used inappropriately as substitutes (Green, 2014; Sitwala, 2014). A

clear distinction can improve the organization, conceptualization and implementation of research and, thereby, improve rigour (Rocco & Plakhotnik, 2009).

Theories serve several purposes in empirical research, e.g., as assumptions to: delineate research problems, guide data collection and interpret results. However, there is no universal definition. Maxwell (2012) described theory as the abstraction of an aspect of a real-world phenomenon, which consists of concepts and their relationships. According to Lynch (2016), it is a description of how the world under investigation, or part thereof, works. Examples include social theories and paradigms. A concept, on the other hand, is a feature of a phenomenon (Green, 2014)— a component of the theory.

Other simple references to theory include proposition, research question, and hypothesis (Lynch, 2016; Walliman, 2018). A proposition is a single statement derived from theory, which expresses a relationship among concepts. A research question is a proposition re-framed as a question. A hypothesis is a testable— falsifiable or confirmable— statement. Although not clear cut, propositions deal with abstract concepts and can take several forms (e.g., to describe distributions, differences among groups, etc.) while hypotheses involve well-defined variables (e.g., independent and dependent) and demonstrate magnitude (e.g., strong-weak; high-low; increase-decrease, etc.) as well as direction (e.g., causes and effect) of relationships among such variables (Bhattacherjee, 2020). Propositions, therefore, are devoid of the strict characteristics of hypotheses (Walliman, 2018).

A framework affords a way of viewing the phenomenon under investigation (Silverman, 2013), and helps organize research to achieve desired outcomes. Theoretical and conceptual frameworks, and models are distinguishable (Green, 2014; Maxwell, 2012; Sitwala, 2014). A model is a visual representation of a framework. A theoretical framework is developed from a theory or part thereof whereas a conceptual framework is a set of interrelated concepts, and may be derived from several sources. Both are constructed by researchers to represent their point of view— to simplify and present what may be a complex abstraction. Therefore, it may be necessary to review frameworks and models to ensure validity. In general, theoretical models are unnecessary where only one theory is involved.

Throughout this chapter, theory refers to social theory and paradigm. Practice theory was used to frame e-government evaluation as a social activity while post-positivism served as the philosophical assumptions to inform research design. The terms concept, proposition, research question, theoretical and conceptual frameworks— and models— are used as explicated above.

The literature review identified several concepts (e.g., e-government, public value, evaluation practice, and development) from several disparate domains relevant to the research problem, which need to be integrated. A conceptual model was, therefore, necessary.

3.2 Practice theory

According to Sitwala (2014), theoretical and conceptual frameworks, as epistemological devices, have implications for research design. Thus, it is pertinent to understand what makes practice theory (and not any other social theory) suitable for this research, the central principles and methodologies, and implications for research.

3.2.1 Why practice theory?

There are several reasons why practice theory was a suitable theoretical framework for this research. Most prominent, it overcomes the limitation of alternate social theories, which treat aspects of social phenomena in isolation (i.e., dualism). The focus on practice (not individuals) as the unit of analysis allows investigation of multiple perspectives, human and non-human, agency and structure, tangible and non-tangible, etc., at the same time (see section 3.2.2).

Maxwell (2012) adjured theoretical congruency; i.e., where two or more theories are integrated, their respective assumptions must be compatible. Browne et al. (2014) observed post-positivism is epistemologically consistent with practice theory. As shown below, both support multiple perspectives and mixed methods and, therefore, lend themselves to the research objectives (i.e., to identify the elements of e-government outcomes evaluation and understand the different patterns of performance).

Practice theory also supports learning and change— two key features of evaluation, as elucidated in section 2.4.2. Practice research provides new insight into current practices (e.g., how things are being done) and offers opportunities for improvement and, therefore, serves the purpose of this research.

Paradigms and their underlying philosophical assumptions are discussed in Chapter 4. There are several versions of practice theory, therefore, it is appropriate to present a general overview and, thereafter, outline which is relevant to this research.

3.2.2 What is practice theory?

Section 2.4.1 described practice. There are many ways to conceptualize and study practices and, therefore, several versions of practice theory— each with its own set of concepts and their

relationships (Sandberg & Tsoukas, 2016; Schatzki, 2017b). The dissensus on the nature of practice permeates theory.

Practice theory— also called social practice theory or theory of practice— is an amalgam of theories, which focus on practice as the main concept (Schatzki, 2017b; Spaargaren et al., 2016). The underlying tenet is that social phenomena (e.g., evaluating, power, marriage, etc.) are caused, and can be explained by, practice (Nicolini, 2017; Schatzki, 2012, 2015). Thus, practices are a means to understand social phenomena.

Practice theory provides concepts and their relationships, the vocabulary and grammar, to re-present a phenomenon under investigation as practice (Hui et al., 2017; Nicolini, 2012, 2017). Such a representation may vary from one practice ontology to another. Thus, a version and its assumptions have implications for how to conduct research (Spaargaren et al., 2016).

A comprehensive analysis of the differences is beyond the scope of this research. However, there are common assumptions, e.g., practices are social, organized, connected, embodied, and embedded (Hui et al., 2017; Welch, 2016). Social suggests collective performance by, but not centred on, individuals. This foregrounds work in organizations as a social practice and explains why practice theory is an appropriate lens with which to examine how e-government outcomes evaluation is done.

Practice is generally viewed as an "entity" with characteristics or elements, and "performance" (i.e., action, doing something). The two are interrelated: the former determines performance while stable patterns emerge from repeated performance (Higginson et al., 2015). For example, "what" to do can suggest "how" to do it; also, patterns of "what" is done emerge and persist through regular performance. Therefore, there have been calls for research to address both (Browne et al., 2014; Higginson et al., 2015; Welch, 2016, 2017). That is, "how combination of elements are enacted" (Higginson et al., 2014, p. 5); "identify the components of practice and their configuration" (Welch, 2017, p. 7).

People interact with material things, guided by an understanding (e.g., what is acceptable, etc.) to perform actions towards an end. Thus, while practices are organized around goals, acceptable behaviour, etc. (Nicolini, 2017; Nicolini & Monteiro, 2017), there is no definite set of such elements.

Practices connect to, depend on, and sustain each other (Nicolini, 2017; Nicolini & Monteiro, 2017). The elements of one practice, e.g., people, rules, desired outcome, venue, etc., can form part of another (Hui et al., 2017; Spaargaren et al., 2016). The interconnection can give rise to

a larger network of practices. However, it is possible to isolate and examine specific aspects (Hui et al., 2017; Nicolini & Monteiro, 2017). Several concepts have emerged to describe the units of practices and their interconnection (Spaargaren et al., 2016); network diagrams are appropriate (Higginson et al., 2015).

Performance of action typically involves mental, physical, and affective human characteristics. The body knows what to do and how, and can learn. "Embodied" implies people perform through their bodies (Feldman & Orlikowski, 2011; Nicolini & Monteiro, 2017; Spaargaren et al., 2016). Knowledge (and its acquisition) is an important part of practice (Schatzki, 2017b). Learning may be formal (e.g., training) or informal (e.g., experience, through participation).

A part of performance is routine actions, i.e., regularities (Nicolini, 2017). For example, routine performance can arise from learning and repeated performance over time by the same people, using the same tools, in the same place, and towards the same outcome. Thus, through repeated performance, elements combine to form stable patterns, i.e., through "practice as performance", "practice as entities" come to exist (Lamers et al., 2017). There is a growing effort to develop methodologies, which account for patterns (see section 3.2.3).

The differences in contexts of time and space mean the elements of a practice may be organized differently and result in variations of the same practice (Nicolini, 2017; Nicolini & Monteiro, 2017). The body can also place limits on what can and cannot be done in a specific material, telic, affective, normative, and other circumstances. Thus, people may elect to perform an activity differently, under different circumstances. For example, appropriate know-how and general understanding can make it easier to perform an activity; the opposite is also true. Thus, practices are embedded in context.

One strongpoint of practice theory is the potential to bridge the dichotomies characteristic of social research (Feldman & Orlikowski, 2011; Gherardi, 2017; Hui et al., 2017). Aspects of practice (e.g., human and non-human, body and mind, collective and individual, objective and subjective, tangible and intangible, dynamic and stable, etc.), are not mutually exclusive. Practice theory is a flat ontology (Lamers et al., 2017; Spaargaren et al., 2016), which treats the world under investigation as a single plane— not separated into social strata. The epistemic implication is the need to capture different perspectives of practices. While this is unequivocal, there is a dissensus on appropriate methodology (see section 3.2.3).

The turn to research on practices (also called practice-based studies) has seen growing adoption of practice theory in a wide range of fields, e.g., work and organizations (Spaargaren et al., 2016), evaluation (see section 2.4.3), and IS (Tavakoli et al., 2017).

Practice theory generates insights, which enable "change in practices, while supporting and reinforcing those practices that are working" (Feldman & Orlikowski, 2011, p. 19). This makes evident why practices, and not individuals, are units of analysis. However, there is no consensus on whether to examine practices as entities or performances to effect change (Higginson et al., 2015). According to Spaargaren et al. (2016), an important factor is the research purpose.

Elements of practices affect performance and can be "re-crafted" to bring about social and behavioural changes (Welch, 2016, p. 17). For example, training can address inadequate skills, change behaviour, and improve practice. This research adopted this approach, i.e., examined how different groups perform e-government outcomes evaluation to identify aspects which need to change and devised strategies for improvement. Table 3.1 summarizes some of the ontological issues relevant to this research.

Table 3.1: Practice theory: summary of key ontological issues

Concept	*Description*
Practice as entity	A "thing", real with elements
Practice as performance	Action; different combinations of elements form different patterns of performance
Organization of practice	How practices are ordered; understanding, rules, goal, etc., organize practices (e.g., what is acceptable)
Practice ontology	Describes the basic structure of practices and their relationships Practices connect to form a larger network of practices Used to re-present a phenomenon under investigation
Embodiment	Performance expressed through the body (physical, mental, etc.)
Embeddedness	Practices as situated in social and material contexts Variants of the same practice can occur according to place and time Practices and context mutually affect each other
Duality	Mutual constitution: body and mind; human and non-human, tangible and non-tangible, objective and subjective, etc.
Change	Practice theory prescribes change, but not how (whether through entity or performance)
Practice research	Practice as the unit of analysis Focuses on both practice as entity and performance Network diagram an adequate representation of practices Possible to isolate and examine specific aspects of the network

Practice theory is not without shortcomings; e.g., theoretical incoherence; conceptual ambiguities; and under-developed methodology (Hopwood, 2010; Hui et al., 2017; Huizing & Cavanagh, 2011; Nicolini, 2012; Warde, 2014). The conceptual muddle makes conceptualization difficult, e.g., there is no agreement on the nature and number of elements of practice, the extent of human and nonhuman involvement (i.e., whether it is entirely a human

activity or not), etc. Furthermore, guidelines and exemplar empirical research, particularly quantitative, are sparse. The criticisms may be attributable to the nascence of practice theory (Spaargaren et al., 2016)

3.2.3 The methodological debate

There are two ways practice may be studied: through the practitioner's account or observation of practice as it happens. Although practice theory does not provide methodological directions, the pervading wisdom is to observe as people do their work (Lamers et al., 2017; Nicolini, 2017; Nicolini & Monteiro, 2017). As Nicolini observed, it is important to witness "the scene of action" (p. 27) to study practice; "practitioners account is undesirable" (p. 29).

The move away from individualism to focus on the "social" has entrenched qualitative methods which address context (Spaargaren et al., 2016). Thus, quantitative methods— amenable to the study of individuals— have long been considered inappropriate. The assertion by Nicolini (2012) that practice theory renders positivism inappropriate for organization studies bears testament. Consequently, qualitative research has been arrogated a pride of place (Browne et al., 2014; Spaargaren et al., 2016). Nicolini (2012) claimed work is tied to performance in a context and, therefore, analytical research with observation is preferable— as people may provide selective accounts of their experiences in descriptive research. Schatzki (2012, p. 25) maintains "researchers have no choice" but to adopt ethnography and collect data by observation and interview.

Others, however, contend practice theory supports both qualitative and quantitative research. According to Halkier (2017), different research objectives call for different research methodologies. Lamers et al. (2017) adjured pragmatic selection of methods. According to Schatzki (2012), in-depth understanding of specific cases is important, and so are general overviews from which to draw lessons to improve future practices.

Descriptive research plays an important role in the social policy environment (Browne et al., 2014; Schatzki, 2012). For example, quantitative surveys of larger samples to describe program outcomes on stakeholders. As Schatzki observed— and discussed above— some aspects of practices (activities, material entities, etc.) are tangible while others (e.g., affect) are not.

Quantitative, descriptive research to capture differences and similarities in practices is gaining attention (Browne et al., 2014). However, as Schatzki (2012) cautioned, quantitative alone cannot provide understanding to solve problems and should be complemented with qualitative.

Nicolini (2017) cautioned against the mere description of work, to the neglect of explanation of how or why.

There is growing acknowledgment that post-positivism provides an epistemic warrant for mixed methods, to understand the different perspectives of practices. Several practice studies (Bartiauxs & Salmón, 2014; Browne et al., 2014, 2015; Kennedy et al., 2013) have adopted this strategy. According to Mertens and Tarsilla (2015), post-positivism with mixed methods is a tradition in evaluation practice. Markiewicz and Patrick (2016) contend mixed methods is central to evaluation.

According to Creswell (2015), mixed methods research admits different theoretical and philosophical frameworks. It is evident that post-positivism with mixed methods is not only compatible with practice theory but also the multiple perspectives of public value and outcomes evaluation (as outlined in the literature review) and with the research purpose.

As aforementioned, practice theory is a group of theories with a common set of assumptions. Thus, the convention in research is to select a specific theory and apply the core principles (Feldman & Orlikowski, 2011). This research adopted Schatzki's site ontology (Schatzki, 2012, 2013, 2015, 2017a, 2017b, 2019). This version is well-developed and well-cited (Lamers et al., 2017; Nicolini, 2012) and widely used in organization and management studies (Loscher et al., 2019).

3.2.4 Schatzki's site ontology

Practice ontology, as discussed above, shows how practices exist and relate to each other. Schatzki's ontology takes practices as an organized set of activities by people (Schatzki, 2012) and is consistent with the conceptualization as outlined above— entities and performances; organized; embodied; routine; connected; and embedded. Furthermore, it underscores duality (over dualism), learning, and change.

Schatzki (2017b) emphasizes the importance, and the interconnection, among agency; structure; learning; and change. Agency is about actions, choices, and changes, e.g., people make choices and act, which may result in changes. Learning can result in knowledge and enable actions and changes, i.e., agency. Structure is about the context within which agents operate. Thus, practices as situated means context or situation (i.e., the state of the world within which the action is performed). Context is, therefore, the *site* in which practice occurs and is referred to as a *bundle*.

A bundle is the basic unit of practice and houses a practice-material arrangement (Schatzki, 2012, 2015, 2017a, 2017b, 2019). This presupposes practice and material arrangements are tied together. The "glue" is *connection* (also referred to as relation, link, etc.) and consists of practical understanding, general understanding, rules, and teleoaffective structure. Thus, a bundle is composed of three basic elements: organized activities, material arrangements, and *connections*. Schatzki arrogates precedence to human over material entities— the reference to practice as human activity bears testament.

Practical understanding refers to "knowing what bodily doings and sayings to perform to carry out a desired action" (Schatzki, 2017b, p. 34). It is this which enables performance. General understanding stands for "abstract senses… of worth, value, nature", which inform people's actions; rules concern "explicitly formulated directives… instructions or edicts" (Schatzki, 2012, p. 16). Teleoaffective derives from "telos", i.e., ends, outcomes, etc.; and affective, i.e., emotions, motivation, etc. (Loscher et al., 2019). Teleoaffective structures, therefore, link outcomes and affect to activities (Welch, 2017).

The activities in practice may connect to each other (by *connections*) to form a network across space and time; e.g., evaluation activities (i.e., assessment and use) connected by purpose. Material arrangement refers to the set-up of material entities, i.e., an arrangement of people, physical spaces, and other nonhumans— an interconnection of entities. *Connections* (e.g., motivation, desired outcomes, etc.) may also form a larger network.

A bundle, as the context or situation, can give rise to different relationships among activities, material entities, and *connections*, thereby, variations in practices. The actions and *connections* combine only during the "doing of something". Thus, people can elect to perform an activity variously. For example, although e-government outcomes evaluation has a general configuration (practice as an entity), the elements (e.g., purpose, stakeholders, physical space, activities, and outcomes) combine during the actual "doing" of such evaluation. It is apparent there can be a myriad of ways to perform a practice (i.e., variants of a practice). However, often, there are a few archetypes (Schatzki, 2017b).

The elements of a bundle shape each other. As stated above, material arrangements can pre-configure acceptable action. *Connections* can explain what, and why, an action may be performed. Practices may bestow roles, e.g., action can "effect, use, react to, bestow meaning on" material arrangement (Schatzki, 2015, p. 1). As Feldman and Worline (2016, p. 311)

succinctly put it, "what a thing does is determined in practice". For example, evaluation stakeholder may be a participant, user or both.

Practices can persist, change, or emerge over time. Patterns of performances occur and stabilize over time. On the other hand, actions can change from time to time, because of changes in *connections* and material arrangements. For example, innovative ways of doing things may emerge and persist, and the old may discontinue. However, stability is a common feature as humans are inclined to do what is "enjoined and acceptable of them" (Schatzki, 2012, p. 22, 2015, 2017b).

Bundles, like their constituent activities, material arrangements and *connections* may connect to each other, persist, change or emerge over time. Bundles can be linked together to form a constellation of bundles. Constellations, in turn, can connect to each other to form larger constellations. Activities in a bundle can connect to others in a constellation, likewise material arrangements and *connections*. This way, bundles share their elements and stretch over time and space in a constellation. That is, practices, material arrangement and *connections* in a specific bundle can also become part of bundles elsewhere in the constellation. Thus, all bundles in practices have the same constitution, each piece bears the same features as the largest constellation and can be isolated and examined.

Learning (i.e., acquiring skill, knowledge, etc.) occurs by doing or as outcomes and shapes (and is shaped by) elements of a bundle within which it occurs (Schatzki, 2017b). This can support agency and perpetuate, change, or spawn new bundles or features thereof.

Schatzki (2005) maintains that differences in practitioners' characteristics (e.g., position, practical know-how, affect, etc.) may result in differences in the performance of practices. *Connections* determine who and what; roles and positions are linked to a way of acting, knowledge, etc., and can give rise to different performances (Schatzki, 2017b, 2019). For example, managers and non-managers may perform different functions and activities.

Schatzki's ontology has strengths and shortcomings (Loscher et al., 2019). It is well-developed and provides a plethora of terms to describe social phenomena. Also, it can be combined with compatible organizational theories, e.g., concepts can be adopted from the site ontology to describe relevant parts of, or re-frame concepts in, another theory. This makes it suitable for multidimensional phenomena such as evaluation and public value. The main challenge, for empirical research, is that some concepts have not been operationalized and there is a dearth of quantitative methods.

The nature of practice and the core tenets of practice theory have specific implications for organizing research. Furthermore, each version has unique implications. Table 3.2 summarizes the relevant concepts of Schatzki's ontology.

Table 3.2: Schatzki's ontology: relevant concepts

Concept	*Description*
Activity	Action performed (i.e., doing and sayings)
Practice	Open ended, interconnected activities
	Viewed as entity (i.e., real with features) and performance (i.e., doing something)
	Human activity
	Different combinations of elements occur in time and space
	Regular performance results in routine (stable) patterns of performances
	Usually few stable patterns of performance
Connections	Practical and general understanding, rules and teleoaffective structures
	Organize activities and, thereby, practices
Material arrangement	The set-up of material entities, i.e., people, physical spaces, tools, etc.
	Pre-configures practices
Agency	Involves options, actions and change
Structure	The context of action
Stability	Activities, practices, bundles, etc., although labile, generally persist (humans are inclined to acceptable practices),
Differences in practice	Characteristics of participants (e.g., expertise, knowledge, general understanding, motivation, roles and positions, etc.) determine performance
	Position (e.g., management and non-management) and roles may imply different end, motivations, knowledge, etc.
Learning	Occurs through, or as outcome, of participation
	Shapes (and is shaped by) the bundle within which it occurs
Bundle	Practice and material arrangement tied together by *connections*
	Mutual constitution i.e., elements (people, activities and other entities) shape each other
Site	Bundle as the locus of practice
	Defines bundle (therefore, practice-material arrangement) as situated in social, physical and temporal location
	Connect to others to form constellation
	Bundles in site share their elements
Constellation	Site of interconnected sites

3.2.5 Implications for this research

The basic step in practice research is to use a specific practice ontology to re-frame the phenomenon under study as practice. So how can Schatzki's version be keyed to e-government outcomes evaluation? What is an appropriate number of practices in a constellation to isolate and examine? The number is unlimited and practice theory does not give guidance. What are the methodological implications?

A bundle, as a practice-material arrangement, was adopted as the unit of practice. E-government outcomes evaluation was isolated as the largest bundle of activities, material arrangement, and connections. Consistent with the main research objectives, practice was viewed as both an entity and performance. This allowed examination of both the elements and stable patterns of performances (i.e., the "what" and "how") of e-government outcomes evaluation. Thus, this research adhered to the adjuration for practice research to go beyond

mere description to provide understanding (Nicolini, 2017). Comparison among groups of practitioners according to their characteristics, e.g., roles, positions, etc., was examined. Recommendations were drawn to enable changes for improvement.

The research objective underscored description and understanding and justified selection of post-positivism and mixed methods. Section 3.1 discussed compatibility in cases where two or more theories are combined. Practice ontologies are flat (see section 3.2.2) and incommensurate with theories which stratify the world under investigation. Section 3.2.3 showed practice theory and post-positivism are congruent. Section 3.2.4 showed Schatzki's theory can accommodate other organizational theories— and is well-suited to the multiple perspectives of evaluation and public value. Chapter 4 explored paradigms and research methods, their strengths and weaknesses, etc., in detail to make a case for their selection.

3.3 Practice theory and IS practice-based research

The benefits of practice to IS are acknowledged, e.g., to connect practice to theory, and bring how things are done to the attention of researchers (Leidner et al., 2017). The introduction of IT in workplaces has generated interest in the study of work practices. Notwithstanding the growing interest (Tavakoli & Schlagwein, 2016), adoption is low. There may be several reasons, e.g., studies which contribute to theoretical development are preferred to those which address practice and practitioner issues (Leidner et al., 2017). Another is the conceptual muddle and scant methodological development (see section 3.2.3).

According to Leidner et al., the IS problems to which practice research is applied are often underdeveloped and, consequently, lend themselves to exploratory research supported by interview, observation and survey. This is consistent with this research: an explanatory mixed methods survey was adopted.

There are several broad ways in which practice theory is applied in IS (Tavakoli & Schlagwein, 2016). Research may: draw on concepts of different versions of practice theory to focus on multiple perspectives of a phenomenon; adopt and do justice to tenets of a specific version (e.g., Schatzki's); apply practice theory specific to IS, e.g., IS adoption practice; adapt practice theory to develop IS practice theory (e.g. technology use); and identify specific concepts (e.g., agency and structure) to inform a study without recourse to practice theory as such. This research corresponded to the second category (i.e., adoption of a specific version of practice theory).

Practice theory has been applied to various areas in IS and organization studies, e.g., to examine: IS development risk (Öbrand et al., 2019) and IT outcomes (Tavakoli & Schlagwein, 2016). There is a longstanding tradition of IS evaluation practice research. However, the literature shows that, of the approaches stated above, adoption of specific frameworks (CCP, DeLone and McLean, etc.— see section 2.5) is predominant. For example, Song and Letch (2014) extended the CCP framework to examine how IS evaluation is used without reference to practice theory. However, Al-Yaseen et al. (2008) investigated practitioners' perceptions of how post-implementation evaluation is done in large business organizations without recourse to practice theory as such.

This research contributes to the development of practice-based IS research. As shown above, Schatzki's version of practice theory was adopted. The next section applies concepts thereof, and from the literature review— to frame e-government outcomes evaluation as practice and construct a conceptual model.

3.4 Conceptual model

Section 3.1 described the relationships between theory, concept, framework, and model. Conceptual frameworks and models (i.e., visual representations) help to structure the research process—to organize concepts and their relationships and, thereby, the data to collect and analyze to ensure consistency with a research purpose, philosophical assumptions, and outcomes.

Literature review and theoretical framework are essential to the construction of a conceptual model. The literature review (Chapter 2) identified concepts from several relevant domains. Figure 2.5 identified the elements of evaluation practice (i.e., context, resources, activities, output, outcomes and theory). Section 3.2 made a case for the adoption of Schatzki's ontology and outlined the main concepts, e.g., activities, material arrangement, *connections*, bundles, sites, agency, structure, learning, and change.

E-government outcomes evaluation, as what people do, can be re-framed as the interaction among people and other resources; activities; practical knowledge; shared assumptions, beliefs and values (policies, public value, etc.); outcomes; and affect in a context (i.e., the public sector). This can be viewed as a complex system (i.e., interconnected elements directed towards outcomes) and lends itself to the logic model (see Figures 2.3 to 2.5). Brandl et al. (2017) also used the logic model to re-frame service operations management as practice— to examine the interaction among resources, performance, and outcomes.

In general, evaluation involves *assessment* and *use* (i.e., the effect of participation, use of output or both). Applying Schatzki's ontology as described in section 3.2.4, the elements of EOEP may be emplaced in activities, material entities and *connections* as illustrated in Table 3.3. *Context* is the site within which practice occurs; *resources*: human and material entities involved; *activities*: the actions performed; *output*: the products of the evaluation; *outcomes*: i.e., effect; and *theory*: the assumptions which inform how evaluation is done, and embodies the values, rules, etc.

Table 3.3: E-government outcomes evaluation practice as Assessment and Use bundles

	Activities (actions performed)	*Material entities*	*Connections*
Assessment	1. Planning: Initiation, selection of stakeholders, training; selection of evaluation strategy 2. Implementation: Data collection and analysis, stakeholder management 3. Development of evaluation reports 4. Management response to recommendations	1. Resources: people, budget, tools and time 2. Context, e.g., program, stakeholder and organizational a. Program: what is evaluated (e.g., e-government performance); responsibility for evaluation (e.g., evaluation structure) b. Stakeholder: characteristics (type, expertise, involvement, etc.) c. Organizational: (management support, evaluation structure, capacity building)	1. Practical know-how (i.e., skills, experience) 2. Rules (e.g., policies) 3. General understanding (i.e. shared assumptions, approaches, beliefs, values, etc.) 4. Teleoaffective structures a. Goal (i.e., the purpose of evaluation: symbolic, instrumental and conceptual uses) b. Affect (e.g., motivation, satisfaction, commitment, etc.)
Use	1. Product use, (e.g., implementation of recommendations in evaluation report) 2. Process use 3. Effect on individuals, groups and organizations		

Evaluation of outcomes of a specific e-government program may be conceptualized as a bundle, composed of *Assessment* of and *Use* bundles. From section 2.4.2 these concern how assessment of e-government outcomes is done and the effect (of participation or use of output) thereof respectively.

From the working definition of e-government (see section 2.2.1), internal efficiency; effective service delivery; interaction among public sector organizations, citizens, and businesses; and improved accountability and transparency may be measured. This is also consistent with the national evaluation policy framework (section 2.7.2).

This demarcation between *Assessment* and *Use*, however, may be superficial as the bundles are interconnected and share their elements, e.g., context (the public sector), purpose (e.g., to improve how evaluation is done), resources and assumptions (e.g., public value). Furthermore, the bundles mutually affect each other. For example, *Use* depends on *Assessment* (i.e., output

forms basis of *Use*, etc.); also *Use* (e.g., the effect of participation and or use of report) may be changes to improve how *Assessment* is done.

E-government outcomes evaluation in a specific public organization is a bundle, a site within which *Assessment* and *Use* occur. Similar evaluations in different contexts of time and space (e.g., at different times and different public organizations) also constitute bundles and may be connected to form a constellation (e.g., of EOEP in a sector— say health). All bundles (e.g., the constellation of all sectors) may be connected to form the constellation of e-government outcomes evaluations in the public sector. The bundles may share their activities, material arrangements, and *connections*. For example, the evaluation output of one department may have implications for another. Thus, parts of the constellation (e.g., bundles) may possess common characteristics and, therefore, can be isolated and investigated (see section 3.2).

Various studies have taken advantage of such conceptualization to investigate specific practices. For example, Saunders (2011) depicted evaluation in higher education in like manner— as a constellation of evaluation in individual higher institutions, linked together to form a constellation of all evaluation practices in higher institutions across a country. Kennedy et al. (2013) delineated transportation practices from daily routines (e.g., eating, working, shopping, etc.) for their study. This research, therefore, conforms to such tradition.

Section 3.2.2 showed network diagrams can appropriately illustrate the nature of practices (Higginson et al., 2015). Figure 3.2 shows e-government outcomes evaluation as a practice-material arrangement bundle— a site in a public organization. This consists of *Assessment* (A_b) and *Use* (U_b) bundles, i.e., assessment of e-government outcomes and use (i.e., effect) thereof respectively. A_b and U_b are interconnected and share their activities, material arrangements, and *connections*.

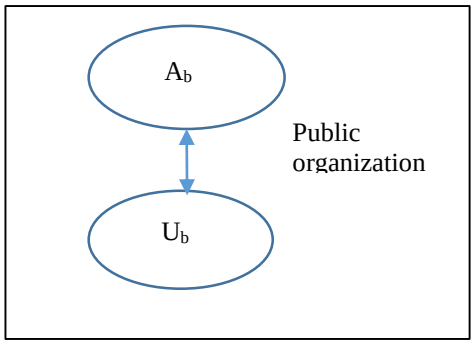


Figure 3.2: E-government outcomes evaluation practice as a site

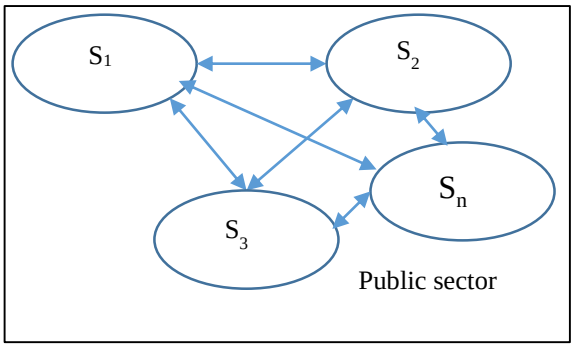


Figure 3.3: E-government evaluation practice in the public sector as "site of sites"

It is possible to examine the whole site (i.e., *Assessment* and *Use* bundles) or an aspect of interest, e.g., a specific bundle (e.g., *Use*), elements (e.g., activities), etc. This is consistent with research on evaluation, e.g., Coryn et al. (2015) and Song and Letch (2014) focused only on use while Kang et al. (2012) investigated both (see section 2.4.5).

Figure 3.3 shows all sites (S_1, S_2, S_3, …, Sn) may be linked together to form the "ultimate" constellation in the public sector. The sites share their elements and, as aforementioned, a part (e.g., a bundle) can be examined to understand the "ultimate" constellation in the public sector. Figure 3.2, therefore, corresponds in structure to the basic unit of Figure 3.3 and constitutes the conceptual model.

A conceptual framework is a tentative description or explanation of a complex phenomenon (Maxwell, 2012). Therefore, this model is shorn of the complexity of e-government outcomes evaluation in real life.

According to Browne et al. (2014) and Higginson et al. (2015), and as explained above, research which focuses on entity investigates the elements of practice (i.e., activities, material arrangements, and *connections*) while performance is about variants of practice (i.e., the different patterns of "doing something", as a result of how the elements cluster in performance).

Several studies have used this approach to understand practices, e.g., transportation practices of households (Kennedy et al., 2013) and water consumption patterns (Bartiauxs & Salmón, 2012, 2014; Browne et al., 2014, 2015). A common feature is the innovative use of quantitative methods to establish and describe the elements of practice and patterns of performance. It is this growing trend (see section 3.2.3), which this research explores.

From the implications in section 3.2.5, the following questions were relevant to the research objectives (i.e., to identify the main elements and patterns of EOEP, and devise strategies for improvement). What are the main elements which describe EOEP (i.e., as an entity)? Are there significant differences among different groups? What are the different ways e-government outcomes evaluation is done (i.e., as performance)? Are there significant differences in patterns of performance among groups? What aspects of practice need improvement? How can these be improved?

Section 1.1 showed the problem under investigation (how e-government outcomes evaluation is done in developing countries) is under-developed and the concepts are not well-defined. This is also true of the concepts from practice theory. The conceptual model, therefore, embodied abstract concepts and relationships. Thus, propositions (not hypotheses) were considered

appropriate (see section 3.1). The main proposition "There is no difference in perception among participants on EOEP" was drawn from the foregoing implications.

Summary

This chapter explicated the roles of theory in this research and outlined the theoretical framework, i.e., practice theory. The methodological debate, including the dominance of qualitative research, was also highlighted. However, as shown above, practice theory allows examination of multiple aspects of practice (i.e., the unit of analysis) at the same time; therefore, there is growing adoption of post-positivism and mixed method research. The flat ontology of practice theory is compatible with this paradigm.

The advantages of Schatzki's ontology were also outlined. It was adopted to re-frame e-government outcomes evaluation as practice, construct a conceptual model and draw implications to organize this research. Context, resources, activities, output, outcomes, and theory were organized into the constituent elements (i.e., activities, material entities and *connections*) of a bundle, which consisted of *Assessment* and *Use* bundles.

The implications for this research included examination of practice as an entity and performance; learning; differences in patterns of performance among practitioners; and change (i.e., improvement). The research problem and concept of practice theory, in general, are not well defined. Thus, propositions were drawn to organize this research.

Research design explicates philosophical assumptions and how the research may be practically conducted to ensure the outcomes are congruent with the purpose and it is to this, which the next chapter turns.

Chapter 4 : Research design

Research involves complex decisions and trade-offs. Research design, *writ large*, is a plan of how to conduct research to achieve the intended purpose. It explicates theoretical and practical assumptions; strategies and techniques; and may serve as an audit trail of choices and their justifications.

Chapter 1 showed the research purpose is descriptive, to identify and describe how e-government outcomes evaluation is done in developing countries. Chapter 2 reviewed the literature from relevant domains to characterize the research problem as practice. Chapter 3 outlined the theoretical framework (i.e. Schatzki's ontology of practice) and epistemic and methodological implications; and presented the conceptual model. Post-positivism and mixed methods were considered suitable for practice research.

Although the unique traditions in domains may be important, how specific research is planned and implemented depends on the purpose, researcher's experience and predilections, available resources, etc. As made clear below, there are several archetypes of mixed methods research. Thus, research design is essential —to describe an adopted strategy; document the design choices and justifications thereof; and, thereby, improve transparency and understanding.

The objective of this chapter is, therefore, threefold: to describe the elements of research design, draw their philosophical and practical implications for this research and, subsequently, integrate all the design decisions into a research strategy.

The plethora of mixed methods to choose from makes a vignette of the strategy adopted herein necessary. Consistent with the research purpose, a two-phased sequential explanatory mixed methods research (quantitative followed by qualitative) was adopted— the latter to explain the findings of the quantitative phase. The implementation of applicable methods (i.e., the logic of inquiry, sampling; data collection, and analysis) are described in detail in the respective phases (quantitative in Chapter 5 and qualitative in Chapter 6). Figure 4.1 illustrates the structure of this chapter.

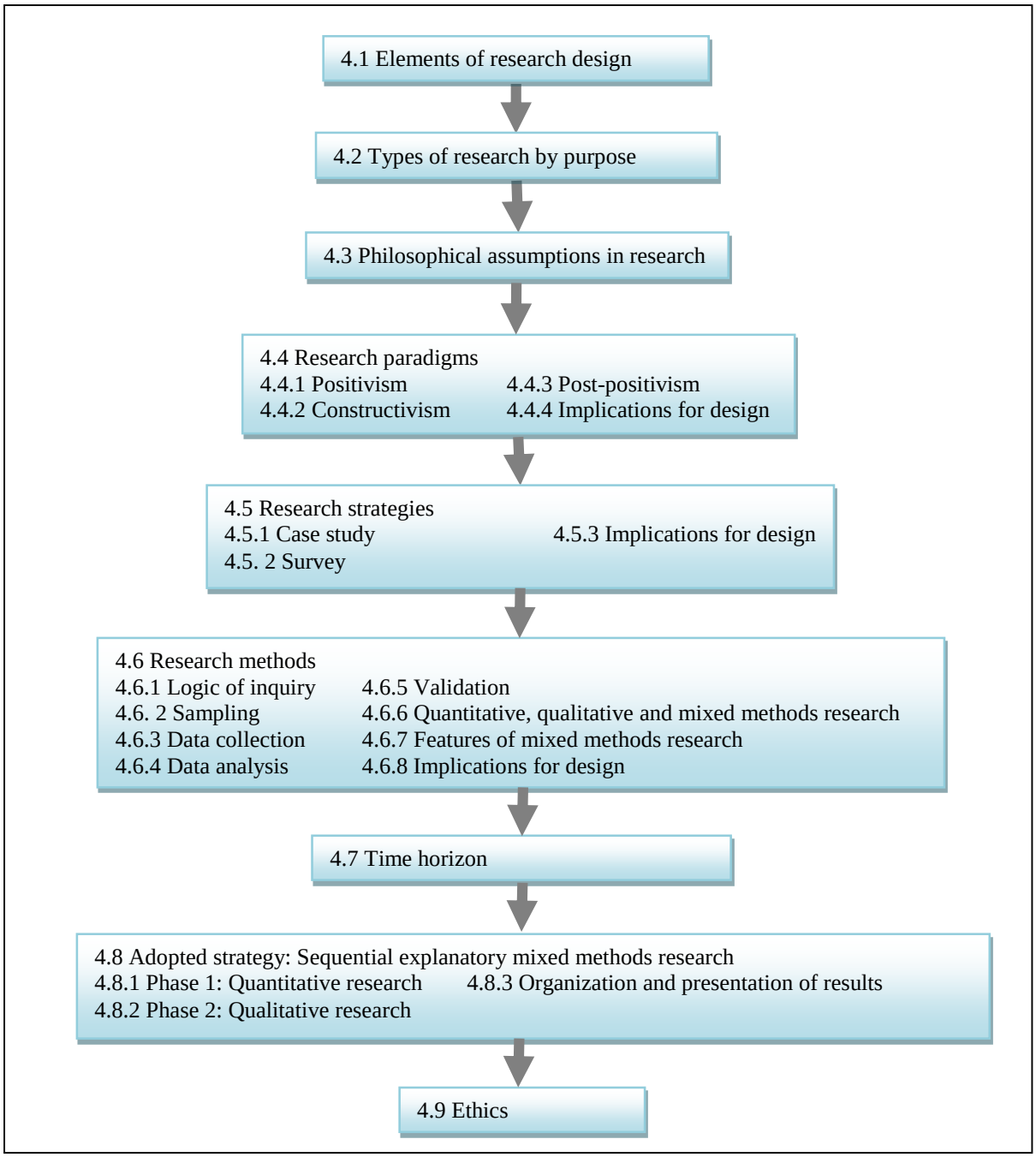


Figure 4.1: Structure of chapter

4.1 Elements of research design

Research design "is the process of making all decisions related to the research before they are carried out" (Blaikie, 2010). It is to empirical research what architectural plan is to a building—an organizing framework to help a researcher plan ahead to ensure the desired outcomes are achieved. It assures congruency among the research type and other elements (Linneberg & Korsgaard, 2019).

There is no agreement on the elements and their definitions (Lynch, 2016). Collis and Hussey (2014) discussed research purpose, process, and outcomes. Purpose is linked to expected outcomes, i.e., contribution— the type of knowledge produced. Process (i.e., how to practically undertake the research) implies a strategy and methods to achieve outcomes (Denscombe, 2014).

Blumer (2017) defined research strategy as how a specific research is designed and conducted. The nature of the problem, philosophical and practical considerations, etc., also influence the process (Saunders et al., 2015). For example, philosophical and paradigmatic assumptions can help define the object under study and suggest the most appropriate means of investigation.

In practice, research is a complex endeavour which involves trade-offs. It is, therefore, important to take steps, throughout the process, to ameliorate potential pitfalls. This requires a strategy and corresponding tactics to address practical aspects of implementation, e.g., suitability, feasibility, and ethics (Blumer, 2017; Denscombe, 2014; Saunders et al., 2015). The place of research design is evident. Purpose, philosophy, strategy, methods, and time horizon provide a framework for classifying research (Collis & Hussey, 2014; Saunders et al., 2015). These are discussed below to draw implications for this research.

4.2 Types of research by purpose

IS research contribution may be broadly categorized into theoretical (e.g., to produce a new— or an improved— knowledge, i.e., evidence, theory, etc.) and practical (e.g., to produce a new or improved solution (e.g., artefact, i.e., algorithm, program, etc.) to an existing problem. The nature of the problem and extant knowledge (i.e., how well the literature, concepts, theories, methods, etc. in a domain are developed) determine the research purpose and, thereby, contribution (Sekaran & Bougie, 2016). Table 4.1 shows the types of research by purpose.

Exploratory research is appropriate for an underdeveloped domain and is often conducted on a small scale to establish characteristics and patterns to develop, but not test, theories. The research process is generally flexible— no restrictions on activities, theories and concepts, and types of data produced. The findings are usually not conclusive or generalizable, but may suggest areas for future research.

Descriptive research provides an account of a phenomenon (e.g., patterns, classifications, relationships among variables, etc.) about which there is ample knowledge and may involve both quantitative and qualitative data and a large number of cases.

While descriptive paints a detailed picture of how things occur, explanatory concerns why things occur, i.e., explains the factors that cause a phenomenon. There can be a myriad of explanations and so, it is common to elide the less plausible causes. This type of research is appropriate in a well-structured domain.

Predictive research forecasts some aspects of the world under study, based on current knowledge (typically, general laws, i.e., generalization) and, therefore, is appropriate in a well-developed domain.

Table 4.1: Types of research by purpose (Collis & Hussey, 2014; Sekaran & Bougie, 2016; Sue & Ritter, 2012)

	Exploratory	*Descriptive*	*Explanatory*	*Predictive*
Goal	Explore	Describe	Explain causes	Predict occurrence
Nature of phenomenon and extent of knowledge development	Unstructured; not well-studied	Structured; not well-studied	Well-structured; well-studied	Well-structured and well-studied
Research questions	What	What and how	Why (and what and how)	What if (and why, how); controls variables
Data type	Qualitative and quantitative	Quantitative and qualitative	Quantitative	Quantitative
Research methods	Qualitative	Quantitative and qualitative	Quantitative	Quantitative
Implications for findings	Insight into a problem	Develop theory	Test theory	Generalization

While the classification above provides a structure, the types are not mutually exclusive. Exploratory and descriptive are similar, and so are explanatory and predictive. Exploratory, description, and explanatory are commonly used (Bhattacherjee, 2020; Saunders et al., 2015). This way, descriptive constitutes a middle ground on a continuum— and can be applied to explore an under-researched problem, describe the object under study and explain relationships which may not be obvious. The preceding chapters showed not much is known about how e-government evaluation is done. Thus, descriptive research (to explore and describe main elements of practice and patterns of performance) was appropriate.

After the research purpose is philosophical assumptions. This can guide selection and justification of what to investigate and how, and thereby, improve the quality of IS research (Hassan et al., 2018).

4.3 Philosophical assumptions in research

The primary concept is realism— claims about objects in the world under investigation on universals, perception, objectivity, morality, etc. (Haig & Evers, 2016). According to Bunge (2006), realism is the assumption that things are real while anti-realism—also nonrealism

(Kovalainen & Anderson, 2014) represents the various positions antithetical to realism. Important features include axiology, semantics, ontology, epistemology, and methodology.

Axiology concerns values (e.g., morals and ethics) and the role of the researcher. Axiological realism assumes values and evidence exist independently of human perception, i.e., an objective, value-free researcher. This is repudiated by anti-realism (Kovalainen & Anderson, 2014), which acknowledges human agency. According to Bunge (2006) while there can be different value judgements, some moral and ethical decisions— subjective or objective— are inviolable. For example, communities, institutions, etc., may have specific considerations, which should be adhered to.

Semantics entail meaning (Niiniluoto, 2014) and concern representation of the world under study (Kovalainen & Anderson, 2014), e.g., as theories, frameworks, models, etc. Semantic realism posits there is a definite structure and, therefore, an ecumenical representation (e.g., mathematical models and formulae) which accurately matches the world. Conversely, semantic anti-realism supports several approximate descriptions of the same world. Common representations include text (e.g., metaphors) and visual models.

Ontology denotes the nature of 'being' or existence, i.e., whether "things" in the world are objects in and of themselves or of human experience. Ontological realism holds that objects are real (i.e., tangible, consist of universals, e.g., properties and relations) and exist of themselves, independently of human perception (i.e., the investigator and objects of investigation are logically independent). Ontological anti-realism, however, conceives entities as objects of human perception, constructed by the investigator; i.e., not real and not logically independent.

Epistemology concerns the nature of knowledge, i.e., what constitutes adequate and valid knowledge, and how such knowledge can be gained. Epistemological realism claims objects, or aspects thereof, can be known completely and objectively, by direct observation. In contrast, epistemological anti-realism admits knowledge is a product of human construction and can be known indirectly, subjectively, and fallibly. Thus, different contexts are likely to produce different forms of valid knowledge.

Some philosophers (e.g., Bunge, 2006) agree that ontology and epistemology are related, e.g., ontological realism presupposes epistemological realism. Others reject such conflation; e.g., Bhaskar (2008) advocated ontological realism supported by epistemological anti-realism.

Methodology concerns how the researcher can go about obtaining knowledge. Methodological realism maintains there is a "scientific method"— modelled on investigations in the physical sciences— to objectively and completely acquire knowledge about objects under study (Bunge, 2006). This assumes, *tout court*, methodology derives from epistemological realism. By contrast, methodological antirealism acknowledges fallibility and rejects the epistemological presupposition of methodology as well as the "scientific method". Table 4.2 summarizes the assumptions above.

Table 4.2: Basic philosophical assumptions (Bunge, 2006; Kovalainen & Anderson, 2014; Niiniluoto, 2014)

Assumptions	*Realism*	*Anti-realism*
Axiology	Value-neutral (i.e., objective values and morals); separates evidence from beliefs, feelings, etc.	Value-laden (i.e., subjective values- and morals); evidence driven by beliefs, feelings, etc.
Semantics	Accurate representation; universal structure	Approximate representation; multiple structures
Ontology	Objective reality; independent of human perception	Subjective reality, attributable to human perception
Epistemology	Knowability; objective observer-phenomena interaction	Fallibility; subjective observer-phenomena interaction
Methodology	"Scientific method"; prescribed methods	"Non-scientific" method; eclectic selection of methods
Implications for paradigm	Pure realism	Pure anti-realism
Goal of science	To discover and establish generalizations	Explanation

Bunge (2006) adjures an integrated framework of assumptions. Notwithstanding the dichotomy, each assumption forms a continuum. In the real world, research may involve both real and perceived objects, varying degrees of value judgement, different forms of representation, etc.

Applied to this research, according to Schatzki's theory (see section 3.2.4), practice is real with characteristics and can exist in different forms (i.e., the elements combine to form different patterns of performance). That is, the problem under investigation can be considered real, represented approximately, and knowledge thereof can be acquired fallibly. The scientific method may, therefore, not be applicable. Research on practice, therefore, may involve both aspects of realism and anti-realism.

Several frameworks for classifying research archetypes have emerged from philosophy. For example, by philosophical standpoints or paradigms (i.e., based on pure realism, pure anti-realism, and "the middle ground") and appropriate research process (i.e., the scientific or

"unscientific method, type of data, type of research (i.e., quantitative, qualitative and mixed methods), and strategy (experiment, survey, case study, etc.).

4.4 Research paradigms

The term paradigm is polysemic; it can refer to a perspective— a set of assumptions and corresponding methods to practically acquire knowledge (i.e., an integrated position, on axiology, semantics, ontology, epistemology, and methodology). As a corollary, different paradigms afford different assumptions and practices and, therefore, research outcomes.

From section 4.3, two broad categories of paradigms may be derived from realism: positivism and the alternatives (referred to as constructivism) which repudiate positivism. As McGregor and Murnane (2010) aptly put it, "there *is* positivism and then there are other paradigms that *deny* positivism as the only way to see the world". Thus, the simplest classification framework consists of positivism and constructivism as the ends of a continuum, with post-positivism as a middle ground. Table 4.3 summarizes the distinguishing features. The following were elided from any further discussion for the reasons stated.

Table 4.3:Research paradigm (Cecez-Kecmanovic & Kennan, 2018; Lincoln et al., 2011; Mertens, 2015)

Element	*Positivism*	*Post positivism*	*Constructivism*
Goal of science	Generalization	Both generalization and explanation	Explanation
Axiology	Value-neutral (i.e., objective values and morals; separates evidence from beliefs, feelings, etc.)	Value-neutral (i.e., objective values and morals; separates evidence from beliefs, feelings, etc.)	Value-laden (subjective values and morals; evidence driven by beliefs, feelings, etc.)
Semantics	Accurate representation of a universal reality	Approximate rather than accurate representation of a universal reality	Approximate representation; multiple realities
Ontology	Objective reality; independent of human perception	Objective reality; independent of human perception	Subjective reality, attributable to human perception
Epistemology	Reality is knowable; objective observer-phenomena interaction	Reality can be known fallibly; total objectivity impractical	Reality can be known fallibly; subjective observer-phenomena interaction
Methodology	The scientific method prescribed methods	"Quasi" scientific methods	"Non-scientific"; eclectic selection of methods
Research strategy	Experiment	Quasi-experiment	Non-experiment (case study)
Implications for methods	Quantitative	Mainly quantitative but admits qualitative	Qualitative
Research quality	Reliability and validity	Reliability and validity; trustworthiness	Trustworthiness
Type of research	Quantitative	Mixed	Qualitative

Pragmatism is generally viewed as a combination of methodologies (Thota et al., 2012). Critical realism supports dualism and, therefore, is incompatible with the flat ontology of practice theory (see section 3.2.2 and 3.2.4). Critical theory is not considered as a paradigm but theory (Mertens, 2015).

4.4.1 Positivism

This assumes a single reality, which exits independent of human perception. The purpose of research, therefore, is to accurately describe the regularities through observer neutrality and the scientific method (i.e., deduction, quantification of observations data, control of empirical variables, and quantitative methods). Rigour, measured by reliability and validity (see section 4.6.5), assures accurate outcomes. Positivism is generally associated with experiment and survey strategies, and quantitative research. The main criticisms include the claim to absolute reality and a "scientific method", outcomes of which are often irrelevant to contexts.

4.4.2 Constructivism

The purpose is to interpret or construct meaning. Constructivism assumes reality is subjective, dependent on human perception and, consequently, admits multiple perspectives and valid meanings of the same reality. Qualitative methods, not "the scientific method", are appropriate to produce a detailed account to inductively generate such meaning. Applicable research strategies include case studies and ethnography. Constructivism is widely criticized for arrogating reality exclusively to human perception and admitting multiple (and often competing) valid knowledge.

4.4.3 Post-positivism

This bridges the chasm between positivism and constructivism. The purpose of research is to discover both patterns and meaning. It admits a single, objective reality which can be known fallibly. Thus, it is important to eliminate alternate explanations or representations of reality (Mertens, 2015).

Post-positivism accepts the influence of context, rejects total neutrality of the researcher, and relies on rigorous research processes to obviate bias and errors. The underlying assumptions support both quantitative and qualitative research. However, it is generally agreed that it is predominantly quantitative research in which a qualitative component is used to gain further insight (Schoonenboom & Johnson, 2017).

Notwithstanding the conflicting views (e.g., from limiting and unnecessary to useful and ineluctable), a paradigm was adopted to afford reflection on strengths, pitfalls, and opportunity costs to justify selection from competing options (Cecez-Kecmanovic & Kennan, 2018; Mertens, 2015).

4.4.4 Implications for design

Although linked with disciplines as well as organizational and individual preferences, it is the research purpose, which determines an appropriate paradigm.

Chapter 2 showed, e-government outcomes evaluation, as practice, is multi-faceted. As explicated in section 4.2, the research purpose and objectives privileged both quantitative and qualitative research and, thus, post-positivism suited the research purpose (i.e., descriptive). While practice is real, it can only be represented approximately and known fallibly and, therefore, the "scientific method" may not be applicable.

Post-positivism has long been considered an appropriate paradigm in IS (Hirschheim, 1985) and there is growing recognition (Mueller & Urbach, 2017). This is also the case in evaluation (Mertens & Tarsilla, 2015) and practice research (Browne et al., 2014)— see chapters 2 and 3. Hartley et al. (2017) called for a combination of quantitative and qualitative methods in public value research (see section 2.2.7). Therefore, post-positivism can give a philosophical warrant for such research.

Section 4.2 established the link between research purpose, process, and outcomes. Sections 4.3 and 4.4 discussed research philosophy and paradigms, and their relationship with research strategies. Various authors (e.g., Cecez-Kecmanovic & Kennan, 2018; Saunders et al., 2015) suggest such a relationship but caution against dogmatic adherence. The main research strategies are discussed below.

4.5 Research strategies

There are several archetypes of how to design and practically conduct research, e.g., experiment, survey, case study, etc. (Twining et al., 2017). These may be linked to methodology and paradigm: each strategy prescribes specific practices and methods for conducting research (Denzin & Lincoln, 2011). For example, a laboratory experiment involves controlled variables and prediction, which suits positivism and the scientific method. However, factors such as the research purpose, problem domain, and practicalities determine appropriate

strategy (Saunders et al., 2015). Case studies and surveys are widely used in IS research (Palvia et al., 2015).

4.5.1 Case study

The main purpose of a case study is to gain an understanding of a phenomenon in a context and is commonly used in exploratory and explanatory research (Saunders et al., 2015). Typically, it involves an intensive investigation of a smaller number of cases to generate rich, textual data from which conclusions may be drawn (Walliman, 2018). Case study lends itself to constructivism, as shown in section 4.4.2. Inductive, qualitative research (Bryman, 2016) with unstructured methods (Walliman, 2018) is usually employed. The characteristic smaller sample size means inferences cannot be drawn to a population.

4.5.2 Survey

The purpose of a survey is to gain broad knowledge about the characteristics, behaviours, opinions, etc., of a larger population by examining a sample in a standard and systematic manner (De Vaus, 2014; Saunders et al., 2015). "Systematic" or structured empirical investigation (De Vaus, 2014) is linked to the deductive approach (see section 4.6.1) and the scientific method. Representative samples, as well as valid and reliable instruments, are essential to credible conclusions. In general, structured methods are employed, and the results may be generalized to the sample itself or the population thereof. Although quantitative methods are generally favoured, the findings cannot explain why patterns and relationships occur. Thus, mixed methods are being applied in surveys to gain more complete perspectives of problems (Check & Schutt, 2012).

A survey is appropriate for exploratory and descriptive research (Saunders et al., 2015) and, as shown in sections 4.4.1 and 4.4.3, well-suited to positivism and post-positivism paradigms. Practical design considerations include response rate, demography and dispersion of a population, time horizon, cost, and instrument and administration thereof.

4.5.3 Implications for design

The research purpose was to describe the main elements of e-government outcomes evaluation and patterns of performance. Evidently, a larger sample was necessary for any meaningful description. Thus, a survey, which involved both quantitative and qualitative data, was considered appropriate. This is also consistent with the post-positivism paradigm. This strategy has been used in practice research, e.g., evaluation (e.g., Coryn et al., 2015), consumption (e.g., Browne et al., 2014, 2015) and IS evaluation (e.g., Al-Yaseen et al., 2008). Furthermore, it

responds to the calls to combine quantitative and qualitative methods in public value (Hartley et al., 2017), practice (Loscher et al., 2019), and IS (Venkatesh et al., 2016). According to Leidner et al. (2017), IS practice-based research is not well-developed and, therefore, a descriptive survey is suitable.

Survey and attendant design choices and justifications are discussed in detail in section 4.8 (which elucidates the overall strategy for this research) and also in Chapters 5 and 6 (which deal with the implementation of the research strategy). As explained in the preceding sections, the primary objective of research design is to ensure the research process leads to the desired outcomes. Paradigms, strategies, methods, and instruments support the implementation of research.

4.6 Research methods

Research design is practically implemented through methods. Methodology and methods are distinguishable (Blumer, 2017; Saunders et al., 2015). Methodology is a set of philosophical assumptions and describes rival methods and how these may be employed (Haig & Evers, 2016) whereas methods are the specific techniques and procedures to gain knowledge (Walliman, 2018). Four main methods are outlined below, *viz*. logic of inquiry, sampling, data collection, and data analysis.

4.6.1 Logic of inquiry

Also called the theory of method (Haig & Evers, 2016), the logic of inquiry refers to assumptions about the logical processes to gain knowledge. An important consideration is whether the research process is based on the scientific method. Thus, there are two main techniques, deductive and inductive (Saunders et al., 2015; Walliman, 2018), each with different implications for outcomes.

Deduction, based on the "scientific method", is a structured approach (De Vaus, 2014; Saunders et al., 2015). There is a general, systematic approach with a set of steps: *a priori* use of theory to frame propositions or hypotheses, systematic observations and description, and conclusions from the data. This shows inference from the general (i.e., theories, laws, etc.) to particular cases. Empirical research involves the operationalization of concepts as well as reliable and accurate measurements. Thus, upfront design, researcher independence, structured methods (i.e., structured sampling, data collection and analysis, and inference), and adequate sample size are necessary— to assure reliability and validity (Saunders et al., 2015). As shown above, deduction is linked to positivism and post-positivism paradigms and survey strategy.

Conversely, induction has no prescribed structure and distils theory from data (De Vaus, 2014; Saunders et al., 2015). Research proceeds from observation to conclusion and theory formulation; the researcher is part of this process (e.g., makes decisions about what is the next step, which findings are relevant or not, etc.). This shows inference is from particular cases towards generalization, e.g., evidence from the data may explain a relationship between concepts. Thus, unlike deductive which applies theory, induction discovers theory, hypothesis, proposition, etc. This approach is emergent and accommodates the vagaries of practicality—new leads may be pursued to reveal unanticipated explanations. As shown above, constructivism and case study are not structured and, therefore, are linked to the inductive approach.

Notwithstanding the distinction, deduction and induction overlap in practice. As DePoy and Gitlin (2016) observed, problem-solving in real life involves aspects of both, e.g., may commence with one and proceed with the other. Logic of inquiry has implications for this research design.

4.6.1.1 Implications for design

While some domains have a predilection for one or the other, e.g., deduction is dominant in IS research (Hassan et al., 2018), an appropriate choice depends on the research context (Saunders et al., 2015). Consistent with the research purpose (i.e., to identify and describe the elements of practice and patterns of performance), paradigm, i.e., post-positivism, and strategy, i.e., survey, this research was primarily deductive, supplemented with inductive, to improve understanding. This research was structured (e.g., upfront design decisions about theories, concepts and their operationalization, and methods) but flexible to accommodate practical considerations.

Central to research is data from which conclusions may be drawn. This brings three logical issues to the fore, namely, sampling, data collection, and data analysis.

4.6.2 Sampling

An important design decision is appropriate data sources and their selection. Data may be collected from several sources, e.g., people, documents, IS artefact, etc. Sampling involves the selection of a subset of members (i.e., data sources) from a population for investigation. The design considerations are an appropriate population, sampling technique (i.e., how the sample may be selected), and adequate sample size (Saunders et al., 2015).

The main criterion for selecting a population and sample is the potential to provide evidence to solve a research problem. Thus, the research purpose defines the characteristics relevant to the investigation and, thereby, the population and the members to select. A well-delineated population is, therefore, important to ensure the selection of a relevant sample.

The factors which determine how a sample may be selected include the type of population, research approach, ethics, and practical constraints, e.g., time, money, and access (Saunders et al., 2015). For example, in the case of a rare-occurring phenomenon, the population may be small, which influences the sample size and selection techniques. There are two main sampling techniques: probability and non-probability sampling (Walliman, 2018). The primary distinction is the extent to which samples are representative and allow the generalization of research outcomes to the entire population.

Probability assumes that each member of the population has an equal chance of being selected. Thus, the population needs to be well-defined, i.e., explicit knowledge about the population is necessary. This requires a sample frame, i.e., an accurate, complete, and up-to-date list of cases in the population. Statistical methods are applied to select a random, representative sample, which allows generalization from the sample to the population. Thus, this technique well-suited to research which seeks to extend knowledge about a population (Uprichard, 2013). In general, the larger the sample, the more representative it may be of the population and, therefore, the more accurate the generalization. Common examples include simple random, systematic, stratified, and cluster sampling.

By contrast, non-probability sampling is suitable for research that explores cases to understand the sample, not the population (Uprichard, 2013). Thus, the findings can provide useful information about the population but cannot be generalized (Sekaran & Bougie, 2016). Examples include convenience, purposive, and snowball.

In convenience sampling, cases are selected based on easy access; purposive sampling allows the selection of members which bear characteristics suitable to the research purpose (Etikan et al., 2016). Uprichard (2013) argued that, in general, sampling is linked to research purpose and data analysis and, therefore, is conducted with a purpose in mind; i.e., purposive is central to all sampling techniques. In snowball, members of a population subsequently introduce others with the requisite characteristics; this is especially useful in cases where it is difficult to identify members of a population (Saunders et al., 2015).

In general, probability sampling is structured and, therefore, linked to the deductive approach, positivism and post-positivism, and survey strategies. By contrast, non-probability sampling is less-structured and associated with induction, constructivism, and case study. However, as shown above, the selection depends on, *inter alia*, suitability to research purpose and practical (e.g., feasibility) and ethical considerations (Saunders et al., 2015). Notwithstanding the distinction, probability and non-probability techniques may be combined in the same research, depending on the purpose.

Non-probability sampling is predominant in social research— sample frames and probability of elements in the population cannot be established, non-response bias, etc. (Rowley, 2014), and is widely employed in management research (Saunders et al., 2015). Data sources and selection have implications for this research design.

4.6.2.1 Implications for design

As stated above, data source, sampling techniques, and adequate sample size depend on the research purpose and characteristics of the population. People (rather than documents, artefacts, etc.) are well-placed to give an account of their practice (i.e., how e-government outcomes evaluation is done— decisions, actions, motivation, attitude, etc.). Thus, consistent with the descriptive research purpose, people (i.e., evaluation stakeholders) were considered an appropriate source of data.

Probability sampling was impractical (e.g., an up-to-date sample frame may not be available); a purposive, snowball sampling technique was adopted. The goal was to enable the selection of people with the requisite e-government outcomes evaluation experience and to assure a larger sample size (see section 5.1).

Although the perspectives of all public sector stakeholder groups are necessary to enrich findings, only public employees were considered. The reason is twofold: public employees are central to service delivery, evaluation, and improvement as enjoined by public value and the *Batho Pele* (see section 2.7.1); and is it easier (e.g., time and budget) to access members of the public with requisite e-government outcomes evaluation experience. Pietersen (2014) selected public employees for similar reasons. Furthermore, Simonofski (et al., 2017) observed evaluation from the perspective of public organizations is uncommon; thus, this research, contributes to the literature.

The sampling strategy above means the findings cannot be representative of all stakeholders and, therefore, not generalizable beyond the sample. Although triangulation of data sources

(e.g., people, documents, artefacts, etc.) may improve the credibility of findings, this was considered impractical due to factors such as time and budget. Sampling decisions are considered in detail in section 4.8 and the respective phases (Chapters 5 and 6). The next section provides an overview of data generation methods and their implications for this research.

4.6.3 Data collection

Document analysis, observation, interview, and questionnaire are applicable to survey (De Vaus, 2014). Each method has strengths and shortcomings and should be evaluated and selected against the backdrop of a unique research context, i.e., research purpose, type of data to collect, method of analysis and practical constraints, etc. (Saunders et al., 2015). The debate on methodology in practice research has been outlined in section 3.2.3.

For this research, the larger sample size required to describe patterns of e-government outcomes evaluation makes analysis of document or "visiting the scenes" of actual performance impractical and are, therefore, elided from further discussion. Questionnaire and interview were considered appropriate. These are suitable for gathering data on characteristics, behaviours, attitudes, experiences, and opinions of people (Saunders et al., 2015; Walliman, 2018)

Questionnaire allows collection of standardized responses from a larger sample over a wide geographical area at a low cost, which can be compared (to establish patterns) and generalized, e.g., to the sample itself or the entire population (Rowley, 2014; Saunders et al., 2015; Walliman, 2018). A higher response rate is important for description. There are various online and offline modes of administration. The shortcomings of structured responses are limited detail and the lack of opportunity to follow-up for further explanation.

Interview yields detailed information to improve understanding, but is time consuming and, therefore, supports a smaller number of cases (Rowley, 2012; Saunders et al., 2015; Walliman, 2018). There are different forms, e.g., structured, semi-structured, and unstructured. A semi-structured interview combines prior organization (e.g., selection and order of questions) and flexibility (e.g., to change questions or ask follow-up questions).

Questionnaire and interview, as self-report methods, suffer from reticence and socially desirable responses. Thus, it is important to consider this in instrument construction and administration (see section 4.6.5).

4.6.3.1 Implications for design

The above shows questionnaire and semi-structured interview lend themselves to descriptive research, post-positivism and survey. The two options are a questionnaire embedded with open-ended questions or a questionnaire followed by an interview. The former is usually lengthy, time-consuming for participants, and prone to a low response rate (Rea & Parker, 2014; Walliman, 2018); the latter affords follow-up on structured responses and is gaining currency in survey research (Check & Schutt, 2012).

A questionnaire followed by an interview was adopted. Instrument construction and validation are discussed in Phases 1 and 2 (see Chapters 5 and 6). Research data is generally meaningless and of no use unless processed, through analysis, into evidence.

4.6.4 Data analysis

The purpose of data analysis and interpretation is to process a large volume of research data to describe, explain or predict patterns. This involves data preparation, organization, and presentation. The methods may be classified as statistical and non-statistical, and an appropriate choice depends mainly on the research purpose and type of data.

Statistical analysis manipulates structured numeric or textual data (De Vaus, 2014). It serves two broad purposes: descriptive analysis can identify relationships in the data and draw conclusions about the sample while inference analysis can examine causal relationships and make predictions (or generalize) to the population. By contrast, non-statistical analysis manipulates unstructured data— developing, comparing, and classifying data into higher level categories or themes from which to draw conclusions.

Notwithstanding the distinction, statistical and non-statistical are often combined, e.g., frequency of specific data in unstructured text; quantitative exploratory analysis methods (such as factor and cluster analyses) involve subjective decisions.

4.6.4.1 Implications for design

Statistical and non-statistical methods were combined, consistent with the descriptive research purpose and data involved, to addresses the concept of practice as "both entity and patterns of performance". Descriptive, exploratory, and cluster analysis served to describe the characteristics of participants and their responses, the important elements (i.e., factors) of EOEP, and how these "clustered" during performance (i.e., patterns of performance). Subsequently, non-statistical data analysis (mainly content analysis) was employed to analyze

the unstructured data to gain a better understanding. Data analysis is discussed in detail in section 4.8 and also in the respective phases, i.e., Chapters 5 and 6. Another design consideration is how to assure the quality of research outcomes.

4.6.5 Validation

It is necessary to address quality in research design to obviate common threats to outcomes, e.g., researcher and participant behaviour and biases, errors in instruments and measurement, etc. (Leedy & Ormrod, 2015; Saunders et al., 2015). The techniques to evaluate quality are different for deductive and inductive research.

Deductive is about quantification, i.e., conceptualization, objective operationalization, and quantitative measurements. Reliability and validity (i.e., consistent and accurate measurement) are important (Leedy & Ormrod, 2015; Saunders et al., 2015).

Internal and external validity (i.e., consistency between empirical observations and the conclusions, and the extent to which conclusions may apply to similar contexts) are commonly used to measure validity (Bryman, 2016). Adoption of standardized measurements, valid operational definition of concepts, review by experts, and appropriate sample may improve validity (Saunders et al., 2015).

On reliability, there are several techniques (Bhattacherjee, 2020; Leedy & Ormrod, 2015). Internal consistency (measured by Cronbach alpha) and expert review are commonly used. Test-retest; interrater; and alternative-form were impractical and, therefore, not discussed further. For example, this research was a requirement for an academic course, thus, only one researcher was involved. Also, instrument administration, multiple times and over a period, was infeasible— time and budget.

There is no yardstick to evaluate the quality of inductive research. However, trustworthiness (i.e., credibility, transferability, dependability, and confirmability— equivalent to internal validity; external validity, reliability, and objectivity respectively) is generally preferred to reliability and validity (Bryman, 2016). Section 4.6.7.6 provides an outline of validation in mixed methods.

In mixed methods research, quality extends beyond the sum of the quantitative and qualitative components (Creamer, 2018). For example, validation of the integrated findings (i.e., findings from the respective phases) is important (Venkatesh et al., 2016).

4.6.5.1 Implications for design

As shown above, the behaviour and biases of the researcher and participants, instruments and their administration, etc., affect the quality of research outcomes. To avoid pitfalls, strategies and tactics were developed upfront: design decisions were based on ethical issues and careful evaluation of available options. This research design bears testament— and also fulfils the call to document mixed methods research to improve transparency (Creamer, 2018). Various guidelines were considered to draw checklists and strategies (e.g., Tables 1.1, 5.1, 6.1, and 6.2) to guide the research process. Validation was addressed in the quantitative and qualitative phases and integration (see Chapters 5, 6, and 7 respectively). This formed part of an overall mixed methods validation strategy formulated in Table 1.1 and corroborated in Table 7.1.

Besides purpose, paradigm, and strategy, research may also be classified by the type of data and applicable methods (Bryman, 2016; Saunders et al., 2015). Three main archetypes of research may be identified.

4.6.6 Quantitative, qualitative and mixed methods research

Research data may be classified variously, *viz*., quantitative (i.e., numeric, about quantity, amounts, etc.) and qualitative (i.e., non-numeric, about meaning); structured (i.e., organized) and unstructured (i.e., unorganized), closed-ended and open-ended, etc. (Creswell, 2014; De Vaus, 2014; Walliman, 2018). In general, quantitative methods are structured and support quantitative data whereas qualitative are unstructured and support qualitative data. As may be evident from the sections above, this classification is linked to the scientific method.

Quantitative research is based on the scientific method and attendant quantification (Saunders et al., 2015). Quantitative methods (e.g., deduction, probability sampling, questionnaire, and structured interview and statistical analysis) are generally employed (De Vaus, 2014). Quantitative research supports description and prediction and, therefore, larger samples; validation involves reliability and validity (Bryman, 2016; Leedy & Ormrod, 2015). According to Creswell (2014), and as shown in sections 4.3 to 4.5, there is a link between positivism and post-positivism, survey and quantitative research.

On the other hand, the purpose of qualitative research is to gain understanding and is, therefore, inconsistent with the scientific method (Leedy & Ormrod, 2015; Wilson, 2014). In general, qualitative methods (e.g., induction, non-probability sampling, unstructured interview and observation, and non-statistical analysis) are employed to support intensive examination of smaller samples; quality of research is measured by trustworthiness (Bryman, 2016; Saunders

et al., 2015). According to Creswell (2014), and shown above (see sections 4.3 to 4.5), there is a link between constructivism, case study and qualitative research.

The distinction between quantitative and qualitative research by data, methods and paradigms is not inviolable (Morgan, 2018). As emphasized above, real-world problems are usually complex, may not admit to one or other; therefore, research may involve some degree of extant theory, deduction and induction, subjectivity and objectivity, numeric and non-numeric data. Thus, as shown in Figure 4.2, quantitative and qualitative research are not dichotomous but represent opposite extremes of a continuum on which a combination of data and applicable methods occur.

Figure 4.2: The quantitative-qualitative methods continuum

There is no consensus on what constitutes mixed methods research, e.g., whether it combines different methods, data sources, paradigms and applies to single or multiple research (Creswell, 2014; Morse & Niehaus, 2016). For the purpose of this research, it denotes a combination of components of qualitative and qualitative research within a single or across multiple research (Venkatesh et al., 2016). That means the combination of quantitative and qualitative data as well as methods from among logic of inquiry, sampling, and data collection and analysis (Schoonenboom & Johnson, 2017).

There is a general agreement, despite the polemics, that mixed methods research not only supports different research purposes and questions simultaneously but also draws on the individual strengths of quantitative and qualitative research and reduces their weaknesses (Jokonya, 2016; Venkatesh et al., 2016). Mixed methods research is gaining prominence in evaluation (Mertens & Tarsilla, 2015), IS (Venkatesh et al., 2016), and practice (Browne et al., 2015).

Notwithstanding the potential benefits, there are challenges; e.g., compared to mono-method research (i.e., either quantitative or qualitative), it is more complex, difficult to implement, and demands more resources (Creswell, 2014; Morse & Niehaus, 2016). Several guidelines have emerged to assist practice (e.g., Creamer, 2018; Schoonenboom & Johnson, 2017; Venkatesh et al., 2013, 2016).

4.6.6.1 Implications for design

As shown above, mixed methods research supports quantitative and qualitative data and methods, and lends itself to the research purpose, problem (i.e., the nature of practice), paradigm, and strategy. Mixed methods research was, therefore, considered appropriate. The next sections outline the characteristic features and corresponding design decisions, and how these were applied in conjunction with the other elements (sections 4.2 to 4.5) to develop an overall strategy for this research.

4.6.7 Features of mixed methods research

A well-designed mixed methods research explicates the following: purpose; underlying philosophy; order (i.e., timing) of components; point of integration; validation techniques; and a visual representation of the strategy (Creamer, 2018; Mckim, 2017; Morse & Niehaus, 2016; Schoonenboom & Johnson, 2017; Venkatesh et al., 2016).

4.6.7.1 Purpose

The purpose determines how the components are linked. There may be several, e.g., to compare different perspectives of a problem, gain in-depth understanding after an extensive investigation, to gain an understanding of an under-developed field to create measurement instrument, etc. Thus, the first step in designing mixed methods research is to explicitly state the purpose.

4.6.7.2 Underlying paradigm

Philosophical assumptions determine the status of components and an appropriate implementation strategy. Three types of mixed methods may be identified by status, therefore, it is important to specify a choice and clarify the philosophical and theoretical positions thereof. Qualitative-dominant combines quantitative component with a predominantly qualitative research process informed by constructivism. In a quantitative-dominant, the qualitative component is combined with predominantly quantitative research supported by post-positivism. Equal status arrogates equal status to the constituent components— and suggests a combination of paradigms.

4.6.7.3 Timing

This concerns the order and dependence among components. There are two main types, according to order. In concurrent, the quantitative and qualitative components are implemented separately and simultaneously, and their findings compared. The underlying assumption is that

each provides a different perspective of a phenomenon and, therefore, evidence from one can be used to confirm or disconfirm the other. In sequential, the findings from one component form the basis of the next, and both are subsequently integrated such that findings in the second phase can help expand on the first.

Creswell (2014) further divides sequential mixed methods research into explanatory and exploratory. In explanatory, an initial quantitative phase is followed by qualitative; the purpose is to apply qualitative findings to explain the quantitative. Exploratory, however, involves an initial, qualitative phase followed by quantitative; the purpose is to explore a phenomenon in detail and apply quantitative techniques to gain a broader perspective. Various strands of mixed methods research, based on order and dependence are also possible (see Schoonenboom & Johnson, 2017; Venkatesh et al., 2016). Evidently, research purpose and philosophical assumptions determine the appropriate strategy (i.e., sequential and explanatory or exploratory; or concurrent).

4.6.7.4 Point of integration

Integration is intrinsic to mixed methods research. There two main design decisions are: what elements (e.g., purpose, data, methods, methodologies, and paradigms) to integrate and the stage (i.e., "interface") in the research process at which the quantitative and qualitative components come together. Methods (e.g., the logic of inquiry, sampling, and data collection and analysis) may be fully or partially integrated across all components (Venkatesh et al., 2016). Often, however, components are integrated at the "results point of integration" (Schoonenboom & Johnson, 2017, p. 115). For example, based on the research purpose, a sequential strategy may separately present quantitative and qualitative results in respective components and subsequently integrate such results to draw conclusions.

4.6.7.5 Prior design and use of typology

The quality of research outcomes derives from both design (e.g., appropriateness of selected methods to purpose and philosophical assumptions) and implementation. As shown above, prior design helps to organize the various elements of research to assure congruency between purpose, process, and outcomes. Schoonenboom and Johnson (2017) aver it is helpful to adopt an archetype of mixed methods research and plan upfront. Typologies (e.g., sequential explanatory, sequential exploratory, embedded, etc.) have been developed to afford common ways of understanding and undertaking mixed methods research. This way, the idiosyncrasies of a specific mixed methods research can be integrated with other general elements of research

design. Sequential explanatory typology (see section 4.8) was adopted and the features incorporated into the design of a survey strategy.

4.6.7.6 Validation

Section 4.6.5 showed it is important to address the quality of mixed methods research. There are various guidelines (e.g., Creamer, 2018; Mckim, 2017; Venkatesh et al., 2013) to support design and implementation, to assure outcomes are consistent with the research purpose. In general, validation in mixed methods research consists of two steps. Each component is validated separately, according to the relevant traditional yardstick as discussed in section 4.6.5— by quantitative (i.e., reliability and validity) and qualitative (i.e., credibility, transferability, dependability and confirmability for qualitative). This is followed by the validation of integrated findings, against the backdrop of the overall research strategy. According to Creamer (2018), visual display devices (e.g., tables and figures) can help document the research process and outcomes thereof, improve transparency and, thereby, credibility.

4.6.7.7 Notation

Mixed methods research is more complex to design and implement, compared to mono-method. According to Creswell (2014) and (Creamer, 2018), visual illustrations help to explain to the research audience. Morse and Niehaus (2016) discuss the notations for designing the common typologies. The symbols QUAN, QUAL, quan and qual represent quantitative dominant, qualitative dominant, quantitative less-dominant, and qualitative less-dominant (the upper and lower case represent dominant and less-dominant respectively). Also, + and → represent concurrent and sequential. So, QUAN → qual represents sequential mixed methods research in which the quantitative component is dominant and the qualitative is less-dominant while QUAL + quan denotes concurrent qualitative dominant and quantitative less-dominant. This notation is used throughout this research. Creamer (2018), Creswell (2014, 2015), Jokonya (2016) and Wu (2012) provide diagrammatic illustrations— applied in section 4.8.

4.6.7.8 Implications for design

This research was primarily descriptive, composed of two phases: Phase 1 (to provide quantitative data for description) followed by Phase 2 (to provide qualitative data for understanding). Consequently, a quantitative dominant, sequential and explanatory mixed methods research was adopted (i.e., QUAN → qual).

Although sequential explanatory mixed methods (SEMM) research is the most dominant social research (Liem, 2018), extant empirical IS research is scant. Some studies (e.g., Al-Yaseen et al., 2008) applied quantitative followed by qualitative but did not situate in the theoretical foundations of mixed methods research while others (e.g., Jokonya, 2016) were conceptual.

The research design integrated the traditional research decisions about purpose, conceptual framework, paradigm, strategy, data, methods and quality into an overall mixed methods strategy (see section 4.8). Research may also be categorized by duration, as discussed below.

4.7 Time horizon

Research may be classified, according to duration, into cross-sectional and longitudinal (Bryman, 2016; Saunders et al., 2015). Cross-sectional research examines a phenomenon as it is, at one point in time to provide a once-off snapshot and, therefore, is well-suited to description. On the other hand, longitudinal research collects data several times over a longer period and, therefore, can track a phenomenon over time to explain changes thereto. Cross-sectional may be less time-consuming and cheaper to conduct but lacks explanatory power longitudinal (Leedy & Ormrod, 2015).

Time horizon is independent of the other elements of research, e.g., purpose; paradigm; strategy; etc. (Bryman, 2016; Saunders et al., 2015) and may be applied to sub-classify research types (e.g., cross-sectional, quantitative case study). However, in general, cross-sectional research is linked to quantitative research (Bryman, 2016) and used in descriptive surveys to describe the occurrence of a phenomenon or correlations among factors (Saunders et al., 2015).

Cross-sectional research suited the research purpose (i.e., to describe the current state of e-government evaluation practice) and was preferable for practical considerations: this research was a requirement for an academic course and the researcher could not afford the time and budget to conduct several investigations over an extended period, as required in longitudinal. The downside was that changes in EOEP could not be tracked.

Sections 4.2 to 4.7 discussed the elements of research, the competing options, design choices, and justification thereof. The following explicates the overall research strategy.

4.8 Adopted strategy: sequential explanatory mixed methods (SEMM) survey

This strategy may be described as a cross-sectional, descriptive, mixed methods survey, underpinned by post-positivism. By convention, the sequential strategy is depicted as vertical (Creamer, 2018). As illustrated in Figure 4.3, this consisted of two phases: quantitative-dominant (Phase 1) followed by qualitative-less dominant (Phase 2), i.e., QUAN → qual. Phase 1 identified the main elements which described e-government outcomes evaluation and the patterns of performance and the results used to design and implement Phase 2. Subsequently, the results of both phases were integrated to draw conclusions.

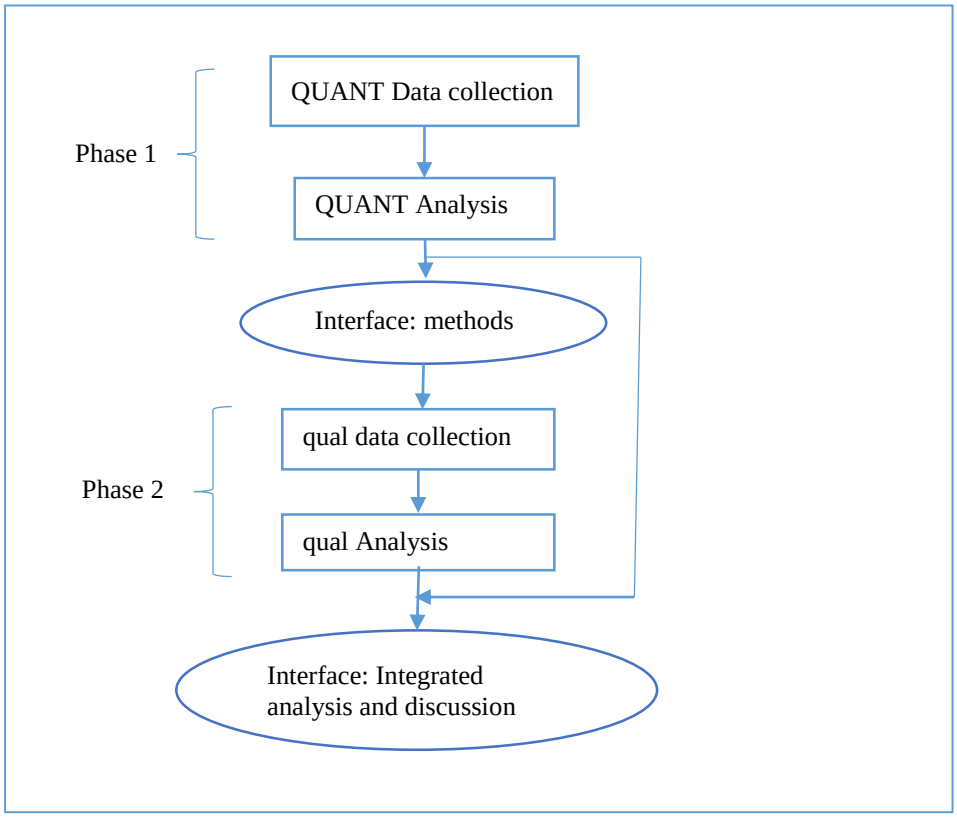


Figure 4.3: Strategy for this research

Table 4.4 summarizes and maps design decisions onto the respective phases in Figure 4.3. This strategy, it may be observed, combined quantitative and qualitative methods and results. The two phases formed the content of the next two chapters (i.e., Chapters 5 and 6), a brief overview of which follows (see sections 4.8.1 and 4.8.3).

This strategy has a longstanding tradition in IS research, notwithstanding the generally low mixed methods adoption rate. For example, Al-Yaseen et al. (2008) adopted a two-phased approach to examine practitioners' perceptions of how post-implementation evaluation of IT

is done in large business organizations (quantitative, to describe current practices followed by qualitative, to understand results of the previous phase). The research contribution was to: add to the body of knowledge (i.e., provide evidence on practice) and understand the results to paint a better picture— similar to this research.

Table 4.4: Summary of design decisions

	Component	
Feature	*Phase 1*	*Phase 2*
Purpose	To describe the main elements of practice and patterns of performance; examine differences in perceptions among groups	To understand the preliminary results of Phase 1
Research type	Quantitative	Qualitative
Paradigm	Quantitative (dominant)	Qualitative (less-dominant)
Data source	Public sector employees with experience	Sub-sample from the sample in Phase 1
Data	Quantitative	Mainly qualitative
Logic of inquiry	Both deductive and inductive	Both deductive and inductive
Sampling	Purposive; snowball	Purposive, stratified sampling
Data collection	Questionnaire	Semi-structured interview
Data analysis	Statistical	Non-statistical
Interpretation and conclusion	Integration of quantitative and qualitative findings in Phases 1 and 2	
Validation	Quantitative (validity and reliability) followed by qualitative (credibility; transferability; dependability; confirmability)	
Limitations	Retrospective (relied on recall, which may not reflect the actual) Self-report (which may result in reticence, socially desirable responses, etc.) Purposive sampling: sample not representative; results not generalizable, etc.	

4.8.1 Phase 1: Quantitative research

Primarily, deduction was employed to integrate concepts from e-government, practice, IT and public sector reform, public value, and evaluation practice (i.e., program, IT, and e-government) to construct a conceptual model. Sampling involved purposive and snowball, to select public employees. Questionnaire was developed, piloted and administered to collect quantitative data on participants' perceptions of EOEP. Statistical analysis involved mainly descriptions and comparisons. Descriptive statistics provided, *inter alia*, overviews (e.g., frequencies) of participants' characteristics and responses. Exploratory data analysis established relationships in the dataset: factor analysis, to identify the main elements which described e-government outcomes evaluation (i.e., as an entity) and cluster analysis, to explore the different ways the factors combined to form the patterns of performance (i.e., practice as performance). Differences among stakeholder groups, e.g., according to their perceptions of elements, which described e-government outcomes evaluation and performance were also explored. Measures were undertaken to ensure reliability and validity.

4.8.2 Phase 2: Qualitative research

Based on the results in Phase 1, an interview protocol was constructed to follow up on what participants viewed as significant factors and patterns of performance. Participants of Phase 1 constituted an appropriate sample frame of public employees with requisite experience. As discussed in Chapter 3, different groups are likely to have different perspectives of how e-government outcomes evaluation is done. Thus, a sub-sample was selected to afford comparison among groups.

Semi-structured, face-to-face interviews were conducted to obtain a rich account of participants' practice. Responses were analysed with a qualitative data analysis technique, namely, content analysis. Steps were taken to ensure trustworthiness. Subsequently, the results of both phases were integrated and discussed, in conjunction with the literature, to draw conclusions.

4.8.3 Organization and presentation of research

The way mixed method research is organized and reported is also important. There is no standard structure; a suitable strategy is one which matches the research design (i.e., the order of the phases) and supports understanding of the audience of this research. There are several guidelines to assist in this endeavour (Creamer, 2018; Fetters & Freshwater, 2015; Jalongo & Saracho, 2016). In SEMM, it is common to report the quantitative and qualitative in the respective phases and integrate the results at a later stage for further analysis; Venkatesh et al. (2016) is an illustration. Implementation of Phase 1 and Phase 2 are reported in detail in Chapter 5 and Chapter 6 respectively; the integration is discussed is in Chapter 7. As stated in section 4.6.7.6, visual display devices help to improve understanding and were, therefore, employed.

Research involves choices and trade-offs, therefore, it is important to make decisions which are not only suitable and feasible but also ethical— especially where human participants are involved. Saunders et al. (2015) adjured ethical considerations throughout the research process. Furthermore, the University demands strict adherence to its code of ethics.

4.9 Ethics

Research ethics concerns a researcher's moral choices and behaviour as well as responsibility to, and relationship with, participants. The two main issues are: integrity of the research process to assure the credibility of outcomes and how to treat human participants (Walliman, 2018).

According to Hassan et al. (2018, p. 269), it is about "right and wrong, good and bad, do's and don'ts"— in brief, how to evaluate the researcher's duties, actions, and consequences. Section 4.3 showed ethics and values are central to axiology and, therefore, philosophical assumptions. Four main categories may be identified: harm, informed consent, confidentiality, and professionalism (Leedy & Ormrod, 2015). According to Wallace and Sheldon (2015), the basic principles of research ethics include autonomy, beneficence, and justice.

Autonomy is about the right to decide, which involves informed consent and encompasses participants' rights to, e.g.: know the nature of the research, participate voluntarily, withdraw at any stage and suffer no consequence, privacy and confidentiality, etc. A cover letter and consent form are two common devices for informed consent (Leedy & Ormrod, 2015; Saunders et al., 2015).

Beneficence is about benefits to, and well-being of, participants; i.e., protection from physical, psychological, social, and all other forms of harm while maximizing potential benefits. Justice concerns how participants are selected and treated throughout the research; e.g., to select a sample without bias and to treat participants with dignity. Professionalism enjoins the researcher, *inter alia*, to exercise integrity in decisions and actions, adhere to institutional and professional ethics codes, and acknowledge the work of others. Confidentiality relates to data accuracy and authorized access; data collection, storage, processing, distribution, and disposal practices may not infringe on the right of participants to privacy and confidentiality.

Ethical considerations informed the research design and implementation. University (and faculty) ethics code was complied with— an ethical clearance was sought and obtained (see Appendix F). Extant research used herein was duly acknowledged. Drawing from guidelines, a cover letter and consent form (Appendix B) were designed. The letter solicited participation and stated the research purpose, rights of the participants, efforts to avoid harm, and contact persons for further enquires. The consent form also documented participants' written consent.

Efforts were made throughout the research process to avoid any acts which could adversely affect participants and to uphold their right to dignity and anonymity. The research topic and objectives posed no harm to society. Sampling, data collection, and analysis were implemented without bias; and the results were presented objectively. No personal, confidential, or sensitive data was collected.

The data was accurately recorded, processed, and used. Physical copies of responses were stored under lock and key; electronic copies were stored in an encrypted, password-protected

folder accessible only to the researcher. When appropriate, the physical copies would be shredded and the electronic, permanently deleted and overwritten. Ethical issues unique to Phases 1 and 2 (i.e., quantitative and qualitative components) are dealt with in Chapters 5 and 6 respectively.

Summary

Research involves complex decisions, trade-offs and pitfalls, therefore, it is important to plan upfront. Research design describes the elements of research and documents competing options, choices, and their justifications; due consideration can assure congruency among the elements, guide the research process and improve the quality of outcomes.

This chapter explicated an overall research strategy. A cross-sectional, SEMM research survey suited the research purpose (i.e., descriptive) and supported the objectives (i.e., to describe the elements of EOEP and the patterns of performance). The purpose of mixed methods was, therefore, to use qualitative research findings to understand the quantitative. Post-positivism was adopted as an appropriate paradigm to support mixed methods research. This is compatible with the nature of, and research on, public value, practice, and evaluation. Public employees with experience of EOEP were considered suitable data sources. A mixed methods validation strategy was implemented to improve research quality.

The research design shows that the research purpose is congruent to the other elements (i.e., paradigm, strategy, type of research, and time horizon). It is to the implementation of the phases to which the next two chapters (Chapters 5 and 6) turn.

.

Chapter 5 : Phase 1 (Quantitative data collection and analysis)

The preceding chapter presented the research design— a plan, which outlined a two-phased sequential explanatory mixed methods research strategy. This chapter is about the implementation and results of Phase 1, i.e., quantitative phase, and answered the following questions:

RQ1: What are the factors which describe how e-government outcomes evaluation is done?
RQ2: What are the patterns of performance among different groups?

Research is practically undertaken through methods. Thus, this chapter described how sampling, questionnaire construction, data collection and analysis, and validation were implemented. The design decisions and applicable implementation tactics are reiterated in Table 5.1 for accessibility.

Table 5.1: Research strategy for Phase 1

Phase	*1*
Rational	To describe the main elements of practice and patterns of performance Examine differences in perceptions among groups
Research type	Quantitative
Research question answered	Question 1 and 2
Type of data	Quantitative
Logic of inquiry	Primarily deductive, supported by inductive
Sampling	Purposive; snowball
Data collection	Questionnaire (mainly closed questions)
Data analysis	Statistical: descriptive, factor, and cluster analyses
Validation	Validity and reliability

A questionnaire was administered to public employees with experience in e-government outcomes evaluation. Of those completed and returned, 106 were usable. The data was analyzed using descriptive, factor and cluster analyses. Six factors that described participants' perceptions of EOEP and three patterns of performance were identified. In both cases, there were significant differences among participants. Figure 5.1 illustrates the structure of this chapter.

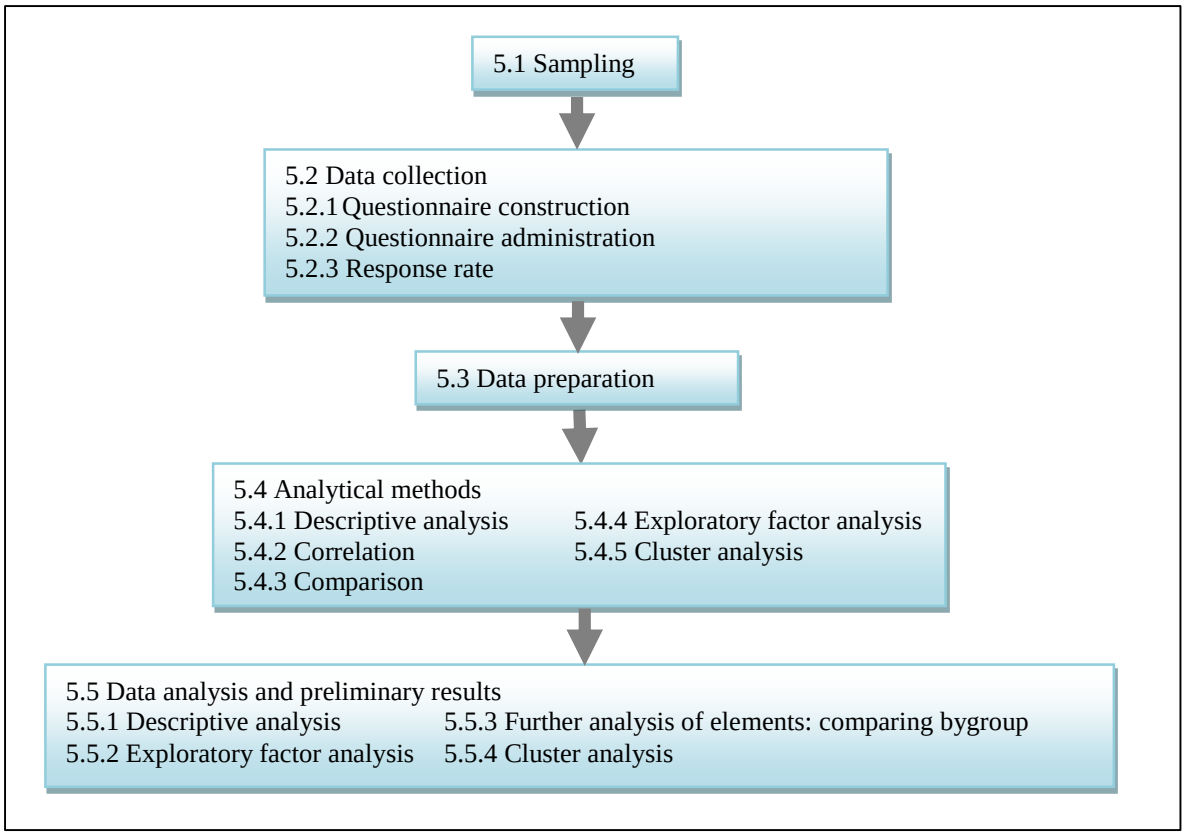


Figure 5.1: Structure of chapter

Chapter 4 showed ethical considerations are woven into the fabric of this research. Efforts were made to uphold the rights of participants, ensure their well-being, address data confidentiality, and assure integrity and transparency of the research process.

5.1 Sampling

As shown in Chapter 2, public value (i.e., public value, public values, and public sector values, e.g., *Batho Pele*) emphasizes the participation of both internal and external stakeholders in the evaluation of public service. However, for practical reasons (as explained in 4.6.2.1), only public employees were considered. A purposive, snowball sampling was adopted to select employees with e-government outcomes evaluation experience.

Some authors (e.g., Bryman, 2016; Rowley, 2014; Saunders et al., 2015) contend reliance on one's network of contacts to select participants can improve response rate. Information Technology (IT) managers in public organizations were more likely to be involved with e-government outcomes evaluation and know other employees with requisite knowledge and experience. Thus, sample selection commenced with contacts by the researcher to a network of colleagues, friends, and acquaintances to identify such IT managers. These were apprised (in personal visits, emails, and phone calls) of the research as well as potential benefits,

requested for their consent for participation, and also asked to introduce other public employees with relevant experience. As aforementioned, quantification involves the operational definition of relevant concepts, translation into variables, and quantitative measurements and analysis.

5.2 Data collection

Operational definition denotes how a variable is measured and encompasses the type of data, level of measurement, and applicable operations or methods (Bhattacherjee, 2020). Validation is important, to assure reliability and validity (see section 4.6.5). Inherent to a questionnaire-based survey is an adequate sample for results to be statistically significant (Rowley, 2014; Walliman, 2018).

Research contexts vary, thus, there is no standardized questionnaire for EOEP; a new one was constructed. Such an endeavour is fraught with threats to reliability and validity. Various questionnaire design guidelines were considered to help reduce the pitfalls.

5.2.1 Questionnaire construction

Some of the features addressed to improve quality and increase response rate included questionnaire length, clarity, acceptability, and comprehensiveness (Bhattacherjee, 2020; Moser & Kalton, 2017; Rea & Parker, 2014; Rowley, 2014; Saunders et al., 2015). The questionnaire items were derived from extant research and theory.

There are three important factors to consider in measurements: the respondent, the object under examination, and the attributes thereof (Dolnicar, 2013). An abstract object or its attribute may be defined by several components. This research concerned public employees, i.e., the respondent; EOEP, i.e., object; and perception of EOEP, i.e., the attribute. Both the object and its attributes are abstract and multidimensional and, therefore, may be defined by components (or dimensions).

The questionnaire construction process followed an outline by De Vaus (2014), to wit, identify and define constructs that represented the relevant concepts, specify their dimensions and sub-dimensions, develop items for each construct, and pilot to test the questionnaire. Thus, the table of relevant concepts identified in the literature (see Table 3.2) and the conceptual model (Figure 3.2) formed the basis of measurements. The questionnaire items were designed to collect data on organizational factors as well as participants' demographics and their perceptions and experiences (Moser & Kalton, 2017; Walliman, 2018). EOEP constituted the unit of analysis and, according to the conceptual model, the primary concept was practice as a bundle.

In brief, a bundle is the site of practice and consists of activities, material entities, and *connection*s. Activities are actions performed; material entities encompass both human and non-human; *connection* (i.e., practical understanding, e.g., technical know-how; general understanding, e.g., shared understanding, values, etc.; rules, e.g., policies; and teleoaffective structures, e.g., ends and affect) ties activities and material entities together to form a bundle.

Thus, e-government outcomes evaluation constitutes a practice bundle. This consists of *Assessment* (i.e., doing the evaluation) and *Use* (i.e., the effect of evaluation) bundles, which may share their elements. There can be different ways in which the elements (e.g., activities, material entities, and connections) may combine. That is, there can be different patterns of performance by different groups of people, e.g., managers and non-managers; professional evaluators and other stakeholders, etc.

Evaluation context encompasses program (i.e., e-government), organizational (i.e., public sector), and stakeholder characteristics. Public value constitutes the evaluation framework and prescribes the outcomes to measure (i.e., efficiency, effectiveness, and democratic values) and how (i.e., process, including stakeholder participation). Stakeholder issues include demographics, the extent of participation, and affective issues. Table 5.2 shows the aspects of practice about which data was collected: practitioners, as evaluation stakeholders (Part I) and their practice (Part II).

Part I concerned demography and measured participants' position in their organization (i.e., management non-management), professional backgrounds (e.g., Information Technology, Evaluation, and Other), and experience (years and knowledge) in EOEP.

Items 1, 2, and 4 are self-explanatory. Item 3 ("Years of experience") was also a screening question— to filter out participants with no experience. Thus, responses for participants with no (i.e., zero) years of experience were not considered for further analysis. Item 5 ("No. of evaluations in the last two years") was to establish activeness and recency of participation in EOEP. Song and Letch (2014) argued that participants may easily recall perceptions and experiences within two years.

Part II measured how e-government outcomes evaluation is done (i.e., EOEP). Thus, items 6 to 11 primarily measured the elements of practice (i.e. activities, material entities, and *connections*). The link between Part I and Part II is apparent: to enable comparisons among participants, e.g., to establish differences in practices, according to demographics. Item 11, further elaborated in Table 5.3, focused on participants' experiences and perceptions.

Table 5.2: Operationalization of items

	Item	*Variable*	*Description*	*Source*	*Question type*	*Scale*
Part 1	1	Position	Participant's position in an organization: management or /non-management	(Coryn et al., 2015)	Single response	Ordinal
	2	Professional background	Participant's professional background	(Brandon & Fukunaga, 2014; Fierro & Christie, 2017; Gagnon et al., 2018)	Single response	Nominal
	3	Experience	Years of experience in e-government evaluation (screening item to exclude participants with no experience)	(Coryn et al., 2015; Cullen et al., 2011; Fierro & Christie, 2017)	Single response	Ordinal
	4	Knowledge	Participants' knowledge of e-government evaluation	(Brandon & Fukunaga, 2014; Christie, 2003; Fierro & Christie, 2017)	Single response	Ordinal
	5	No. of evaluations in the last 2 years	Activeness or recency in e-government evaluation	(Gagnon et al., 2018; Song & Letch, 2014)	Single response	Interval
Part II	6	Evaluation initiation	Responsibility for initiating evaluation	(Brandon & Fukunaga, 2014)	Single responses	Nominal
	7	Evaluation team	Responsibility for conducting evaluation (internal/ external)	(Bourgeois & Naré, 2015)	Single	Nominal
	8	Measurement strategies	The main strategies for conducting evaluation	(Bourgeois & Naré, 2015; Brandon & Fukunaga, 2014; Christie & Fleischer, 2010; Daigneault, 2014)	Multiple responses	Nominal
	9	What is evaluated	Dimensions of e-government outcome (i.e., value) assessed	(DPME, 2011) (Deng et al., 2018; eGep, 2006; Suhardi et al., 2015)	Multiple responses	Nominal
	10	Purpose of evaluation	The main types of evaluation use	(Daigneault, 2014; Gagnon et al., 2018; Song & Letch, 2014)	Multiple responses	Nominal
	11	Perception of e-government evaluation practice	The extent to which participants agreed or disagreed with statements about activities, material entities, and *connections*	See Table 5.3	Single responses	Likert

Table 5.3: Perceptions of e-government outcomes evaluation

	Measurement	*Source*	*Item*
Activities i.e., the actions undertaken in evaluation and use of results	Selection of stakeholders	(Cullen et al., 2011)	11.11
	Stakeholder training	(Brandon & Fukunaga, 2014; Gagnon et al., 2018)	11.12
	Data collection	(Brandon & Fukunaga, 2014; Daigneault, 2014; Gagnon et al., 2018)	11.16, 11.17, 11.18
	Development of reports	(Hansen et al., 2013)	11.20, 11.21
	Dissemination of reports	(Bourgeois & Naré, 2015; Kang et al., 2012)	11.19
	Management response	(Bourgeois & Naré, 2015; DPME, 2011)	11.23
	Follow-up to implement recommendations	(DPME, 2011; Gagnon et al., 2018; Mayne, 2014)	11.24
	Extent of participation in all stages of evaluation	(Chouinard & Milley, 2018; Cousins, 2013; Daigneault, 2014; Fierro & Christie, 2017; Fleischer & Christie, 2009; Szanyi et al., 2013)	11.15
Resources (i.e., resources for evaluation and use)	Human (expertise)	(Brandon & Fukunaga, 2014)	11.6
	Fiscal (budget)	(Gagnon et al., 2018; Wade & Kallemeyn, 2020)	11.5.
	Tools	(Brandon & Fukunaga, 2014; Gagnon et al., 2018)	11.7
Program context	Type of program (e.g., e-government)	(Bourgeois & Naré, 2015)	N/A
Organizational context	Country	(Christie, 2003)	N/A
	Sector (e.g., NGO, public sector, etc.)	(Bourgeois & Naré, 2015; Daigneault, 2014; Kang et al., 2012)	N/A
	Management support	(de Laat & Williams, 2014; Mayne, 2014)	11.1
	Regular evaluation of programs	(DPME, 2011)	11.3
	Regular (i.e., frequent) review of evaluation	(DPME, 2011)	11.4
General understanding	Agreement, i.e., shared view (on what to measure)	(Bittner & Leimeister, 2014)	11.13
	Agreement, i.e., shared view (on use)		11.14
	Evaluation model	(DPME, 2011)	11.8
	Evaluation policies (formal and informal)	(DPME, 2011; Mayne, 2014)	11.2
	Organizational policies (e.g., public sector values)	(DPME, 2011)	11.9
Teleoaffective structure (e.g., motivation; effects or outcomes of the process and product use)	Participation improves satisfaction (e.g., acceptance of recommendations)	(Brandon & Fukunaga, 2014)	11.22
	Participation improves evaluation knowledge	(Fierro & Christie, 2017; Preskill & Boyle, 2008a)	11.25
	Participation increases commitment to evaluation	(Brandon & Fukunaga, 2014)	11.26
	Participation increases willingness to participate in future evaluation	(Brandon & Fukunaga, 2014)	11.27

Items 6 and 7 measured responsibility for evaluation. In the public sector, the structure responsible for initiating (i.e., demand for) evaluation may be internal or external; also the evaluation may be conducted by internal or external teams, or a combination thereof. The literature showed there can be differences in outcomes, according to the responsibility for evaluation (DPME, 2011; Fleischer & Christie, 2009).

Item 8 concerned the main strategies for evaluation; item 9: the dimensions of e-government which are measured; and item 10: the purpose of evaluation (linked to types of use). Item 11 measured the extent to which participants agreed or disagreed with statements about how e-government outcomes evaluation is done. This consisted of 27 sub-items, as shown in Table 5.3, and organized around *Assessment* and *Use*.

Activity is the set of actions undertaken during the evaluation and use of results; e.g., selection of stakeholders, training, data collection, development and dissemination of reports, management response to recommendations and follow-up (to implement the recommendations), etc. Resources include human and non-human (e.g., finance and tools). As stated above, context included program, organizational, and stakeholder characteristics.

General understanding is about shared knowledge, meaning, assumptions, etc. (as may apply to a team) to accomplish work and connotes agreement (Bittner & Leimeister, 2014). Evaluation models, formal (e.g., organizational, evaluation, etc.) and informal policies, etc., provide assumptions to guide how evaluation may be done. General understanding was, therefore, operationalized as agreement, assumptions, etc., and mechanism (i.e., models, policies, knowledge, etc.) by which these may be achieved.

Teleoaffective structure is about ends and affect, which may be measured by effect (i.e., outcomes or changes in knowledge, behaviour and affect), as a consequence of participation in evaluation and use of the results thereof. Affective issues include motivation, commitment, and satisfaction (Brandon & Fukunaga, 2014).

An Agree-Disagree (AD) Likert scale was developed to measure the 27 items which constituted item 11 (Table 5.3). An AD Likert scale elicits direction (i.e., agree-disagree) and magnitude (e.g., strongly-not strongly) of participants' opinions about statements (Chyung et al., 2017). Thus, the scale measured the extent to which participants strongly agreed (or disagreed) with statements, which measured item 11.

The AD scale has strengths and weaknesses (Revilla et al., 2014). It is a popular data collection method because it is simple and practical; e.g., the same scale and instruction may apply to

different concepts and therefore, easier to complete— compared to item-specific scales. For this research, the same instruction applied to the 27 statements. The drawbacks include extreme and acquiescence response biases.

Several aspects of the AD Likert scale are open to debate, e.g., labelling (i.e., whether to label all points on the scale or not), length (i.e., number of point on the scales), even or odd number of points, the level of measurement and, therefore, applicable statistical analysis.

A fully-labelled scale may provide more information, reduce the cognitive effort to interpret options, and remove ambiguity in responses (Moors et al., 2014). Longer scales may provide more information and response options but require higher cognitive effort (e.g., to conceptually differentiate among the labels on an 11-point scale). Four- to seven-point scales are common (Leung, 2011). Saunders et al. (2015) also recommended four to seven-point scales because longer scales are usually collapsed (e.g., into two: agree or disagree) for analysis.

Both even and odd number of points (i.e., presence and absence of midpoint) have pitfalls. Midpoints are often selected for convenience, uncertainty, social desirability, etc., and result in response bias while the absence may foist a choice where neutrality is desirable (Chyung et al., 2017; Haladyna & Rodriguez, 2013; Leung, 2011).

Midpoints may be ambiguous and prone to misinterpretation (Chyung et al., 2017). For example, “neutral”, “undecided”, “not applicable”, “don’t know”, etc., may apply. Therefore, it is important to clearly label the midpoint of opinion as “neither agree or disagree” to express neutrality.

Also contentious is whether the values are ordinal or interval. As above, an abstract concept or construct may be best measured by multiple attributes. Thus, the set of items is generally considered as a composite measure and analyzed as interval with parametric statistics (Bhattacherjee, 2020; Carifio & Perla, 2008; Leung, 2011). Conversely, the individual scores are ordinal, well-suited to non-parametric statistics.

An appropriate scale may depend on the traditions of a field, demography of respondents, etc. (Chyung et al., 2017; Fellows & Liu, 2015; Leung, 2011). Fleischer and Christie (2009) and Gagnon et al. (2018) used 5-point and 7-point scales respectively in evaluation research; similarly, Welch and Feeney (2014) and Deng et al. (2018) used 5-point and 7-point scales respectively to study e-government outcomes. According to Chyung et al., knowledgeable participants may be allowed to express a neutral opinion on a topic.

A fully-labelled five-point scale ("Strongly disagree" "Disagree", "Neither disagree or agree" "Agree" and "Strongly agree") was adopted for this research. The sample was composed of people with requisite e-government outcomes evaluation experience (see section 5.1). Therefore, a fully-labelled five-point scale was considered appropriate, to reduce the cognitive effort and completion time (compared to longer, e.g., seven, nine- and eleven-point) and to afforded neutrality.

Efforts were made to improve both the reliability and validity of responses. The questionnaire (see Appendix C) was divided into two sections (Part I and Part II), corresponding to Table 5.2. and consisted of closed-ended items, except one. Section 4.6.5 outlined techniques (e.g., expert review, adoption of extant items, piloting, etc.) to improve the quality. The research adequately covered the important dimensions of practice and, therefore, EOEP. Schatzki reviewed the table of relevant concepts (Table 3.2) and a draft of the conceptual model to ensure these were aligned to the basic tenets of the site ontology and the feedback was used for improvement.

The questionnaire items were adopted from previous research (and adapted where applicable) or derived from theory, reviewed and revised— to ensure the items accurately measured the intended constructs. The questionnaire was reviewed by three academics (one with expertise in program evaluation and two in e-government) and two IT practitioners (in management positions with experience in e-government evaluation).

To improve general understanding (e.g., clarity of wording, logical arrangement of items, etc.), reduce the cognitive effort, and assure consistent responses, the questionnaire was piloted with eight post-graduate IS students with e-government background. It was duly revised and reduced to two pages and about 25 minutes average completion time (Rea & Parker, 2014; Rowley, 2014). This was subsequently distributed to participants.

5.2.2 Questionnaire administration

Data collection relied on self-reporting: participants responded to questions about their practice. Both offline (i.e., by post and in-person) and online (i.e., by email) were adopted. Contributing factors included time, cost, and sample profile (Phellas et al., 2011; Saunders et al., 2015; Walliman, 2018). The primary mode was e-mail because it is faster, cheaper, and can cover a larger geographical area—most public employees have e-mail addresses.

A link to a Google form version of the questionnaire was e-mailed, together with a Cover letter and Consent form. Participants were asked to submit the Consent form with the questionnaire.

To avoid duplication, Google forms was configured to allow only one response per e-mail address. Each response was numbered and directed to an Excel document, and a copy saved as pdf in a folder on Google Drive. The pdf copies served as backup and were later used for verification.

In-person administration can improve response rate (Rowley, 2014; Walliman, 2018) and complemented email. Survey packages (each numbered and consisting of a Cover letter, Consent form, and Questionnaire) were distributed to public employees in three provinces (i.e., Durban and Pietermaritzburg, in Kwa-Zulu Natal; East London and Bisho, in Eastern Cape; and Pretoria, in Gauteng). To avoid multiple completion (i.e., by email and in-person), the survey package was handed only to participants that had not submitted a response by email. Participants were implored to complete the questionnaire before an appointed date (on which the researcher returned for collection) and asked to introduce other employees with the requisite experience. Further to increasing response rate, in-person administration helped to establish rapport with participants for data collection in the next phase.

Instrument design, cover letter, incentives, and follow-up can also improve response rate (Bryman, 2016; Rowley, 2014). Section 5.2.1 discussed efforts to reduce cognitive effort and completion time. The cover letter assured participants of anonymity and guaranteed their protection from any harm, emphasized the relevance of the research, and offered a copy of findings as incentives for participation. On outstanding questionnaire responses, participants were contacted after ten days, thanked for participation, and reminded to complete and return.

5.2.3 Response rate

There is dissensus on what constitutes adequate cases for analysis. According to Rowley (2014), this is a function of theoretical and practical considerations; the research purpose, questions, and practical constraints (e.g., time and cost, which may limit the number). However, some statistical techniques (e.g., factor analysis, etc.) require larger numbers of cases for validity. Generally, a questionnaire-based survey has a low response rate, e.g., Rowley (2014) observed 20 percent is acceptable; according to Saunders et al. (2015), 35 percent is typical in organizational studies.

In all, 341 questionnaires were distributed (258 by email and 83 in-person), 108 were returned (75 and 33 by email and in-person respectively— 18 of which were returned after reminders to participants). Of this, 106 were valid and usable —a response rate of about 31%, which may

be considered acceptable. Rowley (2014) noted that 100 responses may be considered adequate for statistical analysis.

Non-response bias helps to explain representativeness or otherwise of survey findings (Leedy & Ormrod, 2015). However, section 4.6.2 showed that non-probability sampling was adopted because a sample frame of respondents and their characteristics could not be specified. Consequently, non-response bias could not be estimated as it was impractical to compare the characteristics of respondents and not-respondents. The data was submitted to quantitative analysis, which involved data preparation, analysis, and interpretation.

5.3 Data preparation

This included data screening, coding, capturing, and cleaning. A codebook of relevant variables (each with name, description, data type, and level of measurement) was constructed to code the questionnaire items. Each option was assigned a numeric value to facilitate the quantitative representation of responses. A valid range of values was specified for each code to reduce data entry errors. For example, professional background (i.e., IT, Evaluation, and Other) was assigned a numeric value which ranged from 1 to 3. The codebook was used to set up SPSS to capture responses.

Each questionnaire was screened for completeness and accuracy. This was built into the electronic version, e.g., responses were restricted to only the options provided; thus, submission by participants without requisite experience was not allowed. The questionnaire collected in person were manually screened on receipt.

To ensure data integrity, the "double-entry" approach was adopted. The responses were captured independently by two individuals. In each case, an Excel file which was generated from the email responses was imported into SPSS; subsequently, the responses collect in-person were captured into this file. The two datasets were compared and found to be similar; therefore, the data entry may be considered reliable. Cleaning entailed exploring the data to identify outlier values and effect the necessary corrections.

Analytical methods are based on assumptions, an understanding of which benefits how to use such methods. Thus, a discussion of the methods precedes their application.

5.4 Analytical methods

Appropriate analytic techniques (i.e., permissible operations) turn on assumptions about, level of measurement (i.e., categorical or numeric), sample size, etc. (Corder & Foreman, 2014;

Field, 2017). Parametric techniques assume, *inter alia*, random sampling, numeric (i.e., interval or ratio) measurement, normal distribution, and a large number of cases while nonparametric are well-suited to data which violates such assumptions.

From the research purpose and questions, the primary objectives of the analysis were to describe participants and their responses; as well as the main structure underlying their perceptions of EOEP and the different patterns of performance. Descriptive analysis, correlations, and comparisons were appropriate.

5.4.1 Descriptive analysis

This describes the distribution of data and estimates the central tendency (e.g., mean, mode, and median) and dispersion (e.g., range, variance, and standard deviation). Two ways to measure normality, and determine suitable techniques, are skewness and kurtosis. i.e., the degrees of symmetry and peakedness (Corder & Foreman, 2014). For a normal distribution, both values are equal to zero.

5.4.2 Correlation

It is common to assume relationships among data, e.g., between participants' years of experience and perceptions of how e-government evaluation is done (i.e., alternative proposition). Correlation measures the strength (i.e., an absolute value from 0 to 1— no correlation to perfect correlation), direction (i.e., negative or positive), and statistical significance (i.e., existence) of a relationship.

Statistical significance explores the existence of a relationship in data. The probability of error (expressed as p) shows whether the relationship emanates from the data or not. The rule of thumb is that if p is small (i.e., $p <= .05$), then it is likely the relationship is statistically significant (i.e., derives from the data and not due to chance); however, values from .01 to .10 are acceptable (Hair et al., 2018). An appropriate correlation technique depends on the number of variables involved (i.e., two or more than two), the level of measurement, normality, and sample size.

5.4.3 Comparison

Data may be compared to explore similarities and differences within a single group (or among two or more groups) of people, places, events, etc. (Field, 2017). As with relationships, there is a probability of error; the rule of thumb is p (<=.05), although .01 to .10 is an acceptable

range. An appropriate technique (parametric or nonparametric) depends on the number of groups, level of measurement, normality, and sample size.

Exploratory techniques (e.g., factor analysis and cluster analysis) serve several purposes, e.g., to identify relationships, compare groups to establish differences, etc., in a dataset (Hair et al., 2018).

5.4.4 Exploratory factor analysis (EFA)

Factor analysis is a set of statistical techniques which examine data (e.g., questionnaire responses) to determine the underlying structure, e.g., factors and their relationships (Hair et al., 2018). The techniques may be broadly classified into two: exploratory and confirmatory. Typically, exploratory factor analysis (EFA) "explores" a dataset to discover an underlying structure. Conversely, confirmatory factor analysis (CFA) applies *a priori* assumptions about the problem under investigation (e.g., the exact number of factors) to establish the structure.

Thus, EFA supports theory building and is well-suited to under-developed research problems while CFA is appropriate for theory testing (Kline, 2013). The choice, therefore, depends on the research purpose; EFA is consistent with the purpose of this research (see section 4.1). The main criticisms of factor analysis include factor indeterminacy and naming fallacy (Field, 2017; Kline, 2013). There is no unique solution (i.e., there may be equivalent models which adequately represent the data) and labels are descriptions (and may not be correct) and cannot be treated as constructs.

Other design decisions, apart from type (i.e., whether EFA or CFA), include factorability, extraction method, number of factors, rotation, and interpretation. These decisions are subjective, therefore, factor analysis is considered an inductive technique (Haig & Evers, 2016; Kline, 2013). There are various guidelines, often heuristics without empirical support (Field, 2017).

The quality of results depends on the items and the processes. The questionnaire was designed to include items relevant to the domain of interest and each concept was measured by at least three items. Furthermore, several guidelines (e.g., Beavers et al., 2013; Field, 2017; Hair et al., 2018; Kline, 2013; Osborne, 2015; Yong & Pearce, 2013) were considered. Although presented as sequential, the process is iterative.

5.4.4.1 Factorability

This is about the suitability of factor analysis as a technique. There is no consensus on an adequate sample size. Rules of thumb include, e.g., minimum of 50 but preferably over 100; and sample size (N) to the number of items (*p*) ratio (i.e., N:p), e.g., 3:1; 10:1 and 20:1. It is, however, acknowledged that, as a multivariate technique, EFA requires a larger sample size because smaller samples cause factor instability, i.e., factors with fewer items and low loadings (Beavers et al., 2013).

The following are also recommended. A large number of correlations should be greater than .3. Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (which measures the overall intercorrelation among variables) should be over 0.5. Bartlett's Test of Sphericity (which estimates the presence of correlations) should be significant. Communalities (i.e., the extent to which items correlate with each other) should be higher (greater than 0.4)— evidence that items share a common domain and have little uniqueness. An item loading of 0.2 is considered low.

5.4.4.2 Extraction method

There are several techniques, however, principal axis factor (PAF) and principal component analysis (PCA) are popular (Kline, 2013). PAF identifies common factors (i.e., correlated items, which measure something abstract or latent in common) while PCA determines a smaller set of independent items from a set of correlated items and is, therefore, not considered a true EFA technique. Unlike PAF and PCA, Maximum likelihood (ML) assumes normal distribution (Yong & Pearce, 2013). PCA and PAF are preferable when the purpose of the analysis is not

generalization (Field, 2017).

An appropriate choice depends on which technique can produce an interpretable structure congruent to the purpose of analysis (Field, 2017). However, studies show PAF and PCA solutions are often similar (Field, 2017; Kline, 2013). PCA solutions are generally interpretable and, therefore, may precede PAF to gain a better understanding of the solution (Yong & Pearce, 2013). PAF was considered consistent with the research purpose; however, PCA was explored prior.

5.4.4.3 Number of factors

Commonly used techniques to estimate the number of factors to extract include Kaiser-Guttman Eigenvalue (or K1 rule) and Scree test. Eigenvalue (λ) describes the proportion of variance explained by the items in a factor. The K1 criterion states that only factors with

eigenvalues greater than 1 (i.e., factors composed of two or more items) should be retained. The K1 rule is unreliable, prone to over-and under- factoring; also, factors with eigenvalues just below λ =1 are arbitrarily elided (Kline, 2013; Yong & Pearce, 2013).

The Scree test is a visual representation of the relationship between eigenvalues (Y-axis) and the number of factors to extract (X-axis), which occurs before the "break point" (or point of inflection) and separates important and trivial factors. The main problem is that the "break point" may be difficult to identify, which may result in over-or under- factoring.

Both the K1 rule and Scree test are subjective, different people may come to different conclusions about the number of factors to select (Field, 2017). Over-factoring may be preferable to under-factoring as the latter may introduce substantial errors, e.g., merge distinct factors (Brown, 2014; Kline, 2013). In practice, K1 is used in conjunction with Scree test. Furthermore, it is common practice to explore different numbers of factors to determine an optimal.

It is typical in social research to select successive factors that account for about 60 percent cumulative variance; this can adequately describe the total variance of a dataset and is, therefore, acceptable (Hair et al., 2018).

5.4.4.4 Rotation

A valid solution should be simple to interpret, i.e., factors consist of more than two high loading items (absolute values greater than .4); preferably on a single factor; and are statistically and theoretically significant (i.e., high internal consistency and theoretically related). In general, factors with few (i.e., two), complex (i.e., cross-loading), or low loading (i.e., absolute value <.4) items are not desirable but may be retained based on theoretical, statistical, and practical considerations.

The purpose of rotation is to produce a simple and meaningful structure that explains a high percentage of the total variance. There are two main techniques: orthogonal reduces correlation to ensure factors are independent, while oblique improves correlation. Theoretical considerations determine which is appropriate (Field, 2017). In general, aspects of social problems are interrelated, thus, oblique techniques are considered appropriate in social research (Osborne, 2015). Oblimin and Promax are popular oblique techniques, as Varimax is to orthogonal. It is common to explore the various rotation techniques to select the most interpretable structure.

5.4.4.5 Interpretation

As aforementioned, the main goal of exploratory factor analysis is to identify a structure that represents a dataset. Interpretation examines the factors which parsimoniously, theoretically, and statistically describe the problem under investigation. In an orthogonal solution, the factor matrix is examined. An oblique solution consists of pattern, structure, and factor correlation matrices (which show factor loadings, correlations among factors and items, and correlations among factors respectively). In general, pattern matrices are easier to interpret and, therefore, examined (Field, 2017; Hair et al., 2018).

Items are inspected for cross-loading to determine whether or not to retain (e.g., based on statistical and theoretical significance as well as practical usefulness). The next step is to label the factors, i.e., assign names, which describe the concepts represented by the highest loading items in the factors. As aforementioned, the main criticisms include indeterminacy and naming fallacy—there is no unique solution. Thus, validation is essential to ensure statistical, conceptual, and practical significance.

Factor analysis also assures convergent and discriminant validity: items in a factor are related to each other and, at the same time, different from those in other factors (Bhattacherjee, 2020). It is conventional to determine the reliability of factors (i.e., Cronbach's Alpha (α)).

It is common to reduce a large dataset generated from research into meaningful form; factor analysis can serve this purpose, i.e., reduce a dataset to a smaller set of factors for further analysis (Field, 2017; Hair et al., 2018). For example, to determine the differences across factors, group participants according to their responses, etc. This makes the validation of a solution more important.

5.4.4.6 Further analysis

In general, there are three approaches to data reduction (Hair et al., 2018). The highest loading item on each factor may be selected as the surrogate variable of a factor. A summated (i.e., sum or average) score may be calculated from the highest loading items to replace the factor. A multivariate method may be used to calculate a factor score from all loading items to substitute a factor. Common considerations for selection include research purpose and theoretical assumptions underlying factor analysis.

Each technique has strengths and drawbacks, e.g., a surrogate score is simpler to develop but unidimensional; summated scores are multidimensional and replicable while factor score is orthogonal and not easily replicable (Hair et al., 2018). Section 5.2.1 showed composite

measures reduce errors and are preferable to single items. The summated technique is appropriate for exploratory research (Hair et al., 2018).

The average score was adopted: unlike the sum of scores, it preserves measurement scale and affords comparison across factors (DiStefano et al., 2009) and is widely used (Hair et al., 2018). It is important to establish the reliability, validity, and unidimensionality (i.e., strong relationship among items) of a summated scale (Hair et al., 2018).

Notwithstanding their common purpose (i.e., to identify structure and data reduction), factor analysis groups variables whereas cluster analysis groups objects. The data was submitted to cluster analysis. Factor analysis followed by cluster analysis (referred to as the factor-cluster approach) has found wide application— factor analysis assures unidimensionality (Hair et al., 2018).

5.4.5 Cluster analysis

An important feature of research is classification (Hair et al., 2018). Cluster analysis is an exploratory technique to group similar elements in a dataset into distinct subsets or clusters, according to specific criteria. Together, the clusters constitute the cluster solution.

A cluster analysis strategy includes decisions on the purpose of analysis; sample size; the objects to group, their relevant characteristics and the level of measurement; the measure of similarity; cluster technique; the number of clusters; quality of cluster solution, and interpretation of results. Cluster analysis is subjective— different decisions may result in different cluster solutions. There are various guidelines to improve quality (Balijepally et al., 2011; Hair et al., 2018; Hofstetter et al., 2014; Mooi & Sarstedt, 2014)

5.4.5.1 Purpose and Sample size

Cluster analysis may serve three main purposes, i.e., to: identify groups according to some common characteristics; simplify (e.g., identify a structure) for further analysis; and identify relationships, e.g., differences and similarities otherwise unobservable (Hair et al., 2018).

There is no minimum sample size (Mooi & Sarstedt, 2014), however, an adequate sample size is important to ensure that subsamples (i.e., clusters) are larger for analysis (Hair et al., 2018). According to Mooi and Sarstedt, a minimum sample size of 2^m (where m is the number of items to cluster) is preferable.

The data to analyze may be derived directly, as observations about objects of interest (e.g., people, events, etc.) or indirectly, as responses on variables (e.g., attitudes, perceptions, etc.).

Normality is not a basic assumption, however, representative samples and independent variables are necessary (Hair et al., 2018). The main objectives of cluster analysis make representativeness important for results to be useful. Absolute correlation among variables may distort a cluster solution— Mooi and Sarstedt (2014) suggest less than .9. Factor analysis can reduce multicollinearity among factors and, therefore, is widely used to prepare data for clustering (Hair et al., 2018; Mooi & Sarstedt, 2014).

5.4.5.2 Similarity measurement

A popular method for measuring similarity is to calculate the distance (Hair et al., 2018; Mooi & Sarstedt, 2014). Data with shorter distances between them are likely to be similar and belong to the same cluster. Euclidean distance is used for metric data and applies to scales (e.g., Likert may be treated as equidistant to calculate distances) while log-likelihood distance is used for both metric and non-metric (Mooi & Sarstedt, 2014). Data may be standardized to remove variations, in cases where variables are of different scales (Hair et al., 2018).

5.4.5.3 Cluster techniques

There are two main cluster techniques, hierarchical and non-hierarchical, each with its underlying algorithms (Hair et al., 2018; Mooi & Sarstedt, 2014). Hierarchical is tree-like: each object starts off as a cluster and on each iteration, similar clusters are combined until the cluster solution is obtained. This technique is slow and, therefore, suitable for smaller samples; supports only categorical data; and is sensitive to outliers. Common examples include single, complete and average linkages, centroid, and Ward's. On the other hand, non-hierarchical begins with a single cluster of all data, which is divided into clusters on each iteration. This technique is faster and, therefore, lends itself to larger samples; supports continuous data; and is less sensitive to outliers. An example is k-means.

Non-hierarchical techniques require the number of clusters to be known (e.g., based on theory, previous research, or practical considerations) and specified *a priori* whereas this is induced from the data in hierarchical.

Notwithstanding the distinction, the techniques are often combined in practice, to take advantage of their inherent strengths while ameliorating weaknesses (Hair et al., 2018; Mooi & Sarstedt, 2014). This technique is often referred to as two-step cluster analysis and involves non-hierarchical followed by hierarchical. For example, it can handle larger sets of continuous as well as categorical data, a variety of similarity measurements, and automatic or *a priori* selection of the number of clusters.

There is no consensus on several aspects of cluster analysis, despite the widespread application (Balijepally et al., 2011; Mooi & Sarstedt, 2014). For example, the factor-cluster approach, standardization, appropriate similarity measurements, and cluster techniques. Many decisions are based on judgement and preferences, which may result in different cluster solutions for a specific problem. Validation is, therefore, imperative (Hair et al., 2018).

5.4.5.4 Validation and interpretation of clusters

The quality of cluster solutions may be evaluated from three main perspectives: technical, practical, and interpretability (Mooi & Sarstedt, 2014). Technical concerns the extent to which a solution is statistically reliable and valid. Factors to consider include: cohesion-within and separation-between constituent clusters (i.e., how objects are homogenous within and heterogeneous between clusters), the "correct" number of clusters, the fit between data and solution, and how well clusters reflect grouping criteria (Hair et al., 2018).

There are several validation techniques, e.g., Silhouette measure of cohesion and separation; ANOVA, etc. (to compare clusters means to determine if there are significant differences); and exploration of different algorithms (e.g., hierarchical and non-hierarchical, Euclidean and log-likelihood, etc.) to select the optimal solution (Hair et al., 2018; Mooi & Sarstedt, 2014).

Section 5.4.5.1 showed a cluster solution ought to serve a purpose. In general, simple and interpretable solutions are preferable (Balijepally et al., 2011); complex and difficult solutions may not be practically useful. Interpretation describes the constituent clusters, individually and together. For example, how distinct are clusters, which cluster items delineate clusters, what meaningful labels to assign, which characteristics assign objects to clusters (Hair et al., 2018; Mooi & Sarstedt, 2014). Criterion validity is also important, e.g. items such as age, experience, position in an organization, etc., can provide the practical or theoretical basis to profile clusters. As aforementioned, people may perform a practice differently, according to know-how.

In general, a smaller number of larger clusters is preferable to many small clusters (Norusis, 2012). A cluster is important if it is greater than or equal to 10 percent of the sample size (Hair et al., 2018).

Cluster analysis has found wide application in IS and practice research. For example, a review by Balijepally et al. (2011) showed adoption for a wide range of IS research. Cha and Park (2018) employed factor-cluster analysis (with summated scores) to examine e-government evaluation. Application in practice research includes consumption patterns (Browne et al., 2014, 2015; Pullinger et al., 2013).

The next section applied the assumptions on descriptive, factor, and cluster analyses outlined above to organize, summarize and present the results on how e-government outcomes evaluation is done.

5.5 Data analysis and preliminary results

Common devices to organize and present results include tables, graphs, and statistical estimates. Appropriate choices depend on the level of measurement (e.g., numeric or categorical) and the number of cases to be analyzed. For example, a bar chart is well-suited to categorical and discrete data; histogram: continuous data; pie chart: proportions; and nonparametric statistics: nonparametric data (Foster et al., 2015; Saunders et al., 2015).

5.5.1 Descriptive analysis

This served two objectives, to gain an understanding of the nature of the data and describe participants and their responses. Tables and graphs were mainly employed. Prior to analysis, the data was screened for anomalies. There were no missing values or outliers. Distributions of items (i.e., frequencies, estimates of central tendencies and dispersions, and normality) are presented below.

5.5.1.1 Characteristics of participants

Information about participants, i.e., position in the organization, professional background, and experience in e-government outcomes evaluation (Questionnaire items 1 to 5) are summarised as follows. According to Table 5.4 and Figure 5.2, of the 106 participants 43, 27, and 36 (i.e., about 41, 25, and 34 percent respectively) were from IT, Evaluation, and Other professional backgrounds respectively.

Table 5.4: Distribution of participants by background

	Frequency	*Percent*
IT	43	41
Evaluation	27	25
Other	36	34
Total	106	100

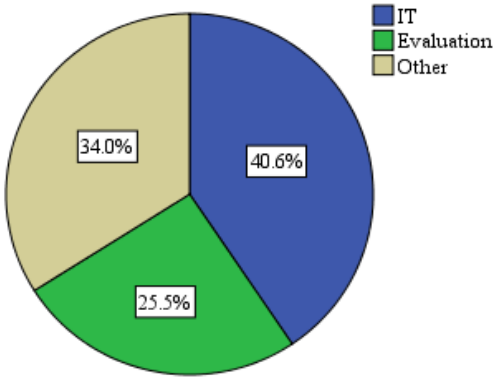


Figure 5.2: Distribution of participants by background

Further analysis of participants by background, management position, years of experience and knowledge, and the number of e-government participated in (in the last two years) are presented in Table 5.5. and Figure 5.3. Forty-eight of the participant (i.e., 45 percent) indicated more than five years of experience; about 80 percent rated their knowledge as average or above average. About 70 percent had been involved in one or more e-government evaluation in the last two years. It may be concluded that participants had requisite experience in e-government evaluation and, therefore, constituted an appropriate sample.

Table 5.5: Distribution of participants by background, experience (years and knowledge) and number of evaluation

	Management		*Years of experience*		*Knowledge*			*No. of evaluation*		
Background	*Yes*	*No*	*1 to 5*	*> 5*	*Min*	*Ave*	*> Ave*	*None*	*1*	*> 1*
IT (N = 43)	14 (32.6%)	29 (67.4%)	26 (60.5%)	17 (39.5%)	7 (16.3%)	21 (48.8%)	15 (34.9%)	15 (34.9%_	24 (55.8%)	4 (9.3%)
Evaluation (N = 27)	15 (55.6%)	12 (44.4%)	5 (18.5%)	22 (81.5%)	0 (0.0%)	6 (22.2%)	21 (77.8%)	2 (7.4%)	18 (66.7%)	7 (25.9%)
Other (N = 26)	12 (33.3%)	24 (66.7%)	27 (75.0%)	9 (25.0%)	15 (41.7%)	15 (41.7%)	6 (16.7%)	16 (44.4%)	18 (50.0%)	2 (5.6%)
Total (N = 106)	43 (100.0%)	27 (100.0%)	58 (54.7%)	48 (45.3%)	22 (20.8%)	42 (39.6%)	42 (39.6%)	33 (31.1%)	60 (56.6%)	13 (12.3%)

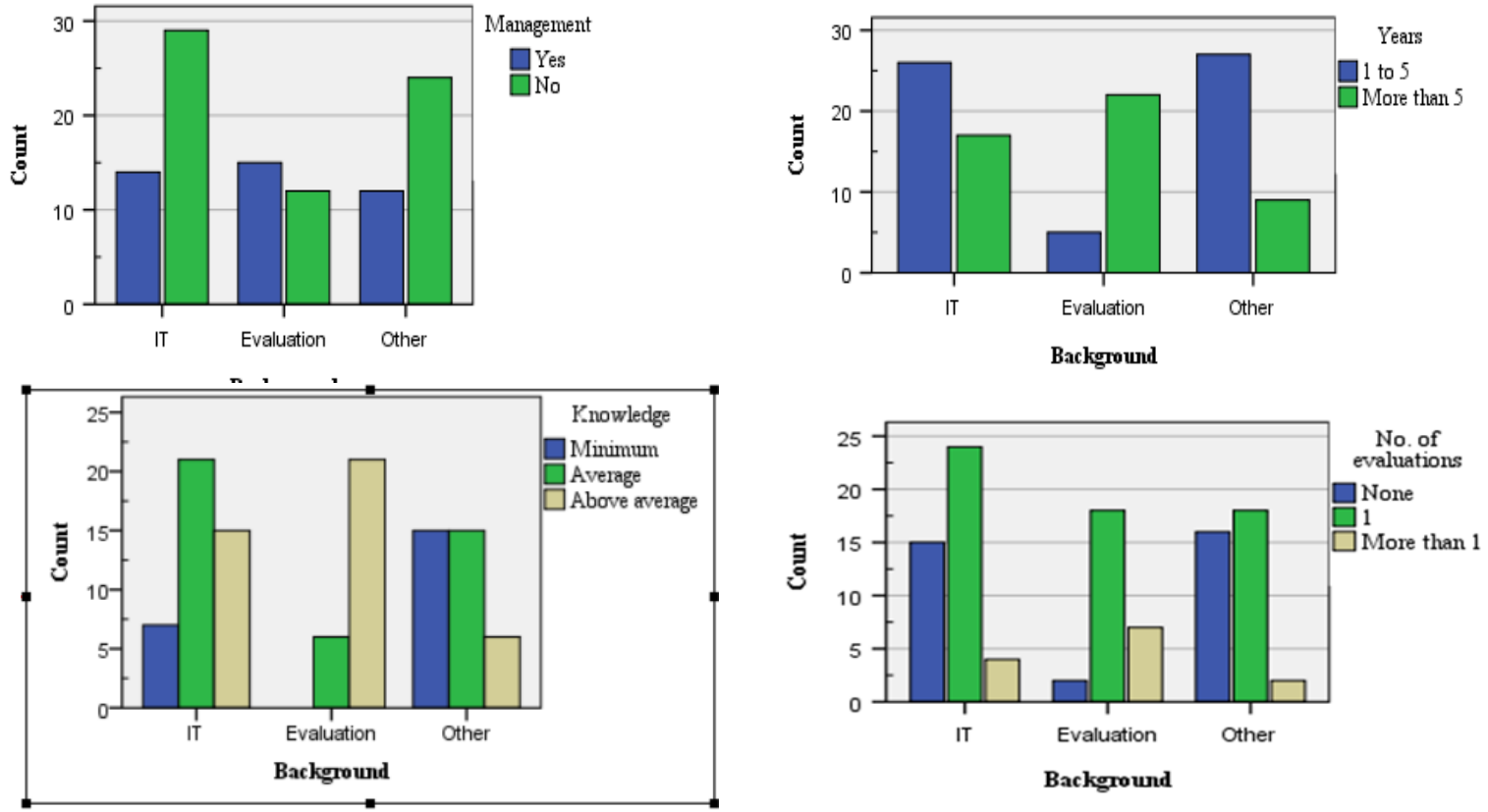


Figure 5.3: Distribution of participants by background, experience (years and knowledge) and number of evaluation

5.5.1.2 Evaluation responsibility, strategies, outcomes, and purpose

Questionnaire items 6 to 11 specifically measured e-government evaluation practice. Figure 5.4 shows about 43 percent of participants indicated that evaluation was initiated within their respective organization, and not from outside. Also, most (about 75 percent) indicated that evaluation teams are constituted by both internal and external stakeholders.

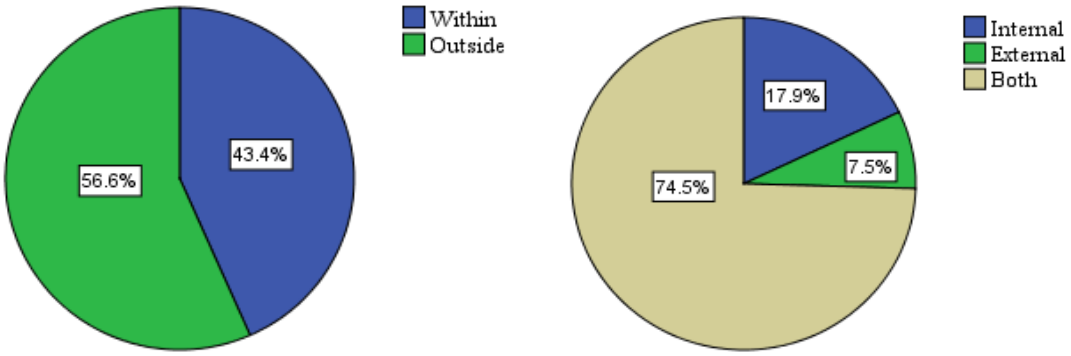


Figure 5.4: Responsibility for evaluation: initiation and composition of teams

Items 8, 9, and 10 concerned the evaluation strategy, evaluation purpose, and e-government outcomes evaluated. These were measured with multiple response items which allowed participants to select one or more options. Each item was analyzed in two stages: to establish the total number of participants and responses to individual options. The latter was further corroborated with evidence from analysis of frequency. All 106 participants responded to these items, the distribution of responses is summarized in Tables 5.6 to 5.8. The most commonly used strategies were survey, focus group, and case study— in that order. Accountability and organizational performance were often evaluated mainly for compliance, decision-making, and learning about e-government. The open-ended items elicited too few responses and, as such, were excluded from further analysis. This may be consistent with the perception that participants are unlikely to answer open-ended questions (Rea & Parker, 2014)— see section 4.6.3.1.

Table 5.6: Evaluation strategies

Strategy	*Responses*		*Percent of Cases*
	N	*Percent*	
Experiment	8	3.3%	7.5%
Survey	101	41.2%	95.3%
Case study	61	24.9%	57.5%
Focus group	72	29.4%	67.9%
Other	3	1.2%	2.8%
Total	245	100.0%	231.1%

a. Dichotomy group tabulated at value 1.

Table 5.7: E-government outcomes evaluated

Evaluand	*Responses*		*Percent of Cases*
	N	*Percent*	
Performance	96	28.3%	90.6%
Participation	39	11.5%	36.8%
Accountability	101	29.8%	95.3%
Transparency	53	15.6%	50.0%
User satisfaction	47	13.9%	44.3%
Other	3	0.9%	2.8%
Total	339	100.0%	319.8%

a. Dichotomy group tabulated at value 1.

Table 5.8: The purpose of evaluation

	Responses		Percent of Cases
Purpose	N	Percent	
Compliance	100	37.5%	94.3%
Learn about e-gov	64	24.0%	60.4%
Learn about process	9	3.4%	8.5%
Decision-making	93	34.8%	87.7%
Other	1	0.4%	0.9%
Total	267	100.0%	251.9%

a. Dichotomy group tabulated at value 1.

5.5.1.3 Likert items

As aforementioned, analysis commenced with screening items. This included frequencies, central tendencies, and dispersion. Table 5.9 shows the absolute values of skewness and kurtosis of all items were greater than 0, i.e., the data was not normally distributed.

It is a common practice to collapse the Likert scale into fewer (e.g., dichotomous and trichotomous) categories to facilitate analysis. The 5-point Agree-Disagree Likert scale for questionnaire item 11 (see Appendix C) was recoded into a trichotomous scale ("Disagree", "Neither agree or disagree" and 'Agree") by combining "Strongly disagree" and "Disagree" into "Disagree"; and "Agree" and "Strongly agree" into "Agree" (see Table 5.10).

Table 5.9: Descriptive Statistics: Likert items (N = 106)

Item	*SD*	*D*	*N*	*A*	*SA*	*Skewness*	*Kurtosis*	*Mean*	*SD*	*Median*	*Mode*	*Range*
Management support	8 (7.5%)	18 (17.0%)	17 (16.0%)	36 (34.0%)	27 (25.5%)	-.53	-.81	3.53	1.251	4.0	4	4
National framework	16 (15.1%)	26 (24.5%)	19 (17.9%)	27 (25.5%)	18 (17.0%)	-.04	-1.23	3.05	1.341	3.0	4	4
Regular evaluation	24 (22.6%)	19 (17.9%)	22 (20.8%)	26 (24.5%)	15 (14.2 %)	-.01	-1.26	2.90	1.380	3.0	4	4
Regular eval review	40 (37.7%)	20 (18.9%)	16 (15.1%)	15 (14.2%)	15 (14.2%)	.50	-1.18	2.48	1.469	2.0	1	4
Adequate budget	12 (11.3%)	29 (27.4%)	27 (25.5%)	23 (21.7%)	15 (14.2%)	.09	-.99	3.00	1.234	3.0	2	4
Adequate expertise	10 (9.4%)	28 (26.4%)	27 (25.5%)	28 (26.4%)	13 (12.3%)	-.01	-.94	3.06	1.186	3.0	2	4
Adequate tools	11 (10.4%)	35 (33.0%)	19 (17.9%)	24 (22.6%)	17 (16.0%)	.15	-1.16	3.01	1.276	3.0	2	4
Batho Pele	13 (12.3%)	18 (17.0%)	26 (24.5%)	36 (34.0%)	13 (12.3%)	-.33	-.84	3.17	1.215	3.0	4	4
Evaluation models	17 (16.0%)	9 (8.5%)	21 (19.8%)	44 (41.5%)	15 (14.2%)	-.62	-.71	3.29	1.280	4.0	4	4
Stakeholder representation	11 (10.4%)	18 (17.0%)	23 (21.7%)	37 (34.9%)	17 (16.0%)	-.39	-.82	3.29	1.226	4.0	4	4
Stakeholder selection	6 (5.7%)	9 (8.5%)	26 (24.5%)	47 (44.3%)	18 (17.0%)	-.76	.24	3.58	1.050	4.0	4	4
Stakeholder training	15 (14.2%)	23 (21.7%)	24 (22.6%)	33 (31.1%)	11 (10.4%)	-.16	-1.03	3.02	1.234	3.0	4	4
Agreement (evaluand)	14 (13.2%)	24 (22.6%)	25 (23.6%)	31 (29.2%)	12 (11.3%)	-.12	-1.01	3.03	1.230	3.0	4	4
Agreement (use)	17 (16.0%)	25 (23.6%)	23 (21.7%	30 (28.3%)	11 (10.4%)	-.05	-1.10	2.93	1.259	3.0	4	4
Stakeholder participation	11 (10.4%)	16 (15.1%)	25 (23.6%)	39 (36.8%)	15 (14.2%)	-.45	-.68	3.29	1.195	4.0	4	4
Evaluation indicators	14 (13.2%)	28 (26.4%)	24 (22.6%)	26 (24.5%)	14 (13.2%)	.04	-1.06	2.98	1.257	3.0	2	4
Financial/non financial	15 (14.2%)	11 (10.4%)	17 (16.0%)	38 (35.8%)	25 (23.6%)	-.62	-.78	3.44	1.339	4.0	4	4
Data quality	11 (10.4%)	33 (31.1%)	25 (23.6%)	23 (21.7%)	14 (13.2%)	.17	-.98	2.96	1.218	3.0	2	4
Reports (distribution)	10 (9.4%)	30 (28.3%)	23 (21.7%)	27 (25.5%)	16 (15.1%)	.02	-1.08	3.08	1.235	3.0	2	4
Reports (recommendations)	12 (11.3%)	20 (18.9%)	22 (20.8%)	36 (34.0%)	16 (15.1%)	-.32	-.94	3.23	1.244	3.0	4	4
Reports (to diff groups)	11 (10.4%)	17 (16.0%)	25 (23.6%)	33 (31.1%)	20 (18.9%)	-.36	-.84	3.32	1.246	3.5	4	4
Management response	17 (16.0%)	25 (23.6%)	19 (17.9%)	31 (29.2%)	14 (13.2%)	-.08	-1.19	3.00	1.309	3.0	5	4
Follow up	36 (34.0%)	26 (24.5%)	16 (15.1%)	17 (16.0%)	11 (10.4%)	.52	-1.02	2.44	1.374	2	1	4
Knowledge	13 (12.3%)	17 (16.0)	13 (12.3%)	41 (38.7%)	22 (20.8%)	-.54	-.91	3.40	1.314	4.0	4	4
Acceptance (results)	11 (10.4%)	14 (13.2%)	22 (20.8%)	42 (39.6%)	17 (16.0%)	-.57	-.58	3.38	1.207	4.0	4	4
Commitment to use	9 (8.5%)	25 (23.6%)	23 (21.7%)	31 (29.2%)	18 (17.0%)	-.16	-1.02	3.23	1.229	3.0	4	4
Future participation	16 (15.1%)	21 (19.8%)	22 (20.8%)	31 (29.2%)	16 (15.1%)	-.18	-1.11	3.09	1.306	3.0	4	4

Table 5.10: Descriptive statistics-collapsed scale (N = 106)

Item	*Disagree*	*Neither*	*Agree*	*Mean*	*Std. Deviation*	*Median*	*Mode*
Mgt support	26 (24.5%)	17 (16.0%)	63 (59.4%)	2.35	.851	3.0	3
National framework	42 (39.6%)	19 (17.9%)	45 (42.5%)	2.03	.910	2.0	3
Regular evaluation	43 (40.6)%	22 (20.8%)	41 (38.7%)	1.98	.894	2.0	1
Regular eval review	60 (56.6%)	16 (15.1%)	30 (28.3%)	1.72	.881	1.0	1
Adequate budget	41 (38.7%)	27 (25.5%)	38 (35.8%)	1.97	.867	2.0	1
Adequate expertise	38 (35.8%)	27 (25.5%)	41 (38.7%)	2.03	.867	2.0	3
Adequate tools	46 (43.4%)	19 (17.9%)	41 (38.7%)	1.95	.909	2.0	1
Batho Pele	31 (29.2%)	26 (24.5%)	49 (46.2%)	2.17	.856	2.0	3
Evaluation model	26 (24.5%)	21 (19.8%)	59 (55.7%)	2.31	.844	3.0	3
Stakeholder representation	29 (27.4%)	23 (21.7%)	54 (50.9%)	2.24	.857	3.0	3
Stakeholder selection	15 (14.2%)	26 (24.5%)	65 (61.3%)	2.47	.733	3.0	3
Stakeholder training	38 (35.8%)	24 (22.6%)	44 (41.5%)	2.06	.882	2.0	3
Agreement (evaluand)	38 (35.8%)	25 (23.6%)	43 (40.6%)	2.05	.877	2.0	3
Agreement (use)	42 (39.6%)	23 (21.7%)	41 (38.7%)	1.99	.889	2.0	1
Stakeholder participation	27 (25.5%)	25 (23.6%)	54 (50.9%)	2.25	.840	3.0	3
Evaluation indicators	42 (39.6%)	24 (22.6%)	40 (37.7%)	1.98	.884	2.0	1
Financial/non-financial	26 (24.5%)	17 (16.0%)	63 (59.4%)	2.35	.851	3.0	3
Data quality	44 (41.5%)	25 (23.6%)	37 (34.9%)	1.93	.876	2.0	1
Report (distribution)	40 (37.7%)	23 (21.7%)	43 (40.6%)	2.03	.889	2.0	3
Report (recommendations)	32 (30.2%)	22 (20.8%)	52 (49.1%)	2.19	.874	2.0	3
Reports (to diff groups)	28 (26.4)	25 (23.6%)	53 (50.0%)	2.24	.846	2.5	3
Management response	42 (39.6%)	19 (17.9%)	45 (42.5%)	2.03	.910	2.0	3
Follow up	62 (58.5%)	16 (15.1%)	28 (26.4%)	1.68	.868	1.0	1
Improves knowledge	30 (28.3%)	13 (12.3%)	63 (59.4%)	2.31	.888	3.0	3
Improves acceptance	25 (23.6%)	22 (20.8%)	59 (55.7%)	2.32	.834	3.0	3
Increases commitment	34 (32.1%)	23 (21.7%)	49 (46.2%)	2.14	.878	2.0	3
Future participation	37 (34.9%)	22 (20.8%)	47 (44.3%)	2.09	.889	2.3	3

While most participants agreed with management support (63 participants, i.e. 59%), few agreed with the adequacy of evaluation resources, i.e., budget, expertise, and tools: 38 (36%), 41 (39%), and 41 (39%) respectively. Also, few agreed with regular e-government outcomes evaluation (i.e., 41, 39%), and regular review of how e-government outcomes is done (i.e., 30, 28%). About three-fifths (i.e., 61, 58%) did not agree that management responded to evaluation reports. About three-quarters (i.e., 78, 74%) did not agree there was follow-up to implement recommendations; a further examination showed over 80 percent were in non-management.

On stakeholder selection, about 61% agreed this was transparent and about half (51%), to representation of all stakeholder groups and participation in all activities respectively. Only two-fifths (41%) agreed with training.

While about 60 percent agreed that participation improved acceptance of recommendation (55.7%) and knowledge (59.4%), it seemed participation did not increase commitment to implement recommendations or willingness to participate in the future (less than half, i.e., 46.2% and 44.3% agreed respectively).

The data was not normally distributed: the absolute values of kurtosis and skewness were greater than zero. This was supported by graphical exploration (i.e., frequency charts) of the data and the Kolmogorov-Smirnov test (p<.001, α=0.05) in Table E1.1 (Appendix E).

Subsequent to analyses of distributions, the data was explored to establish the underlying structure and reduced for further analysis. This commenced with factor analysis to identify the main elements which described participants' perceptions of EOEP. The raw responses (i.e., on the 5-point Likert scale) were analyzed because the trichotomous (i.e., collapsed categories) may introduce distortions.

5.5.2 Exploratory factor analysis

Section 5.4.4 showed exploratory factor analysis suited the purpose of this research and outlined an analysis strategy, which encompassed the following design decisions: purpose, factorability, extraction method, number of factors, rotation, and naming. This was applied to answer the research question:

RQ1: What are the factors which describe how e-government outcomes evaluation is done?

The basic purpose of exploratory factor analysis (i.e., data summarization and data reduction) suited the research question above. Consequently, the technique was applied in two stages, to: establish the structure (i.e., elements and their relationships) which described participants'

perceptions of EOEP and reduce the dataset (i.e., compute average scores to represent the larger dataset) for further analysis.

Preliminary analysis was conducted, to gain insight into the characteristics of the data. Descriptive analysis of items was presented in section 5.5.1 and, therefore, needs no reiteration. Suffice to state the sample size (N=106) was above the minimum of 50 required for factor analysis, there were no outliers and the distribution was non-normal. According to Hair et al. (2018), the effect of non-normality on a sample greater than 50 may be insignificant. As is typical of factor analysis, the Likert scale responses were considered metric.

The underlying structure was estimated from the data itself, the number of factors was unknown, intercorrelation among factors (i.e., elements of practice) was expected and normality was violated. Thus, as discussed in section 5.4.4.2, principal factor analysis (PAF) was preferable.

5.5.2.1 Factor extraction and assessment of overall fit

This establishes whether or not exploratory factor analysis (EFA) was suitable for the dataset. As shown above, the questionnaire was designed to ensure reliability and validity, each concept consisted of at least three items; the sample size was 106. Table 5.11 shows, according to the K1 rule, seven factors accounted for about 58 percent of the total variance of the initial solution— about 60 percent is typical of social research (Hair et al., 2018). Table 5.12 shows the overall KMO was .74 (above the minimum value of .50)— an indication that the sample size was adequate and the underlying structure accounted for variance in the data. The Bartlett's test was significant (χ^2 (351) = 1469.19, $p < 0.001$), which signifies correlation among the data.

Table 5.13 shows the Maximum likelihood test was significant (χ^2 (183) = 267.75, $p < 0.000$)— additional evidence that factor analysis was an appropriate technique. Table E2.1 (Appendix E) shows the communalities after the extraction; most were greater than 0.4 and (18 out of 27 were over 0.5), which means the extracted factors well-represented the items.

From the foregoing, i.e., nature of items, sample size, KMO and Bartlett tests, percentage of total variance, and communalities (and against the backdrop of section 5.4.4.1), it may be concluded that factor analysis was suitable for the dataset.

Table 5.11: Total Variance Explained

	Initial Eigenvalues			*Extraction Sums of Squared Loadings*		
Factor	*Total*	*% of Variance*	*Cumulative %*	*Total*	*% of Variance*	*Cumulative %*
1	6.875	25.462	25.462	6.468	23.955	23.955
2	4.172	15.452	40.915	3.759	13.921	37.876
3	1.969	7.294	48.209	1.613	5.973	43.850
4	1.705	6.315	54.524	1.343	4.973	48.823
5	1.366	5.058	59.582	.941	3.486	52.309
6	1.160	4.295	63.877	.782	2.895	55.204
7	1.056	3.910	67.787	.697	2.583	57.787
8	.996	3.688	71.475			
9	.842	3.117	74.593			
10	.789	2.922	77.515			
11	.686	2.541	80.056			
12	.625	2.316	82.372			
13	.585	2.165	84.537			
14	.547	2.027	86.564			
15	.501	1.854	88.418			
16	.443	1.639	90.058			
17	.409	1.515	91.573			
18	.397	1.471	93.044			
19	.349	1.293	94.338			
20	.290	1.074	95.412			
21	.271	1.003	96.414			
22	.240	.891	97.305			
23	.190	.702	98.007			
24	.172	.639	98.646			
25	.151	.557	99.203			
26	.111	.411	99.615			
27	.104	.385	100.000			

Extraction Method: Principal Axis Factoring.

Table 5.12: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.74
Bartlett's Test of Sphericity	Approx. Chi-Square	1469.19
	df	351
	Sig.	.00

Table 5.13: Goodness-of-fit Test

Chi-Square	*df*	*Sig.*
267.745	183	.000

Figure 5.5 shows the Scree plot was ambiguous: there seemed to be two points of inflection, at factors 4 and 8. As mentioned in section 5.4.4.3, the K1 and Scree plot are not conclusive, therefore, it is common to explore different numbers of factors. Consequently, three to up eight factors were examined. All items loaded below an absolute value of 1.

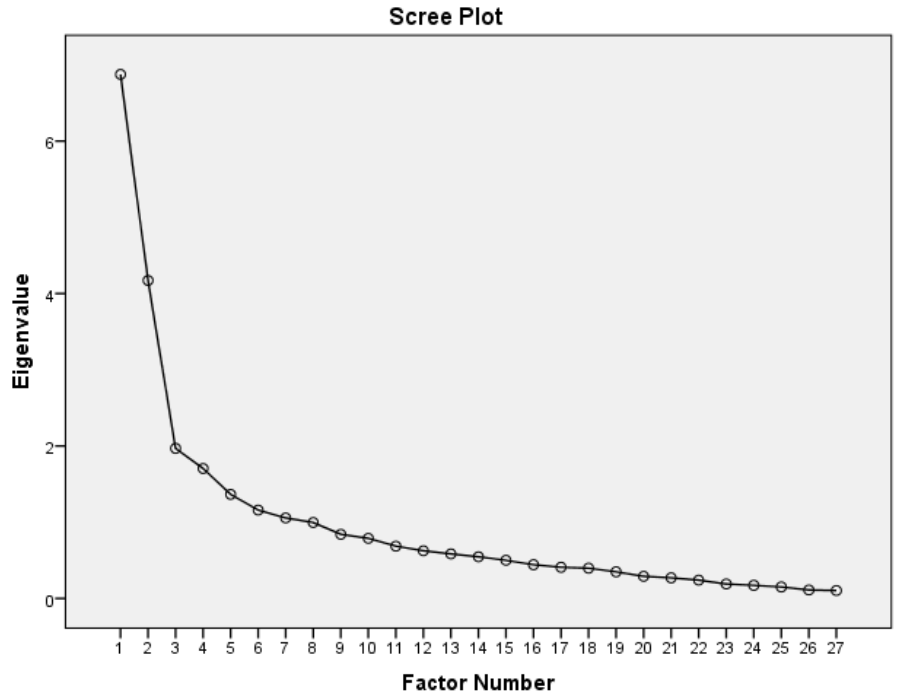


Figure 5.5: Scree plot

5.5.2.2 Factor rotation

Consistent with practice, other techniques (e.g., PCA and Maximum likelihood) with both orthogonal (e.g., Varimax and Equimax) as well as oblique (e.g., Oblimin and Promax) rotations were explored. Items which loaded less than 0.4 (c.f., the recommended minimum absolute value of 0.3) were suppressed; those which did not load on any factor were eliminated and the process repeated. The three-, four- and five-factor solutions were difficult to interpret (i.e., many items from different concepts loaded under factors, which made the factors conceptually and practically insignificant). The six and seven-factor solutions were similar—

and more interpretable. However, the seven-factor contained loadings above .90, a sign of absolute correlation.

The six-factor Promax rotated solution was, therefore, selected as the final solution. The pattern and structure matrices were dissimilar, which confirms correlation among factors and, therefore, suitability of oblique rotation. The pattern matrix was similar to the PCA solution.

While this solution may suffer from over-factoring, it was preferable— consistent with the fundamental purpose of factor analysis (i.e., to ensure that factors accounted for as much of the dataset as possible and provided as many dimensions of the phenomenon under study as possible for further reduction and investigation).

Tables 5.14 to 5.16 show the pattern, structure, and factor correlation matrices of the final solution. It may be noted that the structure matrix is inconsequential to analysis and, thus, cross-loading is acceptable. Three of the 27 items (see Table 5.17) loaded less than .4 and, thus, were excluded.

Table 5.14: Pattern matrix (Principal Axis Factor, Promax rotated)

	Factor						*Communality*
	1	*2*	*3*	*4*	*5*	*6*	
Financial/non financial	.77						0.54
Evaluation models	.77						0.74
Batho Pele	.73						0.44
Reports (to diff groups)	.63						0.42
Knowledge	.61						0.41
Reports (recommendations)	.56						0.39
Stakeholder representation	.50						0.59
Acceptance (results)	.50						0.61
Adequate expertise	.40						0.45
Management response		.89					0.62
Follow up		.61					0.79
Data quality		.54					0.68
Regular eval review		.43					0.41
Regular evaluation			.69				0.49
Adequate budget			.61				0.59
Management support			.55		.41		0.38
Evaluation indicators			.51				0.52
Adequate tools			.49				0.58
Agreement (evaluand)				.85			0.77
Agreement (use)				.77			0.47
Stakeholder selection					.80		0.55
Stakeholder participation					.45		0.55
Future participation						.77	0.56
Commitment to use						.76	0.61
Eigen values	5.61	4.07	1.86	1.69	1.34	1.16	
% of Variance	21.55	15.12	5.91	5.36	3.81	3.07	
Cronbach's Alpha (α)	.86	.77	.75	.85	.59	.74	

Note: loadings < .4 excluded.

Table 5.15: Structure Matrix

	Factor					
	1	2	3	4	5	6
Financial/non financial	.75					
Batho Pele	.72					
Evaluation models	.72					
Reports (to diff groups)	.65					
Reports (recommendations)	.64					.40
Knowledge	.63					
Acceptance (results)	.62			.53		
Stakeholder representation	.58				.45	
Management response		.87				
Follow up		.66	.41			
Data quality		.57	.40			
Regular eval review		.55	.50			
Regular evaluation		.58	.80			
Adequate budget			.61			
Adequate tools			.56			
Management support			.53		.50	
Evaluation indicators		.44	.53			
Agreement (evaluand)				.87		
Agreement (use)	.43			.80		
Adequate expertise	.46			.46		
Stakeholder selection					.77	
Stakeholder participation					.57	
Commitment to use						.73
Future participation						.71

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.

Table 5.16: Factor Correlation Matrix

Factor	*1*	*2*	*3*	*4*	*5*	*6*
1	1.000					
2	-.018	1.000				
3	.010	.438	1.000			
4	.408	-.043	.241	1.000		
5	.283	.264	.106	.133	1.000	
6	.226	.153	-.028	.180	.304	1.000

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.

Table 5.17: Excluded items

The National Evaluation Policy Framework guides how evaluation is done
Stakeholders are trained on how evaluation is done
Evaluation reports are distributed to all stakeholder groups

5.5.2.3 Interpretation

As stated in section 5.4.4.5, pattern matrices are generally easier to interpret. Table 5.14 shows the six factors accounted for over half (about 55%) of the total variance— the first and second factors contributed 21.55 and 15.12 percent respectively. Management support (i.e., Management supports evaluation) loaded on Factors 3 and 5 at .55 and .41 respectively and was retained for Factor 3. Also, three factors (Factors 4, 5, and 6), each of which consisted of

two factors, were retained. These decisions were based on statistical, conceptual, and practical significance (see section 5.4.4.5.) As made clear below, the factors exhibited higher reliability, were conceptually related, and may be useful (e.g., in future research).

Table 5.16 (the factor correlation matrix) showed moderate correlations among some factors, e.g., between Factors 1 and 4 and Factors 5 and 6 were .40. and .30 respectively. This is consistent with the underlying assumptions of practice theory, *viz*., elements of practice are interconnected.

Section 5.4.4.5 showed validation in factor analysis involves statistical, conceptual, and practical significance. Statistical significance of a factor relates to reliability, i.e., internal consistency, estimated by Cronbach alpha (α). For exploratory research, α=0.6 is considered acceptable (Hair et al., 2018); others (e.g., Field, 2017) recommend α=0.5. Tables 5.18 to 5.20 described the factors. Thus, for the purpose of this research (section 4.2), the factors are reliable.

Table 5.18: Descriptive and reliability statistics of factors (N = 106)

	No. of items	*Mean*	*Median*	*Mode*	*Std. Deviation*	*Cronbach's alpha (α)*
Factor1	9	3.29	3.33	3.33	.86	.86
Factor2	4	2.72	2.75	1.75	1.03	.77
Factor3	5	3.08	3.00	2.40	.90	.75
Factor4	2	2.98	3.00	4.00	1.16	.85
Factor5	2	3.44	3.50	3.50	.95	.59
Factor6	2	3.16	3.00	3.00	1.13	.74

Table 5.19: Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.623	.631	6

Table 5.20: Inter-Item Correlation Matrix

	Factor1	Factor 2	Factor3	Factor4	Factor5	Factor6
Factor1	1.000					
Factor2	-.032	1.000				
Factor3	.142	.559	1.000			
Factor4	.482	.185	.259	1.000		
Factor5	.344	.192	.198	.216	1.00	
Factor6	.120	.131	.208	.144	.178	1.000

Interpretation involved, for each extracted factor: a description of the higher loading items, the concept which these represented, an appropriate label, and the statistical significance (see section 5.4.4.5).

Factor 1: Evaluation framework. This was the most important factor, according to participants, and included nine items. Three items, "The indicators measure both financial and

non-financial outcomes", "Evaluation models (*utilization, empowerment, etc.*) guide evaluation" and "Batho Pele principles guide how evaluation is done" all loaded strongly (above .7). The last two related to assumptions about evaluation, relevant to indicators for measurement.

The three items may be conceptually related to **Evaluation framework**, as described by Arbour (2020). Section 2.4.2.6 showed this encompasses institutional values (e.g., Batho Pele), measurement indicators, and may take the form of a model (i.e., a visual representation, see section 3.1). Various governments, organizations, etc., have developed evaluation frameworks. The national evaluation policy framework (DPME, 2011) emphasizes these items.

Table 5.18 shows the mean score of the factor was 3.29 (SD = 0.86). That is, most participants did not agree or disagree with the statements. Notwithstanding the high internal consistency (α = .86), items from several other dimensions of EOEP, e.g., from resources, activities (communicating evaluation results), etc., loaded on this factor. A larger sample may improve the factor structure and, thereby, content validity.

Factor 2: Use of results. This seems to be the second most important factor for participants. The highest loading items (at .89 and .61) measured whether "Management produces strategy on how to implement recommendations" and "There is follow up to implement recommendations". These are conceptually related to the use of recommendations and, thereby, the use of evaluation results, i.e., the *Use* bundle. "How evaluation is done is reviewed regularly" is also relevant (e.g., regular review of how evaluation is done may help improve management response and implementation of recommendations); however, the loading was moderate (.43).

This factor may aptly be named **Use of results** (also called use of findings or product). This is an important part of evaluation (Bourgeois & Naré, 2015; Fleischer & Christie, 2009; Gagnon et al., 2018). Section 2.4.2.4 showed that effect (or impact) occurs after the use of findings or participation. The national evaluation policy (DPME, 2011) stipulates these items.

The mean was 2.72 (SD = 1.03), which suggests that, on average, the participants neither disagreed or agreed on the items. The reliability test (α =.77) showed a high internal consistency. The item on data quality seemed out of place. Improving the factor structure (e.g., by adding more items, increasing the number of participants, etc.) may move this item to an appropriate factor.

Factor 3: Organizational capacity. The most significant items, "Outcomes are regularly evaluated"; "Evaluation budget is adequate" and "Management supports evaluation" loaded at .69, .61, and .55, respectively. "Evaluation tools (e.g., software) are adequate" also loaded, albeit moderately (.49). The literature review showed adequate resources are necessary for evaluation; management support is a complementary resource.

These items relate to **Organizational capacity** to evaluate e-government outcomes. Capacity (i.e., ability to do something) is linked to resources in practice (Erden et al., 2014). Organizational capacity is also important in evaluation, e.g., Gagnon et al. (2018) measured tools and budget; according to Rogers and Gullickson (2018), management support is important, e.g., in resource allocation and use of results.

The mean of this factor was 3.08 (SD = 0.90), which suggests, on average, the participants neither disagreed or agreed on the items. This factor had a high internal consistency (α=.75). The item which measures evaluation indicators may be misplaced and may be moved to an appropriate factor by improving the factor structure (e.g., by increasing the number of participants, adding more items which measure resources, etc.).

Factor 4: Shared understanding. This factor consisted of only two items (i.e., "Stakeholders agree on what to evaluate" and "Stakeholders agree on how to use results"), which derived from the general understanding (see Table5.3). The loadings were significant— .85 and .77 respectively.

The factor may be labelled **Shared understanding**, the central tenet of which is agreement (Bittner & Leimeister, 2014). It seems an understanding of what is to be evaluated, and for what purpose results may be put, was significant to the participants. The preceding chapters showed the importance of stakeholder participation in, e.g., e-government, public value, outcomes evaluation, and practice). The significance of shared understanding among the diverse e-government stakeholders in evaluation is, therefore, apparent.

The mean was 2.98 (SD = 1.16). Table 5.10 showed most participants neither agreed or disagreed to these items. The factor had a high internal consistency (α=.85) and, on face value, the label is appropriate. However, it may be unstable because it consisted of only two items. This may be improved by, e.g., increasing the sample size, increasing the number of items related to general understanding, etc.

Factor 5: Stakeholder participation. The two items (which measured the extent to which "Stakeholders participate in decision making during evaluation" and "Selection of stakeholders

is transparent") loaded significantly at .80 and moderately at .45 in that order. These items belonged to the set which defined Activities (see Table 5.3). Management support, however, cross-loaded highly on Factor 3 (Organizational capacity) and was, therefore, excluded. This factor is linked to **Stakeholder participation**.

As pointed out in the literature review, stakeholder participation is central to various approaches, including development evaluation. Also, public value and practice emphasize stakeholder participation, which may justify the importance of shared understanding.

The mean score was 3.44 (SD = 0.95), i.e., neither agree or disagree. Internal consistency (α=.60) was satisfactory. Furthermore, although conceptually significant, this factor may not be stable and may be improved by adding more items to the questionnaire, increasing sample size, etc.

Factor 6: Affective issues. The two items "Participation increases stakeholders' willingness to participate in future evaluations" and "Participation increases stakeholders' commitment to implement recommendation" loaded significantly at .77 and .76 respectively. These were from the set which measured teleoaffective structures (see Table 5.3). The literature review showed there is growing recognition of affective issues (e.g., attitude, commitment, and motivation) in evaluation practice, public value, and practice theory. It may be concluded that **Affective issues** were significant to the participants.

The mean score for this factor was 3.16 (SD = 1.13), which suggested neither agree or disagree. Internal consistency was high (α = .74). Notwithstanding conceptual significance, the factor may be unstable and may be improved by adding more items which defined Affect to the questionnaire, increasing the sample size, etc.

Tables 5.19 and 5.20, respectively, showed moderate intercorrelation (i.e., α = .6) among the six factors and between specific factors (see, e.g., Factor1 and Factor2 as well as Factor2 and Factor3). These were expected: as explained in the preceding chapters, evaluation practice is a social activity and aspects thereof are not completely independent.

This section established that, in general, the six factors were statistically and conceptually significant, although some may not be stable, i.e., contained fewer items. However, the final solution was practically significant: it provided a useful conceptual framework for understanding e-government outcomes evaluation. The factors were, therefore, submitted to further analysis. It is worth noting, as aforementioned, that the labels are mere descriptions, and may not be taken as correct or to constitute constructs of the EOEP.

5.5.3 Further analysis of elements: comparing by group

Section 5.4.4.6 discussed the importance of data reduction and showed that composite scales are reliable and applicable to exploratory research; also, average scores are preferable. As shown in section 3.2.5, practice theory assumes there are differences in practice among practitioners, according to their characteristics. Thus, further analysis was conducted to compare participants' perceptions on the factors above. The average scores served as substitutes for the factors and the tests were conducted at $p < .05$ and 95 percent confidence interval (CI).

The main proposition "*There is no difference in perception among participants on EOEP*" was drawn in section 3.4. The following were derived to organize the analysis and draw conclusions:

1. There are no differences in perception on Evaluation Framework, according to participants' characteristics
2. There are no differences in perception on Use of results, according to participants' characteristics
3. There are no differences in perception on Organizational capacity, according to participants' characteristics
4. There are no differences in perception on Shared understanding, according to participants' characteristics
5. There are no differences in perception on Stakeholder participation, Evaluation Framework according to participants' characteristics
6. There are no differences in perception on Affective issues, according to participants' characteristics

The characteristics considered were: position, background, experience (i.e., years of experience and knowledge of EOEP). Section 5.4 showed the nature of data determines the appropriate analysis techniques. In consideration of the smaller sample size (N=106) and non-normality, non-parametric techniques were adopted for further analysis. Mann-Whitney U and Kruskal-Wallis tests were conducted. The results are displayed in Tables 5.21 and 5.22. The corresponding decisions to support or reject the propositions above are summarized in Tables E3.1 and E3.2 (Appendix E3).

5.5.3.1 Analysis by position and experience

Both position and experience consisted of two categories, therefore, Mann-Whitney U test was an appropriate technique to explore differences in perceptions of EOEP. Table 5.21 shows the results (test statistics, 2-sided significance, and mean ranks of groups). Table E3.1 shows whether or not the results supported the proposition. The highlighted propositions were rejected.

There were significant differences by position across all factors, except Affective issues. Also, the mean ranks of those in management position (i.e., *Yes*) were higher across all the factors—a suggestion that more participants in this group agreed or strongly agreed with the items which composed the factors.

Similarly, there were significant differences among participants across all factors according to experience, except on Affective issues. The mean ranks showed many participants with more than five years' experience agreed.

5.5.3.2 Analysis by background and knowledge

Both professional background and knowledge consisted of three categories and, therefore, Kruskal-Wallis test was appropriate. Table 5.22 shows the results (Chi-Square, degree of freedom, and significance at p = .05 and CI = .95). Table E3.2 shows whether or not the results supported the propositions. The highlighted propositions were rejected.

The mean rank of participants was highest and lowest for Evaluation and Other backgrounds respectively. That is, more participants within Evaluation agreed or strongly agreed on the items which formed the factors while the opposite applied to Other. Overall, there were significant differences according to background, except for Evaluation framework. However, such observations may not be conclusive (Field, 2017); thus, further analyses (pairwise comparisons) were conducted.

Figure E3.1 (Appendix E3) shows the results. A yellow line indicates a significant difference between respective pairs of factors. There were significant differences among all the groups on Organizational capacity. This was the only factor on which Evaluation and IT differed (i.e., higher perception compared to the latter). On the other five factors, Evaluation had the highest means and while Other had the lowest.

On knowledge, Table 5.22 shows there were significant differences across all factors, except Affective issues. The mean ranks of participants with Above average knowledge were highest across all factors (i.e., agreed or agreed strongly). Figure E3.2 (Appendix E3) shows there were significant differences across all groups on Organizational capacity. On the other five factors, there were significant differences between Above average against Average and Minimum.

Table 5.21: Mann-Whitney test (Differences by management position and experience)

			Elements					
			Framework	*Recommendations*	*Capacity*	*Understanding*	*Participation*	*Issues*
Management	Mann-Whitney U		618.00	958.50	955.00	965.50	1012.00	1202.00
	Z		-4.64	-2.44	-2.46	-2.42	-2.11	-.86
	Asymp. Sig. (2-tailed)		.00	.02	.01	.02	.04	.39
	Mean rank	Yes	70.93	62.62	62.71	62.45	61.32	56.68
		No	42.51	47.75	47.69	47.85	48.57	51.49
Experience (Years)	Mann-Whitney U		1884.50	1935.50	2065.00	1810.00	1967.50	1696.00
	Z		3.13	3.46	4.28	2.70	3.71	1.96
	Asymp. Sig. (2-tailed)		.00	.00	.00	.01	.00	.05
	Mean rank	1 to 5	45.01	44.13	41.90	46.29	43.58	48.26
		> 5	63.76	64.82	67.52	62.21	65.49	59.83

Table 5.22: Kruskal-Wallis test (Differences by background and knowledge, df=2)

			Elements					
			Framework	*Recommendations*	*Capacity*	*Understanding*	*Participation*	*Issues*
Background	Chi-Square		2.18	17.43	41.45	6.13	9.60	7.97
	Asymp. Sig.		.34	.00	.00	.05	.01	.02
	Mean rank	IT	49.78	55.33	55.79	51.93	56.83	57.34
		Evaluation	60.76	70.72	80.52	65.30	64.24	62.39
		Other	52.50	38.40	30.50	46.53	41.47	42.25
Knowledge	Chi-Square		25.97	9.24	30.63	12.59	22.03	2.02
	Asymp. Sig.		.00	.01	.00	.00	.00	.36
	Mean rank	Minimum	36.82	50.41	28.75	41.41	31.64	46.93
		Average	43.74	44.31	48.13	47.24	50.14	52.43
		Above average	72.00	64.31	71.83	66.10	68.31	58.01

An Agree-Disagree Likert scale provides negative and positive directions (Dolnicar, 2013). The following general conclusions may be drawn. There were differences among participants across all factors, except Affective issues. Managers from Evaluation and IT backgrounds with experience (i.e., more than five years of experience and above average knowledge) were likely to have a positive perception about EOEP (i.e., agree or strongly agree about items). On organizational capacity, there were differences across all categories by position, background, years of experience, and knowledge.

The structure which described participants' perceptions of EOEP was further explored with cluster analysis to understand how the elements combined (i.e., patterns of performance) and the differences among groups. The results were compared to the foregoing conclusions.

5.5.4 Cluster analysis

Section 5.4.5 outlined cluster analysis and a strategy for analysis which involved purpose, sample size, the cluster variables, objects to group and their relevant characteristics, measure of similarity, cluster technique, the number of clusters to identify, validation, and interpretation of results.

5.5.4.1 Purpose and sample size

From section 5.4.5.1, cluster analysis was employed to group participants according to the similarity in their perceptions of how e-government outcomes evaluation is done, identify elements of EOEP which defined members in a group, and establish which characteristics of participants predict membership of a group.

The analysis, therefore, addressed the research question: "*What are the patterns of performance among the different stakeholder groups*"? Put differently, "*How do different groups perform their practice*"? Thus, the steps were: identify and describe the different ways e-government outcomes evaluation is done (i.e., the different patterns of performance) and establish the relationships between clusters and the demography of participants.

5.5.4.2 Cluster items and sample size

As aforementioned, the average scores represented the six elements of EOEP and, therefore, constituted the clustering variables or items. The objects to cluster were participants. The sample size (N=106) was adequate, based on the 2^m guideline (see section 5.4.5.1), where m=6, i.e., number of cluster items.

Section 5.4.4.6 showed average scores technique maintains measurement scale, which allows comparison among factors, therefore, standardization was not necessary. Table 5.23 displays the characteristics of the average scores. Table 5.24 shows generally low correlations (the factor-cluster approach reduced the multicollinearity). The cluster items were metric, consequently, a distance measure of similarity was adopted. The analysis and results are presented below.

Table 5.23: Descriptive statistics of elements

	Minimum	*Maximum*	*Mean*	*Std. Deviation*
Framework	1.33	4.67	3.29	.86
Results	1.00	4.75	2.72	1.03
Capacity	1.60	4.80	3.08	.90
Understanding	1.00	5.00	2.98	1.16
Participation	1.00	5.00	3.44	.95
Issues	1.00	5.00	3.16	1.13

Table 5.24: Pearson Correlation among factors

	Framework	*Results*	*Capacity*	*Understanding*	*Participation*	*Issues*
Framework	1					
Results	-.032	1				
Capacity	.142	.559**	1			
Understanding	.482**	.185	.259**	1		
Participation	.344**	.192*	.198*	.216*	1	
Issues	.120	.131	.208*	.144	.178	1

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

5.5.4.3 Clustering technique: Two-step analysis

This was adopted, to combine hierarchical and non-hierarchical analyses. Section 5.4.5.3 outlined the benefits, e.g., according to Hair et al. (2018), it suits many research purposes. Studies (e.g., Besner & Hobbs, 2013; Mooi & Sarstedt, 2014; Norusis, 2012; Pullinger et al., 2013) provided practical illustrations of how the technique may be applied to draw managerial implications. The SPSS Two-step procedure was employed.

Prior to this, hierarchical and non-hierarchical were explored to gain some insight about an optimal solution. In the hierarchical analysis, Squared Euclidean distance was adopted (Likert scale data exhibits both patterns and magnitude (Hair et al., 2018)). Centroid and Ward's methods were explored. Characteristically, two clusters were identified, however, the agglomeration coefficient changes were ambiguous as illustrated in Figure 5.6. Thus, several numbers of clusters were explored. This was followed with non-hierarchical (k-mean) in which the number of clusters from the hierarchical was the seed. This identified three clusters; the relative sizes were 41, 27, and 32 percent.

The two-step analysis commenced with auto-clustering (i.e., default settings). Different solutions were explored with Log-likelihood and Euclidean distance measurements, automatic number of clusters and otherwise, and Bayesian Information Criterion (BIC) as well as Akaike Information Criterion (AIC). The Log-likelihood solutions were easier to interpret; each suggested three clusters. Tables 5.25 shows the BIC auto-clustering results. It may be observed that the optimal solution, indicated by smaller BIC Change combined with larger Ratio of BIC change and Ratio of Distance Measure, was three clusters.

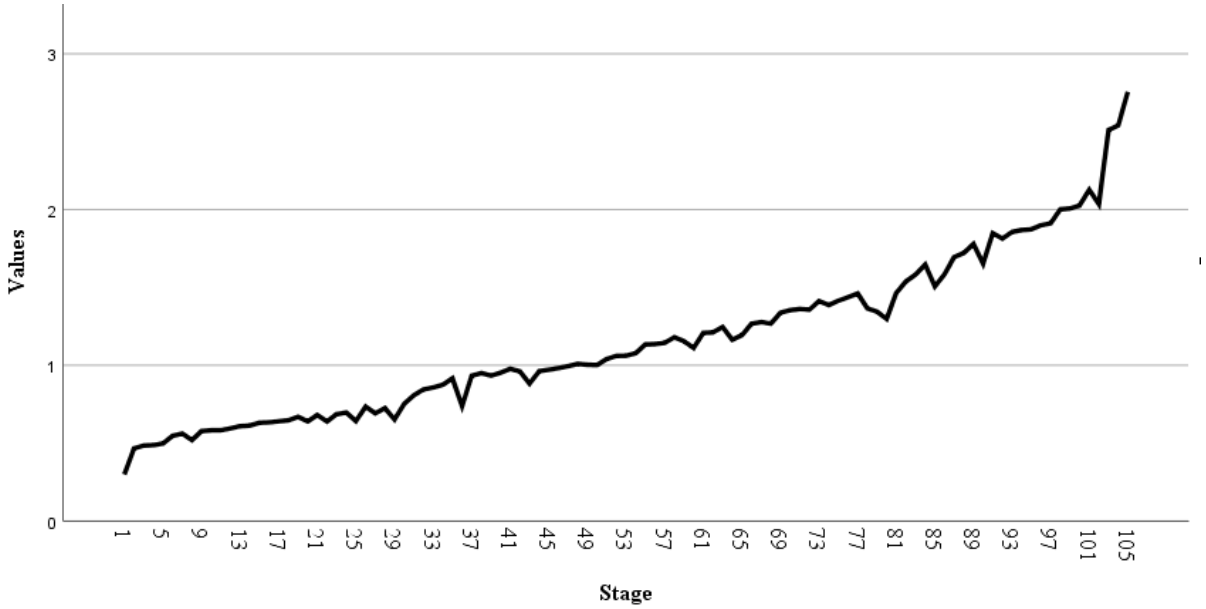


Figure 5.6: Agglomeration schedule coefficients (Centroid method)

Table 5.25: Auto-Clustering

Number of Clusters	*Schwarz's Bayesian Criterion (BIC)*	*BIC Change*[a]	*Ratio of BIC Changes*[b]	*Ratio of Distance Measures*[c]
1	491.306			
2	455.63	-35.67	1.00	1.39
3	445.49	-10.15	.28	2.77
4	477.61	32.12	-.90	1.19
5	513.58	35.97	-1.01	1.32
6	554.35	40.77	-1.14	1.03
7	595.58	41.23	-1.16	1.10
8	638.20	42.62	-1.19	1.01
9	680.90	42.70	-1.20	1.04
10	724.13	43.23	-1.21	1.47
11	771.43	47.30	-1.33	1.02
12	818.86	47.44	-1.33	1.22
13	867.86	49.00	-1.37	1.02
14	917.00	49.15	-1.38	1.07
15	966.61	49.61	-1.39	1.02

a. The changes are from the previous number of clusters in the table.
b. The ratios of changes are relative to the change for the two cluster solution.
c. The ratios of distance measures are based on the current number of clusters against the previous number of clusters.

5.5.4.4 Validation and interpretation

Illustrations (e.g., Besner & Hobbs, 2013; Hair et al., 2018; Mooi & Sarstedt, 2014; Sinclair-Maragh et al., 2015) are instructive. Commonly used techniques include the Silhouette measure of cohesion and separation (i.e., goodness-of-fit) and statistical test of differences. The Silhouette coefficient was 0.4 for each cluster solution, an indication that the cluster quality was fair (Mooi & Sarstedt, 2014).

The four- and five-cluster solutions were elided because their ratio of sizes (smallest clusters to largest) were above 3 (Matsimbe et al., 2018) and some clusters were below 10 percent of the sample size (Hair et al., 2018). The three-cluster solution was selected as optimal.

Interpretation involves examining the cluster centroids (by comparing the lowest and highest) and labelling the clusters (Hair et al., 2018). Table 5.26 shows the centroids and standard deviations of the cluster items as well as the overall solution. As displayed in Figure 5.9, there were significant differences in centroids, an indication that the clusters were distinct. Furthermore, according to Table 5.27, results of the *F* test of ANOVA (with cluster membership and cluster items as independent and dependent variable respectively) showed significant differences (p=.05), which implies the clusters differ on at least two of the cluster items (Mooi & Sarstedt, 2014). Consequently, the three cluster solution may be considered acceptable: parsimonious yet more meaningful, compared to two.

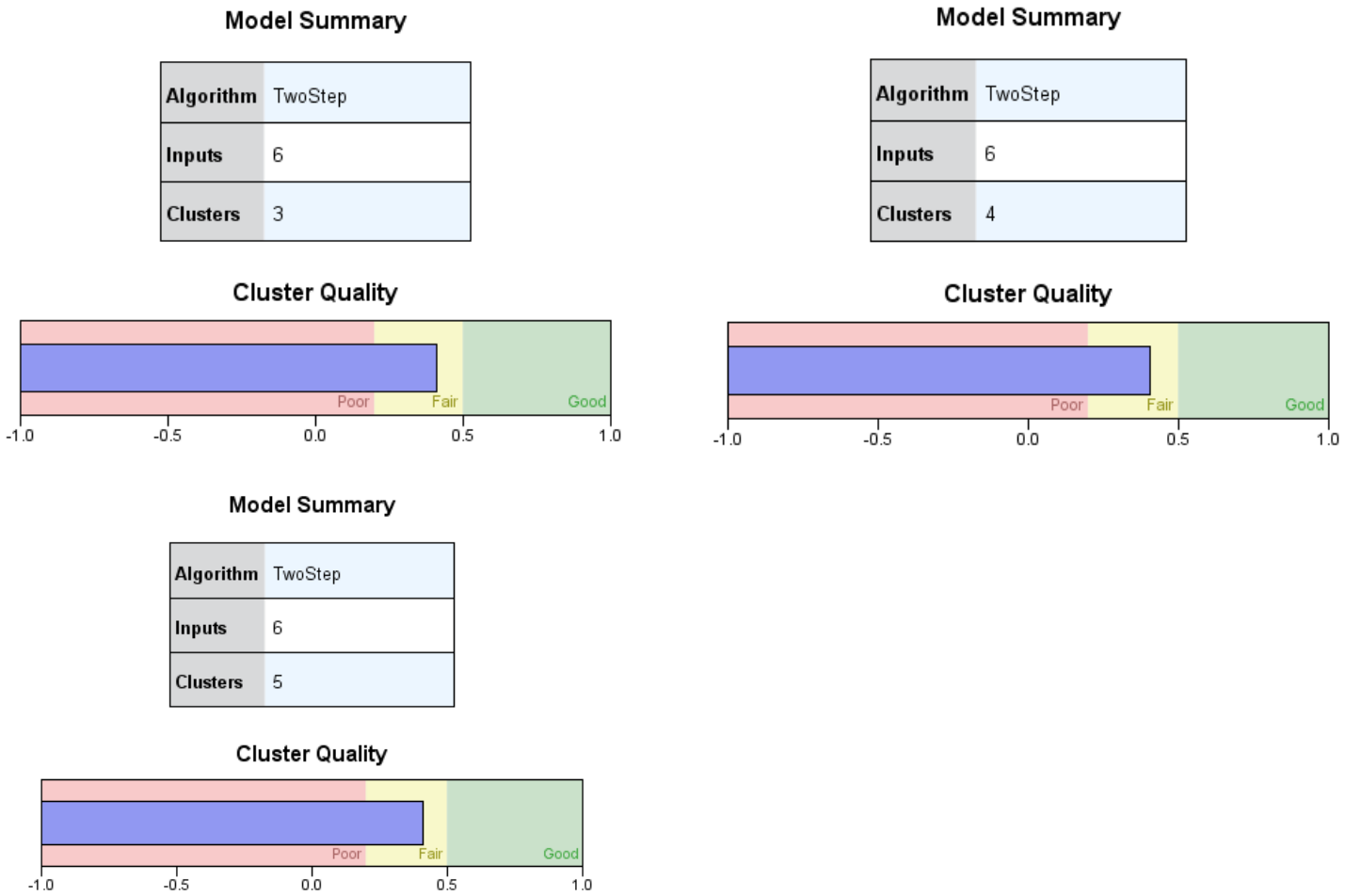


Figure 5.7: Cluster quality

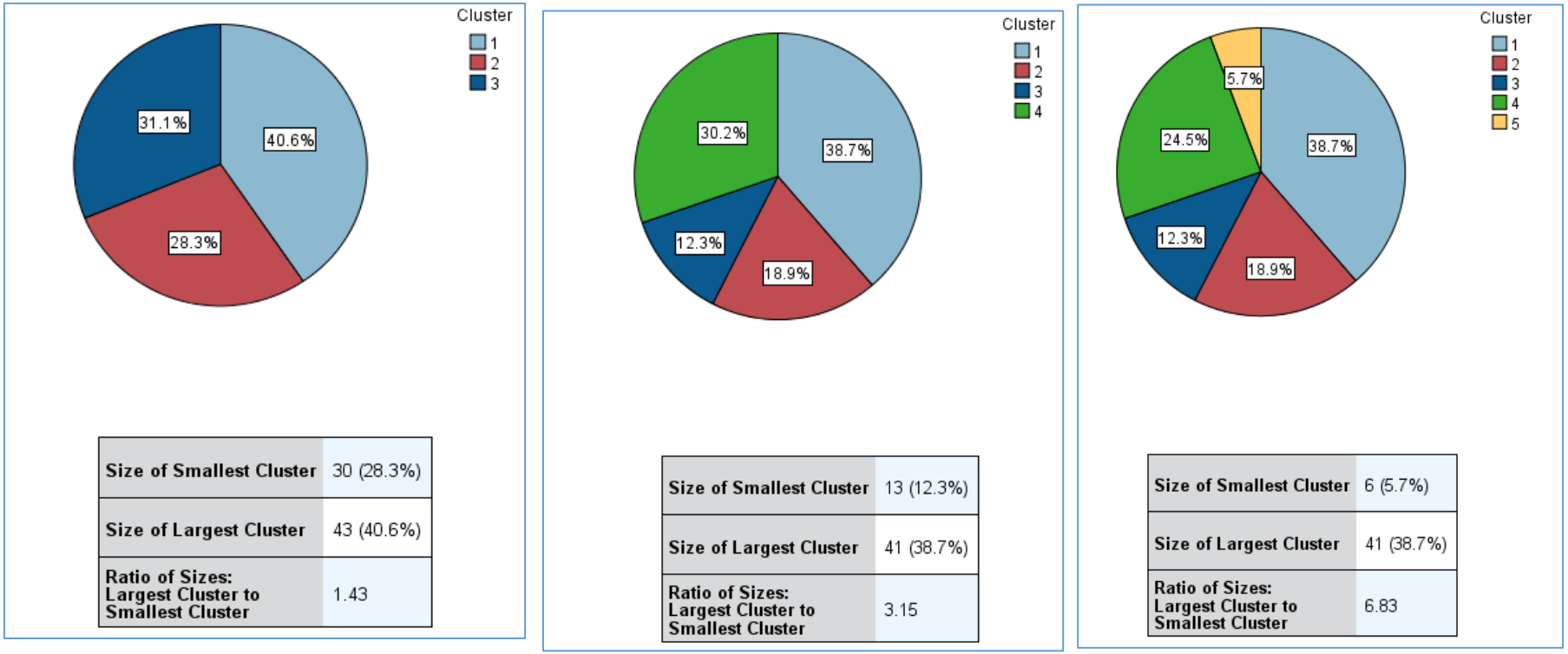


Figure 5.8: Relative sizes of clusters

Table 5.26: Centroids of cluster variables (on a 5-point Strongly disagree to Strongly agree Likert scale)

	Cluster 1 (N=43)		*Cluster 2 (N=30)*		*Cluster 3 (N=33)*		*Combined*	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Framework	3.31	.42	2.53	.87	3.94	.73	3.29	.86
Results	1.81	.54	3.33	.77	3.36	.83	2.72	1.03
Capacity	2.29	.37	3.55	.89	3.69	.59	3.08	.90
Understanding	2.79	.65	1.98	1.06	4.14	.69	2.98	1.16
Participation	3.16	.54	2.98	1.23	4.21	.53	3.44	.95
Issues	2.72	.87	3.23	1.14	3.67	1.22	3.16	1.13

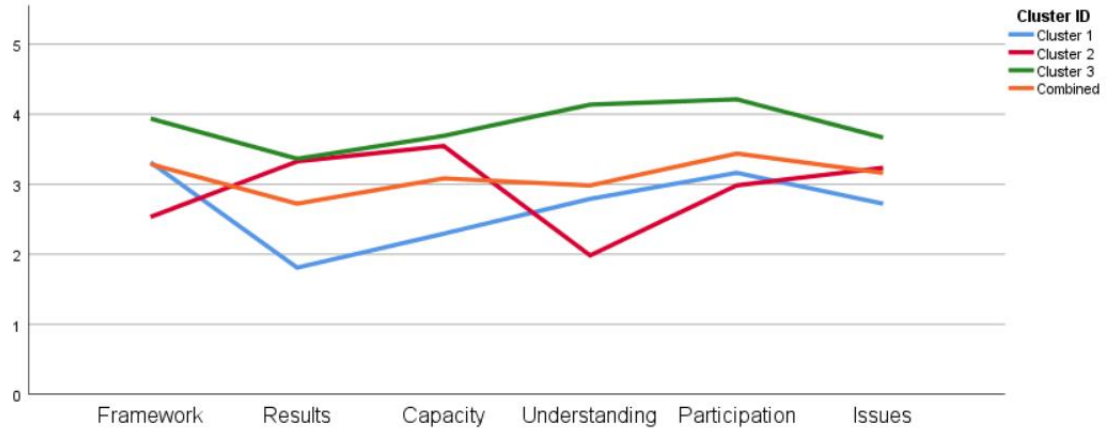


Figure 5.9: Centroids of clusters

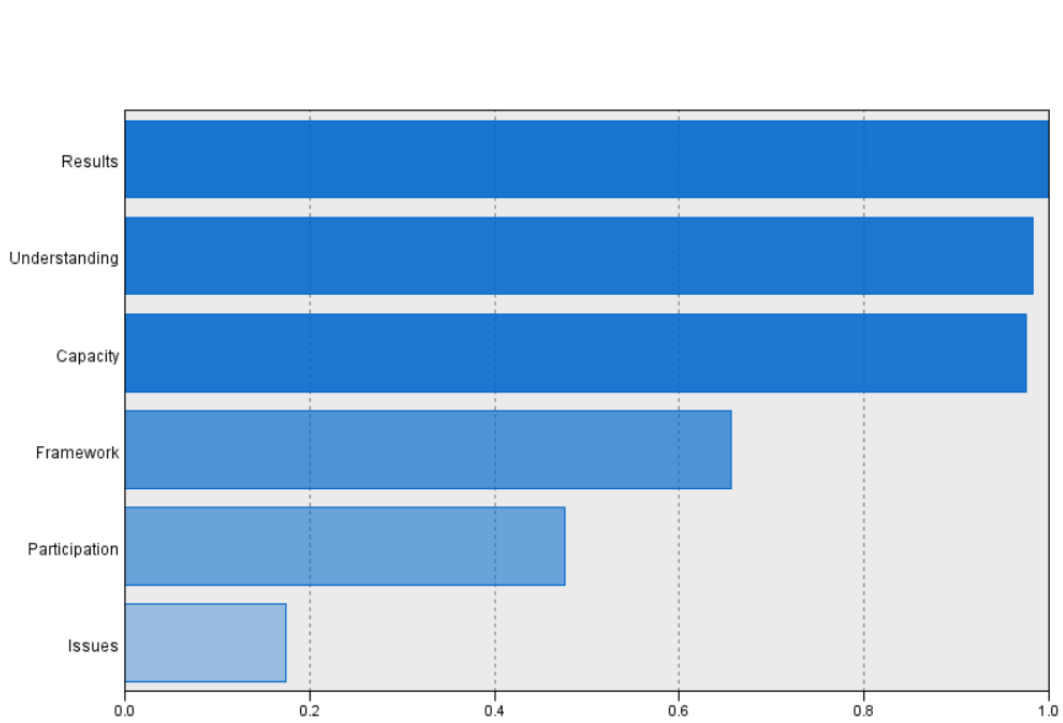
Table 5.27: ANOVA of cluster means (df=2)

	Sum of Squares	*Mean Square*	*F*	*Sig.*
Framework	30.969	15.485	34.438	.000
Results	60.405	30.203	60.837	.000
Capacity	45.480	22.740	58.746	.000
Understanding	75.468	37.734	59.343	.000
Participation	29.234	14.617	23.121	.000
Issues	16.922	8.461	7.458	.001

5.5.4.5 Clusters as variants of performance

As explained in Chapter 3, how elements of practice combine during performance manifest in patterns of performance (i.e., practice as performance). Figure 5.10 shows three different patterns of e-government outcomes evaluation; each represents a group of participants with similar perceptions of performance. Details of the solution include distribution of clusters, described above; importance or contribution of the elements which constitute a cluster; and the corresponding means of such elements.

The clusters and their constituent elements may be examined individually, and together. Figure 5.10 shows importance (i.e. the ability of an element to assign an object to the appropriate cluster), measured on a scale of 0 to 1 and arranged in order of importance, top-down. Thus, the most important (i.e., item with the highest mean) was Use of results; conversely, Affective issues contributed the least to grouping the participants.



Input (Predictor) Importance

1.0 0.8 0.6 0.4 0.2 0.0

Cluster	1	3	2
Label			
Size	40.6% (43)	31.1% (33)	28.3% (30)
Inputs	Results 1.81	Results 3.36	Results 3.33
	Understanding 2.79	Understanding 4.14	Understanding 1.98
	Capacity 2.29	Capacity 3.69	Capacity 3.55
	Framework 3.31	Framework 3.94	Framework 2.53
	Participation 3.16	Participation 4.21	Participation 2.98
	Issues 2.72	Issues 3.67	Issues 3.23

Figure 5.10: Components of clusters

Clusters are characterized by their most important items, e.g., Sinclair-Maragh et al. (2015) described clusters by the two items most important to the participants. As evidenced by the centroids in Figure 5.9, Cluster 3 had the highest means across all items; Cluster 2 was higher than Cluster 1 on three and vice versa, however, the latter had generally low means. On the three important elements, the means of Cluster 2 were higher than Cluster 1 on two.

5.5.4.6 Profiling practice by demography of participants

Profiling establishes the relationship among clusters and external data of interest. For example, demography of a population (age, position, experience, etc.) may be used to "profile", i.e., assign objects to clusters (Hair et al., 2018; Mooi & Sarstedt, 2014; Norusis, 2012). This also helps to achieve criterion validity.

A common approach is to cross-tabulate the cluster membership data with the cluster items (Hair et al., 2018; Sinclair-Maragh et al., 2015). Another is to employ the evaluative field in the two-step SPSS technique (Norusis, 2012).

From the main proposition "*There is no difference in perception among participants on EOEP*", the following were derived to explore and draw conclusions about the patterns of performance according to the demography of participants.

1. There are no differences in perception of performance according to participants' position
2. There are no differences in perception of performance according to participants' background
3. There are no differences in perception of performance according to participants' years of experience
4. There are no differences in perception of performance according to participants' knowledge

For corroboration, both cross-tabulation and evaluative fields were used. Table 5.28 and Figure 5.11 show the results (note the correspondence between the values highlighted in the table and those in the figure). The Chi-square tests showed there were significant differences across all clusters by management position, background, years of experience, and knowledge. It may be concluded that participants' perceptions of how e-government outcomes evaluation is done (i.e., the patterns of performance) differed according to their characteristics. The results support the propositions. The evaluation fields in Figure 5.11, in conjunction with Table 5.28, bring the succinct differences among the clusters to the fore.

Table 5.28: Cluster composition by management position; background, experience and knowledge

		Cluster Number		
		Cluster 1 N=43	*Cluster 2 N= 30*	*Cluster 3 N=33*
Management ($\chi(2)=9.746$, p=.008)	Yes	12 (27.9%)	9 (30.0%)	20 (60.6%)
	No	31 (72.1%)	21 (70.0%)	13 (39.4%)
Background ($\chi(4)=38.515$, p=.000)	IT	14(32.6%)	13 (43.3%)	16 (48.5%)
	Evaluation	1 (2.3%)	11 (36.7%)	15 (45.5%)
	Other	28 (65.1%)	6 (20.0%)	2 (6.1%)
Experience ($\chi(2)= 27.097$, p=.000)	1 to 5	33 (76.7%)	19 (63.3%)	6 (18.2%)
	More than 5	10 (23.3%)	11 (36.7%)	27 (81.8%)
Knowledge ($\chi(4)= 22.536$, p=.000)	Minimum	13 (30.2%)	8 (26.7%)	1 (3.0%)
	Average	22 (51.2%)	11 (36.7%)	9 (27.3%)
	Above average	8 (18.6%)	11 (36.7%)	23 (69.7%)

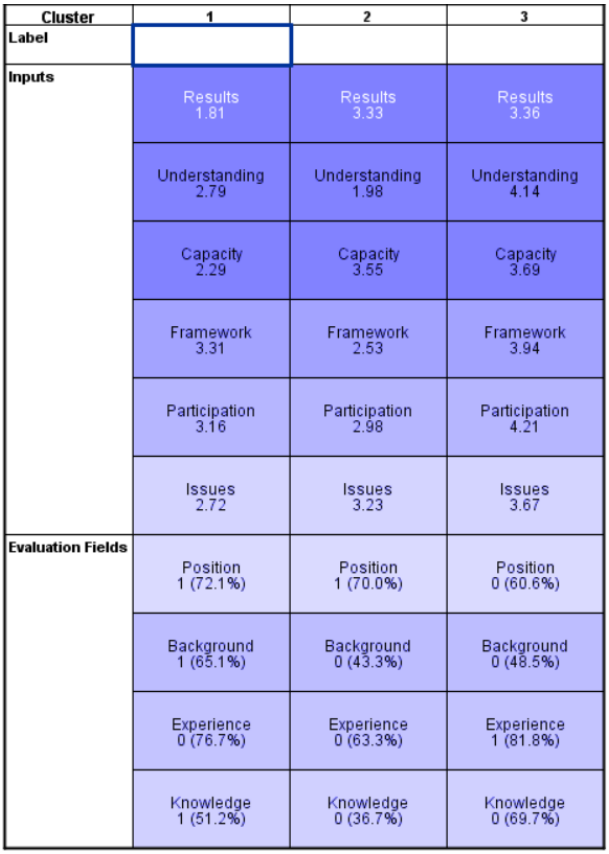


Figure 5.11: Characteristics of participants and variants of practice

It may be observed that Cluster 3 was composed primarily of participants in Management position, almost exclusively from IT and Evaluation backgrounds with Above five years of experience and Above average knowledge of how e-government outcomes evaluation is done.

Cluster 2 comprised mainly of participants in Non-management position from IT and evaluation backgrounds with 1 to 5 years of experience with Average or Above average knowledge. It may be observed that Clusters 2 and 3 differed on management position and experience. Cluster 1 may be differentiated from Cluster 2 mainly by background and

knowledge, the latter mainly consisted of participants of Other background with Average or Minimum knowledge.

The following conclusions may be drawn. Cluster 3, linked to IT and Evaluation Managers with extensive experience; Cluster 2, to IT and Evaluation Non-Managers with experience; and Cluster 1, to Other Non-managers with little experience. Moreover, IT and Evaluation Managers with extensive experience (i.e., participants in Cluster 3) had a positive perception of EOEP; as evidenced by the higher means in Figure 5.11, this group generally agreed (i.e., agreed or strongly agreed) to the items which composed the three most important elements. The foregoing is consistent with the conclusion drawn from the Mann-Whitney and Kruskal-Wallis tests in section 5.5.3.

The cluster solution may, therefore, be considered practically useful: it identified the patterns of performance, linked to the characteristics of participants. This way, participants can be assigned to a pattern of performance according to their characteristics. Thus, the solution can inform a strategy that targets the membership of groups to improve practice. Managerial implications (e.g., how the solution can improve EOEP) are drawn in conjunction with other results, in Chapter 7.

Table 5.29 summarizes the results from the quantitative analyses, i.e.: descriptive, factor and comparisons (Mann-Whitney and Kruskal-Wallis tests) and cluster.

Table 5.29: Summary of quantitative results (N=106)

Descriptive	*Factor analysis and comparisons*	*Cluster analysis*
Evaluation initiation: within an organization **Evaluation teams** consisted of both internal and external stakeholders **Survey** is widely used to measure accountability and organizational performance Evaluation is mainly for c**ompliance and decision-**making and learning about e-government A generally negative perception of EOEP	Differences across all factors, according to position, background, experience and knowledge, except Affective issues. In general, participants did not have positive perceptions of EOEP (neither agreed or disagreed across all factors) Positive perception of EOEP among Management, Evaluation, Over five years' experience, and Above average knowledge In general, no difference between IT and Evaluation against Other; and also Above average against Average and Minimum experience	Three clusters described the different ways e-government outcomes evaluation is done (i.e., patterns of performance) **Cluster 3**: IT and Evaluation managers with extensive experience (years and knowledge); overall: positive perception of EOEP (i.e., highest means across top 3 elements) **Cluster 2**: Non-managers from IT and Evaluation with moderate experience (years and knowledge) **Cluster 1**: Non-managers with little experience (years and knowledge); overall: negative perception of EOEP

Summary

This chapter focused on how Phase 1 was conducted to address the research problem. The primary objectives were to identify the elements which described e-government outcomes evaluation and the different patterns of performance. A questionnaire was designed to collect data from a sample of public employees; 106 responses were analyzed. Table 5.29 summarized the important results.

Descriptive analysis showed that, in general, participants had a negative perception (i.e., did not agree) to the statements about EOEP. Factor analysis identified a structure which consisted of six main elements. Further analysis of the factors established significant differences in perceptions of EOEP, according to the demography of participants. Cluster analysis identified three distinct patterns of performance, linked to participants' demography. The cluster solution seems to support the differences established by Mann-Whitney and Kruskal-Wallis tests.

Notwithstanding the effort at validation, this research, like all others, has limitations. This is addressed in the context of the entire research in Chapter 7. Suffice to state that strategies (e.g., on sampling, questionnaire design and administration, data collection and analysis, and ethics) were devised to obviate biases and errors to assure desired outcomes.

The factor and cluster solutions were statistically and conceptually significant as well as practically useful. Phase 2 was designed as qualitative research to follow up on the results above to gain understanding, and it to this, which the next chapter turns.

Chapter 6 : Phase 2 (Qualitative data collection and analysis)

Section 4.8 outlined a two-phased SEMM survey strategy to answer the research questions below. Phase 1 was quantitative research, to identify the main elements of EOEP and patterns of performance; the preceding chapter described the implementation and presented the results.

RQ1: What are the factors which describe how evaluation of e-government outcomes is done?

RQ2: What are the patterns of performance among different groups? (put differently: "*How do different groups perform e-government outcomes evaluation*"?)

This chapter describes the implementation and results of Phase 2, qualitative research to understand the results of Phase 1. Therefore, both RQ1 and RQ2 were relevant: to understand what is done and how. The design decisions and applicable implementation tactics are reiterated in Table 6.1 for accessibility.

Table 6.1 Research strategy for Phase 2

Phase	*2*
Purpose	To understand the preliminary results in Phase 1
Research type	Qualitative
Research question answered	Question 1 and 2
Data source	Sub-sample drawn from the sample in Phase 1
Type of data	Mainly qualitative
Logic of inquiry	Deductive and inductive
Sampling	Purposive, stratified
Data collection	Semi-structured interview
Data analysis	Non-statistical (content analysis)
Validation	Credibility, dependability, transferability, confirmability

Purposive, stratified sampling was applied to select a sub-sample of 12 from the participants of Phase 1. An interview protocol was constructed and semi-structured interviews were conducted to collect data, which was subsequently submitted to content analysis.

The results suggested inadequate internal capacity to conduct evaluation and implement recommendations for improvement; the effect on participants and their organizations was conceptual (e.g., improved knowledge); participants' characteristics influenced understanding of roles and responsibilities as well as extent of participation; policies (i.e., Batho Pele and the National Evaluation Policy Framework) informed evaluation; organizational performance values predominated. Overall, there were significant differences among the groups of participants. Figure 6.1 illustrates the structure of this chapter.

Research is fraught with pitfalls, thus, further steps were taken to reduce, and where applicable obviate, errors and biases to ensure credible process and outcomes. Section 4.9 showed the importance of ethics to this research, in general.

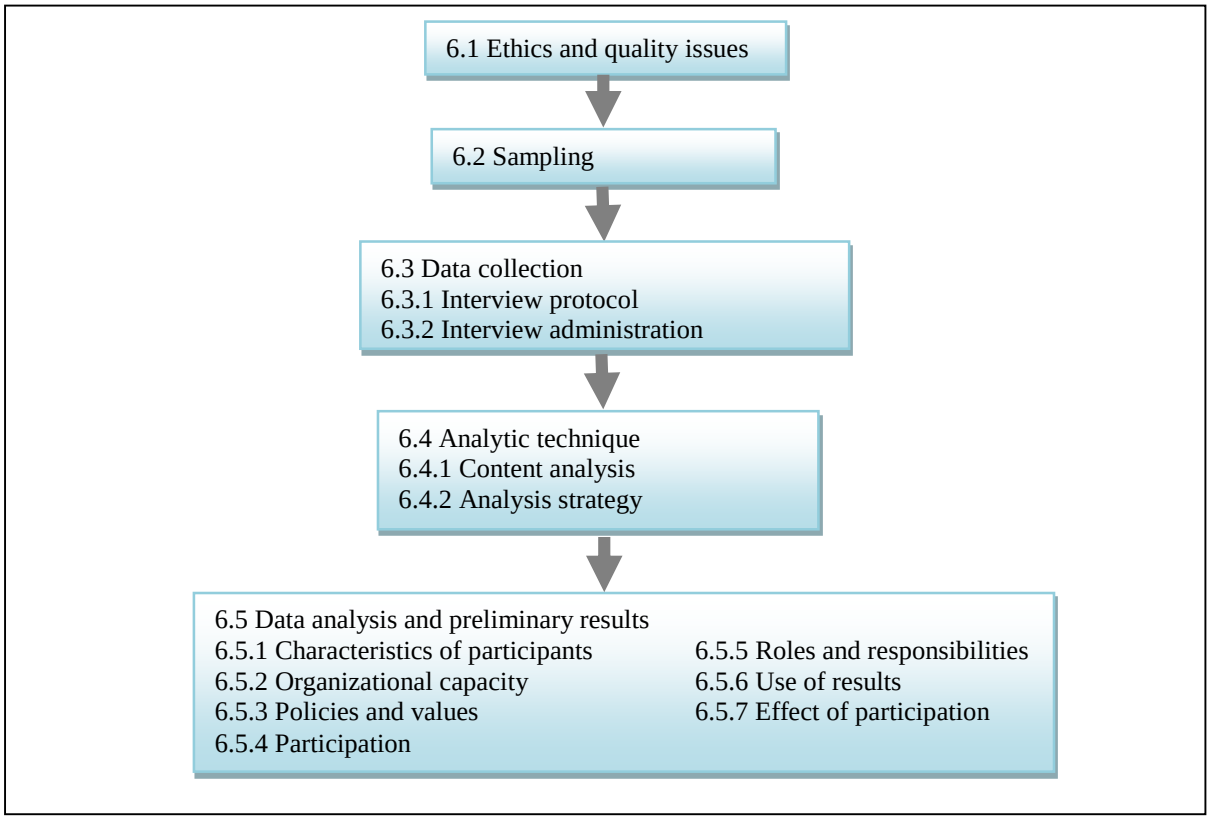


Figure 6.1: Structure of chapter

6.1 Ethics and quality issues

Section 4.3 showed axiology concerns ethics and values and serves as a signpost to research. Ethics and quality are inseparable— more so, in qualitative research. Several decisions depend on the researcher's judgement and there are no criteria to establish quality; furthermore, there are several research models— with different underlying paradigms, strategies, and methods— which can yield different outcomes (Northcote, 2012; Sarker et al., 2013). Addressing ethics and quality issues in advance may improve the understanding of the research audience and increase transparency and credibility.

Section 4.9 discussed the general strategy to address issues of ethics and quality (e.g., institutional requirements, participants' rights, and the researcher's obligations) which are also applicable to this phase and, therefore, need no reiteration. Section 4.6.5 outlined validation strategies in general and the specific aspects to address in qualitative research.

According to Bryman (2016), credibility (i.e., internal validity) is the extent to which research outcomes (data, findings, etc.) are consistent with the research problem; transferability (i.e., external validity) concerns how research can be applied to different contexts; dependability (i.e., reliability) is the degree to which others can come to the same conclusions about outcomes; and confirmability (i.e., objectivity) is the degree to which others can confirm the outcomes as attributable to the research process.

Research purpose, paradigm, contribution, and the field of study are among the factors which determine the quality of qualitative research (Northcote, 2012). Table 6.2 summarizes common techniques to improve quality: self-reflection to obviate bias and errors; relevant research problem; detailed documentation (i.e., audit trail) of decisions and procedures; expert and participant review; requisite and adequate sample; appropriate data collection and analysis methods; representation of participants' views; triangulation (e.g., of sources); interpretations supported by evidence; etc.

Table 6.2: Techniques to improve trustworthiness (Bazeley, 2013; Bryman, 2016; Elo et al., 2014; Noble & Smith, 2015; Sousa, 2014)

Criterion	*Technique*
Credibility	Self-reflection, relevance (research problem); adequate data (sample and sampling technique); expert review (peers, academic); participant validation; instrumentation; appropriate methods; triangulation (methods, people, e.g. reviews); conclusions supported by evidence
Transferability	Adequate information to help determine the applicability of findings to other contexts (e.g., details of sample and sampling technique); audit trail
Dependability	Appropriate theory; accurate instrument (construction and piloting); expert review; audit trail
Confirmability	Reflexion, structured instrument

Trustworthiness commences with a detailed description of the stages in the research process, from the specification of research questions to the interpretation of results (Sousa, 2014). The techniques above were woven into an overall strategy (as shown in Table 6.3) to enhance quality. The research topic was relevant (see sections 1.1 and 1.7) and, in general, posed no harm to participants. Appropriate data sources were selected; an interview protocol was constructed, reviewed, and piloted, and used to conduct semi-structured interviews. Participants' responses were recorded and transcribed, and safely and securely stored. The data was directly analyzed and presented— as provided by the participants. The researcher was reflective and critical throughout the process— to ameliorate, and where possible avoid, bias and improve transparency. Decisions were informed by theory, extant literature, and guidelines, where applicable. Detailed descriptions of decisions and methods were documented at each stage.

Table 6.3:Qualitaive research strategy

Step	*Implementation tactic*
Planning	Relevant research topic
	Specify research objective and question upfront
Sampling	Appropriate data sources, sampling method and sample size
	Specify criterion for inclusion
	Detailed description decisions and actions
Instrumentation	Construct Interview protocol
	Include only valid and relevant items in the protocol
	Review and pilot
	Detailed description of the process
Data collection	Select an appropriate technique: Semi-structured Interview
	Record data (participants' views) "as is"
	Follow-up questions to remove ambiguity
	Detailed description of the process
Data analysis	Transcription to familiarize with interview data
	Appropriate data analysis method: Content analysis
	Initial coding scheme to "focus" analysis
	Review and refine the coding scheme
	Pilot analysis and review results
	Detailed description of the process
Data presentation and	Interpret data "as is" (from participants' point of view)
Interpretation	Use appropriate presentation devices (e.g., tables) to improve understanding
	"Focus" with the research questions, conceptual model, etc.
	Participant validation
	Detailed description of the process
	Compare findings to the conceptual model, theory and extant literature (this is integrated with results of Phase 1)

6.2 Sampling

This was linked to the overall mixed methods research strategy. The results of Phase 1 showed the participants had the requisite knowledge and experience of EOEP and, therefore, were considered an appropriate data source and sample frame for this phase. This can help improve validation because the participants are familiar with the research problem (Jokonya, 2016).

Purposive, stratified sampling strategy was considered appropriate, to allow the selection of participants from different groups and, thereby, comparisons to shed light on the relationship between their characteristics and perceptions of EOEP.

In all, 12 participants were selected for Phase 2. According to Silverman (2013), there is no rule of thumb on sample size for qualitative research; important decisions include adequacy of data to achieve the research purpose, analysis technique, and practical issues, e.g., budget. In general, qualitative research is characterized by small sample sizes. A review by Sarker et al. (2013) found some studies involved six participants; Jokonya (2016) observed 12 to 60 is appropriate; Mckim (2017) noted four to 10 is recommended and selected 11; according to Rowley (2012), 12 interviews of about 30 minutes each may provide adequate data. Thus, 12

was considered adequate for this research; however, the results may not be generalizable beyond this sample.

Participants were selected according to position and professional background; characteristics, e.g. role, etc., are commonly used criterion for inclusion (Palinkas et al., 2015). The distribution is illustrated in Table 6.4. Four (i.e., two managers and two non-managers) were selected from each background, i.e., IT, Evaluation, and Other. The following explicates the data collection and analysis strategies. Although discussed separately, i.e., to facilitate understanding, it is worth noting that these are concurrent.

Table 6.4: Distribution of participants (N=6)

	Position	
Background	*Management*	*Non-management*
IT	2	2
Evaluation	2	2
Other	2	2

6.3 Data collection

As aforementioned, this research was descriptive and the objective of this phase was to understand the results in Phase 1. The research design (see section 4.6.3) showed an interview was an appropriate method to collect detailed information to understand participants' practice —and in their own voices. It allows people to recount their perspectives, perceptions, experiences, attitude, etc. (Flick, 2014) and is widely used as a follow-up in surveys to explore patterns (Gray, 2014).

Also, a semi-structured interview was considered appropriate because it affords both structure and flexibility to the interview process. It is often used to further explore themes identified from questionnaire (Saunders et al., 2015). An interview protocol of precompiled questions was constructed to focus the interview on the specific aspects of EOEP to explore.

6.3.1 Interview protocol

The primary goal of an instrument is to collect relevant data— consistently. Credible, dependable, and unbiased processes and outcomes are essential (Gray, 2014). Several guidelines are available (e.g., Castillo-Montoya, 2016; Jacob & Furgerson, 2012; Leedy & Ormrod, 2015; Rowley, 2012) on how to construct a protocol and administer an interview. Some of the factors to consider include underlying theory, nature and order of questions, and the general structure of the protocol.

The interview protocol focused mainly on the results (i.e., descriptive, exploratory factor, and cluster analyses) of Phase 1 —as summarized in Table 5.29. As may be observed, the main themes thereof are: the factors which described e-government outcomes evaluation practice, the differences in pattern of performance, and participants' characteristics with which to explore such patterns. These are to be examined further to understand what is done and how (from RQ1 and RQ2). The interview protocol was organized around characteristics of participants and the factors (i.e., Evaluation framework, Use of results, Organizational capacity, Shared understanding, Stakeholder participation, and Affective issues). However, as pointed out in sections 5.4.4.5 and 5.5.2.3, labels in factor analysis are not concepts. Thus, the interview questions on factors dwelt on the constituent items. The conceptual model and the questionnaire (Appendix C) formed the bases. Table 6.5 shows the broad aspects of EOEP about which data was collected. As may be noted, these aspects have been discussed in the preceding chapters.

On roles and effects, participants were also asked about their perceptions of other participants and organizations respectively. These are commonly used (e.g., Bourgeois & Naré, 2015; Fierro & Christie, 2017) to provide a perspective on groups and organization.

Table 6.5: Aspects of EOEP examined in Phase 2

	Factor	*Examined…*
1	Characteristics of participants	Background, position, experience and knowledge
2	Organizational capacity (see section 5.5.2.3)	Resources (i.e., expertise and budget) and management support for e-government outcomes evaluation.
3	Policies and values (see section 3.2; related to Evaluation framework)	The extent to which frameworks (i.e., Batho Pele and the National Evaluation Policy Framework (NEPF)) influence evaluation and the dimensions of values to measure
4	Stakeholder participation (see section 2.4.2.7)	The extent of participation in decision making and activities, and disposition to participating in future evaluation.
5	Roles and responsibilities (see section 3.2.4)	Participants' understanding of roles and responsibilities, therefore, the extent of participation
6	Use of results (see section 2.4.2.4 and 3.4)	Management response to, distribution of, and implementation of recommendations
7	Effect of use (see section 2.4.2.4 and 3.4)	Effect of doing and using results of e-government outcomes evaluation on individual and organizations (i.e., participation and use of results)

The interview questions were designed to be simple and clear, to remove ambiguity and elicit appropriate responses from participants, and consisted of open-ended, focused sub-questions (for specific answers) and probes (for elaboration). These were arranged in a logical order (not according to the importance of preliminary results in Phase 1) to provide structure to the interview. Prompts were included to manage the interview process, e.g., introduction to the interview; lead-ins from one question to the next (e.g., *Let's discuss organizational capacity*

for e-government outcomes evaluation); and conclusion, to signify the end of an interview session. An effort was made not to solicit sensitive and personal information.

The interview protocol was reviewed by two post-graduate students and an academic with expertise in qualitative research. The objective was to assess relevance, clarity, and sequence of questions and to remove ambiguity; the feedback was used for improvement. The protocol was piloted (with two post-graduate students with experience in interviewing and one of the participants in Phase 1) and subsequently refined. The final version is shown in Appendix D1.

6.3.2 Interview administration

Leedy and Ormrod (2015) describe important factors to consider, e.g.: identify general questions and follow-up sub-questions; select appropriate data sources; arrange suitable location; request written consent; establish rapport with participants; ask for examples, where necessary; and record responses verbatim to reflect actual responses. Several guidelines have been developed to assist in the interview process (e.g., Jacob & Furgerson, 2012; Rowley, 2012).

Face-to-face interviews were scheduled at participants' convenience; most took place at venues away from their working environment. Each interview lasted about 25 minutes. Three participants were interviewed in Durban; one in Pietermaritzburg, two in East London; one in Bisho, and five in Pretoria. Interviews commenced with an introduction: the researcher expressed gratitude to participants, introduced himself and the purpose of the interview as well as the potential benefits; explained to participants their rights and the researcher's duties and obligations, indicated an approximate duration of the interview, and asked for written consent. Also, participants were encouraged to seek clarity on questions or refrain from responding, where applicable.

There is disagreement on whether or not to record interviews (Punch & Oancea, 2014); however, the benefits (e.g., verbatim and permanent representation of participants' responses) are evident (Sarker et al., 2013). Consent to record was sought, and obtained, from participants. Audio recordings were supplemented by notes; unusual responses and happenings were also written down (this was sporadic, to stay attentive to non-verbal actions and engaged, and not to be engrossed in writing). Follow-up questions, which re-phrased and repeated responses, were used to remove ambiguity.

It was important to establish rapport, reduce apprehension and make participants feel at ease. Two techniques were adopted at the initial stages. After the introduction and exchange of

pleasantries, participants were asked descriptive questions which allowed them to talk freely; responses were rephrased and repeated, where appropriate, to assure participants that the researcher was attentive. Thus, the first item on the interview protocol asked participants to talk about their backgrounds.

Each interview concluded with a clear indication of such; a review of written notes with participants; expression of gratitude for participation; and reassurance that information would not be used in ways to violate their rights. Participants were further informed there may be follow-ups, e.g., to validate results. As aforementioned, the researcher was reflective throughout the process on issues of ethics and quality.

The data collected was prepared, organized, and subsequently submitted to analysis. Consistent with the objective of this phase, a qualitative data analysis technique was adopted. As stated above, the data collection and analysis steps were not completely distinct; e.g., transcriptions and familiarization with content were important for subsequent interviews.

6.4 Analytic technique

Qualitative data analysis organizes and summarizes a large amount of data into a meaningful form. This consists of three basic, interconnected and generally concurrent activities: data reduction, display, and conclusions (Walliman, 2018). Data reduction involves segmenting and summarizing original data, and coding. Data display organizes the output of analysis— often a large volume of data— into meaningful form; common devices include graphs, tables, charts, and diagrams. Drawing conclusions is about generating propositions, a set of general statements, from the data to "paint a picture".

In coding, labels, names, tags, etc., (i.e. codes) are assigned to portions of data (e.g. words or phrases, etc.); codes may be further analyzed, e.g., summarized or grouped into higher-level, abstract codes (Punch & Oancea, 2014; Saldaña, 2015). What may be coded includes activities and perceptions of participants, as well as opinions of the researcher (Saldaña, 2015).

Two main types of codes may be identified: descriptive codes directly express ideas or claims about the data while pattern codes group together codes into meaningful units and may infer beyond the data corpus (Punch & Oancea, 2014). Common techniques in pattern coding include comparison (e.g., to identify similarities and differences), counting occurrences (i.e., frequencies); clustering (e.g., grouping), and correspondence, e.g., with existing theory. There may be many coding cycles; the first cycle often yields descriptive codes while the subsequent generates pattern codes.

Initial codes may be pre-specified (e.g., derived from theory, literature review, previous research, questions in interview protocol, etc.) and applied to coding the data; also, new codes (i.e., emergent codes) may be identified (Saldaña, 2015). The coding process is not linear but iterative— back and forth between data, coding and categorization.

A typical qualitative data analysis strategy involves decisions on appropriate data management, the analysis technique; the level of detail; adoption of software tool, and ethics and quality (Creswell & Poth, 2017; Flick, 2014; Rose et al., 2015; Saldaña, 2015; Walliman, 2018). These factors are not without controversies; Saldaña (2015) averred it is important to acknowledge, but not to reconcile, the contradictions.

Analytic techniques are many and varied (e.g., content analysis; narrative analysis; grounded theory and hermeneutics). However, three main uses may be identified, namely: description, e.g., comparisons to establish similarities and differences among cases; explanation, e.g., to account for such similarities and differences; and theory development, e.g., to generate general statements to develop theory (Flick, 2014). The techniques are not distinct but vary in their capacity to support specific goals of analysis (Punch & Oancea, 2014). For example, while all techniques involve some degree of interpretation to draw conclusions, some may be more amenable to description in participants' voice than explanation which, imposes a researcher's viewpoint.

The selection of an appropriate technique depends on factors such as research purpose; the role of theory; and types and sources of data (Cho & Lee, 2014; Creswell & Poth, 2017; Flick, 2014). Techniques may be broadly classified into two: manifest and latent (Bengtsson, 2016; Rose et al., 2015). The former concerns what is directly observable from the corpus (e.g., exact words, phrases, etc.); is well-suited to description (e.g., reporting participants' experiences—in their words) and applicable to data reduction, e.g., to group data into fewer categories and patterns. Content analysis is a common example. By contrast, latent techniques dig beyond the surface of the data under study to provide explanation; this may require additional information, which may not be readily observable. Examples include hermeneutics, phenomenology, etc. which support explanation.

The primary distinction between deductive and inductive techniques turns on the place of theory (Bengtsson, 2016; Flick, 2014). In the former, theory informs analysis: codes are derived *a priori* from theory and applied to data corpus; the findings are interpreted against the

backdrop of the theory. In inductive, codes emerge from the corpus itself and may be grouped into more general categories to explain the phenomenon of interest.

Some techniques are well-suited to qualitative findings whereas others, e.g., content analysis, can be used for both qualitative and quantitative. The data collection method may also determine an appropriate analysis technique. The manifest technique may be applicable to interview data to describe participants' responses "as is" while latent may be suitable where the researcher's point of view is important (e.g., in participant observation) to generate and interpret data.

There is no generic qualitative data analysis process model. While some techniques prescribe formal guidelines for implementation, in practice, most are eclectic and pragmatic (Flick, 2014). As Punch and Oancea (2014) pointed out, a systematic, disciplined, and transparent process is important in qualitative data analysis to, among other things, assure the trustworthiness of findings. However, others (e.g., Saldaña, 2015) emphasize the unique research context and a creative exploratory process.

Qualitative data analysis may be conducted manually or with software tools. Such tools range from traditional word processors and spreadsheets with support for data storage and management; coding and manipulation (e.g., searching, sorting, displaying, etc.) to a group of specialized software, called Computer Assisted Qualitative Data Analysis Software (CAQDAS), which provide additional features, e.g., quantitative analysis and modelling. Manual analysis may be considered appropriate for small-scale research whereas CAQDAS may be useful for large-scale (Saldaña, 2015). Table 6.6 is a summary of the main decisions on analysis.

Consistent with the descriptive research purpose, *a priori* assumptions were drawn from the preliminary results of Phase 1 and the conceptual model guided this analysis. Thus, descriptive, manifest, deductive analysis was considered appropriate— to represent participants' views "as is". Content analysis was adopted; a detailed discussion follows.

Table 6.6: Qualitative analysis- important factors

Factor	*Analysis decision*
Purpose of analysis (e.g., descriptive, explanatory or theory development)	Descriptive qualitative research
Source (e.g., interviews, observation, etc.)	Interview transcripts
of data and nature findings (i.e., participants' or researcher's view)	Findings to represent participants' views
Level of analysis, i.e., manifest or latent	Manifest, to describe interview responses "as is"
Role of theory, i.e., deductive or inductive	Mainly deductive (*a priori* assumptions from preliminary results of Phase 1)
Use of software tool	Small-scale analysis; mainly manual, supported by a traditional word processor and spreadsheet
Quality/validation	Credibility, dependability, confirmability, and transferability (see Table 6.2)

6.4.1 Content analysis

Content analysis is the systematic examination of data corpus. It supports deduction and induction; manifest, quantitative description as well as latent, qualitative explanation; and various sources of data (e.g., existing documents, interview transcripts, open questions in surveys, etc. (Bengtsson, 2016; Cho & Lee, 2014; Zhang & Wildemuth, 2016). A strongpoint is, therefore, flexibility; the primary weakness is the lack of established procedures.

There are two main types: quantitative and qualitative (Bengtsson, 2016; Zhang & Wildemuth, 2016). However, the two are not dichotomous and may be combined in the same research; e.g., quantitative to establish the frequency of manifest elements of text and qualitative for further understanding. Mayring (2014) characterized it as a mixed-method technique. Content analysis suited the research purpose of this research. Palvia et al. (2017) observed a growing adoption in IS research.

A defining characteristic of content analysis is "systematic". However, there is no standard process; while some aspects are objective, others are based on insight and intuition. Furthermore, there are no specific evaluation criteria for quality. It, therefore, behoves the researcher to ensure outcomes (e.g., codes, categories, and inferences thereof) are accurate, meaningful, and useful (Cho & Lee, 2014; Elo et al., 2014). This has spawned several models and guidelines (e.g., Bengtsson, 2016; Cho & Lee, 2014; Mayring, 2014; Zhang & Wildemuth, 2016).

An effort was made to ensure coding was transparent and the outcomes accurately represented the data corpus and, therefore, participants' viewpoint. An upfront explication of the data

analysis strategy can improve transparency and credibility, and also help the research audience understand the process (Mayring, 2014). The analysis strategy was integrated into the overall research strategy (see Table 6.3).

6.4.2 Analysis strategy

Figure 6.2 shows the analysis strategy. It may be reiterated that data collection and analysis occurred concurrently; furthermore, the steps are iterative— presented as linear only for illustration.

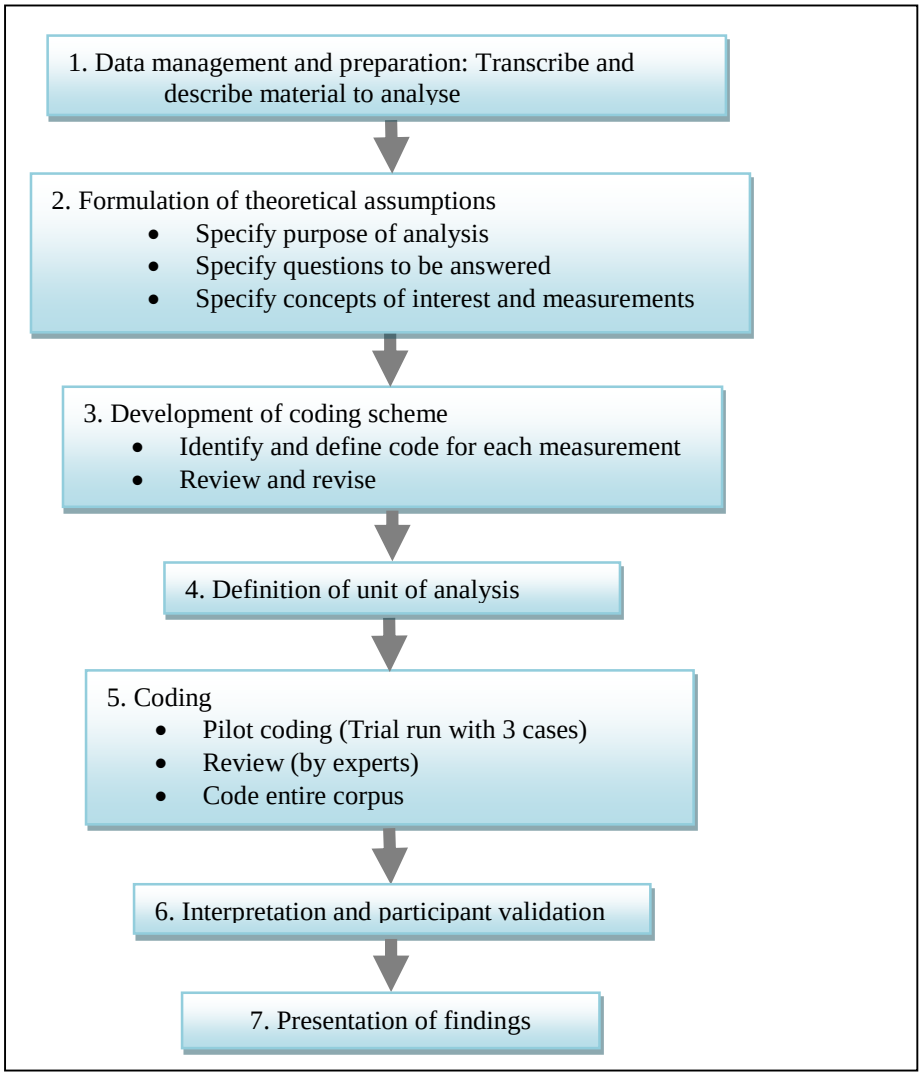


Figure 6.2: Data analysis strategy adapted from (Cho & Lee, 2014; Mayring, 2014; Rose et al., 2015)

6.4.2.1 Data management and preparation

This involved transcription, organization, and familiarization with data (Bazeley, 2013; Rowley, 2012). Soon after an interview, the audio recording for a participant was transcribed and saved in a folder, together with any other relevant artefacts, as a Word document with

password protection. To ensure anonymity, transcripts were stripped of any information which may divulge participants' identity and assigned the corresponding pseudonyms of interview participants. These pseudonyms were used for all subsequent references to transcripts, codes, results, etc. (Creswell & Poth, 2017).

There is an argument against verbatim transcription; i.e., it is time-consuming and arduous and, therefore, may not be worth the effort. Thus, selective transcription, in which salient portions of the interview are transcribed, is preferable (Mayring, 2014; Saldaña, 2015; Zhang & Wildemuth, 2016). This was adopted; each transcript was read and further cleaned of unnecessary detail. The 12 "clean" transcripts constituted the data corpus for content analysis. Transcription provided an opportunity to familiarize with the data to understand, reflect and improve on subsequent data collection and analysis.

6.4.2.2 Formulation of theoretical assumptions

Identifying and specifying the purpose helped to focus analysis. As aforementioned, a descriptive, manifest, and deductive analysis (which combined both quantitative and qualitative) supported the research purpose.

Mayring (2014) observed that content analysis is subjective and that *a priori* theory can improve trustworthiness. Extant theory and previous research may form the conceptual basis (Rose et al., 2015; Schreier, 2012). The conceptual model, preliminary results of Phase 1, and interview protocol provided assumptions about what to look for, guided decisions on data collection and analysis, and improved transparency. The analysis was organized around the central themes of the interview protocol, as shown in Table 6.5.

6.4.2.3 Development of coding scheme

Practical illustrations (e.g., Noble & Smith, 2015; Rose et al., 2015; Saldaña, 2015; Schreier, 2012) proved useful. An initial list of codes was drawn from the interview protocol and extant research, as suggested in the literature (e.g., Miles et al., 2014; Schreier, 2012). Description of the codes, and constituent sub-codes, where applicable, were based on the codebook of Phase 1. The coding scheme was reviewed by an academic and subsequently refined. Appendix D2 shows the final version.

6.4.2.4 Definition of unit of analysis

A transcript (i.e., set of responses from a participant) constituted a coding unit. Each was divided by interview questions, which constituted the units of analysis. These were assigned

codes. This way, a code represented a variable and measured the responses to a question by a participant. This allowed comparisons of responses to each question across participants.

6.4.2.5 Coding

As stated above, Phase 2 was a small-scale research (i.e., only 12 participants). Therefore, manual analysis was preferred to CAQDAS (Saldaña, 2015). The transcripts, with Microsoft Word features (e.g., bold, italic, etc.) for emphasis, were analyzed. Coding was per unit and across participants (i.e., by going through the response to a specific question, one participant after another). The coding scheme was piloted, i.e., applied to three transcripts; the results were reviewed by an academic and found to be consistent. The entire corpus was subsequently coded. The responses for each participant were captured into an Excel sheet, the rows represented participants (i.e., cases) while columns were their respective responses to the question.

Initial coding was descriptive; the second cycle identified patterns (e.g., similarities, differences, and frequencies) according to the demography of participants (i.e., background, position, etc.). Direct quotes which exemplified patterns were also marked. Consistent with the purpose of research and analysis (i.e., descriptive, manifest), an effort was made to represent participants' views "as is". There was little or no inference beyond the text itself.

6.4.2.6 Interpretation and participant validation

The ultimate objective of analysis is to draw conclusions, i.e., make sense of data. This involves further analysis of codes identified (Saldaña, 2015; Zhang & Wildemuth, 2016). Interpretation can improve trustworthiness, e.g., a well-constructed, meaningful set of conclusions is likely to improve credibility and reliability; help others make their own conclusions; increase practical usefulness in relevant contexts (Zhang & Wildemuth, 2016). The qualitative research strategy (see Table 6.3) showed effort was made to ensure conclusions represented participants' views.

The two-step approach, quantitative followed by qualitative (Schreier, 2012), was adopted. In the quantitative, sub-codes were analyzed across participants for descriptive characteristics (e.g., frequency of occurrences) and direction and magnitude of meaning, e.g., strongly disagree, strongly agree, etc. This was followed by analysis (e.g., comparisons) across groups of participants. The qualitative analysis examined participants in each group to identify patterns and draw profiles; this was supported with direct quotations extracted from participants' responses.

Various devices, e.g., rich text features (italic, bold, etc.) and visual displays, e.g., tables, were used to illustrate the structure and hierarchy of categories (Elo et al., 2014; Saldaña, 2015). Strategies, e.g., pilot analysis and review, direct quotations from the empirical data, participant validation of result, etc., (Cho & Lee, 2014) were also adopted to avoid common pitfalls of interpretation (e.g., omissions, misrepresentation, and exaggeration) and to ensure results represented participants' perspectives.

6.4.2.7 Presentation of results

The final research report may serve two primary objectives (Elo et al., 2014), *viz*., to describe the analysis process (i.e., how the results were generated) and help the research audience make an informed decision about the extent to which the outcomes (e.g., codes and conclusions) are valid, accurate and objective and can be applied in different contexts. Thus, the coding process and strategies, outcomes validation techniques, etc., may be reported (Elo et al., 2014; Zhang & Wildemuth, 2016).

The language, tools, etc., employed must be appropriate for the report to be meaningful and useful to its audience. A textual report is generally voluminous and cumbersome to present coherently (Walliman, 2018) while visual is more effective and preferable (Linneberg & Korsgaard, 2019; Verdinelli & Scagnoli, 2013).

Tables augmented by quotes from participants are commonly used and, therefore, were adopted. Conclusions are generally presented as factors, models, propositions, etc., dictated by factors such as the research purpose and audience, etc. The conclusions were drawn as propositions. Validation is discussed in Chapter 7, in conjunction with that of Phase 1. The following sections present the results of the analysis.

6.5 Data analysis and preliminary results

Tables 6.7 to 6.14 summarize participants' responses, coded per interview question. The column headings correspond with the initial coding framework for analysis. Consistent with content analysis, as described in section 6.4.1, both quantitative and qualitative methods were employed to provide an overview of responses (i.e., frequencies) and also understanding (e.g., to establish relationships within and across groups) which may not be observable. The following proposition was drawn from RQ2 to organize the analysis and presentation of results.

> "*There are no significant differences in participants' perception of how e-government outcomes evaluation is done*".

6.5.1 Characteristics of participants

Section 6.2 discussed the sample, i.e., 12 participants: four (two each from management and non-management positions) were drawn from IT, Evaluation, and Other backgrounds. The questions concerned participants' professional background, position, and experience (i.e., years of experience and self-rated knowledge) in e-government outcomes evaluation. These helped to establish rapport with interviewees, as discussed in section 6.3.2. Where appropriate, the values were used to reference quotations extracted from participants, e.g., P1 (IT, Management, <5, Above average) describes participant P1.

Table 6.7 shows about half (7 out of 12) had over five years of experience and most (11 out of 12) rated their knowledge as average or above average. As explained above, the qualitative results were to help understand the quantitative. Thus, the analysis below is standalone, to be subsequently integrated with the quantitative results— to draw and situate conclusions in the literature (see Chapter 7).

Table 6.7: Characteristics of participants

Participant	*Professional Background*	*Positon*	*Experience (Years)*	*Knowledge (Rating)*
P1	IT	Management	>5	Average
P2	IT	Management	>5	Above average
P3	Evaluation	Management	>5	Above average
P4	Evaluation	Management	>5	Above average
P5	Other	Management	<=5	Average
P6	Other	Management	>5	Above average
P7	IT	Non-management	<=5	Average
P8	IT	Non-management	>5	Above average
P9	Evaluation	Non-management	<=5	Average
P10	Evaluation	Non-management	<=5	Above average
P11	Other	Non-management	<=5	Minimum
P12	Other	Non-management	>5	Average

6.5.2 Organizational capacity

This concerned expertise, budget, and management support (i.e., in resource allocation and actual performance). Table 6.8 summarizes the responses. Half (six out of 12) of the participants indicated adequate expertise. Drawing from Table 6.7, it seemed professional background and position were important factors. All from Evaluation background and four from management reported adequacy (c.f. Other and non-management). However, further analysis showed most (eight out of 12, across all background, position, experience and knowledge) drew attention to reliance on external expertise and the need to build internal capacity. For example, P4 (Evaluation; Management; >5; Above average) referred to "*the lack of evaluation skills internally"*. P2 (IT; Management; >5; Above average) indicated a need to

"*build our own capacity because teams consist mostly of consultants*". P12 (Other; Non-management; >5; Average) noted: "*we need people with experience […] because we don't have at the moment*".

Many (i.e., eight of 12) participants indicated adequate budget. As some of the responses show, there may not be dedicated budget; however, the public sector procurement processes applied where needed. For example, according to P3 (Evaluation; Management; >5; Above average): "*Budget is taken for granted because we assume we can secure funds through the public sector procurement process when the decision to evaluate is taken*". Participants P7, P9, P11, and P12 (all in non-management) were not aware if there was adequate budget. For example, P9 (Evaluation, Non-management, <=5, Average) stated: "*I don't have firsthand knowledge of this so I cannot say anything about budget with certainty*". Resource allocation is a management function, which may explain why non-managers were unaware. The lack of dedicated budget conflicts with government policy, as stipulated in DPME (2011).

On management support, five participants, four (i.e., P2, P3, P4, P6; Management, >5, Above average), indicated adequate support (i.e., in resource allocation and evaluation process), enabled by management involvement. For example, according to P6, management support "*removes much of the resources and logistics hurdles... managers on the team facilitate interaction between the team and the organization …that helps to resolve challenges the team may face quickly*." To many (non-management, <5, Average), support was limited to the supply of resources and did not extend to the actual "doing" of the evaluation. As P9 put it: "*It is more of here is the resources you need, make the best of it. It may be because usually, the team is from outside*". Only P12 (IT; Non-management; <=5; Average) disagreed with the adequacy of management support and stated: "*There is no expertise or budget… proof of lack of support from management*". The differences between management and non-management may be explained by differences in interests, expectations, etc., among stakeholder groups, as the literature review on evaluation and practice suggested.

It seems there is inadequate internal organizational capacity to conduct and use evaluation; development of internal expertise, planning, and budgeting in advance, and management support during the evaluation need attention. Also, there are differences among the participants.

Table 6.8: Organizational capacity

Participant	*Expertise*	*Budget*	*Management Support*
P1	Adequate; improved in the last few years	Adequate; not dedicated	Inadequate; for resources not process
P2	Adequate; need to build internal capacity	Adequate; not dedicated; procurement process available	Adequate; management involvement; benefit
P3	Adequate; appropriate experience	Adequate; not dedicated; procurement process available	Adequate; management involvement; benefit
P4	Adequate; external expertise	Adequate; not dedicated	Adequate; management involvement; benefit
P5	Inadequate; a challenge	Adequate; not dedicated; procurement process available	Inadequate; resources and process
P6	Inadequate; mainly consultants	Adequate; not dedicated; procurement process available	Adequate; management involvement; benefit
P7	Inadequate; need for internal expertise	Not aware	Inadequate; resources not process
P8	Inadequate; external expertise	Adequate; not dedicated	Inadequate; resources not process
P9	Adequate; external expertise	Not aware	Inadequate; resources not process
P10	Adequate	Adequate; not dedicated; process available	Adequate; management involvement; benefit
P11	Inadequate; external expertise	Not aware	Inadequate; resources not process
P12	Inadequate	Not aware	Inadequate; resources and process

6.5.3 Policies and values

Table 6.9 summarizes the responses. It seems policies and values are important factors to consider; various explanations were proffered. Only P11 (Other, Non-management, <5; Minimum) was not aware of the direct influence of Batho Pele principles. The other participants viewed the principles as a framework with implications for how to assess public services— such as e-government. For example, Batho Pele principles: "*direct service delivery and also how service can be assessed from the perspective of stakeholders …. and improved for all those that access it*" (P1: IT, Management, >5, Average); "… *are the values that guide service delivery… this means ICT performance must also be measured according to the principles...* (P7: IT, Non-management, <=5, Average).

All participants were aware of the influence of NEPF (National Evaluation Policy Framework) and generally described it as a framework. For example, as P3 pointed out, it is the "*framework within which we operate… it prescribes outcomes evaluation and also describes what to*

evaluate, evaluation purpose, activities, participants and their roles, and so on. It talks to what to do and how".

The responses showed that efficiency (e.g., in resource allocation, operation, management, reduced cost, etc.); effectiveness (effect on participants, improved services, satisfaction with IT services, etc.) and democratic values (e.g., public participation) as discussed in the literature (see section 2.6.3) influenced how evaluation is done. For example, according to P3 (Evaluation, Management, >5, Above average):

> "*Batho Pele and other policies clearly highlight efficient public service without waste. That is why it is important to know ICT contribution. Also, we need to know how people think of the services we deliver so we need to assess the service delivery and any impacts from their perspective.*"

As may be observed, efficiency (i.e., organizational performance) values predominated (11 occurrences, compared to effectiveness (i.e., impact): six and democratic (i.e., public participation: three). Furthermore, only those in Evaluation (except one), referred to two or more values. This may be explained by their background and, therefore, a better understanding of outcomes evaluation.

Participants' description of Batho Pele and NEPF above is consistent with the definition of a framework as policy, which prescribes how something is done (section 2.4.2). It may be concluded that Batho Pele and NEPF frameworks influenced how e-government outcomes evaluation is done; organizational performance values are predominant. In general, there are no significant differences among participants, according to their characteristics.

Table 6.9: Policies and values (Batho Pele and NEPF)

Participant	*Batho Pele prescribes…*	*Nepf prescribes…*	*Benefits*
P1	How to assess; stakeholder participation; service improvement	How assess; stakeholder participation	Efficient use of resources
P2	How to assess; stakeholder participation	How to assess, stakeholders	Efficient operation; public participation
P3	Service improvement; public participation	Outcomes evaluation; how to access, purpose, stakeholders, roles, etc.	Efficient public service; improved service; public participation; public satisfaction
P4	Service improvement; values to assess; public participation	Outcomes evaluation, stakeholders, values to assess, etc.	Reduced cost of service delivery; public participation; public satisfaction
P5	How to assess	Purpose, activities, stakeholders	Efficient public sector operation
P6	Public participation; customer satisfaction	How to access (dos and don'ts), stakeholders,	Public satisfaction
P7	How to assess	How to assess	Reduced cost of service delivery
P8	Service improvement; stakeholder participation	Activities; stakeholder participation	Efficient use of resources
P9	How to assess; service improvement	How to assess; evaluation purpose, activities	Operational and managerial efficiency; public satisfaction
P10	How to assess; public involvement	How to assess; purpose, activates, stakeholders	Efficient service delivery; impact on the public
P11	Not aware of a direct link	How to assess	Efficient operation
P12	How to assess	How to assess	Efficient service delivery

6.5.4 Participation

Table 6.10 summarizes the responses on the extent of involvement in decision making and activities, and disposition to participating in future evaluation. All participants were involved to some degree in decision making— eight: throughout; and three: somewhat. A link may be established between the participants' characteristics and the degree of participation. Seven of the eight that participated throughout were from IT and Evaluation backgrounds; six had over five years' experience and average or above-average knowledge. Furthermore, seven participants linked the extent of participation to their positions, roles, or both. For example, P4 (Evaluation, Management, >5, Above average) referred to "*as a manager [and] my expertise in evaluation*" to explain participation throughout; and P8 (IT, Non-management, >5, Above average): "*Part of my job description is to review ICT implementation*". On the other hand, those that participated somewhat were mainly (three out of four) from Other and deferred to experts, due to lack of expertise. For example, according to P5 (Other, Management, <=5, Average), "*those of us who are not experts in the field rely on the opinion of experts*".

On activities, all participated to some degree. However, only five throughout the process: four from Evaluation, with average or above average knowledge. Thus, professional background

(and the expertise and role it affords) is important. Some responses underscored this; e.g., according to P4 (Evaluation, Manager, >5, Above average): "*I am obliged to participate in all activities [because] of my position and as lead evaluator*". On the other hand, almost all those that participated somewhat (which included all four from Other) indicated a lack of expertise; e.g., P11: "*not very much involved in the technical side of things because I am not an expert*".

Only two out of 12 were not willing to participate in future evaluation. These were from Other; again, this may be explained by a lack of expertise. According to P11 (Other, Non-management, <=5, Minimum), "*It requires technical skills to participate fully and I don't have*". For those willing to participate, the primary motivation was learning (from participation) and their role or position. As P7 (IT, Non-management, <=5, Average) pointed out, "*it is part of my duty but more importantly, it is also for my own career development.*".

It seems expertise and role (afforded by professional background and or position) influence the extent of participation in decision-making and activities. There are significant differences: participants from Evaluation are likely to participate throughout both. Furthermore, learning and role (afforded by background or position) may be the motivation for future participation.

Table 6.10: Participation

Participant	*Decisions*	*Activities*	*Future participation*
P1	Throughout; position	Somewhat; no expertise	Yes; for experience
P2	Throughout; position;	Somewhat; other commitment	Yes; better understanding
P3	Throughout; position/role	Throughout, position/role	Yes; role; learning
P4	Throughout; position/role	Throughout; position/role	Yes; role
P5	Somewhat; reliance on expert advice	Somewhat; lacks expertise	Yes; experience; better understanding
P6	Throughout	Somewhat; lacks expertise	Yes; role (unit responsible evaluation)
P7	Somewhat: lacks expertise	Somewhat; lacks expertise	Yes; role, career development
P8	Throughout; role	Throughout; role	Yes; for experience
P9	Throughout; role	Throughout; role	Yes; for experience
P10	Throughout; role	Throughout; role	Yes; role
P11	Somewhat; up to management and experts	Somewhat; lacks expertise	No; lack of expertise
P12	Somewhat; up to experts	Somewhat; lacks expertise	No; lack of expertise

6.5.5 Roles and responsibilities

This examined the link between roles and how practitioners do their work. Table 6.11 summarizes the responses. Most participants (nine out of 12, and including all the eight from IT and Evaluation backgrounds) understood their roles and responsibilities and attributed them to experience and preparation (i.e., workshop, orientation, daily briefs, etc.). Some of the

narratives include: "*because of my expertise … previous experience in e-government evaluation and also being part of the structure which coordinates the evaluation*" (P3: Evaluation, Management, >5, Above average) and "*My role and responsibilities were clear to me…. workshop at the beginning helped to define the tasks I performed in the team*" (P2: IT, Management, >5, Above average). Three (P5, P11, and P12) out of the four from Other found it a challenge and attributed this to lack of expertise. According to P11: "*I'll not say I understand although the daily briefs go a long way … in my view training can ensure meaningful skills development*".

Eight out of 12 (all from IT and Evaluation) indicated others understood their roles and responsibilities. For example, according to P4 (Evaluation, Management, >5, Above average): "*The criteria for selecting participants is knowledge of evaluation and the context […] We also have strategies* (*like workshops before we start the evaluation) to make sure that team members know what to do*". Those that found it challenging (all from Other) reported a lack of expertise. For example, P11 (Other, Non-management, <5, Minimum): "*Not everyone is clear… some of us do not have the experience and cannot contribute*".

It seems professional background (and the expertise it brings to the roles) and preparation are important factors, which enable understanding of roles and responsibilities: participants from IT and Evaluation are likely to understand their roles, compared to Other.

Table 6.11: Understanding of roles

Participant	*Participant's understanding*	*Other participants' understanding*
P1	Yes (expertise; preparation)	Yes (expertise; preparation)
P2	Yes (expertise; preparation)	Yes (expertise; preparation)
P3	Yes (expertise; part of evaluation structure)	Yes (preparation
P4	Yes (expertise; preparation)	Yes (expertise; preparation)
P5	No (challenging)	No (challenging)
P6	Yes (preparation)	Yes (preparation)
P7	Yes (expertise)	Yes (expertise)
P8	Yes (preparation)	Yes (expertise)
P9	Yes (preparation)	Yes (expertise)
P10	Yes (expertise; preparation)	Yes (expertise)
P11	No (challenging)	Yes (challenging
P12	No (challenging)	No (challenging)

6.5.6 Use of results

This looked at the following about reports: management response; dissemination to stakeholders; and implementation of recommendations. Table 6.12 summarizes the responses. Most (i.e., nine out of 12) participants indicated management responded to the evaluation report which set out improvement plans and proceeded to describe the processes. It seems there is no

difference according to participants' background. However, most in management positions linked the process to public sector policies (NEPF, Batho Pele, etc.). According to P6 (Other, >5, Above average), "*Management studies the report and draws up their response and plans to implement the recommendations as set out in the policies like the national evaluation framework*".

It may also be observed that, in general, there are processes to distribute reports to the various stakeholder groups. Also, there seem to be no differences according to participants' characteristics. According to P1, the distribution of reports complies with policy and "… *we distribute to the relevant stakeholder groups*". However, two participants P7 and P11 have never been involved in and, therefore, are not aware of such processes; again, P12 stated none existed.

In general, recommendations may not be implemented; evaluation is mainly viewed as a formality, not as a strategic tool. According to P7: "…*it appears evaluations are for the sake of evaluation. It is something to look at going forward*" and P12: "…*there is no strategic use for improvement".* Two participants were not aware if this was done or not. However, according to most managers, recommendations may be implemented— informed by management response. According to P3, management response "*becomes a contract of improvement*"; P4: "*there is no mechanism to find out whether the improvement is implemented*".

It may be concluded that management responds to reports and there are processes to distribute reports to stakeholder groups. However, the recommendations may not be implemented. Also, on management response and implementation, there seem to be differences between management and non-management.

Table 6.12: Use of reports

Participant	*Management response*	*Distribution*	*Implementation of recommendations*
P1	Process available; linked to policies	To all stakeholders	Implemented; not aware if assessed
P2	Process available; linked to policies	Process available; linked to policies; to all stakeholders	Implemented; based management response
P3	Process available; linked to policies	Process available; linked to policies; to stakeholders; different versions	Implemented; based on management response
P4	Process available; linked to policies	Process available; to all stakeholders; different versions	Implemented; not assessed
P5	Process available; linked to policies	Process available; to all stakeholders	Not implemented; opportunity to improve lost
P6	Process available	Process available; linked to policy; to all stakeholders	Not implemented; evaluation is to tick boxes
P7	Process available	Not aware; no process	Not implemented; evaluation for its sake
P8	Not aware of a process	Process available; to all stakeholders; different versions	Not implemented; no accountability for this
P9	Process available	Process available; to all stakeholders	Hardly implemented; formality
P10	Process available	Process available; to all stakeholders; different versions	Not aware
P11	Not aware of a process	Not aware	Not aware;
P12	No process	Not distributed; no strategy	Not implemented

6.5.7 Effect of participation

This measured changes in participants' perceptions, and their organizations', about e-government outcomes evaluation. Table 6.13 summarizes the responses. In general, the participants reported changes, i.e., improved awareness of the benefits of evaluation; understanding of processes; and experience. These are forms of learning— the effect of doing evaluation, using findings, or both (Alkin & King, 2016). Most participants indicated awareness, e.g., P1 (IT, Management, >5, Average) learned that outcomes evaluation can "*improve the level of service delivery*" and P9 (Evaluation, Non-Management, <5, Average) "…*about the e-government evaluation process. I've gained knowledge about my career*".

Also, many participants noted changes in perceptions (i.e., increased awareness of benefit) in their organizations. P7 (IT, Non-Management, <5, Average) stated there is a "*view that e-government evaluation can benefit service delivery so effort and resources are being invested in that [the evaluation]"*. Three participants from Other background opined not much has changed because there is a lack of awareness. According to P6, e-government outcomes evaluation has "*not* [become] *part of the organizational fabric"* and P12: "*many people are not aware* [*also*] *evaluations are not used for strategic decisions so there is no tangible effect on the organization*".

It seems the effect on participants and their organizations was conceptual (i.e., improved knowledge) not improvement of e-government outcomes evaluation *per se*. There may be no significant differences among groups. As shown in section 6.5.4, the motivation for future participation may be learning (i.e., conceptual).

Table 6.13: Effect of participation

Participant	*Effect on participant*	*Effect on organization*
P1	Aware of benefits;	Aware of benefits
P2	Aware of benefits; gained experience	Aware of benefits
P3	Understand the process and challenges	Improved ICT services
P4	Understand the process; gained experience	Experience (from conducting evaluation)
P5	Aware of benefits and challenges	Awareness of benefits
P6	Aware of benefits	No change (not part of organization fabric)
P7	Aware of benefits	Aware of benefits
P8	Understand the process	Aware of benefits
P9	Understand the process; gained experience	Improved ICT services
P10	Aware of benefits; gained experience	Aware of benefits
P11	Aware of benefits	No change (no awareness)
P12	Aware of benefits	No change (no awareness; no tangible effect)

Section 6.4 showed it was appropriate to present the results as propositions. Table 6.14 summarizes the preliminary results as general statements. This summary was reviewed by participants P3 (Evaluation, Management, >5, Above average) and P8 (IT, Non-Management, >5, Above average). As emphasized in the overall research strategy (see section 4.8), this phase was to help understand the quantitative. Thus, the two were integrated and discussed in conjunction, and overall conclusions (i.e., meta-inference) were drawn therefrom.

Table 6.14: Summary of preliminary results

Aspect of EOEP	*Preliminary finding (proposition)*
Organizational capacity	Inadequate internal organizational capacity to conduct and use evaluation: • Lack of internal expertise, dedicated budget; and management support throughout the evaluation process • Need to develop internal expertise, plan and budget in advance, and management support the evaluation need attention • There are differences in perceptions among groups by position and background
Policies and values	Policies (i.e., Batho Pele and NEPF) influence EOEP • Organizational performance as a predominant value • No significant differences among groups
Participation	Expertise and role (from professional background and position) influence the extent of participation • Differences by background (Evaluation more likely to participate throughout both decision-making and activities) • Learning and role as motivation for future participation
Roles and responsibilities	Professional background (and the expertise it brings to the role) and preparation are important factors • Differences among groups (IT and Evaluation are likely to understand)
Use of report	Some degree of use • Management response; distribution of reports • Recommendations may not be implemented • Differences by position (on management response and implementation of recommendations)
Effect	On both participants and their organization is learning (i.e., conceptual) • Improved awareness, understanding and experience • No significant differences among groups Learning is the motivation for participation

Summary

The primary objective of this chapter was to understand the preliminary results of the quantitative phase, i.e., the elements of e-government outcomes evaluation practice and the patterns of performance. A subsample of 12 public employees, drawn from the sample of Phase 1, were interviewed. Content analysis was conducted to explore participants' perceptions of how e-government outcomes evaluation is done. Conclusions (summarized in Table 6.14) were drawn as propositions (i.e., general statements).

Notwithstanding the effort to improve quality, this phase had limitations— discussed in Chapter 7 in the context of the entire research. Suffice to state that strategies were implemented to assure trustworthiness. The sample and selection techniques were appropriate; the protocol was supported by the results of the previous phase, reviewed by experts, piloted, and revised; the analytic method (i.e., content analysis) suited the research purpose. The results (i.e., propositions) were derived from the data “as is” to represent participants' views.

As explicated in the SEMM research strategy in section 4.8, the results of the respective phases were integrated to draw and discuss the overall conclusions and it is to this, which the next chapter turns.

Chapter 7 : Discussions and conclusion

The overall objective of empirical research is to produce evidence to help solve a problem. This research set out to explore how e-government outcomes evaluation is done in developing countries. The preceding chapters have explicated the problem and relevance; reviewed literature to understand and situate the research in extant knowledge; developed assumptions and specified an overall plan; described the implementation thereof; and presented the results.

The next step is to establish the extent to which the results contributed to solving the problem. The main objective of this chapter is, therefore, to integrate results of the respective phases to draw overall conclusions; discuss correspondence to theory, the managerial implications, weaknesses, and contribution; and advance an agenda for future research on EOEP. Figure 7.1 illustrates the structure of this chapter.

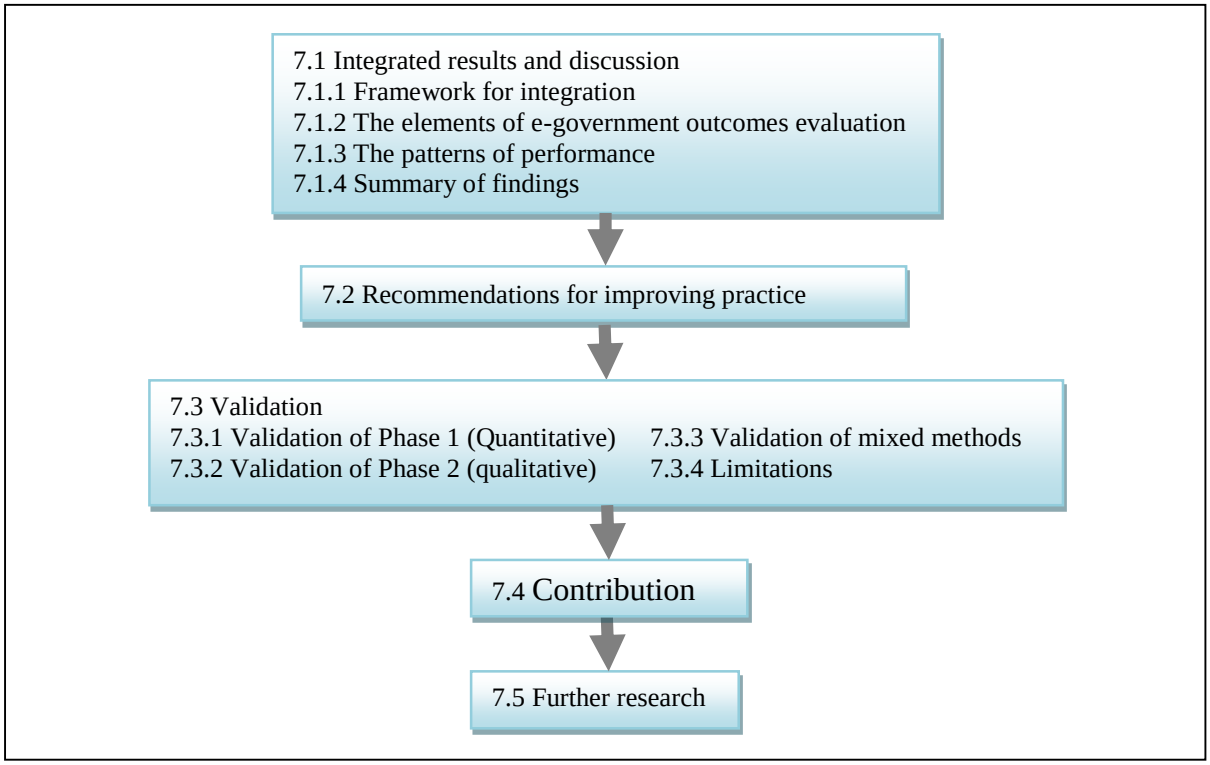


Figure 7.1: Structure of chapter

7.1 Integrated results and discussion

The conceptual model (see Figure 3.4) re-framed how e-government outcomes evaluation is done as a bundle of practice-material arrangement in the public sector, *viz*., an interaction among entities (people and other things, e.g., resources), activities, and connections (e.g.,

practical knowledge, shared assumptions and values; and expected outcomes). Doing evaluation (i.e., assessment) and the effect (of participation and use of results) were represented as *Assessment* and *Use* bundles respectively. Context, resources, activities, assumptions, and outcomes were examined.

As with mixed methods research, the SEMM research strategy (section 4.8) implied the results of the quantitative and qualitative phases were reported separately (see Chapters 5 and 6 respectively). These results were integrated at this stage to paint a complete picture and draw conclusions.

7.1.1 Framework for integration

Joint display devices (e.g., tables and figures) are effective for integration and further analysis, e.g., to bring together to compare, contrast, and identify patterns in data from various phases, which otherwise may not be evident (Fetters & Freshwater, 2015). Berman (2017) and Venkatesh et al. (2016) provide illustrations.

The purpose of this research was descriptive. The following research questions (specified in section 1.1) informed the integration and analysis:

RQ1: What are the factors which describe how evaluation of e-government outcomes is done?

RQ2: What are the patterns of performance among the different groups? Put differently, "*How do different groups perform e-government outcomes evaluation*"?

RQ3: How can e-government evaluation be improved?

Phase 1 (quantitative research) described participants' characteristics and responses; the main factors which characterized how e-government outcomes evaluation is done (i.e., the elements of practice as object) and the different ways in which these factors combine during performance (i.e. patterns of practice as performance). Phase 2 (qualitative research) provided insight into the elements of practice and the performance (i.e., what is done and how).

Thus, consistent with practice research, the objective of this analysis was to "identify the components of practice and their configuration" (Welch, 2017, p. 7), i.e., elements and patterns of performance. Table 7.1 is the framework of analysis and summarizes the results of the respective phases.

Table 7.1: Integration framework for analysis: combines results of the quantitative and qualitative phases (Tables 5.29 and 6.14)

Preliminary quantitative result (N=106)			*Preliminary qualitative results (N=12)*
Descriptive	*Exploratory factor analysis and comparisons*	*Cluster analysis*	*Content analysis*
Purpose of evaluation: compliance, decision-making, not learning about e-government or the evaluation process Evaluation initiated from within Evaluation teams consist of both internal and external stakeholders Main strategies: survey Main outcomes measured: organizational performance and accountability Generally negative perceptions of EOEP	Differences across all factors, according to position, background, experience and knowledge, except Affective issues. Positive perception of practice among Management, Evaluation, Over five years' experience, and Above average knowledge In general, no difference between IT and Evaluation against Other; and also Above average against Average and Minimum experience	**Cluster 3**: IT and Evaluation managers with extensive experience (years and knowledge); **Cluster 2**: IT and Evaluation non-managers with moderate experience; **Cluster 1**: Other non-managers with little experience **Implications**: clusters represent differences in participants' perceptions of how e-government outcomes evaluation is done according to their characteristics Positive perception of practice (highest mean) in Cluster 3 The cluster solution seems to support results from the comparisons of groups across factors	
	Organizational capacity: neither agreed or disagreed (on "Outcomes are regularly evaluated"; "Evaluation budget is adequate" and "Management supports evaluation" "Evaluation tools (e.g., software) are adequate")		Inadequate internal organizational capacity to conduct and use evaluation: • Lack of internal expertise, dedicated budget; and management support throughout the evaluation process • There are differences in perceptions among groups by position and background
	Evaluation framework: neither agreed or disagreed (on "The indicators measure both financial and non-financial outcomes", "Evaluation models (*utilization, empowerment,*		Policies (i.e., Batho Pele and NEPF) are important

etc.)" and "Batho Pele principles guide how evaluation is done")	• Organizational performance as a predominant value • No significant differences among groups
Shared understanding: neither agreed or disagreed (on "Stakeholders agree on what to evaluate" and "Stakeholders agree on how to use results")	Roles and responsibilities: background (and, expertise and it brings to the role) and preparation are important factors • Differences among groups (IT and Evaluation are likely to understand)
Stakeholder participation: neither agreed or disagreed (on "Stakeholders participate in decision making during evaluation" and "Selection of stakeholders is transparent")	Participation: background (and expertise therefrom) influences the extent of participation • Differences by background (Evaluation more likely to participate throughout both) Learning and role as motivation for future participation • No differences among groups
Use of results: neither agreed or disagreed (on "Management produces strategy on how to implement recommendations"; "There is follow up to implement recommendations"; "How evaluation is done is reviewed regularly")	Use of results: Some degree of use • Management response; distribution of reports • Recommendations may not be implemented • Differences by position and management (on management response and implementation)
Affective issues: neither agreed or disagreed (on "Participation increases stakeholders' willingness to participate in future evaluations" and "Participation increases stakeholders' commitment to implement recommendation")	Effect: Learning (i.e., conceptual) on both participants and their organization • Improved awareness, understanding and experience) • No differences among groups Learning is the motivation for future participation

7.1.2 The elements of e-government outcomes evaluation

The descriptive analysis showed the main purpose of evaluation was compliance and decision-making, survey was widely used to measure accountability and organizational performance outcomes, and evaluation was initiated by external entities (i.e., external demand for evaluation). The qualitative phase seemed to support this; i.e., organizational performance outcomes were dominant and recommendations were not implemented for improvement. The literature shows evaluation in the development context is for compliance (Chouinard & Cousins, 2015), dominated by objective measurements, e.g., survey (section 2.4.6). Song and Letch (2014) found evidence of symbolic IT evaluation use (i.e., to comply with demand). In e-government, positivism is predominant (Hébert, 2013) and organizational performance, compared to democratic values, predominates (Sterrenberg, 2017).

Factor analysis identified inadequate organizational capacity; the qualitative results seemed to suggest inadequate internal organizational capacity. Evaluation in developing countries is characterized by a lack of resources (section 2.4.6). The differences in perceptions (in both quantitative and qualitative analyses) were expected, e.g., managers (c.f. non-managers) were likely to know about resourcing, a strategic function. Also, those from Evaluation (c.f. Other) were likely to know about internal expertise. Fierro and Christie (2017) found differences in evaluation capacity among stakeholder groups.

The descriptive analysis showed the main outcomes measured were organizational performance. Factor analysis identified Batho Pele principles and measurement of financial and non-financial outcomes were important in **evaluation framework**. The qualitative results showed policies (i.e., Batho Pele and NEPF) and organizational performance outcomes were important. **Policies** define roles and responsibilities, outcomes, acceptable actions, etc.—consistent with the tenets of public value and public service values, evaluation framework, and practice theory (2.2.5, 2.4.2.6, and 3.2.4). The qualitative results showed no significant differences. This seems plausible: policies in the public sector are prescriptive and punitive and are likely to elicit shared practice.

Factor analysis showed **understanding** (of what to evaluate and how to use results) was important. The qualitative results suggest background and preparation enabled understanding of **roles and responsibilities**. Both quantitative and qualitative results established differences in perceptions. These are supported by literature: understanding involves knowledge about what to do and how and is facilitated by learning (Bittner & Leimeister, 2014). Section 3.2.4

showed roles and positions in organizations prescribe "a way of being" (Schatzki, 2017b, p. 36) and bear on knowledge, actions, etc. Evaluation frameworks prescribe functions to assure appropriate actions and outcomes (Arbour, 2020). Understanding, professional background, and preparation (e.g., training before evaluation commenced) are linked. For example, professional background may determine positions and roles, and responsibilities; preparation can promote, understanding, expertise, thereby, professional development.

Factor analysis identified **participation** in decision-making and stakeholder selection as important. The qualitative results showed participants were selected by professional background; the expertise this (i.e., background) brings to a role determined the extent of participation in decision-making and activities. The differences were expected: participants from evaluation background were likely to possess the necessary expertise and participate throughout. Chouinard and Cousins (2015) observed expertise determined the participants to select and the extent of participation. According to Fielding (2013) other than evaluators, stakeholders hardly participate beyond specifying the scope of an evaluation, due to lack of expertise. Learning and participants' role seemed to influence future participation.

On **use of results**, factor analysis showed management response and follow-up to implement recommendations were important. The qualitative results suggested reports were considered by management and distributed to stakeholders, but recommendations were not implemented. This supported the symbolic use discussed above. Consideration of reports is a form of learning (Alkin & King, 2017). However, the non-implementation of recommendations may suggest learning is not for improvement. This may be supported by the descriptive analysis: the purpose of evaluation was neither learning to improve e-government nor the evaluation process itself. Inadequate internal organizational capacity may explain why recommendations may not be implemented (Gagnon et al., 2018). Decisions, e.g., to distribute reports, implement recommendations, etc., are management responsibility (Arbour, 2020). This may explain the differences between managers and non-managers.

The link between purpose and use was discussed above (e.g., section 2.5.8). The qualitative results showed the **effect** of evaluation on participants and their organizations was learning to improve knowledge (i.e., conceptual use); this also suggests learning was not for improvement. Mayne (2014) found this (i.e., conceptual use, not improvement) was generally the case.

The quantitative results showed there were no differences on **Affective issues** (i.e., participation increases disposition to future evaluation and participation increases commitment

to implement recommendation). The qualitative showed the motivation for future participation was learning.

7.1.3 The patterns of performance

The quantitative comparisons showed differences in perceptions of what is EOEP. The distinct clusters also suggested differences in patterns of performance among the groups. The qualitative results seemed to confirm these differences. Therefore, it may be concluded there were differences among the participants on the elements of e-government outcomes evaluation and the patterns of performance (i.e., EOEP as both an entity and performance). The characteristics of participants, i.e., professional background, position, and experience (i.e., years of experience and knowledge) were important factors. Overall, participants had negative perceptions of their practice; however, IT and Evaluation managers with extensive experience were likely to have positive perceptions.

The difference in perceptions is an important topic in evaluation (Brandon & Fukunaga, 2014). Discordance among groups may result from differences in everyday experiences (Fierro & Christie, 2017); e.g., from resourcing, involvement, expectations, etc. Agreement may signify a shared view, evidence of an established evaluation culture (Preskill & Boyle, 2008b). Thus, this research leveraged the differences observed among groups (e.g., IT and Evaluation managers with expertise differed from the other groups) to develop a strategy for improvement.

7.1.4 Summary of findings

Sections 7.1.2 and 7.1.3 shed light on how e-government outcomes evaluation is done, i.e., the main elements and patterns of performance. Table 7.2 draws and juxtaposes conclusions against findings in the literature review (Chapters 2 and Chapter 3).

As aforementioned, mixed methods research enables further conclusions (i.e., meta-inference) otherwise not possible from either quantitative or qualitative research. The following overarching propositions are drawn from the conclusions:

1. There is inadequate internal organizational capacity to do e-government outcomes evaluation and use results
2. There is a lack of a shared view of how e-government outcomes evaluation is done
3. E-government outcomes evaluation is for compliance; learning is not for continuous improvement
4. Organizational performance outcomes are evaluated while democratic and development (e.g., public participation, anti-corruption, etc.) are ignored

5. Evaluation is not atheoretical but informed by public sector policies

Table 7.2: Summary of results

	Conclusion	*Consistent with literature on…*
1	Evaluation is symbolic (i.e., for compliance) and conceptual (e.g., improved knowledge), not improvement	Development evaluation (see section 2.4.2.7); IT evaluation (Song & Letch, 2014); evaluation (Mayne, 2014)
2	Organizational performance values are widely measured, compared to democratic	E-government evaluation (Sterrenberg, 2017); public value (Bryson et al., 2014)
3	Expertise (from professional background, roles and positions) and preparation (i.e., training, workshop, etc.) enable understanding of roles	Practice theory (Schatzki, 2017b); evaluation (Preskill & Boyle, 2008b)
4	Expertise (from professional background, roles and positions) enable participation	Practice theory (Schatzki, 2017b); evaluation (Preskill & Boyle, 2008b)
5	Recommendations are not implemented	Evaluation (Wholey, 2015); e-government (Assar & Boughzala, 2016; Cha & Park, 2018; Kasimin, Aman, & Noor, 2013)
6	There is some degree of use, i.e., learning (not for improvement)	Evaluation (Mayne, 2014); development evaluation (Chouinard & Cousins, 2015); e-government (Aman & Kasimin, 2011; Coombs, 2015)
7	Inadequate internal organizational capacity to do evaluation and use results (lack of resources)	E-government in developing countries (section 2.3); evaluation in development (2.4.6)
8	Differences in perceptions (lack of a shared view)	Practice (Schatzki, 2005); evaluation practice (Fierro & Christie, 2017)
9	E-government outcomes evaluation is informed by policies	Contrary to the literature, i.e., IT evaluation not supported by theory (section 2.5.9)

The symbolic and conceptual use as well as the lack of implementation and learning for improvement may be attributable to lack of expertise and external demand for evaluation (Mayne, 2014). The literature review on South Africa as context (section 2.7.2.2) identified evaluation challenges, e.g., a lack of expertise, limited evaluation capacity, and evaluation for compliance. This is consistent with the conclusions of this research.

Consistent with the objective of this research, the next question to answer was: "*RQ3: How can e-government evaluation be improved*?

7.2 Recommendations for improving practice

Among the implications drawn from practice theory and specified in section 3.4 were: what aspects of practice need improvement and how can these be improved? It is not uncommon to compare findings to the literature to identify what is positive (i.e., effective) to uphold and negative or mixed (e.g., challenges) to improve. Brandon and Fukunaga (2014) reviewed the research findings on stakeholder participation in evaluation for the same purpose. This is consistent with the approach to improve practice, which was adopted for this research; i.e., to identify and recraft the parts of practice which need improvement (see section 3.2.2). Table 7.3 compares the positive and negative findings.

Table 7.3: Positive and negative or mixed findings

	Positive	*Negative or mixed*
Purpose		Compliance, not improvement
Evaluand		Organizational outcomes (e.g., performance); democratic and development ignored
Evaluation strategy		Survey; strategies for explanation ignored
Organizational capacity		Inadequate internal expertise No dedicated budgeted Limited management support
Policies	Framework to inform evaluation and what to evaluate	As stated above, democratic values ignored
Participation	Throughout in decision-making expertise and roles are important	Somewhat in activities due lack of expertise
Roles and responsibilities	Understand roles Expertise and preparation are important	Not translated into participation throughout
Use of report	Management consideration Distribution to groups Learning	Recommendations not implemented
Effect	Learning about EOEP	Not for improvement
Affective issues	Learning and role as motivation for participation	
Difference in perceptions	By participants' characteristics IT and evaluation managers with extensive evaluation experience: positive perception; likely to participate	Other non-managers with little expertise: negative perceptions; likely not to participate

Recommendations may be drawn for e-government evaluation stakeholders, both within the public sector (e.g., evaluation practitioners and managers) and outside (e.g., consultants, academic institutions, and professional organizations).

As shown in sections 3.2 and 3.4, the bundle is central to practice research. Thus, any strategy for improvement needs to target the bundle itself. That is, activities, material entities (e.g., people and non-human resources), and connections (i.e., practical knowledge; shared assumptions and values; rules and teleoaffective structures, e.g., goals and affective issues).

Section 3.2.4 identified agency, structure, learning, and change as important factors which influence practice. For example, individual and organizational motivation and attitudes may influence agency (i.e., choices, actions and, thereby, learning and change); organizational structures can also pre-configure agency, learning, and change; and learning can improve knowledge, and engender agency and change.

To be effective, strategies for evaluation and use need to be purposeful and intentional about interactions among people, learning, positive attitude, etc. (Chouinard & Cousins, 2015; King & Stevahn, 2016; Preskill & Boyle, 2008a, 2008b). Section 2.4.4 discussed some of the factors

which can improve evaluation and use. For example, organizational learning, formal structures, management participation, capacity building, evaluator facilitation, and evaluation champion. These may be leveraged in consideration of Table 7.3 to create an enabling organizational environment for agency, learning, change, and thereby, improvement.

Shared view of evaluation. Participants were aware of frameworks (i.e., Batho Pele and NEPF), which inform how evaluation is done. As pointed out in section 2.7.1, public organizations are enjoined to develop specific institutional policies applicable to their context. Thus, an institutional evaluation framework which specifies evaluation purpose (which presupposes what to evaluate, by whom and for what use), enables shared understanding, and aligns practices towards the achievement of such purpose are important. Organizational structures and activities, e.g., communication; forums; competitions; etc., to improve awareness and help stakeholders internalize the purpose and value of evaluation may be useful. The development of an evaluation framework is a strategic activity and may be undertaken by management in conjunction with evaluation structures.

In addition to purpose, what to evaluate as well as legitimate processes to do evaluation and assure recommendations are relevant to the organizational context, are also important. Institutional evaluation frameworks can ensure evaluation accounts for both organizational and democratic values. Appropriate processes and methods, e.g., mixed methods (as explained in section 2.2.7) can capture both tangible organizational and intangible public values; support compliance; and produce relevant information to promote learning for improvement.

Inadequate internal organizational capacity. This is a barrier to doing evaluation and using the results. The findings show there is a need for improvement. A strategy to improve internal expertise can include intentional participation and learning, i.e., "learning how to do" and "learning how to learn" (Chouinard & Cousins, 2015). For example, mentorship and teamwork facilitated by professional evaluators; formal training, e.g., courses by academic institutions and professional bodies, etc. (Podems et al., 2014; Preskill & Boyle, 2008b). Other mechanisms, e.g., clearly defined roles and responsibilities; recognition and reward schemes, can promote participation and learning (Gagnon et al., 2018; Preskill & Boyle, 2008a).

Planning and budgeting, i.e., evaluation as a budget item (Gagnon et al., 2018; Wade & Kallemeyn, 2020) can improve internal demand for evaluation. Management support, e.g., evaluation champions (especially, managers with expertise) can facilitate evaluation and use: identify opportunities for, and advocate the value of, evaluation; encourage participation and

buy-in; develop and allocate resources; manage interactions, etc. (Preskill & Boyle, 2008a; Rogers & Gullickson, 2018; Wade & Kallemeyn, 2020).

On **use of report**, the current practice (i.e., management response to, and distribution of, report) can improve learning and is worth upholding. However, the actual implementation of recommendations for improvement is important for evaluation to take root (Preskill & Boyle, 2008b). Some strategies have been outlined above; e.g., improved understanding of the value of evaluation; appropriate evaluation strategy (e.g., mixed methods) to assure useful reports; stakeholder participation (e.g., in doing evaluation and developing reports) can increase buy-in; structures to support learning, and internal expertise to use, reports; etc. (King & Stevahn, 2016; Wade & Kallemeyn, 2020).

The results showed the main **effect** was learning about e-government outcomes evaluation. An effective learning strategy needs to set out goals (Preskill & Boyle, 2008b). Characteristics, e.g., knowledge, skills, and affective issues (i.e., motivation, attitude, expectations, etc.) of individuals, groups, and organizations influence choices and actions (Preskill & Boyle, 2008b; Wade & Kallemeyn, 2020) and, therefore, are important determinants of learning and change.

Strategies by management (e.g., participation, teamwork, training, and mentoring— discussed above) need to set goals in consideration of these factors to build on learning about the evaluation process to include implementation of recommendations. Learning for improvement can be coupled to compliance (Mayne, 2014); the two are not mutually exclusive. Other mechanisms, e.g., reward systems, capacity development (e.g., expertise), etc., can improve motivation and attitudes.

This research showed evaluators and managers with experience were likely to have positive perceptions (i.e., agree with the statements) on EOEP. Overall, evaluators, managers, and evaluation structures may be central to the recommendations above. For example, evaluators are experts and can facilitate teamwork and mentorship to develop internal expertise; managers are likely to allocate resources and act as evaluation champions; and evaluation structures may be responsible for organizing evaluation.

However, as discussed in preceding chapters, the influence of development agencies permeates e-government and development evaluation in general. Traditionally, these agencies have spearheaded evaluation capacity building in developing countries and, therefore, can bring their efforts to bear on strategies to improve e-government outcomes evaluation practice.

While development agencies may not be the panacea, strategies such as allocation of part of the grant money to evaluation; clear specification of the role of stakeholders; and review of methodology before an evaluation can go a long way (Fielding, 2013). Table 7.4 summarizes the recommendations for stakeholders; development agencies may be involved in all aspects.

Table 7.4: Summary of recommendations:

Aspect of EOEP	*Stakeholder group*	*Recommendation*
Shared view of evaluation	Management, evaluation structures and development agencies	Develop specific evaluation policies applicable to organizational context, which: • Leverage organizational structures (e.g., communication) and activities (e.g., competitions) to improve awareness among stakeholders • Emphasize processes to promote evaluation and use • Encompass organizational and democratic values • Promote learning for improvement
Internal organizational capacity	Management, evaluation structures, evaluators, and development agencies	• Implement structures and activities (e.g., mentorship, teamwork, academic and professional courses, etc.) to facilitate intentional participation and learning to improve expertise • Plan and budget to improve internal demand for evaluation • Identify and install evaluation champions to advocate for resources, manage interpersonal relationships, evaluation and use, etc. • Include managers in evaluation and structures to improve management support
Use of report	Management, evaluation structures and stakeholders and development agencies	Involve stakeholders in development of reports to improve: • Learning about results • Buy-in and implementation of recommendations
Participation	Management, evaluation structure and stakeholders and development agencies	• Selection of participants with expertise to improve participation throughout the evaluation • Intentional learning (e.g., mentorship, teamwork, training, etc.) to facilitate participation and learning
Roles and responsibilities	Management, evaluation structures, stakeholders and development agencies	• Clearly define roles • Frequent training, workshops, etc., to leverage understanding of roles to do and use evaluation
Effect	Management, evaluation structures and development agencies	to improve intentional learning and implementation; couple improvement with compliance
Affective issues	Management and evaluation structures and development agencies	Reward systems, mentorship, etc. to reduce anxiety, improve expertise and spawn positive motivation, attitudes, etc.

It may be observed that the strategies for improvement are interrelated. For example, addressing organizational capacity can impact on participation and improve use and learning; similarly, implementing recommendations can improve learning and, thereby, organizational capacity. Thus, the foregoing strategies need to be considered together.

Notwithstanding best efforts to ameliorate the threats to this research, the following weaknesses and problems are worth noting, alongside the findings and benefits.

7.3 Validation

As discussed in sections 4.6.5, the research process (from conceptualization through operationalization to empirical measurements and drawing conclusions) is fraught with threats to quality. For example, inappropriate sample, researcher and participant biases and errors, instrumentation and measurement errors, and invalid conclusions.

As is typical, validation involved evaluating the respective quantitative and qualitative phases and, subsequently, the mixed methods research itself, according to their traditional yardsticks (see section 4.6.7.6).

The first level of validation is to establish whether the other elements are congruous with the research purpose and outcomes. “This is why a good research design is important” (Saunders et al., 2015). Chapter 4, which embodied the research design, considered the alternatives of each research element, documented the choices, and justified their suitability to the research purpose.

Chapters 3 showed, practice— as a real empirical object— admits to quantitative and qualitative methods of inquiry to examine both tangible and intangible aspects. Chapter 4 showed the research purpose and questions concerned what is practice and how it is performed, which made post-positivism and mixed methods suitable candidates. Post-positivism is compatible with the flat ontology of practice theory. The review on mixed methods showed sequential explanatory typology was appropriate: quantitative, to identify the elements of practice and patterns of performance; qualitative, to follow-up on the quantitative for further insight. A survey allowed adequate data for description. It may be concluded that the theories and research strategy were appropriate for the research purpose (i.e., descriptive), questions, and expected outcomes. Validation of quantitative, qualitative, and integrated phases are described below.

7.3.1 Validation of Phase 1

Validation of quantitative research involves validity (i.e., internal, external, construct, content, and face validity) and reliability (internal consistency). Chapter 5 described the stages in the phase, relevant methods and their application, and efforts to improve quality. Table 7.4 shows a summary.

Table 7.5: Validation of Phase 1 (Quantitative)

Element of research	*Implementation tactic*	*Comment*
Purpose	Descriptive (to describe current practice)	
Research question	RQ1 ("what" type) and RQ2 ("how" type)	Adequate; consistent with the purpose
Research strategy	Survey	Adequate; consistent with the purpose
Sampling	Population (stakeholders with experience) Sample: public employees with experience Sampling: purposive and snowball Sample size (N=106)	Appropriate sample, sampling technique; adequate sample size. Conclusions, however, not generalizable beyond the sample
Instrument	Questionnaire	Consistent with strategy and purpose
	Construct validity (constructs from theory or extant research)	Initial list (of concepts and corresponding constructs) and questionnaire reviewed by experts
	Content validity (items for constructs derived from theory or extant research) Face validity	
	Reliability	Piloted and revised to improve understanding and consistent responses
	Detailed description of decisions and actions	Documented (section 5.2.1)
Data collection	Questionnaire	Appropriate
	Email, in-person, and follow-up	To improve response rate
	Coding and data storage to improve data integrity	"Double-entry"
	Detailed description of decisions and actions	Documented (sections 5.2.2 and 5.2.3)
Data analysis methods	Characteristics of participants and responses	Descriptive analysis appropriate
	Factor analysis	Appropriate for research purpose (to identify the structure of responses, i.e., elements of practice) Appropriate data (factorability) Statistically, conceptually and practically adequate solution
	Comparisons for differences among groups (Mann-Whitney and Kruskal-Wallis)	Non-parametric methods appropriate for small sample size and non-normal distribution (sections 5.4.3 and 5.5.3)
	Data reduction (average scores)	Appropriate method; maintains measurement scale
	Cluster analysis	Appropriate (to identify the patterns of performance and group people according to the patterns) Data appropriate (factor-cluster) Statistically, conceptually and practically adequate solution
	Detailed description of decisions and actions	Sections 5.4 and 5.5
Reporting	Tables and appropriate figures	

7.3.2 Validation of Phase 2

Various strategies and tactics (e.g., Tables 6.2 and 6.3; Figure 6.2) were devised to help improve quality. Table 7.6 serves as a checklist for validating the qualitative research.

Table 7.6: Validation of Phase 2 (Qualitative)

Step	*Implementation tactic*	*Comment*
Planning	Relevance of research topic	Specified in section 1.1
	Research objective and question upfront	Specified in section 1.1
	Research strategy upfront	Specified in sections 4.8 and 6.1
Sampling	Appropriate data sources, sampling method and sample size	Sample drawn from Phase 1; purposive; N=12
	Specify criterion for inclusion	People with experience, from Phase 1
	Detailed description of decisions and actions	Adequate description (section 6.2)
Instrumentation	Interview protocol construction	Protocol for the semi-structured interview
	Valid items	Relevant concepts (from Phase1)
	Review and pilot	Reviewed by experts and piloted
	Detailed description	Adequate description (section 6.3.1)
Data collection	Semi-structured Interview	Semi-structured interview appropriate for understanding Structured to reduce researcher bias (similar questions to elicit valid responses) Anonymity to reduce participant bias
	Follow-up questions to remove ambiguity	Follow-up to reduce participant errors
	Data recorded "as is"	Responses recorded with permission
	Detailed description of decisions and actions	Adequate description (section 6.3.2)
Data analysis	Analysis strategy	Specified upfront (Figure 6.2)
	Content analysis method	Appropriate for the research purpose
	Transcription to familiarize with data	Transcribed
	Initial coding scheme to "focus" analysis	Appendix D2
	Review and refine coding scheme	Initial coding scheme reviewed and revised
	Pilot analysis and review results	Sample transcripts analysed and reviewed
	Detailed description of decisions and actions	Adequate description (section 6.5)
Data presentation and Interpretation	"Focus" with propositions	Organized by propositions
	Tables to improve meaning	Tables (sections 6.5 and 7.1)
	Interpret "as is", from participants' viewpoint	Summaries (supported by quotes); conclusions drawn from data
	Participant validation	Participant review of summary and conclusions
	Detailed description of decisions and actions	Adequate description (section 6.5)

7.3.3 Validation of mixed methods

The checklist for mixed methods research validation in section 1.1 is reproduced in Table 7.7 to show that the tenets of SEMM design were appropriately implemented in this research.

Table 7.7: Validation: mixed methods research

	Aspect of research	*Guideline*	*Tactic*
1	Title	Mention the topic and type of mixed methods	Both mentioned in title
2	Abstract	Outline the research topic, type of mixed methods, data sources, data integration, and what was learnt	All discussed
3	Introduction	Establish the need for research	Formulated in problem statement (section 1.1)
		Establish overall rationale for mixed methods	Stated: to examine the different aspects of practice
		Where necessary, formulate both quantitative and qualitative questions to collects quantitative and qualitative data	RQ1: quantitative, about "what" (i.e., to identify elements) RQ2: qualitative, about how (i.e., to understand patterns of performance— how something is done)
		Outline research strategy	Explicated in section 1.3
4	Literature review	Review of mixed methods literature for audience	Reviews of literature, basic features, main typologies, etc. (section 4.6.7)
		Establish deficit of mixed methods in the area of study	Chapter 2 (Literature review) established paucity in IS, public value and practice research
5	Research design	Overall rationale for mixed methods	Explanatory (stated in Abstract, Introduction, and section 4.8
		Detailed description of applicable typology	Overview of SEMM (section 4.8)
		Visual illustration of quantitative and qualitative methods in the order of phases	Illustrated by diagrams (Figure 4.3)
		Detailed description of methods according to phases	Overview of quantitative (section 4.8) Purpose: to describe practice Sampling— purposive, snowball Data collection— questionnaire Data analysis— descriptive, exploratory factor and cluster analyses Overview of qualitative (section 4.8) Purpose— to understand results of Phase 1 Sampling— purposive, stratified; sampling frame: Phase 2 sample Data collection: semi-structured interview Data analysis—content analysis
		Mixing	Sampling; results
6	Implementation	Phase 1: Quantitative (Chapter 5) Phase 2: Qualitative; follow-up on Phase 1 (Chapter 6)	
7	Discussion and conclusion	Integrated results of phases to draw meta-inferences (Chapter 7)	
8	Contribution	To domain, i.e., e-government outcomes evaluation (Chapter 7) To methodology, i.e. practice research and mixed methods (Chapter 7)	
9	Reporting	Write for the audience; presented implementations in respective phases and integrated results; used headings, visual display, etc.	

Sections 7.3.1 and 7.3.3 showed several strategies were employed to reduce threats and assure the quality of outcomes throughout this research.

7.3.4 Limitations

Notwithstanding its value (e.g., support for quantitative and qualitative perspectives of a problem) and suitability to the purpose of this research (i.e., to describe current practice), a sequential explanatory mixed methods strategy has its shortcoming. For example, the questionnaire and factor and cluster analyses in Phase 1 pre-structured which aspects of EOEP to examine. Also, cross-sectional research strategy may only paint a picture at a certain point in time and, therefore, not capable of explaining substantive changes to practice over a period.

The sample and context of the investigation were described in detail. While e-government evaluation stakeholders extend beyond the public sector, this research considered only public employees. Non-probability (i.e., purposive, snowball, and stratified) sampling techniques were employed in both phases, and the sample sizes were small. Therefore, the findings may not be generalized beyond the sample, as it is neither representative of the stakeholders nor the South African public sector. Furthermore, both public value and practice depend on context; consequently, the insights generated herein may be transferable to similar contexts, but not generalizable (Rowley, 2012).

New instruments were constructed for data collection; most of the items were either new or adapted to suit this research. Although validated, the questionnaire and interview protocol are not standardized. The data collection techniques relied on self-reporting to elicit retrospective data. This may introduce instrument bias; e.g., participants may not recall their experience; give socially desirable responses; etc. Factor, cluster, and content analyses involve several subjective decisions and, therefore, different researchers could produce different outcomes.

Regardless of the foregoing limitations, the adopted strategies were adequate, and the results provided conceptual and empirical bases to advance the under-developed domain of e-government outcomes evaluation. This research makes several contributions to EOEP and the literature.

7.4 Contribution

As discussed in section 1.7, research is valuable when the outcome is relevant to practice and extends an existing body of knowledge in a field. Although there are no standard criteria for evaluation, the following may be considered: the importance of the topic, the difference the findings can make to practice, originality of research design and implementation, the extent to which outcomes engender a new understanding of an old problem, identification of new concepts and their relationships, etc. (Gill & Dolan, 2015; Hassan, 2014).

This research makes contributions by addressing the gaps identified in the literature review. That is, the paucity of literature on e-government (and IS) outcomes evaluation, particularly, in developing countries; and the under-developed theoretical, methodological, and empirical research on practice in IS. Furthermore, it outlines implications for practice and research.

The findings provide evidence of current practice to advance an understanding of how e-government outcomes evaluation is done in a developing country and draws attention to aspects which are effective and worthy of maintaining— or are not, and need improvement. This can heighten awareness among stakeholders (public organizations, practitioners, researchers, etc.) about enablers, challenges, etc., and spur changes in attitude and help improve future practice. Furthermore, the recommendations derived from findings may provide a framework to assist organizations to put in place strategies for improvement. These can result in improvements in outcomes evaluation, use of results to effect change, and, thereby, help alleviate the otherwise high e-government failure rate.

The research strategy employed an innovative theoretical and methodological approach. Evaluation, public value, and Schatzki's version of practice theory were explicated and drawn on to frame EOEP as how work is done in public organizations and to construct a conceptual model. This research, thus, answers the calls to adopt evaluation as a reference domain for IS evaluation research (Goldkuhl & Lagsten, 2012) and address the paucity of research on IS and e-government outcomes evaluation (Bannister & Grönlund, 2017; Pekkola & Päivärinta, 2019) as well as IS practice (Leidner et al., 2017). Moreover, it serves as an illustrative example of how IS phenomenon can be conceptualized as practice and examined as both entity and performance. The application of mixed methods responds to calls to employ such methodology to advance empirical research on public value (Hartley et al., 2017); practice, based on Schatzki's theory (Loscher et al., 2019); and IS (Venkatesh et al., 2016).

This research, therefore, makes original contributions (Gill & Dolan, 2015; Hassan, 2014): new insight into how e-government evaluation is done; recommendations to improve practice; a novel theoretical and methodological development (i.e., adoption of practice theory and SEMM); and an agenda for future research on e-government and, more broadly, IS outcomes evaluation practice research.

7.5 Further research

The growing adoption of new technologies in the public sector means how e-government outcomes evaluation is done may continue to preoccupy researchers and practitioners for years to come. From the findings and limitations, there are several ways to improve and extend this research. The following draws agenda for future research on the topic under examination and the methodology employed.

The foregoing findings show there were differences in perceptions, e.g., between managers and non-managers on organizational capacity. It may be important to examine and compare perceptions of different groups in organizations to determine consensus and discord: to help understand and target strategies to improve evaluation and use. Fierro and Christie (2017) and Tallon (2014) examined differences between groups in evaluation practice and IS business value respectively.

The literature review underlined the importance of affective issues in public value, practice, and evaluation. There is growing emphasis on research on affective issues in IS (Zhang, 2013) and evaluation (Brandon & Fukunaga, 2014). The findings show positive attitudes towards future participation. Research can examine how affective issues (e.g., practitioners' motivation) influence how e-government outcomes evaluation is done; e.g., to understand the relationship between participation, improved expertise, incentives, etc., on affective issues (attitude, motivation, satisfaction, buy-in, etc.).

This research identified various challenges to current practice (i.e., negative or mixed findings). The literature on evaluation often ignores research on challenges (Brandon & Fukunaga, 2014). It may, therefore, be worth investigating the organizational capacity for e-government evaluation, lack of implementation of recommendations, etc. Also, an interesting observation was the near-absence of the democratic (e.g., public participation) and development (e.g., poverty alleviation) values of e-government and its implication on evaluation. These may be research problems to pursue.

Replication of this research can improve on the instruments and methods as well as validate the findings. A larger sample size, which includes evaluation stakeholders external to the public sector (i.e., businesses and citizens) may paint a more representative picture of practice. Longitudinal research may examine how e-government outcomes evaluation has been done over time, e.g., to understand changes and whether there has been improvement. Other

strategies, e.g., document analysis; observation, etc. may be adopted to corroborate findings from retrospective self-reporting.

Comparative studies— the staple of public sector research— can reveal similarities, differences, and idiosyncrasies as well as support learning from initiatives of others to improve practice. As evidenced in the literature review, evaluation, public value, and practice are shaped by context. This research concerned practitioners and their practice, not organizations *per se*. There are differences among public organizations (Hodgkinson et al., 2017), which may permeate practices. Hammerschmid et al. (2013) reported differences in evaluation use across subsectors (e.g., health, education, etc.). It would be instructive to know how different organizations, subsectors, and developing countries evaluate e-government outcomes and use the results thereof.

7.6 Summary

This research set out to explore how e-government outcomes evaluation is done in a developing country, with South Africa as context. Notwithstanding its importance to the delivery of public service, this problem has attracted little attention. The research purpose was, therefore, descriptive; the objectives were to identify the main elements of EOEP, describe the patterns of performance, and recommend ways to help improve future practice.

Schatzki's version of practice theory was employed to conceptualize e-government outcomes evaluation as practice, construct a conceptual model, and draw implications for investigation. A sequential explanatory mixed methods (SEMM) survey research underpinned by post-positivism was adopted; this consisted of two phases.

In the quantitative phase, data was collected with questionnaire from 106 participants; factor and cluster analyses identified six elements of practice and three patterns of performance respectively. The qualitative phase was a follow-up on the quantitative results; this involved the interview of 12 participants and content analysis. The instruments were validated by experts and piloted.

Five overarching conclusions may be drawn about EOEP: inadequate organizational capacity for evaluation; lack of implementation of recommendations; democratic and development e-government outcomes (e.g., public participation, anti-corruption) need attention; evaluation is informed by public sector policies; and significant differences in perceptions among participants (i.e., the characteristics of participants influenced their perceptions of how e-government outcomes evaluation is done).

This research makes four main contributions, i.e., helps to fill the gaps in the literature on e-government outcomes evaluation; provides recommendations to help improve practice; draws an agenda to advance research, and adds to the effort to develop mixed methods methodology for empirical research on practice as well as IS.

References

Abrahams, M., A. (2014). A review of growth of monitoring and evaluation in South Africa: Monitoring and evaluation as profession, an industry and governance tool. *African Evaluation Journal*, *3*(1), 1–8.

Al Rawahi, K., Coombs, C., & Doherty, N. F. (2016). The realization of public value through e-government: A structuration perspective. *Investigation of E-Government Role in Public Value Production*. Thirty Seventh Conference on Information Systems, Dublin.

Aladwani, A. M. (2016). Corruption as a source of e-government projects failure in developing countries: A theoretical exposition. *International Journal of Information Management*, *36*(2016), 105–112.

Alkin, M. C., & King, J. A. (2016). The historical development of evaluation use. *American Journal of Evaluation*, *37*(4), 568–579.

Alkin, M. C., & King, J. A. (2017). Definitions of evaluation use and misuse, evaluation influence, and factors affecting Use. *American Journal of Evaluation*, *38*(3), 434–450.

Al-Yaseen, H., Eldabi, T., Paul, R. J., & El-Haddadeh, R. (2008). Post-implementation evaluation of IT systems: A close review of practice. In Z. Irani & P. E. D. Love (Eds.), *Evaluating nformation systems- Public and private sector* (pp. 134–152). Elsevier/Butterworth-Heinemann.

Aman, A., & Kasimin, H. (2011). E-government evaluation and organisational learning. *International Journal of Digital Society*, *2*(2), 469–477.

Arbour, G. (2020). Frameworks for program evaluation: Considerations on research, practice, and institutions. *Evaluation*. https://doi.org/10.1177/1356389020920890

Armytage, L. (2011). Evaluating aid: An adolescent domain of practice. *Evaluation*, *17*(3), 261–276.

Assar, S., & Boughzala, I. (2016). Evaluation approaches to e-government applications. In *Génierie et management des systèmes d'information-mélanges en l'honneur de Jacky Akoka* (pp. 71–79). https://hal.archives-ouvertes.fr/hal-01414865/document

Balijepally, V., Mangalaraj, G., & Iyengar, K. (2011). Are we wielding this hammer correctly? A reflective review of the application of cluster analysis in Information Systems research. *Journal of the Association for Information Systems*, *12*(5), 375–413.

Bannister, F., & Connolly, R. (2014). ICT, public values and transformative government: A framework and programme for research. *Government Information Quarterly*, *31*(2014), 119–128.

Bannister, F., & Grönlund, Å. (2017). Information Technology and government research: A brief history. *50th Hawaii International Conference on Systems Science*, 2943–2952.

Bartiauxs, F., & Salmón, R. (2012). Are there domino effects between consumers' ordinary and "green" practices? An analysis of quantitative data from a sensitisation campaign on personal carbon footprint. *International Review of Sociology*, *22*(3), 471–491.

Bartiauxs, F., & Salmón, R. (2014). Family dynamics and social practice theories: An investigation of daily practices related to food, mobility, energy consumption and tourism. *Nature and Content*, *9*(2), 204–224.

Bazeley, P. (2013). *Qualitative data analysis: Practical strategies*. Sage.

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation*, *18*(6).

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open*, *2*(2016), 8–14.

Benington, J., & Moore, M. H. (2010). *Public value: Theory and practice*. Palgrave Macmillan.

Benington, J., & Moore, M. H. (2011). Public value in complex and changing times. In J. Benington & M. H. Moore (Eds.), *Public value: Theory and practice* (pp. 1–20). Palgrave Macmillan.

Berghout, E., & Powell, P. (2010). IS/IT Benefits management and realization. In K. Grant, R. Hackney, & D. Edgar (Eds.), *Strategic information systems management* (pp. 273–299). Cengage Learning.

Berman, E. (2017). An exploratory sequential mixed methods approach to understanding researchers' data management practices at UVM: Integrated findings to develop research data services. *Journal of EScience Librarianship*, *6*(1), e1104.

Besner, C., & Hobbs, B. (2013). Contextualized project management practices: A cluster analysis of practices and best practices. *Project Management Journal*, *44*(1), 17–34.

Bhaskar, R. (2008). *A Realist theory of science*. Routledge.

Bhatia, D., Bhatnagar, S. C., & Tominaga, J. (2009). How do manual and e-government services compare? Experiences from India. In *Information and communications for development 2009: Extending reach and increasing impact* (pp. 67–82).

Bhattacherjee, A. (2020). *Social science research: Principles, methods, and practices*. Global Text Project.

Bittner, E. V. C., & Leimeister, J. M. (2014). Creating shared understanding in heterogenous work groups: Why it matters to achieve it. *Journal of Management Information Systems*, *31*(1), 111–144.

Blaikie, N. (2010). *Designing social research*. Polity Press.

Blumer, M. (2017). Introduction: Problems, theories and methods in sociology- (How) do they interrelate? In M. Blumer (Ed.), *Sociological research methods: An introduction* (pp. 1–37). Routledge.

Bolívar, M. P. R., Muñoz, L. A., & Hernández, A. M. L. (2016). Scientometric study of the progress and development of e-government research during the period 2000-2012. *Information Technology for Development*, *22*(1), 36–74.

Bourgeois, I., & Naré, C. (2015). The "usability" of evaluation reports: A precursor to evaluation use in government organizations. *Journal of MultiDisciplinary Evaluation*, *11*(25), 60–67.

Bozeman, B. (2012). Public values concepts and criteria: The case for "progressive opportunity" as a criterion. *Center for Integrative Leadership. Paper Presented at the Creating Public Value in a MultiSector, SharedPower World, Minneapolis, MN*.

Brandl, K., Moi, M. J., & Petersen, B. (2017). The reconfiguration of service production systems in response to offshoring: A practice perspective. *International Journal of Operations and Production Management*, *37*(9), 1246–1264.

Brandon, P. R., & Fukunaga, L. L. (2014). The state of the empirical research literature on stakeholder involvement in program evaluation. *American Journal of Evaluation*, *35*(1), 26–44.

Breidahl, K. N., Gjelstrup, G., Hansen, H. F., & Hansen, M. B. (2016). Evaluation of large-scale public sector reforms: A comparative analysis. *American Journal of Evaluation, 38*(2), 226–245.

Brown, M. M. (2015). Revisiting the IT productivity paradox. *American Review of Public Administration, 45*(5), 565–585.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.

Browne, A. L., Anderson, B., & Pullinger, M. (2015). Methods as intervention: Intervening in practices through qualitative and mixed methods. In Y. Strenger & C. Maller (Eds.), *Social practices intervention and sustainability: Beyond behaviour change* (pp. 179–185). Routledge.

Browne, A. L., Pullinger, M., Medd, W., & Anderson, B. (2014). Patterns of practice: A reflection on the development of quantitative/mixed methodologies capturing everyday life related to water consumption in the UK. *International Journal of Social Research Methodology, 17*(1), 27–43.

Bryman, A. (2016). *Social research methods*. Oxford University Press.

Bryson, J. M., Crosby, B. C., & Bloomberg, L. (2014). Public value governance: Moving beyond traditional public administration and the New Public Management. *Public Administration Review, 74*(4), 445–456.

Bryson, J. M., Crosby, B. C., & Bloomberg, L. (2015). Discerning and assessing public value: Major issues and new directions. In J. M. Bryson, B. C. Crosby, & L. Bloomberg (Eds.), *Public value and public administration* (pp. 1–24). Georgetown University Press.

Bunge, M. (2006). *Chasing reality: Strife over realism*. University of Toronto Press.

Burton-Jones, A., McLean, E., R., & Monod, E. (2015). Theoretical perspectives in IS research: From variance and process to conceptual lattitude and conceptual fit. *European Journal of Information Systems, 24*(6), 664–679.

Bwalya, K. J., & Mutula, S. M. (2014). *E-government: Implementation, adoption and synthesis in developing countries*. De Gruyter/Saur.

Cameron, R. (2015). Public service reform in South Africa: From apartheid to new public management. In A. Massey & K. Johnston (Eds.), *The international handbook of public administration and governance* (pp. 135–157).

Carden, F., & Alkin, M. C. (2012). Evaluation roots: An International perspective. *Journal of Multidisciplinary Evaluation, 8*(12), 102–118.

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education, 42*(12), 1150–1152.

Castillo-Montoya, M. (2016). Preparing for interview research: The interview protocol refinement framework. *The Qualitative Report, 21*(5), 811–831.

Cecez-Kecmanovic, D., & Kennan, M. A. (2018). The methodological landscape: Information systems and knowledge management. In K. Williamson & G. Johanson (Eds.), *Research methods: Information, systems, and contexts*. Elsevier.

Ceric, A. (2015). Bringing together evaluation and management of ICT value: A systems theory approach. *Electronic Journal of Information Systems Evaluation, 18*(1), 19–35.

Cha, Y. U., & Park, M. J. (2019). Evaluating and planning Information Systems in the public sector: The case of Korea. *Information Development, 35*(4), 655–665.

Chaysin, P., Daengdej, J., & Tangjitprom, N. (2016). Survey on available methods to evaluate IT Investment. *Electronic Journal of Information Systems Evaluation, 19*(1), 71–82.

Check, J. W., & Schutt, R. K. (2012). *Research methods in education*. Sage.

Chelimsky, E. (2013). Balancing evaluation theory and practice in the real world. *American Journal of Evaluation*, *34*(1), 91–98.

Chen, H. T. (2015). *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective*. Sage.

Cho, J. Y., & Lee, E.-H. (2014). Reducing confusion about grounded theory and qualitative content analysis: Similarities and differences. *The Qualitative Report*, *19*(32), 1–20.

Chouinard, J. A., Boyce, A., Hicks, J., Jones, J., Long, J., Pitts, R., & Stockdale, M. (2017). Navigating theory and practice through evaluation fieldwork: Experiences of novice evaluation practitioners. *American Journal of Evaluation*, *38*(4), 493–506.

Chouinard, J. A., & Cousins, J. B. (2015). The journey from rhetoric to reality: Participatory evaluation in a development context. *Educational Assessment, Evaluation and Accountability*, *27*(1), 5–39.

Chouinard, J. A., & Miley, P. (2016). Mapping the spatial dimensions of participatory practice: A discussion of context in evaluation. *Evaluation and Program Planning*, *54*(2016), 1–10.

Chouinard, J. A., & Milley, P. (2018). Uncovering the mysteries of inclusion: Empirical and methodological possibilities in participatory evaluation in an international context. *Evaluation and Program Planning*, *67*(2018), 70–78.

Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation*, *2003*(97), 7–36.

Christie, C. A. (2012). Advancing empirical scholarship to further develop evaluation theory and practice. *Canadian Journal of Program Evaluation*, *26*(1), 1–18.

Christie, C. A., & Fleischer, D. N. (2010). Insight into evaluation practice: A content analysis of designs and methods used in evaluation studies published in North American evaluation-focused journals. *American Journal of Evaluation*, *31*(3), 326–346.

Chyung, S. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement*, *56*(10), 15–23.

Cloete, F. (2012). E-Government lessons from South Africa 2001—2011: Institutions, state of progress and measurement. *The African Journal of Information and Communication*, *2012*(12), 128–142.

Coldwell, M. (2019). Reconsidering context: Six underlying features of context to improve learning from evaluation. *Evaluation*, *25*(1), 99–117.

Collis, J., & Hussey, R. (2014). *Business research: A practical guide for undergraduate & postgraduate students*. Palgrave Macmillan.

Connelly, S., & Vanderhoven, D. (2018). The craft of evaluative practice: Negotiating legitimate methodologies within complex interventions. *Evaluation*, *24*(4), 419–437.

Coombs, C. (2015). When planned IS/IT project benefits are not realized: A study of inhibitors and facilitators to benefits realization. *International Journal of Project Management*, *33*(2015), 363–379.

Cordella, A., & Bonina, C. M. (2012). A public value perspective for ICT enabled public sector reforms: A theoretical reflection. *Government Information Quarterly*, *29*(4), 512–520.

Corder, G. W., & Foreman, D. I. (2014). *Nonparametric sampling: A step-by-step approach*. John Wiley & Sons.

Córdoba-Pachón, J. R. (2014). Systemic patterns of practice to improve e-government evaluation. In M. P. Rodríguez-Bolívar (Ed.), *Measuring e-government efficiency* (pp. 9–24). Springer New York.

Coryn, C. L., Noakes, L. A., Westine, C. D., & Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation, 32*(2), 199–226.

Coryn, C. L., Ozeki, S., Wilson, L. N., Greenman, G. D., Schröter, D. C., Hobson, K. A., Azzam, T., & Vo, A. T. (2015). Does research on evaluation matter? Findings from a survey of American Evaluation Association members and prominent evaluation theorists and scholars. *American Journal of Evaluation*, *37*(2), 159–173.

Cousins, J. B. (2013). When does a framework become a theory? Reflections from an accidental theorist. *Evaluation and Program Planning*, *38*(2013), 67–70.

Creamer, E. G. (2018). *An introduction to fully integrated mixed methods research*. Sage.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage.

Creswell, J. W. (2015). *A concise introduction to mixed methods research*. Sage.

Creswell, J. W., & Poth, C. N. (2017). *Qualitative inquiry and research design: Choosing among five approaches*. Sage.

Cullen, A. E., Coryn, C. L., & Rugh, J. (2011). The politics and consequences of including stakeholders in International Development Evaluation. *American Journal of Evaluation, 32*(3), 345–361.

Dahlberg, T., & Kivijärvi, H. (2018). Role of IT Management and the dynamics behind IT Business Value Creation – A longitudinal assessment. *51st Hawaii International Conference on System Sciences*, 5136–5145.

Dahler-Larsen, P. (2018). Theory-based evaluation meets ambiguity: The role of Janus variables. *American Journal of Evaluation*, *39*(1), 6–23.

Daigneault, P. M. (2014). Taking stock of four decades of quantitative research on stakeholder participation and evaluation use: A systematic map. *Evaluation*, *45*(2014), 171–181.

de Graaf, G., Huberts, L., & Smulders, R. (2016). Coping with public value conflicts. *Administration & Society*, *48*(9), 1101–1127.

de Laat, B., & Williams, K. (2014). Evaluation use within the European Commission (EC). In M. Läubli Loud & J. Mayne (Eds.), *Enhancing evaluation use: Insights from internal evaluation units* (pp. 1–14). Sage.

De Vaus, D. A. (2014). *Surveys in social research*. Routledge.

DeLone, W. H., & McLean, E., R. (2016). Information Systems success measurement. *Foundations and Trends® in Information Systems, 2*(1), 1–116.

Deng, H., Karunasena, K., & Xu, W. (2018). Evaluating the performance of e-government in developing countries: A public value perspective. *Internet Research, 28*(1), 169–190.

Denscombe, M. (2014). *The good research guide: For small-scale social research projects*. Open University Press.

Denzin, N. K., & Lincoln, Y. S. (2011). Introduction: The discipline and practice of qualittative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 1–19). Sage.

DePoy, E., & Gitlin, L. N. (2016). *Introduction to research: Understanding and applying multiple strategies*. Elsevier.

Dighe, S., & Sarode, T. (2019). Evaluation in the Global South: Practices, problems and prospects. *Reconsidering Development*, *6*(1). https://pubs.lib.umn.edu/index.php/reconsidering/article/view/2077/1641

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation, 14*(20), 1–11.

DOJ. (1996). *Constitution of the Republic of South Africa*. Department of Justice (DOJ), South Africa.

Dolnicar, S. (2013). Answering good survey questions. *Journal of Travel Research, 52*(5), 551–574.

DPME. (2011). *National Evaluation Policy Framework*. Department of Performance, Monitoring and Evaluation (DPME), South Africa.

DPSA. (1997). *Batho Pele- "People first": Whitepaper on transforming pubic public service delivery*. Department of Public Service and Administration (DPSA), South Africa.

DPSA. (2001). *Electronic Government- The Digital Future*. Department of Public Service and Administration (DPSA), South Africa.

DTPS. (2016). *National Integrated ICT Policy White Paper*. Department of Telecommunications and Postal Services (DTPS), South Africa).

DTPS. (2017). *National e-government Strategy and and Roadmap*. Department of Telecommunications and Postal Services (DTPS), South Africa.

eGep. (2006). *EGovernment Economics Project, Measurement framework final version*. European Commision, eGovernment Unit, DG, Information Society.

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *Sage Open*, *4*(1), 1–10.

Erden, Z., Schneider, A., & von Krogh, G. (2014). The multifaceted nature of social practices: A review of the perspectives on practice-based theory building about organizations. *European Management Journal*, *32*, 712–722.

Estevez, E., & Montoya, M. S. R. (2015). Analyzing the impact of Information and Communication Technology for development: Lessons learnt from SIRCA II program. *The Electronic Journal of Information Systems in Developing Countries*, *69*(1), 1–17.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1–4.

Feldman, M., & Orlikowski, W. J. (2011). Theorizing practice and practicing theory. *Organization Science*, *22*(5), 1240–1253.

Feldman, M., & Worline, M. (2016). The practicality of practice theory. *Academy of Management Review*, *15*(2), 304–324.

Fellows, R., & Liu, A. (2015). *Research methods for construction*. John Wiley & Sons.

Fetters, M. D., & Freshwater, D. (2015). Publishing a methodological mixed methods research article. *Journal of Mixed Methods Research*, *9*(3), 203–213.

Field, A. P. (2017). *Discovering statistics using IBM SPSS statistics*. Sage.

Fielding, S. L. (2013). Mediating program evaluations: The role of funding agencies. *Journal of Applied Social Science*, *7*(1), 110–114.

Fierro, L. A. & Christie. (2017). Evaluator and program manager perceptions of evaluation capacity and evaluation practice. *American Journal of Evaluation*, *38*, 376–392.

Fleischer, D. N., & Christie, C. A. (2009). Evaluation use: Results from a survey of US American Evaluation Association Members. *American Journal of Evaluation*, *30*(2), 158–175.

Flick, U. (2014). Mapping the field. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 3–18). Sage.

Foster, L., Diamond, I., & Jefferies, J. (2015). *Beginning statistics: An introduction for social scientists*. Sage.

Frisk, J. E., Bannister, F., & Lindgren, R. (2015). Evaluation of Information Systems investments: A value dials approach. *Journal of Information Technology*, *30*(3), 276–292.

Frost, D., & Lal, B. (2019). E-government project design in developing countries. In A. Elbanna, Y. Dwivedi, & D. Wastell (Eds.), *International Working Conference on Transfer and Diffusion of IT* (pp. 155–178). Springer.

Fukumoto, E., & Bozeman, B. (2019). Public values theory: What is missing? *The American Review of Public Administration*, *49*(6), 635–648.

Gagnon, F., Aubry, T., Cousins, J. B., Goh, S. C., & Elliott, C. (2018). Validation of the evaluation capacity in organizations questionnaire. *Evaluation and Program Planning*, *68*(2018), 166–175.

Gherardi, S. (2012). *How to conduct a practice-based study: Problems and methods*. Edward Elgar.

Gherardi, S. (2017). One turn ... And now another one: Do the turn to practice and the turn to affect have something in common? *Management Learning*, *48*(3), 345–358.

Gill, P., & Dolan, G. (2015). Originality and the Phd: What is it and how can it be demonstrated. *Nurse Researcher*, *22*(6), 11–15.

Goldkuhl, G., & Lagsten, J. (2012). Different roles of evaluation in information systems research. *Workshop on IT Artefact Design and Workpractice Intervention, 10 June 2012, Barcelona*.

Goldman, I., Mathe, J. E., Jacob, C., Hercules, A., Amisi, M., Buthelezi, T., Narsee, H., Ntakumba, S., & Sadan, M. (2015). Developing South Africa's national evaluation policy and system: First lessons learned. *African Evaluation Journal*, *3*(1). http://www.aejonline.org/index.php/aej/article/view/107

Gray, D. E. (2014). *Doing research in the real world*. Sage.

Green, H. (2014). Use of theoretical and conceptual frameworks in qualitative research. *Nurse Researcher*, *21*(6), 34–38.

Guerreiro. (2016). Impact of IS/IT investments on firm performance: Does stakeholder orientation matter? *The Electronic Journal Information Systems Evaluation*, *19*(2), 99–111.

Haig, B. D., & Evers, C. W. (2016). *Realist inquiry in social science*. Sage.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate data analysis*. Cengage Learning.

Hajli, M., & Sims, J. M. (2015). Information Technology (IT) productivity paradox in the 21st scentury. *International Journal of Productivity and Performance Management*, *64*(4), 457–478.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.

Halkier, B. (2017). Methods and methods' debates within consumption research. In M. Keller, B. Halkier, T. A. Wilska, & M. Truninger (Eds.), *Routeledge handbook on consumption* (pp. 36–46). Routeledge.

Hammerschmid, G., Van de Walle, S., & Stimac, V. (2013). Internal and external use of performance information in public organizations: Results from an international survey. *Public Money & Management*, *33*(4), 261–268.

Hansen, M., Alkin, M. C., & Wallace, T. L. (2013). Depicting the logic of three evaluation theories. *Evaluation and ProgramPlanning*, *38*(2013), 34–43.

Hartley, J., Alford, J., Knies, E., & Douglas, S. (2017). Towards an empirical research agenda for public value theory. *Public Management Review*, *19*(5), 670–685.

Hassan, N. R. (2014). Value of IS research: Is there a crisis? *CAIS*, *34*(41), 801–816.

Hassan, N. R., Mingers, J., & Stahl, B. C. (2018). Philosophy and Information Systems: Where are we and where should we go? *European Journal of Information Systems*, *27*(3), 263–277.

Hébert, M. K. (2013). Anthropological thinking about e-government evaluation. In J. R. Gill-Garcia (Ed.), *E-government success factors and measures: Theories, concepts and technologies* (pp. 866–885). IGI Global.

Heeks, R. (2008). Benchmarking e-government: Improving the national and international measurement, evaluation and comparison of e-government. In Z. Irani & P. Love (Eds.), *Evaluating information systems: Public and private sector* (pp. 257–301). Elsevier/Butterworth-Heinemann.

Henry, G. T., & Mark, M. M. (2003). Toward an agenda for research on evaluation. *New Directions for Evaluation*, *2003*(97), 69–80.

Higginson, S., Mckenna, E., Hargreaves, T., Chilvers, J., & Thomson, M. (2015). Diagramming social practice theory: An interdisciplinary experiment exploring practices as networks. *Indoor and Built Environment*, *24*(7), 950–069.

Higginson, S., Mckenna, E., & Thomson, M. (2014). Can practice make perfect (models)? Incorporating social practice theory into quantitative energy demand models. *Behave 2014 - Paradigm Shift: From Energy Efficiency to Energy Reduction through Social Change*, 1–17.

Hirschheim, R. (1985). Information Systems epistemology: An historical perspective. *Research Methods in Information Systems*, *1985*, 13–35.

Hiziroglu, A., Yildiz, M., & Akpinar, M. T. (2017). Measuring the public value of e-government: Evidence from Turkey. *International Journal of Management and Applied Science*, *3*(3), 121–128.

Hodgkinson, I. R., Hannibal, C., Keating, B. W., Chester Buxton, R., & Bateman, N. (2017). Towards a public service management: Past, present, and future directions. *Journal of Service Management*, *28*(5), 998–1023.

Hofstetter, H., Dusseldorp, E., van Empelen, P., & Paulussen, T. W. G. M. (2014). A primer on the use of cluster analysis or factor analysis to assess co-occurrence of risk behaviors. *Prventive Medicine*, *27*(2014), 141–146.

Hopwood, N. (2010). *Practice theory, ethnography and the body*. Centre for Research Learning and Change Seminar Series, University of Technology, Sydney.

Hughes, O. E. (2012). *Public management and administration*. Palgrave Macmillan.

Hui, A., Schatzki, T. R., & Shove, E. (2017). Introduction. In A. Hui, T. R. Schatzki, & E. Shove (Eds.), *The nexus of practice: Connections, constellations, practitioners* (pp. 1–6). Routledge.

Huizing, A., & Cavanagh, M. (2011). Planting contemporary practice theory in the garden of information science. *Information Research*, *16*(4).

Idoughi, D., & Adbelhakim, D. (2018). Developing countries egovernment services evaluation: Identifying and testing antecedents of satisfaction- Case of Algeria. *International Journal of Electronic Government Research*, *14*(1), 63–85.

Irani, Z., & Love, P. E. D. (Eds.). (2008). Information Systems evaluation: A crisis of understanding. In *Evaluating Information Systems- Public and private Sector*. Elsevier/Butterworth-Heinemann.

Ivaldi, S., Scaratti, G., & Nuti, G. (2015). The practice of evaluation as evaluation of practices. *Evaluation*, *21*(4), 497–512.

Jacob, S. A., & Furgerson, S. P. (2012). Writing interview protocol and conducting interviews: Tips for students new to the field of qualitative research. *The Qualitative Report*, *17*(42), 1–10.

Jalongo, M. R., & Saracho, O. N. (2016). *Writing for Publication*. Springer International Publishing.

Jansen, A. (2012). The understanding of ICTs in public sector and its impact on governance. In H. J. Scholl (Ed.), *International Conference on Electronic overnment* (pp. 174–186). Springer Berlin Heidelberg.

Jokonya, O. (2016). Builiding and validating Information Systems theory using a case study sequential explanatory mixed methods research. *PACIS 2016 Proceedings*, *218*. http://aisel.aisnet.org/pacis2016/218

Jørgensen, T. B., & Bozeman, B. (2007). Public values: An inventory. *Administration & Society*, *39*(3), 354–381.

Joseph, R. (2013). A structured analysis of e-government: Trends and opportunities. *Government Information Quarterly*, *30*(4), 435–440.

Kallermeyn, L. M., Hall, J., Friche, N., & McReynolds, C. (2015). Cross-continental refelctions on evaluation practice: Methods, use and valuing. *American Journal of Evaluation*, *36*(3), 339–357.

Kamau, G., & Wausi, A. (2015). Evaluating the public value of eGovernment Services. In P. Cunningham & M. Cunningham (Eds.), *2015 IST-Africa Conference: Lilongwe, Malawi, 6—8 May 2015* (pp. 1–9). IEEE.

Kang, J., Andersen, S. G., & Finnegan, D. (2012). The evaluation practices of US international NGOs. *Development in Practice*, *22*(3), 317–333.

Kappelman, L., McLean, E., Johnson, V., & Torres, R. (2016). The 2015 SIM IT issues and trends study. *MIS Quarterly Executive*, *15*(1).

Kasimin, H., Aman, A., & Noor, Z. M. (2013). Using evaluation to support organizational learning in e-government system: A case of Malaysia government. *International Journal of Electronic Government Research*, *9*(1), 45–64.

Kearns, I. (2004). *Public value and e-government*. https://www.ippr.org/files/images/media/files/publication/2011/05/public_value_egovernment_1522.pdf?noredirect=1

Kelly, G., Mulgan, G., & Muers, S. (2002). Creating Public Value: An analytical framework for public service reform. *London: Strategy Unit, Cabinet Office*. http://unpan1.un.org/intradoc/groups/public/documents/un-dpadm/unpan044950.pdf

Kennedy, E., Krahn, H., & Krogman, T. N. (2013). Taking social practice theories on the road: A mixed-methods case study of sustainable transportation. In M. J. Cohen, H. S. Brown, & P. J. Vergragt (Eds.), *Innovations in sustainable consumption: New economics, socio-technical transitions and aocial practices* (pp. 252–276). Edward Elgar.

Keramati, A., Behmanesh, I., & Noori, H. (2018). Assessing the impact of readiness factors on e-government outcomes: An empirical investigation. *Information Developent*, *34*(3), 222–241.

Kettani, D., & Moulin, B. (2014). *E-government for good governance in developing countries: Empirical evidence from the EFez Project*. Anthem Press.

King, J. A., & Alkin, M. C. (2018). The centrality of use: Theories of evaluation use and influence and thoughts on the first 50 years of user research. *American Journal of Evaluation*, *40*(3), 431–458.

King, J. A., & Stevahn, L. (2016). Facilitating interactive evaluation practice: Engaging stakeholders constructively. In R. S. Fierro, A. Schwartz, & D. H. Smart (Eds.), *Evaluation and facilitation. New directions for evaluation* (Vol. 149, pp. 67–80). Wiley Periodicals.

Kline, R. B. (2013). Exploratory and confirmatory factor analysis. In S. Schatsschneider & Y. Petscher (Eds.), *Applied quantitative analysis in the social sciences* (pp. 171–207). Routledge.

Kovalainen, H. A., & Anderson, D. R. (2014). Majesty of truth and the moral sentiment: Emerson's and Pierce's ethico-ontological realism. In K. Westpal (Ed.), *Realism, science and pragmatism* (pp. 221–242). Routledge.

Kroll, A. (2015). Drivers of performance information use: Systematic literature review and directions for future research. *Public Performance & Management Review*, *38*(3), 459–486.

Lal, B., Dwivedi, Y. K., & Rana, N. P. (2018). Understanding 'Development' From the perspective of e-government, digital divide and ICT4D literature: A research agenda. In A. Zuiderwijk & C. C. Hinnant (Eds.), *19 Annual International Conference on Digital Research*.

Lamers, M., Van Der Duim, R., & Spaargaren, G. (2017). The relevance of practice theories for tourism research. *Annals of Tourism*, *62*(2017), 54–63.

Leedy, P. D., & Ormrod, J. E. (2015). *Practical research: Planning and design*. Pearson.

Leeuw, F. L., & Donaldson, S. I. (2015). Theory in evaluation: Reducing confusion and encouraging debate. *Evaluation*, *21*(4), 467–480.

Leidner, D. E., Kettinger, B., Gonzales, E., & Milovich, M. (2017). Introduction to the HICSS-50 Minitrack on practice-based IS Research. *50th Hawaii International Conference on Systems Science*. HICSS 2017.

Leung, S.-O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-Point Likert scales. *Journal of Social Service Research*, *37*(4), 412–421.

Leviton, L. C. (2015). Evaluation practice and theory: Up and down the ladder of abstraction. *American Journal of Evaluation*, *36*(2), 238–242.

Liem, A. (2018). Interview schedule development for a sequential explanatory mixed method design: Complementary-alternative medicine (CAM) study among Indonesian psychologists. *International Journal of Social Research Methodology*, *21*(4), 513–525.

Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research*. Sage.

Linneberg, M. S., & Korsgaard, S. (2019). Coding qualitative data: A synthesis guiding the novice. *Qualitative Research Journal*, *19*(3), 259–270.

Loscher, G., Splitter, V., & Seidl, D. (2019). Theodor Schatzki's theory and its implications for organization studies. In S. Clegg & M. Pina e Cunha (Eds.), *Management, organizations and contemporary social theory* (pp. 115–134). Routledge.

Lynch, S. M. (2016). *Using statistics in social research: A concise approach*. Springer.

Lyytinen, K., & Rowe, F. (2017). Increasing relevance in IS research: Contextualizing knowledge in networks. *25th European Conference on Information Systems (ECIS), June 5-10*, 2557–2566.

Madzivhandila, T. P., Griffith, G. R., Fleming, E. M., & Nesamvuni, A. E. (2010). Meta-evaluations in government and government institutions: A case study example from the Australian Centre for International Agricultural Research. *2010 Conference (54th), February 10-12, 2010, Adelaide, Australia*.

Marcon, G. (2014). Public value theory in the context of public sector modernization. In J. Guthrie, G. Marcon, S. Russo, & F. Farneti (Eds.), *Public value management, measurement and reporting* (pp. 323–352). Emerald Group.

Mark, M. M. (2008). Building a better evidence for evaluation theory: Beyond general calls to a framework of types of research on evaluation. In N. L. Smith & P. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 111–131). Guilford Press.

Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, *10*(1), 35–57.

Markiewicz, A., & Patrick, I. (2016). *Developing monitoring and evaluation framework*. Sage.

Matsimbe, C., Mbanza, A. A., Gelb, K., Jonas, J. A., & Nhapuala, G. (2018). Use of two-step cluster analysis to understand how emotions affect night shift teenage students. *Journal of Education, Society and Behavioural Science*, *24*(3), 1–15.

Maxwell, J. A. (2012). *Qualitative research design: An interactive approach*. Sage.

Mayne, J. (2014). Issues in enhancing evaluation use. In M. Läubli Loud & J. Mayne (Eds.), *Enhancing evaluation use: Insights from internal evaluation units* (pp. 1–14). Sage.

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173

McCoy, C. (2017). Addressing the technical in socio-technical systems research: A review of the literature. *23rd Americas Conference on Information Systems*. AMCIS 2017, Boston.

McGregor, S. L. T., & Murnane, J. A. (2010). Paradigm, methodology and method: Intellectual integrity in consumer scholarship. *International Journal of Consumer Studies*, *34*(4), 419–427.

Mckim, C. A. (2017). The value of mixed methods research: A mixed methods study. *Journal of Mixed Methods Research*, *11*(2), 202–222.

McLaughlin, J. A., & Jordan, G. B. (2015). Using logic models. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (pp. 62–87). John Wiley & Sons.

Mertens, D. M. (2015). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Sage.

Mertens, D. M. (2018). *Mixed methods design in evaluation* (Vol. 1). Sage.

Mertens, D. M., & Tarsilla, M. (2015). Mixed methods evaluation. In S. N. Hesse-Biber & B. R. Johnson (Eds.), *The Oxford handbook of multimethod and mixed methods research inquiry*. Oxford University Press.

Mertens, D. M., & Wilson, A. T. (2018). *Program evaluation theory and practice: A comprehensive guide*. Guilford Press.

Meynhardt, T. (2009). Public Value Inside: What is Public Value Creation? *International Journal of Public Administration*, *32*(3–4), 192–219.

Meynhardt, T. (2015). Turning a conceptual framework into a scorecard. In J. M. Bryson, B. C. Crosby, & L. Bloomberg (Eds.), *Public value and public administration* (pp. 147–169). Georgetown University Press.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. Sage.

Miller, R. L. (2010). Developing standards for empirical examinations of evaluation theory. *American Journal of Evaluation*, *31*(3), 390–399.

Misuraca, G., Codagnone, C., & Rossel, P. (2013). From practice to theory and back to practice: Reflexivity in measurement and evaluation for evidence-based policy making in the information society. *Government Information Quarterly*, *30*, S68–S82.

Mooi, E., & Sarstedt, M. (2014). Cluster analysis. In M. Sarstedt & E. Mooi (Eds.), *A concise guide to market research: The process, data and methods using IBM SPSS statistics* (pp. 273–324). Springer.

Moon, M. J., Lee, J., & Roh, C. Y. (2014). The evolution of internal IT applications and e-government studies in public administration: Research themes and methods. *Admininstration & Society*, *46*(1), 3–36.

Moore, M. H. (1995). *Creating public value: Strategic management in government*. Harvard University Press.

Moore, M. H. (2013). *Recognizing public value*. Harvard University Press.

Moors, G., Kieruj, N. D., & Vermunt, J. K. (2014). The effect of labelling and numbering of response scales on the likelihood of response bias. *Sociological Methodology*, *44*(1), 369–399.

Morgan, D. L. (2018). Living within blurry boundaries: The value of distinguishing between qualitative and quantitative research. *Journal of Mixed Methods Research*, *12*(3), 268–279.

Morse, J. M., & Niehaus, L. (2016). *Mixed method design: Principles and procedures*. Routledge.

Moser, C. A., & Kalton, G. (2017). *Survey methods in social investigation*. Routledge.

Mueller, B., & Urbach, N. (2017). Understanding the why, what and how of theories in IS research. *Communications of the Association for Information Systems*, *41*(17), 349–388.

Nabatchi, T. (2012). Four frames for understanding public values in administration and governance. *Public Management Research Conference, Syracuse University. June*, 2–4.

Nag, N. S. (2018). Government, governance and good governance. *Indian Journal of Public Administration*, *64*(1), 122–130.

Naidoo, G., Singh, S., & Levine, N. (2011). An overview of Internet developments and their impact on e-government in South Africa. In A. Al Ajeeli & Y. A. L. Al-Bastaki (Eds.), *Handbook of research on e-services in the public sector: E-government strategies and advancements*. Information Science Reference.

Naidoo, V. (2015). Changing conceptions of pubic "management" and public sector reform in South Africa. *International Public Management Review*, *16*(1), 23–42.

Nam, T. (2018). Examining the anti-corruption effect of e-government and the moderating effect of national culture: A cross-country study. *Government Information Quarterly*, *35*(2), 273–282.

Nguyen, T. D., Nguyen, T. M., & Cao, T. H. (2015). Information Systems success: A literature review. In T. K. Dang, R. Wagner, J. Küng, N. Thoai, M. Takizawa, & E. Neuhold (Eds.), *Future data and security engineering* (Vol. 9446, pp. 242–256). Springer International Publishing. https://doi.org/10.1007/978-3-319-26135-5_18

Nicolini, D. (2012). *Practice theory, work, and organization: An introduction*. Oxford University Press.

Nicolini, D. (2017). Practice theory as a package of theory, method and vocabulary: Affordances and limitations. In M. Jonas, B. Littig, & A. Wroblewski (Eds.), *Methodological Reflections on Practice Oriented Theories* (pp. 19–34). Springer.

Nicolini, D., & Monteiro, P. (2017). The practice approach: For a praxeology of organizational and management studies. In A. Langley & H. Tsoukas (Eds.), *The Sage Handbook of process organization*. Sage.

Niiniluoto, I. (2014). Scientific realism, independence, causation and abduction. In K. Westpal (Ed.), *Realism, science and pragmatism* (pp. 159–172). Routledge.

Noble, H., & Smith, J. (2015). Issues of validity and reliability in qualitative research. *Evidence Based Nursing*, *18*(2), 34–35.

Norgbey, E. B. (2016). Debate on the appropriate methods for conducting impact evaluation of programs within development context. *Journal of MultiDisciplinary Evaluation*, *12*(24), 59–66.

Northcote, M. T. (2012). Selecting criteria to evaluate qualitative research. In M. Kelly (Ed.), *Narratives of Transition: Perspectives of Research Leaders, Educators & Postgraduates* (pp. 99–110). The Center for Higher Education, Learning and Teaching. http://research.avondale.edu.au/edu_papers/38

Norusis, M. J. (2012). Cluster Analysis. In M. J. Norusis (Ed.), *IBM SPSS Statistics 19 Statistical Procedures Companion*. Pearson.

Öbrand, L., Augustsson, N. P., Mathiassen, L., & Holmström, J. (2019). The interstitiality of IT risk: An inquiry into Information Systems development practices. *Information Systems Journal*, *29*(2019), 97–118.

OECD. (2005). *E-government for better government* (E. Lau, Ed.). OECD Publishing.

OECD. (2009). *Rethinking e-government services: User-centred approaches*. OECD Publishing.

OECD. (2013). *OECD E-government studies: Egypt*. OECD Publishing.

OECD. (2014). *Recommendation of the Council on Digital Government Strategies*. OECD Publishing.

OECD (Ed.). (2016). *Digital government in Chile: Strengthening the institutional and governance framework*. OECD Publishing.

Ofir, Z., & Kumar, S. (2013). Evaluation in developing countries: What makes the difference? In S. I. Donaldson, T. Azzam, & R. F. Conner (Eds.), *Emerging Practices in International Development Evaluation* (pp. 11–24). Information Age.

Osborne, J. W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment, Research & Evaluation*, *20*(2), 1–7.

Otieno, I., & Omwenga, E. (2015). Towards the development of a citizen-centric framework for evaluating the impact of e-government: A case study of developing countries. In P. Cunningham & M. Cunningham (Eds.), *2015 IST-Africa Conference: Lilongwe, Malawi, 6—8 May 2015* (pp. 1–9). IEEE.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*(5), 533–544.

Palvia, P., & Baqir, N. (2015). ICT policies in Developing Countries: An evaluation with the extended design-actuality gap framework. *The Electronic Journal of Information Systems in Developing Countries*, *71*(1), 1–34.

Palvia, P., Chau, P. Y. K., Kakki, M. D., Ghoshal, T., Uppala, V., & Wang, W. (2017). A decade plus long introspection of research published in Information & Management. *Information & Management*, *54*(2017), 281–227.

Palvia, P., Kakki, M. D., Ghoshal, T., Uppala, V., & Wang, W. (2015). Methodological and topic trends in Information Systems research: A meta-analysis of IS journals. *Communications of the Association for Information Systems*, *37*(30), 630–650.

Pandey, V., & Gupta, S. (2018). A comprehensive four-stage framework for evaluation of Information Communication Technologies for Development interventions. *Information Technology for Development*, *24*(3), 511–531.

Pang, M.-S., Lee, G., & DeLone, W. H. (2014). IT resources, organizational capabilities, and value creation in public-sector organizations: A public-value management perspective. *Journal of Information Technology*, *2014*(29), 187–205.

Pather, S. (2017). Contextualising Information Systems evaluation research: Towards a classification of approaches. *11th European Conference on Information Systems Management*, 252–261.

Patton, M. Q. (2018). Evaluation science. *American Journal of Evaluation*, *39*(2), 183–200.

Pekkola, S., & Päivärinta, T. (2019). Information Systems benefits and value: A literature review. *51st Hawaii International Conference on System Sciences*, 6155.

Phellas, C. N., Bloch, A., & Seale, C. (2011). Structured methods: Interviews, questionnaires and observation. In C. Seale (Ed.), *Researching society and culture* (pp. 182–202). Sage.

Phillips, S., Goldman, I., Gasa, N., Akhalwaya, I., & Leon, B. (2014). A focus on M&E of results: An example from the Presidency, South Africa. *Journal of Development Effectiveness*, *6*(4), 392–406.

Phillips, T., & de Wet, J. P. (2017). Towards rigorous practice: A framework for assessing naturalistic evaluations in the development sector. *Evaluation*, *23*(1), 102–120.

Pietersen, C. (2014). Implementation of Batho Pele principles in an educational district office. *Mediterranean Journal of Social Sciences*, *5*(3), 252–261.

Podems, D., Goldman, I., & Jacob, C. (2014). Evaluator competencies: The South African government experience. *The Canadian Journal of Program Evaluation*, *28*(3), 71–85.

Preskill, H., & Boyle, S. (2008a). Insights into evaluation capacity building: Motivations, strategies, outcomes, and lessons learned. *The Canadian Journal of Program Evaluation*, *23*(3), 147.

Preskill, H., & Boyle, S. (2008b). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, *29*(4), 443–459.

Pullinger, M., Browne, A. L., & Medd, W. (2013). *Patterns of water: The water related practices of households in southern England, and their influence on water consumption and demand management*. Lancaster University. https://www.escholar.manchester.ac.uk/uk-ac-man-scw:187780

Punch, K. F., & Oancea, A. (2014). *Introduction to research methods in education*. Sage.

Radnor, Z., Osborne, Stephen, & Glennon, R. (2016). Public management theory. In C. Ansell & J. Torfing (Eds.), *Handbook of theories of governance* (pp. 46–60). Edward Elgar.

Ramli, R. M. (2017). E-government implementation challenges in Malaysia and South Korea: A comparative study. *Electronic Journal of Information Systems in Developing Countries*, *80*(7), 1–26.

Rea, L. M., & Parker, R. A. (2014). *Designing and conducting survey research: A comprehensive guide*. Wiley brand.

Reckwitz, A. (2002). Toward a theory of social practices: A development in culturalist theorizing. *European Journal of Social Theory*, *5*(2), 243–263.

Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2014). Choosing the Number of Categories in Agree–Disagree Scales. *Sociological Methods & Research*, *43*(1), 73–97.

Reynolds, M., Gates, E., Hummelbrunner, R., Marra, M., & Williams, B. (2016). Towards systemic evaluation. *Systems Research and Behavioral Science*, *2016*(33), 662–673.

Richmond, A. (2015). The public sector as an enabler of national development. In K. Appiah-Adu & M. Bawumia (Eds.), *Key determinants of national development: Historical perspectives and implications for developing economies* (pp. 66–78). Gower.

Rocco, T. S., & Plakhotnik, M. S. (2009). Literature reviews, conceptual frameworks, and theoretical frameworks: Terms, functions, and distinctions. *Human Resource Development Review*, *8*(1), 120–130.

Rog, D. J. (2015). Infusing theory into practice, practice into theory: Small wins and big gains for evaluation. *American Journal of Evaluation*, *36*(2), 223–238.

Rogers, A., & Gullickson, A. (2018). Evaluation champions: A literature review. *Journal of MultiDisciplinary Evaluation*, *14*(30), 46–63.

Rose, J., Flak, L. S., & Sæbø, Ø. (2018). Stakeholder theory for the e-government context: Framing a value-oriented normative core. *Government Information Quarterly*, *35*(2018), 362–374.

Rose, S., Spinks, N., & Canhoto, A. I. (2015). *Management research: Applying the principles*. Routledge.

Rowe, F. (2014). What literature review is not: Diversity, boundaries and recommendations. *European Journal of Information Systems*, *23*(3), 241–255.

Rowley, J. (2012). Conducting research interviews. *Management Research Review*, *35*(3/4), 260–271.

Rowley, J. (2014a). Designing and using research questionnaires. *Management Research Review*, *37*(3), 308–330.

Rowley, J. (2014b). Designing and using research questionnaires. *Management Research Review*, *37*(3), 308–330. https://doi.org/10.1108/MRR-02-2013-0027

Rubin, H. J., & Rubin, I. (2012). *Qualitative interviewing: The art of hearing data*. Sage.

Rutgers, M. R. (2015). As good as It gets? On the meaning of public value in the study of policy and management. *The American Review of Public Administration*, *45*(1), 29–45.

Ryan, M. (2015). Framing student evaluations of university learning and teaching: Discursive strategies and textual outcomes. *Assessment & Evaluation in Higher Education*, *40*(8), 1142–1158.

Sabani, A., Deng, H., & Thai, V. (2018). *A conceptual framework for the adoption of e-government in Indonesia*. Australian Conference on Information Systems, Sydney, Australia.

Sabani, A., Deng, H., & Thai, V. (2019). Evaluating the Performance of E-Government in Indonesia: A Thematic Analysis. *Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance - ICEGOV2019*, 435–440.

Saldaña, J. (2015). *The coding manual for qualitative researchers*. Sage.

Sandberg, J., & Tsoukas, H. (2016). Practice theory: What is it, its philosophical base, and what it offers organization studies. In R. Mir, H. Willmott, & M. Greenwood (Eds.), *The Routledge Companion to philosophy in organization studies* (pp. 184–198). Routledge.

Santarek, K., & Obluska, I. (2012). Process approach to the evalaution of Information Systems effectiveness. *Information Systems in Management*, *1*(2), 148–159.

Sarker, S., Xiao, X., & Beaulieu, T. (2013). Qualitative studies in Information Systems: A critical review and some guiding principles. *MIS Quarterly*, *37*(4), iii–xviii.

Saunders, A., & Brynjolfsson, E. (2016). Valuing Information Technology related intangible assets. *MIS Quarterly*, *40*(1), 83–110.

Saunders, M. (2011). Setting the scene: The four domains of evaluative practice in higher education. In M. Saunders, P. Trowler, & V. Bamber (Eds.), *Reconceptualising evaluation in higher education: The practice turn* (pp. 1–17). McGraw-Hill.

Saunders, M. (2012). The use and usability of evaluation outputs: A social practice approach. *Evaluation*, *18*(4), 421–436.

Saunders, M., Lewis, P., & Thornhill, A. (2015). *Research methods for business students*. Pearson.

Savoldelli, A., Codagnone, C., & Misuraca, G. (2014). Understanding the e-government paradox: Learning from the literature and practice on barriers to adoption. *Government Information Quarterly*, *31*(2014), 563–571.

Savoldelli, A., Misuraca, G., & Codagnone, C. (2013). Measuring the public value of e-government: The eGEP2.0 model. *Electronic Journal of E-Government*, *11*(1), 373–388.

Schatzki, T. R. (2005). Peripheral vison: The sites of organizations. *Organization Studies*, *26*(3), 465–484.

Schatzki, T. R. (2012). A Primer on practices: Theory and research. In J. Higgs, R. Barnett, S. Billet, M. Hutchings, & F. Trede (Eds.), *Practice-Based Education: Perspectives and Strategies* (pp. 13–26). Sense.

Schatzki, T. R. (2013). The edge of change. In E. Shove & N. Spurling (Eds.), *Sustainable practices: Social theory and climate change* (pp. 31–46). Routledge.

Schatzki, T. R. (2015). Practices, governance and sustainability. In Y. Strenger & C. Maller (Eds.), *Social practices, intervention and sustainability: Beyond behaviour change* (pp. 15–30). Routledge.

Schatzki, T. R. (2017a). Multiplicity in social theory and practice ontology. In M. Jonas & B. Littig (Eds.), *Praxeological politcal analysis* (pp. 31–46). Routledge.

Schatzki, T. R. (2017b). Practices and people. *TPA- Teoria e Prática Em Adminstração*, *7*(1), 26–53.

Schatzki, T. R. (2019). *Social change in a material world*. Routledge.

Schoonenboom, J., & Johnson, B. R. (2017). How to construct a mixed methods research design. *Köln Z Soziol*, *69*(2), 109–131.

Schreier, M. (2012). *Qualitative content analysis in practice*. Sage.

Schryen, G. (2013). Revisiting IS business value: What we already know, what we still need to know, and how to get there. *European Journal of Information Systems*, *2013*(22), 139–169.

Schwandt, T. A. (2015). *Evaluation foundations revisited: Cultivating a life of the mind for practice*. Stanford University Press.

Schwandt, T. A. (2018). Evaluative thinking as a collaborative social practice: The case of boundary judgment making. *New Directions for Evaluation*, *2018*(158), 15–137.

Scott, M., DeLone, W. H., & Golden, W. (2016). Measuring eGovernment success: A public value approach. *European Journal of Information Systems*, *25*(3), 187–208.

Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill-building approach*. John Wiley& Sons.

Shafritz, J. M., Russell, E. W., & Borick, C. (2015). *Introducing public administration*. Routledge.

Silverman, D. (2013). *Doing qualitative research: A practical handbook*. Sage.

Simonofski, A., Snoeck, M., Vanderose, B., Crompvoets, J., & Habra, N. H. (2017). Reexamining e-participation: Systematic literature review of citizen participation in E-government service delivery. *23rd Americas Conference on Information Systems*. AMCIS 2017, Boston.

Sinclair-Maragh, G., Gursoy, D., & Vieregge, M. (2015). Residents' perceptions towards tourism development: A factor-cluster approach. *Journal of Destination Marketing and Management, 4*(2015), 36–45.

Sitwala, I. (2014). Is there a conceptual difference between theoretical and conceptual frameworks. *Journal of Social Sciences, 38*(2), 185–195.

Smith, N. L. (1993). Improving evaluation theory through the empirical study of evaluation practice. *Evaluation Practice, 14*(3), 237–242.

Smith, N. L., & Brandon, P. R. (Eds.). (2008). *Fundamental issues in evaluation*. Guilford Press.

Song, X., & Letch, N. (2012). Research on IT/IS Evaluation: A 25 year review. *The Electronic Journal Information Systems Evaluation*, *15*(3), 276–287.

Song, X., & Letch, N. (2014). Understanding the utilization of Information Systems evaluation: Four case studies. *PACIS 2014 Proceedings, 321*. http://aisel.aisnet.org/pacis2014/321

Sousa, D. (2014). Validation in Qualitative Research: General Aspects and Specificities of the Descriptive Phenomenological Method. *Qualitative Research in Psychology*, *11*(2), 211–227.

Spaargaren, G., Lamers, M., & Weenink, D. (2016). Introduction: Using social practice to research social life. In G. Spaargaren, D. Weenink, & M. Lamers (Eds.), *Practice theory and research: Exploring the dynamics of social life* (pp. 3–27). Routledge.

Stanimirovic, D., & Vintar, M. (2013). Conceptualization of an integrated indicator model for the evaluation of e-government policies. *Electronic Journal of E-Government*, *11*(2), 293–307.

Stern, E., & Saunders, M. (2015). Standing back and looking forward: Editor's reflections on the 20th Anniversary of Evaluation. *Evaluation*, *21*(4), 380–390.

Sterrenberg, G. (2017). A conceptual framework for evaluating e-government systems success: A service ecosystem approach. *50th Hawaii International Conference on Systems Science*, 2529–2538.

Stockdale, R., Irani, Z., & Love, P. E. D. (2008). Revisiting the content, context and process of IS evaluation. In Z. Irani & P. E. D. Love (Eds.), *Evaluating Information Systems- Public and private Sector*. Elsevier/Butterworth-Heinemann.

Stockdale, R., & Standing, C. (2006). An interpretive approach to evaluating information systems: A content, context, process framework. *European Journal of Operations Research*, *173*(2006), 1090–1102.

Sue, V. M., & Ritter, L. R. (2012). *Conducting online surveys*. Sage.

Suhardi, S., Sofia, A., & Andriyanto, A. (2015). Evaluating e-government and good governance correlation. *Journal of ICT Research and Applications*, *9*(3), 236–262.

Sundberg, L., & Larsson, A. (2017). The impact of formal decision processes on e-government projects. *Administrative Science*, *7*(14), 1–12.

Szanyi, M., Azzam, T., & Galen, M. (2013). Research on evaluation: A needs assessment. *Canadian Journal of Program Evaluation*, *27*(1), 39–64.

Tallon, P. P. (2014). Do you see what I see? The search for consensus among executives' perceptions of IT business value. *European Journal of Information Systems*, *2014*(23), 306–325.

Tavakoli, A., & Schlagwein, D. (2016). A review of the use of practice theory in Information Systems research. *PACIS 2016 Proceedings*, *267*. http://aisel.aisnet.org/pacis2016/267

Tavakoli, A., Schlagwein, D., & Schoder, D. (2017). Open strategy: Literature review, re-analysis of cases and conceptualization of practice. *Journal of Strategic Information Systems*, *267*(2017), 163–184.

Thota, N., Berglund, A., & Clear, T. (2012). Illustration of paradigm pluralism in computing education research. In M. de Raadt & A. Carbone (Eds.), *14th Australasian Computing Education Conference* (Vol. 123, pp. 103–112). Australian Computer Society, Inc.

Toffel, M. (2016). Enhancing the practical relevance of research. *Production and Operations Management*, *25*(9).

Töhönen, H., Itälä, T., Kauppinen, M., & Männistö, T. (2015). Towards systemic evaluation of the business value of IT. *Fifth International Symposium on Business Modelling and Software Design (BSMD 2015)*, 163–170.

Turner, A. H. (2015). Instilling public service values and professionalism through Information Literacy. *Journal of Public Affairs Education*, *21*(1), 41–54.

Turner, M., Hulme, D., & McCourt, W. (2015). *Governance, management and development: Making the state work*. Palgrave.

Twining, P., Heller, R. S., Nussbaum, M., & Tsai, C. C. (2017). Some guidance on conducting and reporting qualitative studies. *Computers & Education*. https://doi.org/10.1016/j.compedu.2016.12.002

UN. (2014). *United Nations e-government survey 2014: E-government for the future we want*. United Nations.

UN. (2016). *United Nations e-government survey 2016: E-government in support of sustainable development*. United Nations.

UN. (2018). *United Nations e-government survey 2018: Gearing e-government to support sustainable and resilent societies*. United Nations.

UN. (2020). *United Nations e-government survey 2020: Digital government in the decade of action for sustainable sevelopment*. United Nations.

Uprichard, E. (2013). Sampling: Bridging probability and non-probability designs. *International Journal of Social Research Methodology*, *16*(1), 1–11.

van der Wal, Z., Nabatchi, T., & de Graaf, G. (2015). From galaxies to universe: A cross-disciplinary review and analysis of public values publications from 1969 to 2012. *The American Review of Public Administration*, *45*(1), 13–28.

Vandenabeele, W., Leisink, P., & Knies, E. (2013). Public value creation and strategic human resource management: Public service motivation as a linking mechanism. In P. Leisink, P. Boselie, M. van Bottenburg, & D. M. Hosking (Eds.), *Managing social issues: A public values perspective* (pp. 37–54). Edward Elgar.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in Information Systems. *MIS Quarterly*, *37*(1), 21–54.

Venkatesh, V., Brown, S. A., & Sullivan, Y. W. (2016). Guidelines for conducting mixed-methods research: An extension and illustration. *Journal of the Association for Information Systems*, *17*(7), 435–495.

Verdinelli, S., & Scagnoli, N. I. (2013). Data display in qualitative research. *International Journal of Qualitative Methods*, *2013*(12), 359–381.

Vo, A. T., & Christie, C. A. (2015). Advancing research on evaluation through the study of context. *New Directions for Evaluation*, *2015*(148), 43–55.

Wade, J., & Kallemeyn, L. (2020). Evaluation capacity building (ECB) interventions and the development of sustainable evaluation practice: An exploratory study. *Evaluation and Program Planning*, *79*(101777), 1–9. https://doi.org/10.1016/j.evalprogplan.2019.101777

Wahid, F. (2013). Themes of research on e-government in developing countries: Current map and future roadmap. *46th Hawaii International Conference on Systems Science*. HICSS 2013.

Wallace, M., & Sheldon, N. (2015). Business research ethics: Participant observer perspectives. *Journal of Business Ethics*, *128*(2), 267–277.

Walliman, N. (2018). *Research methods: The basics*. Routledge.

Warde, A. (2014). After taste: Culture, consumption and theories of practice. *Journal of Consumer Culture*, *14*(3), 279–303.

Weerakkody, V., Irani, Z., Lee, H., Osman, I., & Hindi, N. (2015). E-government implementation: A bird's eye view of issues relating to costs, opportunities, benefits and risks. *Information Systems Frontiers*, *17*(4), 889–915.

Welch, D. (2016). Social practices and behaviour change. In F. Spotswood (Ed.), *Beyond behaviour change: Key issues, interdisciplinary approaches and future directions* (pp. 237–255). Policy Press.

Welch, D. (2017). Behaviour change and theories of practice: Contributions, limitations and developments. *Social Business*, *7*(3–4), 241–261.

Welch, E. W., & Feeney, M. K. (2014). Technology in government: How organizational culture mediates information and communication technology outcomes. *Government Information Quarterly*, *31*(4), 506–512.

Welch, J., Rimes, H., & Bozeman, B. (2015). Public value mapping. In J. M. Bryson, B. C. Crosby, & L. Bloomberg (Eds.), *Public value and public administration* (pp. 1–24). Georgetown University Press.

Wholey, J. S. (2015). Use of evaluation in government: The politics of evaluation. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (pp. 798–815). John Wiley & Sons.

Wilson, J. (2014). *Essentials of business research: A guide to doing your research project*. Sage.

Witesman, E. (2016). *From public values to public value and back again*. http://cord.asu.edu/wp-content/uploads/2015/02/2015123001-Public-value-to-public-values-and-back-for-PVC1.pdf

Wu, P. F. (2012). A mixed mthods approach to Technology Acceptance research. *Journal of the Association for Information Systems*, *12*(4), 299–322.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94.

Zafarullah, H., & Sarker, A. E. (2016). Public management reforms in developing countries: Towards a synthesis. In N. Ahmed (Ed.), *Public policy and governance in Bangladesh: Forty years of experience*. Routledge.

Zahran, I., Al-Nuaim, H., Rutter, M., & Benyon, D. (2015). A critical analysis of e-government evaluation models at national and local municipal levels. *Electronic Journal of E-Government*, *13*(1), 28–42.

Zhang, M., & Gable, G. G. (2014). *Seeking better IT value conceptualization: Some meta-theoretical considerations*. http://eprints.qut.au/79288/

Zhang, P. (2013). The Affective Response Model: A theoretical framework of affective concepts and their relationship in the ICT context. *MIS Quarterly*, *37*(1), 247–274.

Zhang, Y., & Wildemuth, B. M. (2016). Qualitative analysis of content. In B. M. Wildemuth (Ed.), *Applications of social research methods to questions in information and library science* (Vol. 318, pp. 318–329).

Appendices

Appendix A: Cover letter

Department of Information Systems
Leslie Commerce Building Engineering Mall,
Upper Campus
Or Private Bag X3
Rondebosch, 7701
Tel: +27650-2261
Fax No: (021) 650-2280
23 November 2016

To whom it may concern

Invitation to participate in research

Research purpose: I am conducting research for PhD thesis on e-government outcomes evaluation practice (i.e., *how e-government outcomes evaluation is done and the results used*). This may help improve the use of e-government in public service delivery.

Participation: Your participation is appreciated. You are requested to complete the *Consent form* and *Questionnaire*. This should take about 25 minutes. You may also be requested to participate in an interview at a later stage.

Participants' rights and safety: This research has been approved by the Commerce Faculty Ethics in Research Committee. Participation is voluntary and you can choose to withdraw at any time. No adverse consequences shall arise from your participation. Your rights to dignity, anonymity and confidentiality shall be upheld. Data collected shall be handled and, on completion of the research, disposed of safely and securely. Publications shall not make direct reference to you or your organization.

Incentive: Recommendations shall be drawn for improving how evaluation of e-government outcomes is done and the results used. Participants may request for copies.

Enquiries: For further enquiries, please contact the researcher by email (details below).

Yours sincerely

C Boamah-Abu (Researcher)
Charles.Boamah@uct.ac.za

Prof Michael Kyobe (Supervisor)
Michael.Kyobe@uct.ac.za

Appendix B: Consent form

Department of Information Systems
Leslie Commerce Building Engineering Mall
Upper Campus
Or Private Bag X3
Rondebosch, 7701
Tel: +27650-2261
Fax No: (021) 650-2280
23 November 2016

Research title: A sequential explanatory mixed methods investigation of e-government outcomes evaluation practice in developing countries: South Africa as context.

Purpose: To understand e-government outcomes evaluation practice (i.e., *how e-government outcomes evaluation is done and the results used*).

Consent: By signing this form, you agree to participate in this research. You may decide to withdraw your participation at any stage of the research.

Signature _____

Date: _____

Appendix C: Questionnaire

The purpose is to collect data on your perceptions of e-government outcomes evaluation practice (i.e., *how e-government outcomes evaluation is done and the results used*). Before you proceed, please read the *Cover letter* and sign the *Consent form*.

Part I: Background: *In each of the following, please select the option which best describes your background*

1. Current position: ☐ Management ☐ Non-management

2. Professional background: ☐ IT ☐ Evaluation ☐ Other

3. Experience in e-government evaluation: ☐ None ☐ 1 to 5 years ☐ More than 5 years

4. Knowledge of e-government evaluation: ☐ Minimum ☐ Average ☐ Above average

5. Evaluations you have participated in (*in last 2 years*): ☐ None ☐ 1 ☐ More than 1

Part II: Perceptions of most recent experiences in evaluation of e-government outcomes

In each of the following, please select the option which best describes your experiences

6. Evaluation initiated by public organization responsible for e-government ☐ Yes ☐ No

7. Evaluation conducted by: ☐ Internal staff only ☐ External staff only ☐ Both

8. Strategies used (**select one or more**): ☐ Experiment ☐ Survey
☐ Case study ☐ Focus group
☐ Other:_____

9. What is evaluated (**select one or more**):

- ☐ Organizational performance (e.g., improved processes, etc.)
- ☐ Participation (e.g., in policy processes by citizens, businesses, etc.)
- ☐ Accountability (e.g., reporting on resource allocation to government and public)
- ☐ Transparency (e.g., access to timely and accurate information by the public)
- ☐ User satisfaction (e.g., improved access to services by citizens and businesses)
- ☐ Other (**specify**)

10. The main purpose of evaluation (**select one or more**)

- ☐ To comply with public sector regulations
- ☐ To learn to improve e-government
- ☐ To learn to improve how e-government outcomes evaluation is done
- ☐ To support organizational decision-making (e. g., planning and budgeting)
- ☐ Other (**specify**)

11. Indicate to what extent you agree with each of the following statement:

1 = Strongly disagree; 2 = Disagree; 3 = Neither agree nor disagree;

4 = Agree; 5 = Strongly agree

		1	2	3	4	5
1	Management supports evaluation	1	2	3	4	5
2	The National Evaluation Policy Framework guides evaluation	1	2	3	4	5
3	Outcomes are evaluated regularly	1	2	3	4	5
4	How evaluation is done is reviewed regularly	1	2	3	4	5
5	Evaluation budget is adequate	1	2	3	4	5
6	Evaluation expertise is adequate	1	2	3	4	5
7	Evaluation tools (e.g., software) are adequate	1	2	3	4	5
8	Batho Pele principles guide how evaluation is done	1	2	3	4	5
9	Evaluation models (*utilization, empowerment, etc.*) guide evaluation	1	2	3	4	5
10	All the main stakeholder groups are represented	1	2	3	4	5
11	Selection of stakeholders is transparent	1	2	3	4	5
12	Stakeholders are trained on how evaluation is done	1	2	3	4	5
13	Stakeholders agree on what to evaluate	1	2	3	4	5
14	Stakeholders agree on how to use results	1	2	3	4	5
15	All stakeholders participate in decision making throughout the evaluation	1	2	3	4	5
16	Reliable indicators for measuring e-government benefits are available	1	2	3	4	5
17	The indicators measure both financial and non-financial benefits	1	2	3	4	5
18	Appropriate data on e-government is available	1	2	3	4	5
19	Evaluation reports are distributed to all stakeholder groups	1	2	3	4	5
20	Evaluation reports clearly state recommendations on how to use results	1	2	3	4	5
21	Evaluation reports specify recommendations for each stakeholder group	1	2	3	4	5
22	Stakeholder participation improves acceptance of recommendations	1	2	3	4	5
23	Management produces strategy on how to implement recommendations	1	2	3	4	5
24	There is follow up to implement recommendations	1	2	3	4	5
25	Participation improves stakeholders' knowledge on evaluation	1	2	3	4	5
26	Participation increases stakeholders' commitment to implement recommendations	1	2	3	4	5
27	Participation increases stakeholders' willingness to participate in future evaluations	1	2	3	4	5

****End of Questionnaire****

Appendix D: Interview protocol and initial coding frame

Appendix D1: Interview protocol

Date: Time:	Venue:	Participant:
Comments:		

1. Let's start with your **background**.
   - Can you please tell me about your background in e-government outcomes evaluation?
   - **Professional background**?
   - **Current position** (*Management/Non-management*)?
   - **Years of experience**
   - How would you rate your **knowledge:** Below average, Average, Above average?

2. Now let's discuss **organizational capacity**.
   - How would you describe the following in your organization?
   - **Expertise** for e-government outcomes evaluation?
   - **Budget** for e-government outcomes evaluation?
   - **Management support** for e-government outcomes evaluation?

3. Let's look at the influence of **policies** and **values**.
   - To what extent is e-government outcomes evaluation influenced by:
   - **Batho Pele** principles?
   - **National Evaluation Policy Framework**?
   - Can you describe the **values** (i.e., *outcomes,* etc.) that influence how e-government outcomes evaluation is done?

4. Let's focus on **participation** in e-government outcomes evaluation.
   - Describe the extent of your own participation in:
   - **Decision making**
   - **Activities**
   - Would you like to participate in e-government evaluation in the **future**? *Yes/No?* Why?

5. Now let's look at **roles** in evaluation. How would you describe:
   - Your understanding of **your role** and responsibilities?
   - Other participants' understanding of **their roles** and responsibilities?

6. Let's talk about the **use of evaluation results**. Can you please tell me:
   - How **management** respond to evaluation report?
   - How evaluation reports are **distributed**?
   - To what extent are recommendations **implemented**?

7. Lastly, let's focus on the **effect** of e-government outcomes evaluation (i.e., *changes in perceptions*). Can you describe the effect of:
   - Participation on **you** as an individual?
   - Conducting evaluation on **your organization**?

Appendix D2: Initial coding scheme

Question		Code (**Values**)
1	Participant's background	Professional background (**IT/Evaluation/Other**)
		Position (**Management/Non-management**)
		Experience **(<=5 years/> 5 years**)
		Knowledge(**Minimum/Average/Above average**)
2	Organizational capacity	Expertise (**Adequate/Inadequate**)
		Budget (**Dedicated/Not; Adequate/Inadequate**)
		Management support (**Adequate/Inadequate**)
3	Policies and values	Batho Pele (**Influence/No influence**)
		NEPF (National Evaluation Policy Framework) (**Influence/No Influence**)
		Values (**Efficiency/Effectiveness/Democratic**)
4	Participation	Decision-making (**Somewhat/Throughout**)
		Activities (**Somewhat/Throughout**)
		Future (Yes/No); Motivation (**Open-ended**)
5	Roles and duties	Participant's understanding (**Understand/Don't understand**)
		Other's understanding (**Understand/Don't understand**)
6	Use of report	Management response (**Yes? Process?/No**)
		Distribution (**Yes? Process/No**)
		Implementation of recommendations (**Implemented/Not implemented**)
7	Effect	Individual (**Type**)
		Organizational (**Type**)

Appendix E: Additional quantitative results

E1: Descriptive statistics

Table E1.1: One-Sample Kolmogorov-Smirnov Test (N=106)

		Most Extreme Differences				
	Item	Absolute	Positive	Negative	Test statistic	Asymp. Sig. (2-tailed)
1	Management support	.241	.134	-.241	.241	.000c
2	National framework	.186	.179	-.186	.186	.000c
3	Regular evaluation	.175	.148	-.175	.175	.000c
4	Regular eval review	.221	.221	-.157	.221	.000c
5	Adequate budget	.178	.178	-.150	.178	.000c
6	Adequate expertise	174	.172	-.174	.174	.000c
7	Adequate tools	.219	.219	-.168	.219	.000c
8	Batho Pele	.215	.125	-.215	215	.000c
9	Evaluation models	.266	.149	-.266	.266	.000c
10	Stakeholder representation	.227	.128	-.227	.227	.000c
11	Stakeholder selection	.267	.176	-.267	.267	.000c
12	Stakeholder training	.202	.154	-.202	202	.000c
13	Agreement (evaluand)	.191	.157	-.191	.191	.000c
14	Agreement (use)	.188	.167	-.188	.188	.000c
15	Stakeholder participation	.233	.135	-.233	.233	.000c
16	Evaluation indicators	.179	.179	-.169	.179	.000c
17	Financial/non financial	.256	.122	-.256	.256	.000c
18	Appropriate tools	.200	.200	-.152	.200	.000c
19	Reports (distribution)	.187	.187	-.176	.187	.000c
20	Reports (recommendations)	.223	.140	-.223	.223	.000c
21	Reports (to diff groups)	.207	.120	-.207	207	.000c
22	Management response	.202	.174	-.202	.202	.000c
23	Implementation	.211	.211	-.147	.211	.000c
24	Improves knowledge	.271	.139	-.271	.271	.000c
25	Improves acceptance	.254	.143	-.254	.254	.000c
26	Increases commitment	.198	.162	-.198	.198	.000c
27	Future participation	.199	.148	-.199	.199	.000c

c. Lilliefors Significance Correction

E2: Explanatory factor analysis

Table E2.1: Communalities (Initial solution)

	Initial	Extraction
Management support	.605	.610
National framework	.625	.507
Regular evaluation	.726	.757
Regular eval review	.582	.441
Adequate budget	.535	.409
Adequate expertise	.492	.388
Adequate tools	.537	.385
Batho Pele	.669	.603
Evaluation models	.780	.661
Stakeholder representation	.550	.406
Stakeholder selection	.549	.497
Stakeholder training	.514	.405
Agreement (evaluand)	.739	.769
Agreement (use)	.718	.719
Stakeholder participation	.405	.420
Evaluation indicators	.467	.498
Financial/non financial	.702	.590
Data quality	.573	.380
Reports (distribution)	.698	.633
Reports (recommendations)	.697	.586
Reports (to diff groups)	.651	.587
Management response	.664	.744
Follow up	.582	.471
Knowledge	.635	.559
Acceptance (results)	.644	.547
Commitment to use	.627	.485
Future participation	.596	.567

E3: Propositions and decisions

Table E3.1 Position and Experience (Mann-Whitney Test)

	Proposition	Sig.	Decision
1	The distribution of Framework is the same across categories of Position.	.00	Reject the null hypothesis.
2	The distribution of Results is the same across categories of Position.	.01	Reject the null hypothesis.
3	The distribution of Capacity is the same across categories of Position.	.01	Reject the null hypothesis.
4	The distribution of Understanding is the same across categories of Position.	.02	Reject the null hypothesis.
5	The distribution of Participation is the same across categories of Position.	.03	Reject the null hypothesis.
6	The distribution of Issues is the same across categories of Position.	.39	Retain the null hypothesis.
7	The distribution of Framework is the same across categories of Experience.	.00	Reject the null hypothesis.
8	The distribution of Results is the same across categories of Experience.	.00	Reject the null hypothesis.
9	The distribution of Capacity is the same across categories of Experience.	.00	Reject the null hypothesis.
10	The distribution of Understanding is the same across categories of Experience.	.01	Reject the null hypothesis.
11	The distribution of Participation is the same across categories of Experience.	.00	Reject the null hypothesis.
12	The distribution of Issues is the same across categories of Experience.	.05	Retain the null hypothesis.

Asymptotic significances are displayed. p=.50

Table E3.2. Background and Knowledge (Kruskal-Wallis Test)

	Proposition	Sig.	Decision
1	The distribution of Framework is the same across categories of Background.	.34	Retain the null hypothesis.
2	The distribution of Results is the same across categories of Background.	.00	Reject the null hypothesis.
3	The distribution of Capacity is the same across categories of Background.	.00	Reject the null hypothesis.
4	The distribution of Understanding is the same across categories of Background.	.05	Reject the null hypothesis.
5	The distribution of Participation is the same across categories of Background.	.01	Reject the null hypothesis.
6	The distribution of Issues is the same across categories of Background.	.02	Reject the null hypothesis.
7	The distribution of Framework is the same across categories of Knowledge.	.00	Reject the null hypothesis.
8	The distribution of Results is the same across categories of Knowledge.	.01	Reject the null hypothesis.
9	The distribution of Capacity is the same across categories of Knowledge.	.00	Reject the null hypothesis.
10	The distribution of Understanding is the same across categories of Knowledge.	.00	Reject the null hypothesis.
11	The distribution of Participation is the same across categories of Knowledge.	.00	Reject the null hypothesis.
12	The distribution of Issues is the same across categories of Knowledge.	.36	Retain the null hypothesis.

Asymptotic significances are displayed. p=.50

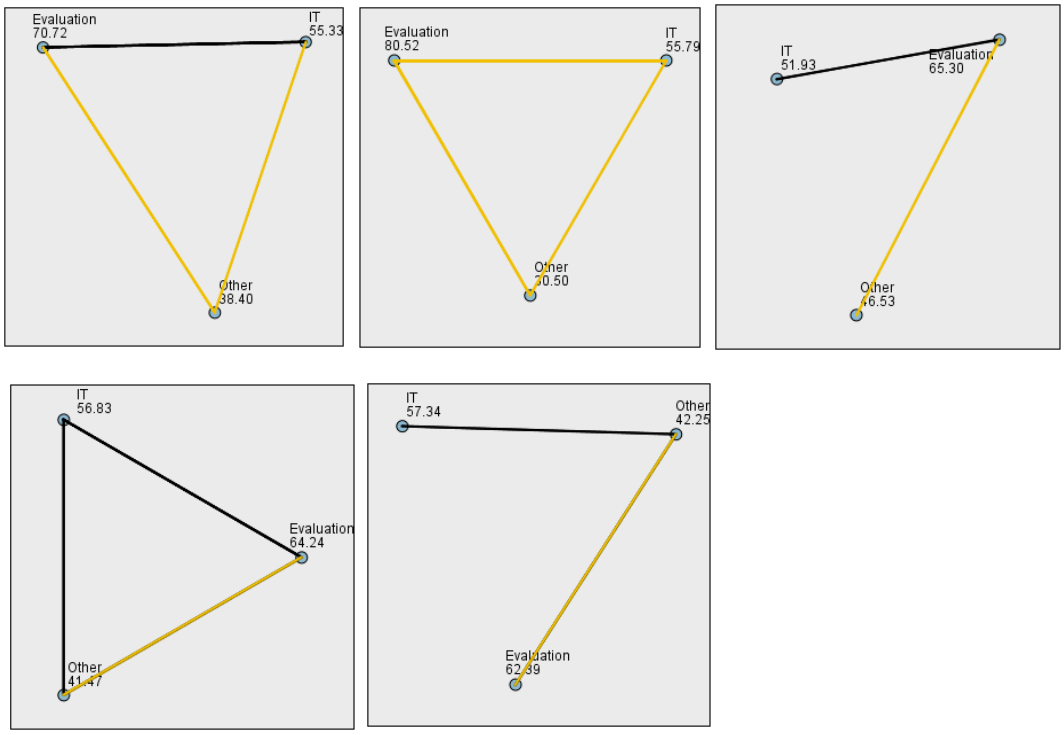


Figure E3.1: Differences by background on Factors 2 to 6 (Left-Right; Top-Down)

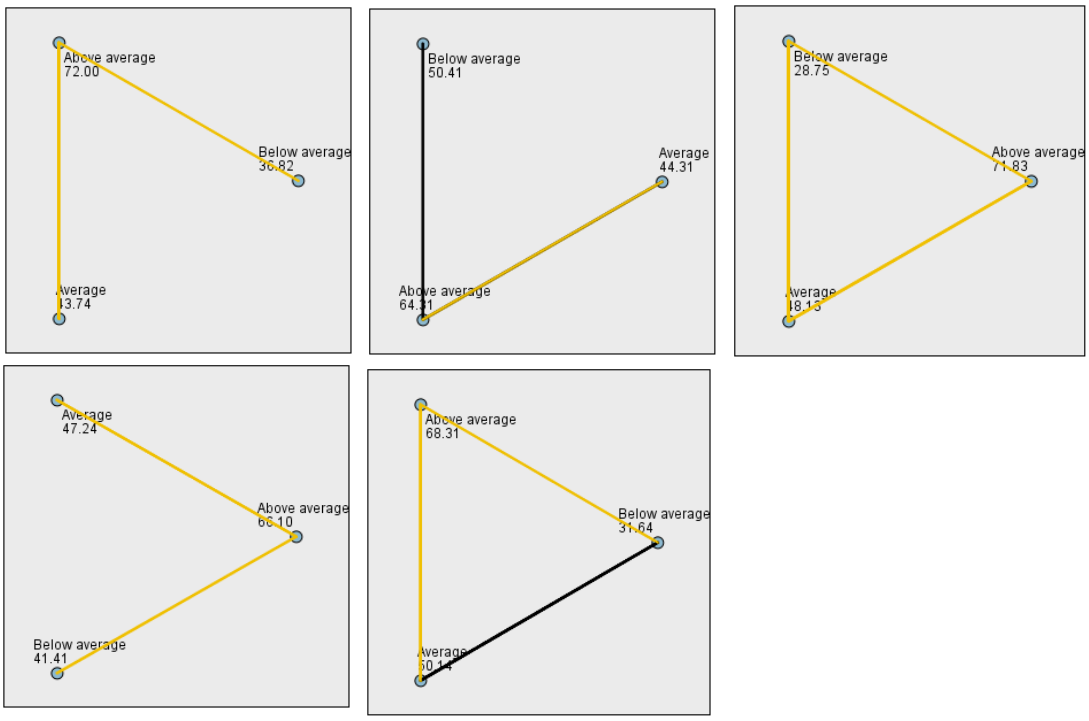


Figure E3.2: Differences by knowledge on Factors 1 to 5 (Left-right; top-down)

Appendix F: Ethics approval

UNIVERSITY OF CAPE TOWN
FACULTY OF COMMERCE
Igniting Knowledge and Opportunity

Ethics Approval Request for the Study entitled:

E-government outcomes evaluation practices in developing countries: South Africa as context

Signed :

	Full name and signature	Date
Principal Researcher/Student:	Charles Boamah-Abu	20/11/16

This application is approved by:

Supervisor	Prof Michael Kyobe	18-Nov 2016
Co- Supervisor		

Approved.
Prof U. Rivett 4.01.2017
Chair
Ethics in Research Committee
Commerce Faculty
University of Cape Town

Com Ethics_V4