THE INFLUENCE OF TRANSACTIONAL DATA QUALITY ON MONETARY SAVINGS OF ESKOM DISTRIBUTION, FREE STATE.

By

CHARL JOHANNES BESTER

MTech in Business Administration

In the Faculty of Management Sciences

CENTRAL UNIVERSITY OF TECHNOLOGY, FREE STATE

SUPERVISOR

PROF PATIENT RAMBE

SEPTEMBER 2018

DECLARATION

I, Charl Johannes Bester, Student Number: , declare that this research paper I submit to the Central University of Technology, Free State for M-TECH: Business Administration is my own independent work and has not previously been submitted by me at another university.

21 Sept 2018

SIGNATURE OF STUDENT DATE

ACKNOWLEDGEMENTS

I could not have completed this dissertation successfully without support from various parties to whom I extend my appreciation. Praise to God for showering me with the ability, people and resources to conclude this research. To my wife, children and family, a big thank you for your continued love and support throughout this endeavour. My dear parents, my deepest appreciation for your unconditional love and for the life skills you taught me. Prof Rambe, without your guidance, patience, support and continued encouragement, I would not have been able to complete my research. Lastly, I extend my appreciation to Eskom for allowing me the time, resources and opportunity to complete this study.

ABSTRACT

Eskom, South Africa's major electricity generation, transmission and distribution utility, is currently in a very difficult financial position. Electricity sales are declining, debt and primary energy costs are soaring, existing customers owe Eskom billions in debt while tariff increases on electricity sales are too low to recover costs for the generation, transmission and distribution of electricity. Consequently, Eskom is examining all its operations in order to identify areas where possible savings can be realised.

Given the centrality of data in both informing prudent corporate decision making and advancing cost saving mechanisms, it is undeniable that inaccurate data can lead to inappropriate decisions and costly corporate blunders. There is sufficient evidence to demonstrate that an improvement in data quality can increase a company's turnover by approximately 15%. Nevertheless, the reality is that data quality improvement strategies tend to focus on master data. As a result, the researcher sought to establish the exact effects that improvements in transactional data quality could have on monetary savings of Eskom Distribution Free State.

Drawing on the aforementioned electricity utility, a survey was conducted on technical field staff's perspectives regarding transactional data quality of customer calls relating to electricity supply problems (ESP) received from its call centre. In addition to the survey, historical transactional data on the entity's ESP customer calls were also analysed to establish the influence of data quality on cost savings of this entity. The survey was conducted on 303 Eskom technicians during 2017. The historical data sets for the period April 2012 to March 2017 were also analysed. Since the assessments on monetary impact of the mentioned transactions are carried by Eskom rather than the customer, the perceptions of customers were not considered in this study. It was contended that the individuals directly involved in assessing the monetary effects of data quality would be ideally positioned to have logical and credible opinions on this subject rather than customers who were considered to have limited knowledge on this subject.

The results from the historical data analysis using mean, frequency distribution, cross tabulation, correlation analysis of survey data, and mean distribution, regression analysis and correlation analysis for historical data, revealed potential monetary savings of 17.18% arising from avoidable costs on transactions related to ESP customer calls. These monetary savings were dependent on Eskom's ability to increase its transactional data quality on ESP customer calls from 81.31% to 100%. While it was acknowledged that avoidable costs could only be calculated from quantifiable operational costs, savings would potentially increase if the effects of improved customer service, faster supply restoration times and work hours saved to perform preventative maintenance to reduce overall fault volumes were quantified in monetary terms. It was also noted that if the costs of increasing data quality were lower than the 17.18% monetary savings potential established in the study, then such data quality improvement strategies would improve Eskom's financial position. Furthermore, descriptive analysis on survey results revealed that an improvement in customer call transactional data quality at the source has the potential of creating savings of up to 47.7% for transactions related to customer calls requesting service for an ESP. This finding was however not supported by inferential analysis. Nonetheless, the study recommends that Eskom should continually identify and investigate high value transactional data quality as it offers significant savings potential through cost avoidance.

TABLE OF CONTENTS

DECLARATION ii

ACKNOWLEDGEMENTS iii

ABSTRACT iv

TABLE OF CONTENTS vi

LIST OF ACRONYMS AND ABBREVIATIONS xi

LIST OF FIGURES xii

LIST OF TABLES xiii

1 CHAPTER 1: ORIENTATION OF THE STUDY 1

1.1 Introduction 2

1.2 Problem background 3

1.3 Problem statement 4

1.4 Research aim 6

1.4.1 Research objectives 6

1.4.2 Research questions 7

1.5 Research methodology 8

1.6 Research ethics 9

1.7 Significance of the study 9

1.8 Proposed conceptual framework 10

1.9 Outline of the dissertation 12

1.10 Chapter summary 12

2 CHAPTER 2: CONCEPTUALISATION AND OPERATIONALISATION OF TRANSACTIONAL DATA QUALITY AND COST SAVINGS 13

2.1 Introduction 14

2.2 Transactional data quality 14

2.2.1 Defining data and data categories 14

2.2.2 Definition of data quality and transactional data quality 15
2.2.3 Perspectives on transactional data quality... 15
2.2.4 Data quality improvement methodologies/strategies 21
2.2.5 Global state of data quality 23
2.2.6 Data quality in South Africa 24
2.2.7 Transactional data within Eskom 25
2.2.8 Master and transactional data quality within Eskom Distribution 26
2.2.9 Impact and cost of poor data quality.. 27
2.2.10 Impact of poor transactional data quality on Eskom Distribution 29
2.3 Feedback 30
2.3.1 Defining feedback.. 30
2.3.2 Perspectives on feedback.. 30
2.3.3 Application of feedback in organisations.. 31
2.3.4 Feedback on transactional data quality within Eskom Distribution 32
2.4 Costs.. 32
2.4.1 Defining costs 32
2.4.2 Types of costs 33
2.4.3 Importance of costs 34
2.4.4 Cost calculation 35
2.4.5 Eskom Distribution costs on ESP faults.. 36
2.5 Monetary savings 37
2.5.1 Defining monetary savings 37
2.5.2 Monetary savings through cost reduction 37
2.5.3 Eskom's monetary saving strategies 39
2.6 Chapter summary.. 39
3 CHAPTER 3: METHODOLOGY 41
3.1 Introduction 42

3.2 Research paradigm 42

3.3 Research design 42

3.4 Research approach 43

3.5 Population 43

3.6 Sampling 44

3.7 Data collection 44

3.8 Data preparation 45

3.9 Data analysis 46

3.10 Validity and reliability 48

3.10.1 Validity 48

3.10.2 Reliability 49

3.11 Ethical considerations 50

3.12 Chapter summary 50

4 CHAPTER 4: PRESENTATION OF FINDINGS 52

4.1 Introduction 53

4.2 Survey results 53

4.2.1 Overview of descriptive data and analysis of survey results 53

4.2.2 Inferential statistics of survey results 62

4.2.3 Summary of survey results 68

4.3 Results on historical data 68

4.3.1 Overview of data summary 68

4.3.2 Comparison of transactional data quality: source system with feedback 71

4.3.3 Relationship between costs and transactional feedback received and transactional data quality: source system 73

4.3.4 Compare cost impact of incorrect transactions to transactional data quality: feedback received and transactional data quality: source system 74

4.3.5 Relationship between monetary saving, total cost, transactional data quality: source system and transactional data quality: feedback received 77

4.3.6 Summary of historical dataset 78

4.4 Chapter summary 79

5 CHAPTER 5: DISCUSSION OF FINDINGS 80

5.1 Introduction 81

5.2 Objectives and outcomes 81

5.2.1 Quality of customer call transactional data captured at Eskom Contact Centres from source system data quality measurements and feedback 81

5.2.2 Costs related to ESP transactions at Eskom Distribution Free State 83

5.2.3 Impact transactional data quality at source system has on feedback received for transactions 84

5.2.4 Ascertain whether transactional data quality affects transaction costs 85

5.2.5 Influence of the calculated costs on monetary cost savings. 86

5.2.6 Influence of transactional data quality of customer calls on monetary cost savings 86

5.3 Chapter summary 88

6 CHAPTER 6: CONCLUSION AN RECOMMENDATIONS 90

6.1 Introduction 91

6.2 Recapping research objectives and key findings 91

6.3 Recommendations 93

6.3.1 Recommendations for policy 93

6.3.2 Recommendations for practice 94

6.3.3 Implications for further research 95

6.4 Limitations of this study 95

6.4.1 Costs for transactions 95

6.4.2 Validity of the feedback indicator 95

6.5 Concluding remarks 96

7 REFERENCES... 97

8 APPENDICES... 107

8.1 APPENDIX A – Questionnaire 107

8.2 APPENDIX B – Permission to conduct research....................................... 109

8.3 APPENDIX C – Extension of permission to conduct research 110

8.4 APPENDIX D – TURNITIN report for plagiarism 112

8.5 APPENDIX E – Confirmation of language editing 113

LIST OF ACRONYMS AND ABBREVIATIONS

BPP	Business Productivity Programme
CDQM	Complete Data Quality Methodology
CNC	Customer Network Centre
DTC	Design to Cost
DQ	Data Quality
EDA	Enterprise Digitial Assistant
ESP	Electricity Supply Problem
FSOU	Free State Operating Unit
GPS	Global Positioning System
IQ	Information Quality
KPI	Key Performance Indicator
MDM	Master Data Management
NDP	National Development Plan
NERSA	National Energy Regulator of South Africa
OU	Operating Unit
QA	Quality Assurance
SAIDI	System Average Interruption Duration Index
SAP	Systems Applications and Products in data processing
SPSS	Statistical Package for the Social Sciences
SQL	Sequential Query Language

LIST OF FIGURES

Figure 1.1: ESP capturing, execution and transaction data quality measurement areas 5
Figure 1.2: Conceptual framework: The relationship between transactional data quality and monetary savings 12
Figure 2.1: Job order cost sheet example 36
Figure 4.1: Distribution of years of experience at Eskom CNC 54
Figure 4.2: Distribution of years of EDA usage 54
Figure 4.3: Distribution of single customer dispatch work orders received from dispatch per month 55
Figure 4.4: Source system and feedback based data quality per financial year 72
Figure 4.5: Transactional data quality: source system and transactional data quality: feedback received compared to total cost impact of feedback indicating incorrect transactions 75

LIST OF TABLES

Table 2.1: Popular data quality dimensions 16
Table 2.2: Eskom applying CARAT principle 17
Table 2.3: Root cause of poor quality data 20
Table 2.4: Complete Data Quality Methodology (CDQM) steps 23
Table 2.5: Perceptions on DQ 24
Table 2.6: Data quality improvement benefits categories 28
Table 3.1: Cronbach's alpha coefficients on questionnaire 50
Table 4.1: Background variables 55
Table 4.2: Source system data quality 56
Table 4.3: Feedback analysis 57
Table 4.4: Correct data transactions during normal work time/overtime 58
Table 4.5: Incorrect data transactions during normal work time/overtime 58
Table 4.6: Total travel distance and additional costs incurred 59
Table 4.7: Usage of feedback to determine cost 60
Table 4.8: Monetary savings underlying the prevention of incorrect transactional data quality 60
Table 4.9: Monetary cost contribution of work orders with correct and incorrect data 61
Table 4.10: Correlations between Source system data quality and feedback 63
Table 4.11: Correlation between source system data quality and costs: Travel 63
Table 4.12: Correlation between source system data quality and costs: Percentage work orders 64
Table 4.13: Correlation between feedback and travel distance 64
Table 4.14: Correlation between feedback and ESP, customer side faults during work hours and overtime 65
Table 4.15: Correlation between costs on kilometres travelled and monetary savings 66
Table 4.16: Correlation between percentage work order costs and monetary savings 66
Table 4.17: Correlation between source system data quality and monetary savings 67
Table 4.18: Correlation between feedback and monetary savings 67
Table 4.19: Total work order volume distribution 70

Table 4.20: Source system transactional data quality vs feedback on correct transactions per financial year 72
Table 4.21: Logistic regression of transactional data quality on transactional feedback 73
Table 4.22: Regression of travel cost on transactional data quality and transactional feedback 74
Table 4.23: Regression of total time cost on transactional feedback and transaction data quality 74
Table 4.24 : Transactional data quality: source system and transactional data quality: feedback received compared to total cost impact of transactions with feedback indicating incorrect transactions 76
Table 4.25: Correlation of monetary savings with transactional data quality: source system, transactional data quality: feedback received and total transaction cost 77
Table 4.26: Regression of monetary saving with transactional data quality: source system, transactional data quality: feedback received and total cost 78
Table 5.1: Literature on data quality compared to historical dataset and survey results 83

1 CHAPTER 1: ORIENTATION OF THE STUDY

1.1 Introduction

Most organisations, irrespective of their size, depend on data to conduct their daily business activities and utilise data as a key asset to create and sustain competitive advantage over their competitors. Similarly, global electricity utilities and organisations that generate, transmit and distribute electricity, utilise data in various formats and from different sources as their enablers to generate and sell electricity to their customers. In the United States alone, there are 3 200 utilities competing for a share of the electrical sales market worth around $400 billion (Martin, Chediak & Wells, 2013). In the South African context, the sole public electricity utility, Eskom Holding SOC Ltd, hereinafter referred to as Eskom, sold electricity for a total value of R175 billion (around $12.1 billion US) in the 2017-2018 financial year (Eskom Holdings SOC Ltd, 2018a). Eskom is a complex organisation that comprises an electricity business, investments and subsidiaries. The electricity business has 3 core units and these are:

- Eskom Generation: Responsible for the generation of electricity.
- Eskom Transmission: Responsible for the transmission of generated electricity over long distances to distribution sites.
- Eskom Distribution: Responsible for the distribution and sales of electricity to the bulk resellers and end consumers (Eskom Holdings SOC Ltd, 2018a).

Despite the exceptional operational performance of its electricity business concerning generation plant availability and network performance during the 2017-2018 financial year, Eskom is experiencing severe financial challenges attributed to multiple causes. Some of these challenges are, as articulated in Eskom's 2018 Integrated Report (2018a), a consequence of:

- Irregular expenditure and senior executive financial mismanagement.
- Increased debt related to the building of new generation capacity.
- Increased costs of repayment of debt due to credit rating downgrades.
- An effective 2.2% price increase from the National Energy Regulator of South Africa (NERSA), which was too low to recover costs for electricity generation, transmission and distribution.
- Decreased electricity sales.
- An increase in primary energy costs.

- The existence of a Municipal arrear debt of R13.6 billion.

Regardless of Eskom's financial predicament, the South African government as the sole shareholder still expects the corporation to fulfil its strategic objectives as highlighted in the National Development Plan (NDP) (National Planning Commission, 2012). These strategic objectives are encapsulated in Eskom's mandate of "providing a stable electricity supply in a sustainable and efficient manner, in order to assist in lowering the cost of doing business in South Africa and enabling economic growth" (Eskom Holdings SOC Ltd, 2018a:6).

Eskom is scrutinising its operations for opportunities to realise savings and fulfil its mandate of improving its internal efficiency and remaining financially viable. Existing literature highlights that an increase in data quality has the potential to increase revenue (Batini & Scannapieco, 2016; Experian Data Quality, 2017). Nevertheless, while existing strategies of data quality improvement have a strong focus on the improvement of the quality of master data and its business benefits, less attention is devoted to the quality of transactional data, due to its volatile nature and perceived lower business value (Entity group, 2016). This study considers the aforementioned literature on the significance of data quality in the operations of corporations and cost reductions in promoting efficient business operations in the investigation of the influence of transactional data quality on monetary savings of Eskom Distribution Free State in South Africa, which is one of the nine Distribution Operating Units (OU) of Eskom's electricity business. Eskom Distribution's OU is located in one of South Africa's nine provinces, Eastern Cape, Free State, Gauteng, Kwazulu Natal, Limpopo, Northern Cape, North West, Mpumalanga and Western Cape. The choice of this electrical utility is informed by the reality that this institution is confronted with deteriorating liquidity and profitability challenges that threaten its ability to remain viable in the long run (Eskom Holdings SOC Ltd, 2018a).

1.2 Problem background

When a business is confronted with financial difficulties, it has to implement measures to increase its income and lower expenses by reducing unnecessary spending and laying off excess staff (Sadgrove, 2015). Although it is an inexorably complex business operation endeavour, obtaining correct data concerning business operations and interpreting it appropriately is critical to implementing best financial decisions during times of economic downturn or recession (Bate, 2009). The same applies to Eskom, which due to its dire financial predicament, was compelled to implement various initiatives such as the Business Productivity Programme (BPP), Design to Cost (DTC) strategy and the prioritisation of capital expenditure within all sectors of the organisation. However, these strategies did not sufficiently curb operational costs and Eskom has recently started to review its organisational design. The engagements with senior executives regarding possible dismissals at their specific level have started. The objective of all this and other initiatives is to ensure that the corporation remain financially viable by reducing costs and improving liquidity (Eskom Holdings SOC Ltd, 2018a; Eskom Holdings SOC Ltd, 2018c).

As one of the top 20 electrical utilities in the world, Eskom supplies approximately 95% of South Africa's electricity and exports 45% of the electricity to many parts of Africa (Topco Media, 2014; Eskom Holdings SOC Ltd, 2015b). Electricity access and pricing play a major role in supporting industrialisation, economic growth and the overall improvements in people's living standards (Stern, 2010; Hu, 2013). While Africa has the lowest electricity supply in the world, its economies are mainly dependent on electricity (McDonald, 2009) thus, making electricity supply and access, sources of contention in communities and the broader society. Therefore, it is critical that Eskom remains economically viable and is able to supply electricity at affordable prices to support and grow the economies of South Africa and Africa.

1.3 Problem statement

Companies throughout the world face many difficulties in quantifying the costs of poor quality data due to tangible and intangible components that need to be considered in the calculation process (Wang et al., 2015). Although only operational costs can be reliably calculated (Redman, 2013a), which is an understatement of total costs, it is advisable to determine the totality of costs whenever high value data is concerned. The savings potential ingrained in the improvement of data quality can be quite significant as poor quality data has been estimated to cost companies up to 20% of their revenue (Wang et al., 2015). Whilst considerable emphasis has been placed on the quality of master data within the methodology and application of Master Data Management (MDM) (Loshen, 2010), transactional data quality issues are often ignored due to its high volume and inherent volatility (Entity group, 2016).

A typical area uncovered by the researcher offers saving potential in faults caused by customers (rather than faulty Eskom equipment) that Eskom Distribution technical staff had to attend to. The origin of these faults is from customers who contact the Eskom contact centre to report an electricity supply problem (ESP). The Eskom contact centre agent (i.e. data creator) would verify the validity of the request, categorise the fault as an ESP and transactional data is captured in the contact centre system (i.e. source system) to invoke a process to fix the ESP. The process involves the creation of a workorder (based on the transaction data captured) in a downstream system, which triggers a technician (i.e. the data consumer) to execute the workorder by responding to the call. Such a response happens when the technician travels to the customer to fix the ESP. The financial cost of such workorders is solely carried by Eskom. Unfortunately, mistakes or misinformation causes incorrect categorisation of these calls and results in unwarranted cost implications on Eskom Distribution when manpower and transport are dispatched to fix a problem that is not the companies' responsibility. Fortunately, a feedback mechanism exists for Eskom Distribution technical staff as data consumers to indicate a problem they experience with a transaction, by marking such a transaction as a customer side fault.

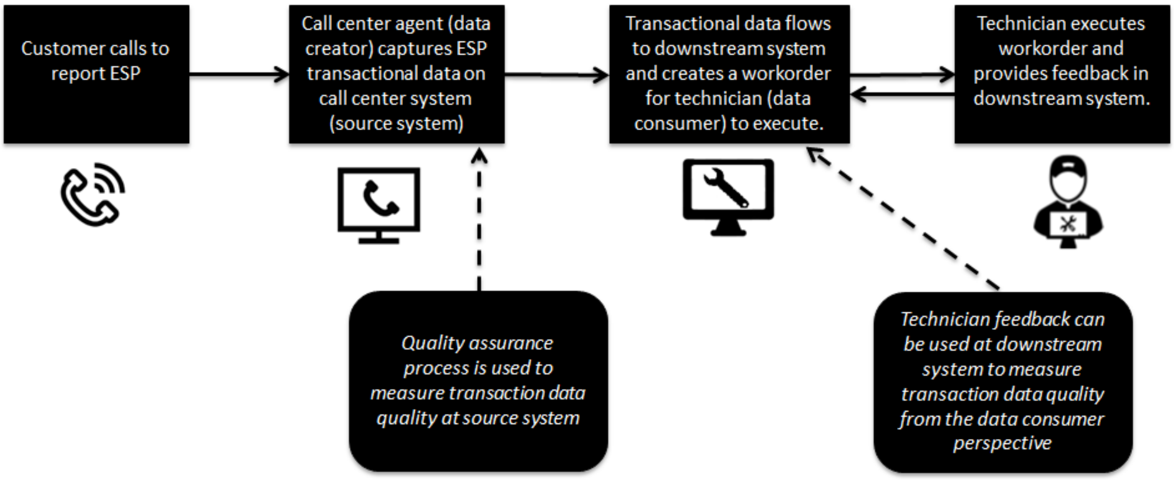


Figure 1.1: ESP capturing, execution and transaction data quality measurement areas

The problem, therefore, is the lack of knowledge on the exact impact of transactional data quality embodied in customer calls on the financial position (especially monetary savings) of Eskom Distribution. This stems from the fact that a detailed quantification of operational costs arising from problematic transactional data from customer calls has not yet been conducted.

1.4 Research aim

The aim of the study is to contribute to organisational efficiency, and resource and data management literature, through a broad understanding of how transactional data quality contributes to financial cost saving in a cash squeezed public electricity utility (Alrayes, 2015; Batini & Scannapieco, 2016). Based on the researcher's knowledge, no empirical research has been conducted in South Africa on the effects of improvements in transactional data quality on monetary savings.

1.4.1 Research objectives

The following objectives were formulated to fulfil the aim of the study:
The main research objective is to determine the influence of transactional data quality of customer calls on the monetary savings of Eskom Distribution Free State (see detailed conceptual framework in Figure 1.2).

The secondary objectives of this study are to:

1. Establish the quality of customer call transactional data captured at Eskom Contact Centres [V] from source system data quality measurements.
2. Determine customer call transactional data quality based on technical field staff's feedback [X] on transactions.
3. Identify the costs related to ESP transactions at Eskom Distribution Free State [Y].
4. Determine the impact transactional data quality at the source system (Contact Centre) has on feedback produced by technical field staff for transactions executed based on source system data (that is [V] → [X]).

5. Analyse how transactional data quality from the source system (that is [V]) and feedback received from technical field staff (that is [X]), impact transaction costs (that is [Y]).
6. Establish the influence of the calculated costs [Y] on the monetary savings [Z].

1.4.2 Research questions

1.4.2.1 Main question

1. What is the influence of transactional data quality of customer calls (measured from source system and field technicians' feedback on transactions executed based on source system data) on monetary cost savings of Eskom Distribution Free State (that is [V] → [Z] and [X] → [Z])

1.4.2.2 Sub questions

1.1. What is the quality of transactional data captured from customer calls at Eskom Contact Centres based on source system measurement?

1.2. How many downstream system transactions have field technicians' feedback which indicates incorrect transactional data?

1.3. What are the costs, which impact on ESP transactions at Eskom Distribution Free State?

1.4. How does data quality measured at the source system influence feedback on transactions by field technicians (that is [V] →[X])?

1.5. How do data quality measured at the source system and feedback received from field technicians affect costs on transactions (that is [V] → [Y] and [X] → [Y])?

1.6. Overall, how do the calculated costs impact on monetary cost savings (that is [Y] → [Z])?

1.5 Research methodology

A positivist paradigm was adopted for the execution of this research and as a result, a survey was considered as the appropriate research design. Even though the study examines a specific case involving one organisation in order to gain broad knowledge, the actual research approach and process of data collection process was based on a survey design. The survey was designed based on Eskom Distribution Free State, which is one of nine operating units in Eskom Distribution that all have similar standardised operating procedures. A quantitative approach was used to determine:

- Transactional data quality measured at the source.
- Transactional data quality from the receiver's (field technicians) perspective.
- Costs of labour hours and the kilometres travelled.

Data for this study was collected from two sources. The first source, a historical dataset of all ESP transactions, was extracted by means of sequential query language (SQL) from 1 April 2012 to 31 March 2017. Data for the second source was gathered by means of a questionnaire which was administered during 2017 to the entire population of 303 technical staff members operating at the Eskom Distribution Free state. Both datasets were coded and captured into Microsoft Excel and exported to the Statistical Package for Social Sciences (SPSS) for detailed analysis. Both descriptive and inferential statistics were used to analyse the data. The specific descriptive statistical tools employed were mean, standard deviation and cross tabulation/contingency table analysis, whilst the inferential tools focussed on correlation and regression analysis.

The validity and reliability of data had to be ensured to maintain the credibility of the research. Face validity was ensured by carrying out a careful scrutinisation of the questionnaire items to ensure their correct wording for the target audience, and that each question was related to the objectives of this study. Content validity was ensured by making an in depth literature review the basis for the design of the questionnaire. It was also ensured by requesting the supervisor and the statistician to scrutinise the questionnaire to establish if the breadth of the concepts had been covered sufficiently.

Historical data validity was ensured by adhering to business rules during data extraction and excluding incomplete records that could not be normalised and thus prevent the generation of skewed results. The reliability of the secondary data was ensured through the performance of multiple extractions for the same time period. Record counts and completeness were checked and compared to ensure that a reliable dataset had been extracted. Finally, the reliability of the questionnaire was confirmed by adhering to external and internal consistency procedures such as:

- A representative population.
- Standardised testing conditions.
- Consistent scoring.
- Low test difficulty.
- Calculated Cronbach's alpha with a generally acceptable but moderate overall internal consistency of 0.638.

1.6 Research ethics

Clearance was received to access and analyse data from the Eskom's Talent and Skills Management department. Survey participants were informed of the project goals, and that participation was voluntary and their anonymity guaranteed. The sensitive environment in which Eskom operates in compelled the researcher to aggregate historical results on financial information, technicians and customers, just as the anonymity of customers as well as technician information was assured. Finally, Data was stored on the Eskom network to ensure that it was safe and yet accessible to the researcher and statistician.

1.7 Significance of the study

This research contributes to an increased knowledge of how transaction data quality can yield an effective resource utilisation and monetary savings in the South African electrical Eskom Holdings SOC Ltd. It highlights the importance of quantifying the quality of transaction data from the data creator and the data consumer's perspective in order to generate a balanced view of data quality. A feedback mechanism is identified as a critical element in determining data quality from data consumers' perspective.

Insights gained from this study can assist Eskom Holdings SOC Ltd management to identify untapped areas of transaction datasets which offer potential for monetary savings. For business in general, it highlights the importance of not only focussing on master data quality to sustain or increase profits, but the need to also scrutinise transaction data quality in order to ensure that resource wastage does not occur during transaction execution. In addition, the provision of a feedback mechanism for data consumers and the closure of the feedback loop to data creators can assist in identifying transaction data issues, determine their costs and increase the ability to prevent future occurrences of data quality challenges.

1.8 Proposed conceptual framework

According to Cargan (2007:29), a conceptual framework "provides a clear concept of the areas in which meaningful relationships are likely to exist" and "works in conjunction with [the researcher's] goals to justify the study". Therefore, the subsequent description and the illustration in Figure 1.2 present three conditions and the five relationships, which constitute the gist of this study:

- Condition 1: Quality of transactional data is assessed by scrutinising quality measures at the source system [V] from a data creator's (contact centre) perspective.
- Condition 2: Quantifying feedback [X] will provide a perspective on transactional data quality but from a data consumer (field technician) perspective.
- Condition 3: Costs[Y] will be determined by analysing the amount of kilometres travelled and work hours consumed that cause costs to be incurred during transactions.

- Relationship 1 – [V] → [X]: The one directional relationship between quality of transactional data of the source system [V] and quality of transactional data from feedback [X] will provide valuable insight into how [V] impacts [X].
- Relationship 2 – [X] → [Y]: A one directional relationship between quality of transactional data from feedback [X] and costs[Y] incurred will determine how many transactions had feedback on customer faults and what the costs of these transactions are.
- Relationship 3 – [V] → [Y]: Quality of transactional data of the source system [V] one directional relationship with costs [Y] will uncover how monetary cost calculation by means of identified costs is impacted by transactional data quality measurements at the source.
- Relationship 4 – [Y] → [Z]: Costs [Y] will have a one directional relationship to monetary savings [Z] as the calculated monetary costs from identified costs will determine the amount of monetary savings.
- Relationship 5 – [V] → [Z]: The one directional relationship between the quality of transactional data from the source system [V] and monetary savings [Z] will be investigated to determine the impact transactional data quality [V] has on monetary savings [Z].
- Relationship 6 – [X] → [Z]: The one directional relationship between quality of transactional data from the Feedback [X] and monetary savings [Z] will be investigated to determine the effect feedback on monetary savings [Z].

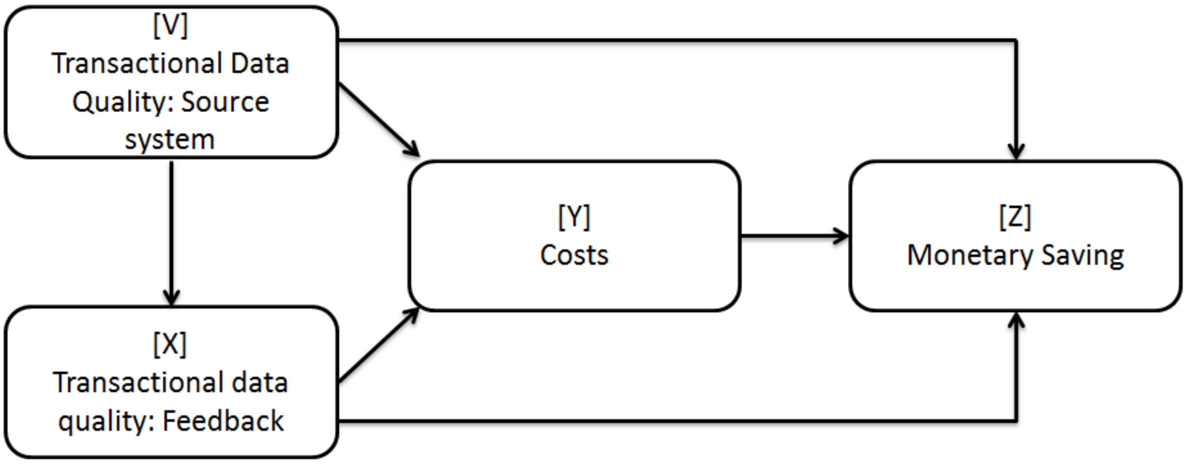


Figure 1.2: Conceptual framework: The relationship between transactional data quality and monetary savings

1.9 Outline of the dissertation

This dissertation comprises six chapters and explores the influence of transactional data quality on the monetary savings of Eskom Distribution Free State. Below is a short description of each chapter:

Chapter 1 outlines an orientation to the study. It explains the background, problem statement and research aim, as well as outline of the research methodology and limitations of the study.

Chapter 2 reviews the literature on transactional data quality, feedback, costs, monetary savings and their application in Eskom.

Chapter 3 outlines the methodology employed to undertake this research.

Chapter 4 presents the research findings based on the data analysis conducted.

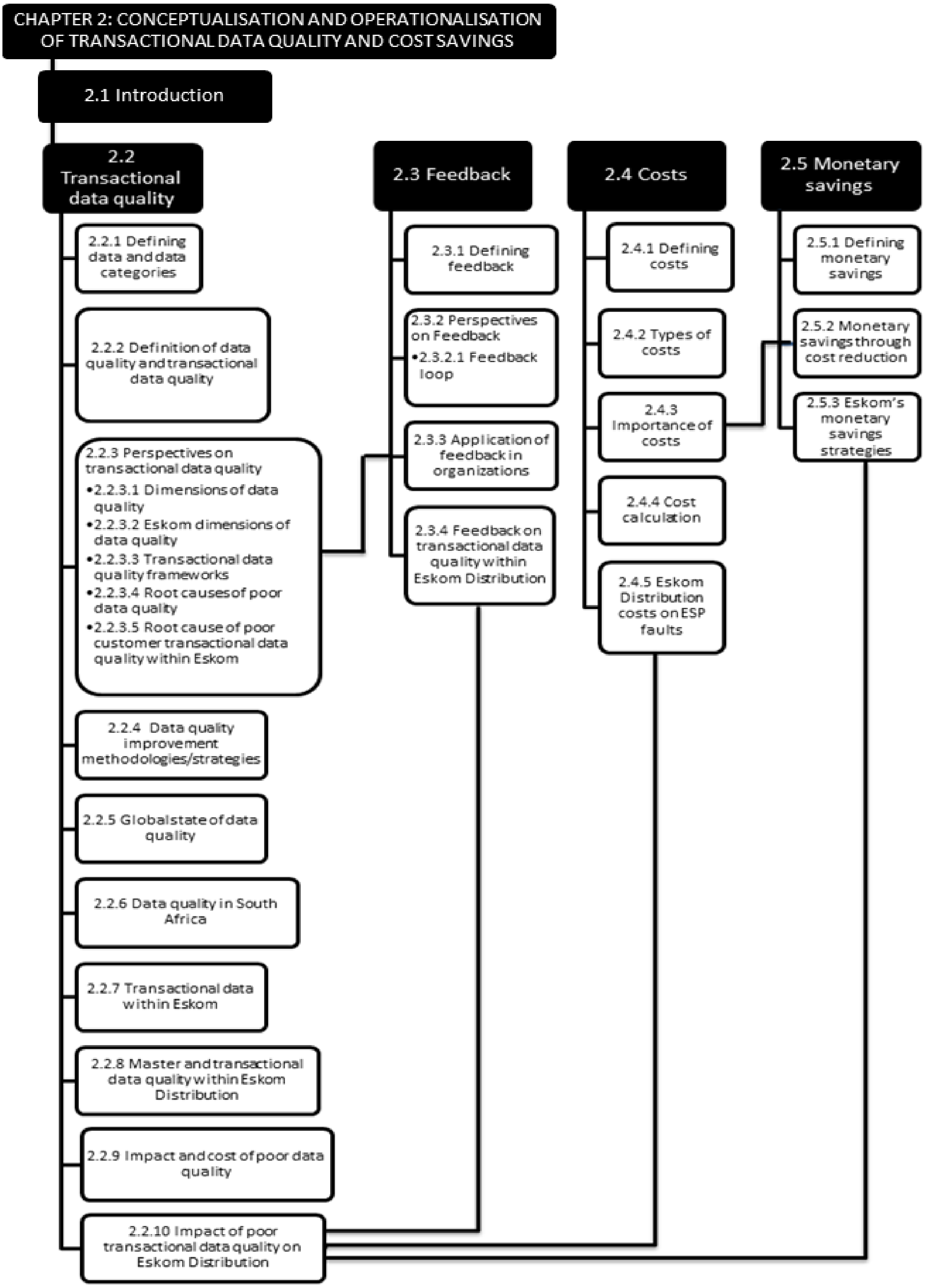
Chapter 5 discusses and interprets the research findings.

Chapter 6 offers the conclusion and recommendations from this study.

1.10 Chapter summary

This orientation of the study commenced with an introduction, followed by a research background and an unpacking of the problem. The chapter also outlined the research aim, formulation of objectives and questions, conceptual framework, the methodology and the study limitations. The next chapter reviews the literature on the four variables highlighted in the conceptual framework.

2 CHAPTER 2: CONCEPTUALISATION AND OPERATIONALISATION OF TRANSACTIONAL DATA QUALITY AND COST SAVINGS



2.1 Introduction

The previous chapter rendered the motivation for conducting this study and outlined the main variables the study is anchored in. The constitution and the relationships of these concepts are illustrated in Figure 1.2, the conceptual framework. This chapter reviews current literature focusing on the four variables highlighted in the conceptual framework namely, transactional data quality (predictive variable), feedback (predictive variable), costs (mediating variable) and monetary savings (outcome variable), which form the foundation for this study.

2.2 Transactional data quality

Transactional data quality is a composite concept that integrates transaction and data quality. A full comprehension of this term can be achieved through referring to data and data quality, concepts that require definition. As such, the term data is foundational to unpacking the construct, transaction data quality.

2.2.1 Defining data and data categories

Data is defined as input into a computer programme/computer-based application, of unprocessed items such as text, images, numbers, video and audio, whereas information is processed/organised data (i.e. output) that has a coherent meaning in a specific context. Although data and information are often problematically used interchangeably in existing literature (Epstein, 2012; Vermaat, 2014), the aforementioned systems theory-based distinction implies that these terms are not synonymous. Data is referred to as an important but unique business asset that can be replicated and shared easily and cheaply (Redman, 2013b; McCafferty, 2016).

Data can be categorised into master, transactional, reference, historical, temporary and metadata. Master data refers to vital business information about products, suppliers and customers, which normally has a low change frequency. Transactional data, the gist of this study, is that data, which is captured during an interaction that is combined with master and/or reference data to form a transaction at a specific time. Transactional data changes frequently and is hence, highly varied. Reference data serves to define a business entity such as a customer, product or supplier and does not change frequently. Historical data is data that relates to previous transactions, which includes master, reference and transactional data and is retained for compliance purposes (Loshen, 2010; Borek et al., 2014).

2.2.2 Definition of data quality and transactional data quality

A generic definition related to overall data quality (DQ) and information quality (IQ) will be employed as it applies to all data types, including transactional data quality. Although there is no precise universally acceptable definition of data quality, Fürber (2015:21) states that DQ is defined as "the degree to which data fulfils requirements" set by "(1) individuals or groups of individuals, (2) standards, (3) laws and other regulatory requirements, (4) by business policies, or (5) even by expectations of data processing applications". As such, the fulfilment of institutional and global specifications and conformity to set standards is integral to the quality of data. Similarly, Redman (2013a:18) renders a comprehensive informative definition of DQ as "exactly the right data and information in exactly the right place at the right time in the right format to complete an operation, serve a customer, make decisions or set and execute strategy".

While Redman's definition seems to conflate information and data, it is conceptually rich to the extent it demonstrates that location, time, service delivered and data form are collectively fundamental to the determination of the acceptability of data quality. Data quality is also articulated via determined levels of data. Here, DQ is normally rated on 2 levels, where level 1 refers to high quality/good data and level 2 to low quality/bad/dirty data (Moskwa, 2015). An application of both definitions to transactional data quality means that data quality will be high when it meets the expectations of its stakeholders and fulfils certain expected requirements. Furthermore, data quality will be defined as low if it fails to conform to certain expectations and requirements.

2.2.3 Perspectives on transactional data quality

A thorough comprehension of data quality dimensions is necessary in order to achieve a solid grasp of transactional data quality. The quality dimensions form a critical component within data quality frameworks. In addition, the frameworks and the root causes of poor transactional data as well as data quality improvement methodologies provide valuable insights into how data quality issues can be identified and corrected to mitigate its debilitating effects on business operations and performance. The subsequent sections of this study explore each facet in greater detail.

2.2.3.1 Dimensions of Data quality

Data quality dimensions are attributes, criteria and facets of quality that allow a business to specify, measure and quantify the quality of its data. Numerous data dimensions exist and these lend themselves to measure a specific aspect of data quality. Some of the popular dimensions related to data values and presentation are depicted in Table 2.1. A certain amount of resources and effort is required to facilitate the appropriate measurement of each dimension. As a result, it is important to understand each dimension sufficiently in order to ensure that it is measurable, applicable to the data being measured, and worth the expenditure of resources required to bring about the assessment (Wang & Strong, 1996; McGilvray, 2008; Cai & Zhu, 2015).

Table 2.1: Popular data quality dimensions

Dimension	Description	Measurement example
Uniqueness	Only one instance of a data entity exists, thereby preventing duplication.	Duplicate analysis testing.
Accuracy	How accurate does data represent the object it models.	Comparison to an existing source of correct information.
Consistency	Different contexts for consistency exist. For example, data in two or more data sets must be consistent with one another or data formatting must be the same for all records in the database.	Assessment on: Items across multiple datasets or on formatting within the same dataset or database.
Completeness	Mandatory data items should not have blank/null values.	Measure the amount of blank/null values for mandatory items in a dataset.
Timeliness	Data must be available within a specified period.	Time difference between when data is expected vs when it is available must be considered.
Validity	Data must conform to its defined formatting, specified in terms of its format, type and range.	Conformance to the specified formatting.

Source: (Loshen, 2010; DAMA UK Working Group, 2013; Rantala, 2016).

Whenever a business employs these dimensions to measure data quality, it should devote its attention to the following aspects:

- Identification of high value data, especially the ones that support business processes and decisions.
- Agree on the quality assessment rules, based on:

- Organisational requirements for the identified data.
- The most appropriate dimensions and the weight each dimension will carry towards measuring the total quality of identified data.
- Value or range per dimension signifying high or low quality data.
- Impact of non-compliance with the assessment rules (McGilvray, 2008; Loshen, 2010; DAMA UK Working Group, 2013).

The identification of the appropriate data and quality assessment rules is followed by the development of measurement instruments based on the dimensions seeking to facilitate quality measurements. An application of the measurement instruments and subsequent analysis of their results will reveal if data complies with the assessment rule's values/ranges. Any data with quality that is inadequate can be addressed by either applying short-term fixes such as data cleaning or finding a permanent solution by tackling the root cause of the poor data quality (McGilvray, 2008; Loshen, 2010; DAMA UK Working Group, 2013).

2.2.3.2 Eskom Dimensions of data quality

Eskom subscribes to specific principles whenever it deals with data and information. These are: "Principles relating to the quality of data or information requiring data/information to be: C – complete; A – accurate; R – relevant; A – accessible; and T – Timely", which are known as the CARAT principles)(Eskom Holdings SOC Ltd, 2015a:6,7). The CARAT principles are actually data quality dimensions and three of these are mentioned in Table 2.1 that depicts some of the popular dimensions. An example of how Eskom applies these principles is found in Table 2.2 that is an extract from its record keeping and management bulletin.

Table 2.2: Eskom applying CARAT principle

C	Completed records.
A	Accurate reporting.
R	Relevant and reliable records for all our activities in order to expedite decision-making.
A	Availability and easy access to records is critical.
T	Timely completion of associated documentation is critical in preventing unnecessary delays.

Source: (Eskom Holdings SOC Ltd, 2016:6,7)

2.2.3.3 Transactional data quality frameworks

Some popular frameworks for determining data quality are:

1) Wang and Strong's (1996) framework of data quality that was developed to aid Information Security professionals in comprehending and meeting data quality requirements from the viewpoint of data consumers. The framework is expressed in a two level hierarchy consisting of four main categories where each category possesses its relevant data quality dimensions. The four categories and their individual dimensions are:
   a) *Intrinsic data quality*- Believability, accuracy, objectivity and reputation.
   b) *Contextual data quality*- Value addition, relevance, timeliness, completeness and appropriate amount of data.
   c) *Representational data quality*- Interpretability, ease of understanding, representational consistency and concise representation.
   d) *Accessibility data quality*- Accessibility and access security.

2) The modified Information Quality Model developed by Bovee, Srivastava and Mak (2003), which seeks to improve on earlier data quality models by addressing obscurities contained within certain categories, redefining the categories, adding a user specified dimension and moving dimensions to their appropriate categories. The model is presented as a three-tier hierarchy, consisting of four main attributes with their relevant elements. Two of the four attributes' elements are expanded by means of sub elements on the third level of the hierarchy. The attributes are:
   a) *Accessibility*- Time and cost. No third level.
   b) *Interpretability*- Intelligibility and meaningfulness. No third level.
   c) *Relevance*
      i) User specified- As many sub-elements as required by the user of the model can be used.
      ii) Timeliness- Currency and volatility.
   d) *Credibility*
      i) Accuracy- Known, assigned and measured.
      ii) Completeness- No third level.
      iii) Consistency- Discrete and continuous.

iv) Non-Fictitiousness- Records, attributes and values.

3) The Big Data Quality Framework formulated by Cai and Zhu (2015) aspires to address the data quality issues inherent in huge amounts of data mined by businesses on a daily basis. The model proposes, in a way similar to the previous framework, a three level hierarchal framework that contains five dimensions, each with its sub level of elements. The third level contains indicators that are definitions for each element. Further offered, is an assessment process and its feedback mechanisms, which facilitate practical framework implementation. The first two levels of the hierarchy are presented below.
   a) *Availability*: Accessibility and timeliness.
   b) *Usability*: Credibility.
   c) *Reliability*: Accuracy, consistency, integrity and completeness.
   d) *Relevance*: Fitness.
   e) *Presentation quality*: Readability.

A common thread within Eskom's CARAT principles and the mentioned frameworks is that they strive to provide a list of data quality dimensions to use within specified boundaries, which render a data practitioner the ability to define the complete quality of data under evaluation.

2.2.3.4 Root causes of poor data quality

Data has a life cycle, which suggests that the root causes of poor data quality can affect data during various stages of its life cycle. A typical life cycle of data can consist of the following stages: Collection, pre-processing, processing, post-processing, sharing, storage and archival, and data destruction (Arora, 2016). The life cycle, which can possess various formats depending on the type of organisation, adds significant complexity to the definition and comprehension of the causes of poor data quality. Fortunately, research on the root causes of poor data quality continues to develop and progress. Table 2.3 presents some of the common root causes of poor quality data.

Table 2.3: Root cause of poor quality data

Root cause	Description
Access quality	Large data volumes can make on-time access difficult.
Aging quality	Too old information cannot be trusted.
Entry quality	Data not entering the system correctly at origin can be problematic.
Identification quality	When similar objects are identified differently.
Integration quality	All information not integrated correctly to provide an accurate representation of an object can be a challenge.
Interpretation quality	Subjectivity of data creator during data production can lead to incorrectly captured data values.
Organisational quality	When data cannot be reconciled between different systems within the organisations.
Post creation testing quality	Inadequate quality testing of data after its creation.
Process quality	Failure to maintain data integrity during system processing.
Source quality	When multiple sources exist with different versions of the same data.
Training quality	Poor training of data creators on data entry processes and procedures.
Usage quality	Information used and interpreted incorrectly by users.

Source: (McKnight, 2009; Singh & Singh, 2010; Loshen, 2011; Wang et al., 2015)

A business can shift its focus, after identifying the root causes of data quality problems, to improve or eliminate the causes or to minimise their impact. Nonetheless, the above examination of the general causes of data quality problems critically leads to the focusing on poor transactional data quality, which is the focus of this study.

2.2.3.5 Root cause of poor customer transactional data quality within Eskom

The customer transactional data considered in this study is related to calls made by customers to the Eskom Contact Centre while reporting loss of electricity supply. A customer's call to report an electricity supply problem (ESP) is followed by the Eskom contact centre agent's posing certain questions to the customer from a predefined call script (also called case based reasoning) in order to determine the cause of the problem. The agent's interpretation of the customer's feedback as an indication that the loss of electricity supply originates from Eskom's equipment, results in the generation of a dispatch request to a technical person tasked with resolving the issue. The dispatch request results in an automatic creation of a work order. The work order consists of master data that includes the customer's information, the related equipment and the geographical location of the customer as well as transactional data such as the fault symptom and additional directions to the fault.

The receivers of the work order will perform quality assurance (QA) on it, in order to determine its validity and whether it would be containing the minimum amount of information needed to execute the work. Unfortunately, the QA process does not verify whether the agent interrogated or interpreted the customer's feedback correctly. This lack of agent interrogation verification gives rise to occasional scenarios where technical resources (.e.g. technical personnel, transport, and time invested) are devoted to sites where the ESP would not have been a consequence of the malfunctioning of Eskom's equipment but customer's erroneous configuration or equipment failure. These scenarios involving the generation of poor quality transactional data could be the result of the contact centre agent's failure to follow the call script correctly, failure to interpret the customer's feedback correctly or emanate from the customer's provision of false or inadequate information. Table 2.3 indicates that the agents generate poor DQ,s often due to training quality or interpretation quality. However, the false information supplied by the customer is not related to data quality as there is no way to verify the truth of the customer's feedback before a technician and resources are dispatched to the site.

2.2.4 Data quality improvement methodologies/strategies

The measurement and improvement of DQ can be a very daunting task. Redman (2013a:16,24) rightly suggests that the best place to manage and consequently improve data quality is "at the points of data creation" and that the best practice for producing accurate results is "getting it [data capturing] right the first time." A business has to follow a logical and methodical approach in order to improve its DQ and achieve accurate data capturing. Based on his vast industry experience, Loshen (2011) introduced a logical approach to improving DQ that emphasises the importance of building a business case that highlights the current state of DQ, the value a business derives from different data sources, the benefits the business will realise from improving its DQ and the cost of improving DQ. He proposes the following steps for the implementation of his approach:

1. Identification of and distinguishing DQ with high impact from that with a low impact on the business.
2. Determination of the root cause of the DQ defects.
3. Correction of faulty processes.
4. Correlation of business value with source data quality, thereby establishing the value that a business derives from a specific data source.
5. Institution of the best practices to address flawed data production.

Batini and Scannapieco (2016) compared and analysed thirteen data quality methodologies created between 1998 and 2006. They created a new methodology based on the study results called the Complete Data Quality Methodology (CDQM), whose steps are listed in Table 2.4. Even though CDQM does not specifically mention the building of a business case, which Loshin (2011) emphasises, the focus from steps 6 to 10 highlights the importance of reducing overall costs when implementing techniques for improving data quality. The determination of improvement costs in relation to benefits realised can assist in the building of a business case in order to get the necessary approval and financial backing for the implementation of a data quality improvement programme.

Table 2.4: Complete Data Quality Methodology (CDQM) steps

Phase	Step
1. State reconstruction	1. Reconstruct the state and meaning of most relevant databases and data flows exchanged between organisations. Thereafter, build the database and dataflow/organisation matrices. 2. Reconstruct most relevant business processes performed by organisations and build the processes /organisation matrices. 3. Reconstruct the norms and organisational rules for each process or group of processes related to a macro process that regulate the macro process and the service provided.
2. Assessment	4. Check problems with users: Identify the most prevalent causes of poor DQ and the effect they have on data consumers. 5. Identify relevant DQ dimensions and metrics, measure quality of databases and data flows, and identify their critical areas.
3. Improvement	6. Fix the new DQ levels for each database and data flow, to improve process quality and reduce costs under a required threshold. 7. Conceive process re-engineering activities and choose DQ activities that may lead to DQ improvement targets set in step 6 that relate to data/activity matrix to clusters of databases and data flows involved in DQ improvement targets. 8. Choose optimal techniques for the DQ activities. 9. Find improvement processes in the data /activity matrix. 10. Compute approximate costs and benefits for each improvement process defined in the previous step and choose the optimal one, checking that the overall cost-benefit balance meets the targets of step 6.

Source: (Batini & Scannapieco, 2016:353-386)

2.2.5 Global state of data quality

Röthlin (2004) studied the perceptions on data quality (DQ) of information systems managers from 500 large Swiss companies using a questionnaire with a scale of -3 (very bad) to 3 (excellent). He found out that master data from human resources was perceived to be of the highest quality compared to other forms of data whilst transactional data was of the lowest quality. A modification of his scale to percentages, reflected in Table 2.5, reveals that the quality of master data ranged from 74.5% to 84.3% (with a mean of 78.3%) whilst the quality of transactional data was 72%.

Similarly, Experian's (2013) global data quality research examined the opinions of executives, management and administrative staff who were intimately involved with data management from a wide range of industry sectors across Europe and the

United States. The findings reveal that these companies perceived their customer and potential client data quality to be 78%, which is very close to the mean of Röthlin's (2004) study. Another important observation from this study is that big leaps in data volumes tend to erode data quality as quality levels measured over a 12 month period signalled a 5% drop. In addition, 1400 data practitioners within various industries spread across 8 countries around the globe were surveyed in Experian's (2017) research. Here, customer and potential client data quality was perceived to be 73%, implying a further 5% drop in data quality from 2013 to 2017.

Table 2.5: Perceptions on DQ

Data Description	Data type	Perception scale: -3(very bad) to +3 (excellent)	Convert to %
HR Data	Master data	2.06	84.3%
Costing data	Master data	1.69	78.2%
Supplier data	Master data	1.57	76.2%
Customer data	Master data	1.47	74.5%
Transaction Data	Transaction data	1.32	72.0%

Source: (Röthlin, 2004:256)

2.2.6 Data quality in South Africa

The section explores the quality of South African data before finally shifting the focus to Eskom's. According to Neil Thorns, Informatica's territory manager for sub-Saharan Africa, the state of South African customer data quality is a big concern that hampers the quality of companies' decision-making (Burrows, 2014). Thorns believes that the majority of South African companies have low data quality with an average accuracy of 50% or less. This is significantly lower than the 73% global mark as indicated by Experian's (2017) latest survey on master customer data. The main reason for the low quality of data is contributed to the lack of business rules and automation during data capturing especially in some government departments and small, micro and medium enterprises (SMMEs) due to constrained high-technology adoption and resource limitations (Burrows, 2014; Dlova, 2017). This low data accuracy subsequently makes it difficult to make proper decisions based on the analysis of customer data.

Inversely, World Economics compiled a Data Quality Index on the Gross Domestic Product (GDP) of 154 countries and employed 5 indicators, which are: base year, system of national accounts, the informal economy, quality of statistics and corruption. South Africa was ranked 49th with an overall score of 77.1% (World Economics, 2017), which is close to the 73% global mark as indicated by Experian's (2017) latest survey on customer data. Taking into consideration that GDP is an important financial indicator within South Africa, it is understandable that its data quality will be t higher than the customer data of an average company as more emphasis will be placed on correct data capturing at national level.

Statistics South Africa performed another measure of the quality of master data for the Data Quality Report in its 2016 Community Survey. Data quality was measured based on the percentage imputation performed, which is the replacement of missing values on the captured data. Overall imputation rates measured at 5%. Furthermore, precision of estimated key variables was determined by means of confidence intervals, which measure the uncertainty associated with a sample statistic. A 95% confidence level that was recorded expressed that estimates lay within the calculated lower and upper limit intervals (Statistics South Africa, 2016). The low imputation rates and high confidence level can be associated with rigorous automation, data validation and feedback mechanisms built into the data capturing stage by means of electronic surveys and quality assurance performed post capturing.

The above indicates that data quality in South Africa varies widely and is dependent on the industry, importance of the master data and the capturing methods employed. Electronic capturing and the automation of business rules, data validation, feedback to the capturer and quality assurance performed on data post capturing, seem to deliver data of highest quality.

2.2.7 Transactional data within Eskom

The transactional data under focus here is that of Eskom, a national public company mandated to generate, transmit, and distribute electricity to industrial, mining, commercial, agricultural, redistributors, and residential customers (Rambe & Modise, 2016). Eskom's size and complexity yields a large volume of transactions and these give rise to daily occurrences of transactional data within the company. Some transactions of a financial nature, such as procurement transactions or overtime hours, undergo a rigorous QA process to ensure that the data generated from these transactions is captured correctly before the transaction is processed further. The purpose of QA is to prevent Eskom from incurring unwarranted costs arising from data capturing errors or an incorrect interpretation and application of business rules. However, the QA process is expensive, as it requires additional hours of labour to verify the data. As such, it is not feasible to subject all transactions to a rigorous QA process. However, Eskom relies, for the generation of correct transactional data on non-QA transactions, on its information processing systems, business rules applied to the captured data and the transactional data creator's ability to capture the data correctly. A technical glitch in Eskom's data capturing systems, non-adherence to business rules or technical misjudgement of data capturers may compromise the quality of the transactional data and lead to serious financial implications for the organisation.

2.2.8 Master and transactional data quality within Eskom Distribution

The section focuses on the quality of data at Eskom Distribution in general, which are equally applicable to Eskom Distribution Free State. Consistent with its distribution environment, which involves the distribution and sale of electricity to the end consumer, Eskom Distribution also supplies monthly feedback on the state of its high value master data records to various internal stakeholders (i.e. employees, supervisors and senior management). Such customer data has a high business value since clientele transactions form a core constituent of Eskom Distribution's transaction data. The customer data is, however, followed in importance by plant data, which is data concerning Eskom Distribution's equipment.

The reality that selected portions of Eskom Distribution's transactional data are used to calculate key performance indicators (KPIs) of the organisation, suggests that poor transactional and master data can compromise the business operations. Eskom Distribution's staff in the QA department and Data Officers from various departments actively monitor transactional data portions used to calculate KPIs, because the KPIs must adhere to national and institutional regulatory standards. An example of such a KPI is the System Average Interruption Duration Index (SAIDI) score, which is a measure used to determine the average duration of service interruptions of the system. A SAIDI score of 39 or lower must be met as determined by the NERSA (Eskom Holdings SOC Ltd, 2017a). The transactional data related to the SAIDI KPI undergoes a rigorous QA process to ensure that each interruption is measured correctly. Nonetheless, the large volumes of master and transactional data flowing through Eskom Distribution’s information systems on a daily basis, undermines the feasibility of testing all data sets for their quality. Consequently, only data perceived to have high economic, operational and financial value is tested and verified for its quality.

2.2.9 Impact and cost of poor data quality

One key question that begs for an answer is: *Why is data quality such an important issue*? It can be argued that data supports decision making within businesses. The use of poor quality data results in poor operational and strategic decisions that negatively influence productivity and increase operational costs. Poor quality data also affects customer satisfaction and perceptions about the business, which results in loss of revenue. The increase in operational costs stems from the rework arising from the misuse of resources, such as people, time and equipment, and the costs incurred while making corrections to reported errors (Samitsch, 2014; Fürber, 2015).

The Eppler Helfert classification (Batini & Scannapieco, 2016), categorises the monetary costs caused by poor data quality as either direct or indirect. Direct costs have an immediate measurable impact such as costs related to the re-entry of data, verification and compensation. In addition, indirect costs have a delayed impact such as, lower reputation, wrong decisions and sunk investments. Other costs involved in ensuring and/or improving data quality include prevention, detection and repair costs. Prevention costs focus on the creation phase of data and involve the training of data capturing staff, monitoring of compliance to institutional standards and the development and deployment of standards. Finally, detection costs encompass activities such as analysing data after its creation and reporting on findings, whereas repair costs cover planning for data repairs as well as the implementation of repair plans.

Findings from the 2015 Global Data Quality Research indicate that 92% of the companies surveyed experienced challenges in generating data of high quality mainly due to the role played by tangible and intangible objects (Experian data quality, 2015). In fact, research (Eckerson, 2002; Samitsch, 2014) shows that US businesses lose over $600 billion per year due to poor DQ. Therefore, improving data to a higher quality can result in good decision making and lower operational costs through increased operational efficiency, increased organisational performance and high revenue (Samitsch, 2014; Zhang, 2014). Batini and Scannapieco's (2016) data quality improvement benefits classification in Table 2.6 proposes three categories where benefits can be realised when improving data quality and these are monetisable, quantifiable and intangible.

Table 2.6: Data quality improvement benefits categories

Category	Description	Example
Monetisable	Values can be expressed in monetary terms.	Increased revenue.
Quantifiable	Cannot be expressed in monetary terms, but in other numeric terms.	Hours saved to perform more productive work.
Intangible	Cannot be expressed in any numeric term.	Improved reputation.

Source: (Batini & Scannapieco, 2016:324)

A further discovery from the Global Data Quality Research focusing on the monetisable benefit of improved data quality notes that organisational profits could increase by as much as 15% if data of the highest quality is generated (Experian data quality, 2015). It is important, when attempting to increase data quality, to identify data with high value within the context of decision making or business processes (DAMA UK Working Group, 2013). A cost-benefit analysis is also required in order to ensure that the envisaged savings will offset the costs involved to improve the data quality (Haug, Zachariassen & Van Liempd, 2011; Batini & Scannapieco, 2016). Redman (2013a) proposes that, in the context of causal costs stemming from poor data quality, only operational costs can be estimated or determined with a high level of accuracy. Thus, despite the fact that factors like disgruntled customers, poor decision-making, and inability to manage risks can drive operational costs higher than those calculated or anticipated, there are currently no reliable methods to determine the costs of such factors.

The impact and cost of poor data quality present a valuable opportunity for Eskom to scrutinise its operations for saving opportunities. Much needed monetary savings and the consequent increased profits can be unlocked if high value data with low quality can be identified and improved. However, the costs incurred by Eskom for improving data quality should be less than calculated operational cost savings, otherwise the benefit of improved data quality will not be realised.

2.2.10 Impact of poor transactional data quality on Eskom Distribution

Eskom Distribution technicians are only responsible for reconnecting a customer's power supply if the interruption was caused by a fault on Eskom's electrical network. As mentioned in 1.3, incorrectly categorised customer calls by Eskom's contact centre triggers unnecessary dispatching of technicians to faulty customer equipment and not faults related to Eskom's electrical network. This incorrect categorisation is a direct result of poor quality of the transactional data captured from customer calls to Eskom's contact centre.

Poor quality transactional data on customer calls within Eskom Distribution, as the example above, impacts on the business negatively in various ways. These ways include:

- Reduced customer satisfaction as customers requiring genuine and reliable service from a technician have longer waiting times.
- Decreased ability to meet regulatory compliance (NERSA, 2002).
- Staff morale is impacted negatively as technicians drive to resolve invalid faults.
- Cost implications as each request for service results in the spending of resources like the paid hours that technicians' spent attending to invalid faults and vehicle kilometres driven to address these matters.
- Lost opportunity cost as planned maintenance could have been performed.

2.3 Feedback

Feedback is a very helpful tool that can be utilised to the benefit of an organisation if used correctly. Therefore, understanding what feedback is and the value that is derived from it is explored in the subsequent section.

2.3.1 Defining feedback

Feedback is "Information about reactions to a product, a person's performance of a task which is used as a basis for improvement" and "the modification or control of a process or system by its results or effects" (Stevenson, 2010:640). Feedback is a very important mechanism that exists within both the natural and engineered world and can be utilised to improve performance, correct errors and achieve a desired result. A practical example of feedback is evident within a biological system, such as a human being, where the action of placing a hand on a hot object will result in feedback via the pain sensors that inform the brain that the hand is burning, resulting in an action that lifts the hand from the hot object. Feedback control is used within the context of science and engineering to correct or normalise a system's output when the measured feedback indicates a deviation from an expected output (Abramovici & Chapsky, 2000). Feedback within Eskom is generated from multiple sources such as staff observations, customer comments, results from monitoring equipment and analysis algorithms applied to certain datasets. This feedback offers potential benefits if it is applied to improve the business facets it is linked to.

2.3.2 Perspectives on feedback

2.3.2.1 Feedback loop

The utilisation of feedback to improve performance via feedback control is achieved by means of feedback loops. The mere presence of feedback does not necessarily provide any benefit to a system. Instead, the processing of the feedback and reaction towards it, can be used to adjust a system to respond appropriately to future needs. Lidwell, Holden and Butler (2010:92,93) define a feedback loop as "a relationship between variables in a system where the consequences of an event feeds back into the system as input, modifying the event in the future". Feedback loops are categorised into two systems and these are; closed loop and open loop.

Closed loop systems are interconnected in a cycle whereas the cycle is absent in an open loop system. An example of a closed loop system related to data quality is where users of data provide feedback to data creators as soon as a problem with the data is detected. Data creators will then use the feedback to identify the cause of a data issue and correct the issue to prevent future reoccurrence (Biehl, 2016). In addition, an open loop system is cheaper, simpler and more stable than a closed loop as it operates on set parameters during the input stage and therefore, the output will not have any effect on the input stage. Unfortunately, an open loop does not take advantage of any benefits that feedback may provide. Nevertheless, one of the main advantages of a closed loop system over an open loop is its ability to provide improved quality of control as open loop systems are often inaccurate due to a lack of error correcting ability. The single loop is the simplest form of a closed loop system, having only a single feedback channel, whereas multi loop systems are more complex and have two or more feedback channels (Chesmond, 2014).

2.3.3 Application of feedback in organisations

As stated earlier, feedback in itself does not necessarily bring about change in a system, but rather change is facilitated by means of feedback loops. The use of feedback via feedback loops within the data environment can create data, improve data quality, advance data processes and increase value generated from data (Alexopoulos, Loukis & Charalabidis, 2014). Human feedback is an important requirement when seeking to improve DQ by means of feedback loops. Human feedback can influence direct, indirect data or automatic DQ improvements which are described in more detail below (Wang et al., 2015; Brodie & Palmer, 2016).

- *Direct* DQ improvements can occur if a human being identifies errors on a dataset. The identified error (feedback) can be used to correct the data immediately if the system or business rules allow it. The loop is thus closed when data is corrected due to an identified error (feedback).
- *Indirect* DQ corrections can occur when a human identifies an error, but is prevented from performing a data correction due to system or business rules. Consequently, the feedback will be passed to another user with sufficient system and business privileges to address the problem.

* *Automatic* DQ improvement can occur if human feedback informs computer system algorithms responsible for correcting data/preventing data mistakes (Wang et al., 2015; Brodie & Palmer, 2016).

2.3.4 Feedback on transactional data quality within Eskom Distribution

Eskom Distribution has created a feedback channel for its technicians who deal with transactional data on customer faults related to an ESP. When a technician arrives on site and realises that the fault is on the customer's side and not caused by Eskom's electrical network or equipment, the option is available to mark the transaction as a customer side fault. This feedback will then be stored against the specific transaction and can be extracted later for further investigation. Some investigations are done sporadically but due to current system design restrictions, the feedback is not automatically looped back to the data originators for action, correction and improvement. Consequently, the value from a feedback loop is not optimally realised.

2.4 Costs

An understanding of the costs that are applicable to business processes can prove invaluable for any business that would be trying to optimise its operations and increase profits. The types of costs, their importance and calculation will be discussed in the following section.

2.4.1 Defining costs

It is important to have knowledge of the cost of a product or service in order to assess its impact on a business. Law (2016:156) defines costs as "An expenditure, usually of money, incurred in achieving a goal such as producing certain goods, building a factory or closing down a branch." (Law, 2016:156). From an electrical utility perspective, Willies and Schrieber (2016:162) regard cost as "the total sacrifice that must be expended or traded in order to gain some desired product or end result. It can include money, labour, materials, resources, real estate effort, lost opportunity, and anything else that is given up to gain the desired end." To measure costs from a common perspective, all the aspects mentioned are converted usually to money.

Some costs cannot be converted into money, which calls for complex evaluation methods in order to assess their impact on the business.

2.4.2 Types of costs

Various types of costs can exist within a business and usually a classification system is applied to categorise costs. Costs on a product/service are classified normally as direct or indirect. Direct costs, on the one hand, are tracked straightforwardly to the product/service. These direct costs, which include material and labour consumed during the manufacturing of a product, vary with the amount of product produced. Indirect costs, on the other hand, cannot be easily associated with the end product/service such as labour and material required to clean the premise where a service is delivered (Keller, 2015).

Three types of costs can fall under the direct/indirect cost classification:

1. Material: Cost of raw materials for product/service, spare parts, consumables, and office supplies.
2. Labour: Remuneration costs such as salaries, overtime, bonuses and commission.
3. Other expenses: These costs, which include premise rates and taxes, office furniture, general building repair costs, insurance, lighting and heating, do not fall under the material or labour categories (Mehta, 2016).

Apart from the direct and indirect classification, a few other cost concepts need to be considered and these are:

- Avoidable costs: Such costs can be avoided if certain strategies are employed. For instance, the discontinuation of a certain activity results in no direct costs related to the activity being incurred.
- Controllable costs: Costs that a unit within an organisation has the authority to incur and can be influenced by a specific person within the lower and middle management of an organisation.
- Uncontrollable costs: These costs cannot be controlled by a person in charge of a specific unit but rather by top management.

- Opportunity cost: This refers to the value of a sacrifice made due to a company executing one action at the expense of another.
- Operational costs: Routine costs related to the day-to-day running of a business, like labour and equipment for operations, maintenance and service.
- Overhead costs: This is the aggregate of all indirect costs and consists of all costs that cannot directly be linked to product or service grouped together such as premise rates and taxes, labour costs related to cleaning of the premise and insurance.
- Postponable costs: Costs that can be deferred to a later point in time.
- Fixed costs: These costs, such as insurance, monthly rent, monthly instalments and inspections, do not change when more or less goods or services are produced.
- Variable costs: These are cost increases or decreases in items such as raw materials, product packaging and electricity in relation to the output of products/services.
- Replacement cost: The cost incurred to replace an existing asset such as replacing outdated furniture (Holloway, 2016; Mehta, 2016; Willis & Schrieber, 2016).

2.4.3 Importance of costs

In order for a business to stay profitable, it must ensure that the costs involved in making a product or service are lower than the selling price of the product or service. If a business is unable to achieve this feat, it will suffer a loss, which will eventually lead to bankruptcy. Therefore, having intimate knowledge of the operational costs enables a business to determine the correct selling price for its product/service. Furthermore, cost information assists businesses to determine the minimum acceptable order size, product profitability, most preferred suppliers, budgetary planning, product delivery options, conduct a cost benefit analysis, identify cost drivers and keep a project within budget (Clough et al., 2015; Ernst & Young LLP, 2015; Shim, 2016).

2.4.4 Cost calculation

The continuous making of the same type of product or service delivery makes the cost per unit of the product/service to stay constant. As a result, the volume of the output is employed as the prime parameter to determine the overall cost over a specific period. In such an instance, a process cost system will be utilised to determine the costs involved. However, if a variety of products is produced or each service delivered is unique, a process cost system cannot be used for cost calculation. In such circumstances, job order costing is a preferred method to calculate product costs (or cost of service), as the cost for each job completed will be different (Vanderbeck & Mitchell, 2015). Job order costing involves a job order cost sheet that will contain a unique number, the job details, as well as direct and overhead costs, and these are used to calculate the cost for each job (Vanderbeck & Mitchell, 2015). Figure 2.1 illustrates an example of a job order cost sheet. The sheet indicates the process cost method used by Brava Boards Co to manufacture a quantity of 1000 skateboards for its customer, Riders' Warehouse. It contains a breakdown of direct materials and direct labour consumed during manufacturing. As 35% of total factory utilisation is spent on manufacturing of skateboards, that percentage of factory overheads is charged against this specific job. In instances where indirect cost is difficult to calculate, the resource consumed the most, such as labour hours, is used as the cost allocation base for indirect cost assignment. This method implies that the overhead is proportionally divided to the highest consumed resource (Warren, Reeve & Duchac, 2015).

	A	B	C	D	E	F	G	H	I
1	Brava Boards Co.								
2	Job Cost Sheet								
3	Customer Name:	Riders' Warehouse					Job No:	101	
4	Address:	5239 Arce Street					Date Started:	2/18/16	
5		Phoenix, AZ 85339							
6	Quantity:	1,000					Date Completed:	2/28/16	
7	Product:	skateboards							
8	Description:	6-ply							
9									
10	DIRECT MATERIALS			DIRECT LABOR			FACTORY OVERHEAD		
11	Date	Mat'l Req. No.	Amount	Date	Time Tkt. No.	Amount	Date	Basis Applied	Amount
12									
13	2/18	2203	18,980	2/18	1065	5,310		35% of	
14	2/28	5483	7,760	2/25	3259	3,320		total	
15				2/28	4442	4,250	2/28	over-head	10,780
16									
17	Total		26,740			12,880			10,780
18	SUMMARY						Remarks:		
19	Direct materials		$26,740	Selling price		$75,000			
20	Direct labor		12,880	Mfg. cost		50,400			
21	Factory overhead		10,780	Gross profit		$24,600			
22	Total cost		$50,400						
23	Unit cost		$50.40						

Figure 2.1: Job order cost sheet example
Source: (Vanderbeck & Mitchell, 2015:33)

2.4.5 Eskom Distribution costs on ESP faults

Whenever an Eskom Distribution field technician is dispatched to an ESP fault, he is issued with a work order. A work order contains information such as a unique number, date of fault reporting, description of the fault, customer detail, location of the customer and the assigned technician. The real-time information related to the progression of the fault is captured during work order execution. The completion of the work linked to a work order is followed by the capturing of the actual labour hours and these are approved for financial processing. Thus, the major direct costs incurred during the execution of a work order are labour hours, distance travelled and material consumed.

Furthermore, the work order can serve as a proxy to a job order cost sheet and the direct labour costs can be used to calculate the cost of a work order. Data enrichment on available geographical coordinates linked to a work order can be applied to calculate distance travelled and the cost associated with it. Unfortunately, material consumed is captured in the enterprise resource planning system known as Systems Applications and Products in data processing (SAP) software and a reliable link back to the work order system called Maximo does not exist, therefore performing material costing per work order is not feasible. The overhead related to operations are shared between many operational functions within a Customer Network Centre (CNC), which also means that the overhead cannot be applied easily to an allocation base such as hours worked.

2.5 Monetary savings

Different strategies can by employed to achieve monetary savings. This section will unpack the meaning of monetary savings and the cost reduction strategies that can be implemented to achieve the savings.

2.5.1 Defining monetary savings

Savings are defined as the “avoidance of overspending, reduction in expenditure or cost, amount of money saved” (Editors Of Webster's II Dictionaries, 2005:1006). The terms savings and cost savings are often used interchangeably as evidenced in the overlap within the definitions. Cost saving is defined as "the fact of saving money or of spending less money than was planned" (O'Shea, 2011:185) and “a reduction in expenses, especially in business” (Oxford Online Dictionary, 2017:1). Savings are either hard dollar or soft dollar savings. On the one hand, hard dollar savings refer to planned costs that are saved (compared to a previous baseline) which affect the bottom line, such as increased productivity, reduction in overtime, less defective work and operational savings that include shorter processing times. Soft dollar savings on the other hand, describe cost avoidance savings such as process improvements and increased output without increased resource utilisation that is facilitated by improved efficiencies in production and process systems, customer retention and growth, improved space utilisation and capacity expansion (Kubiak, 2013; Protzman et al., 2016).

2.5.2 Monetary savings through cost reduction

Cost reduction strategies are the primary method utilised to achieve monetary savings. Cost reduction is defined as “the achievement of real and permanent reduction in unit cost of goods manufactured or services rendered without impairing their suitability for the use intended or diminution in the quality of the product" (Chakraborty, 2004:666). Effective utilisation of resources is of paramount importance to achieve cost reduction. Plans to reduce costs should be evaluated carefully and implemented strategically in order to achieve long-term savings and business benefits. The implementation of cost reduction strategies can have financial cost implications such as automation tactics where people are replaced by machines. The cost savings achieved from the reduced salary bill should offset the cost of implementing automation. Furthermore, cost reduction strategies, such as the outsourcing of key functions to external entities and laying off staff members that used to perform such functions can also damage company value by undermining some of its core abilities. Consequently, only strategies in which the benefit outweighs the overall cost and business impact should be considered when trying to achieve savings through cost reduction strategies. Some of the areas that can be scrutinised for potential cost savings are re-work, wastage, salaries, travel, administrative expenses, legal expenses, overproduction and impaired effectiveness (Murragapan, 2014; Mehta, 2016).

A typical cost reduction strategy geared at achieving monetary savings is a reduction in salary expenses by reducing staff numbers. This strategy is, however, risky as it can erode some of a business' core abilities and competencies. A combination of different strategies that retains critical skills can serve businesses well in the long run. Fabiani et al., (2015) in their study on the European firms' response to the global economic crisis of 2009 highlight the way the firms made their human capital as important, such that they did not dismiss permanent employees during those trying times. The firms focused on freezing wages, cutting flexible pay components, reducing work hours, laying off temporary staff and reducing non-labour costs as much as possible. Thus, cost reduction strategies can achieve monetary savings and boost company earnings. This is exemplified by the case of Kenyan tea farmers who were able to increase profits in spite of a reduction in output, by focusing their cost reduction efforts on human resources (correct staffing, staff retention and proper training), optimisation of tools and technology, cost reduction during procurement and finding alternative and cheaper energy sources (Namu et al., 2014).

It is clear that, quality improvement can reduce costs by lessening reworks and recalling of products, but proper decision-making is dependent on quality data, hence this study. Therefore, whenever a business seeks to employ quality improvement strategies as a cost reduction mechanism, it has to ensure the data supporting its decisions is of good quality (DeOreo & Wish, 2015).

2.5.3 Eskom's monetary saving strategies

Eskom currently employs various strategies to achieve its targeted monetary savings (Minister of Finance, 2015) and financial sustainability by means of cost reduction and increasing its sales. The five focus areas for savings and increased sales up to the year 2022 are to:

1. Increase demand for electricity locally by 2.1% and exports by 8%.
2. Reduce primary energy expenses by R53 billion.
3. Use advanced analytics to deliver savings of R14 billion.
4. Optimise capex by R65 billion.
5. Release government guarantees of R105 billion by 2020.
6. Workforce optimisation strategies to realise savings of R11.8 billion (Eskom Holdings SOC Ltd, 2017b; Eskom Holdings SOC Ltd, 2018b).

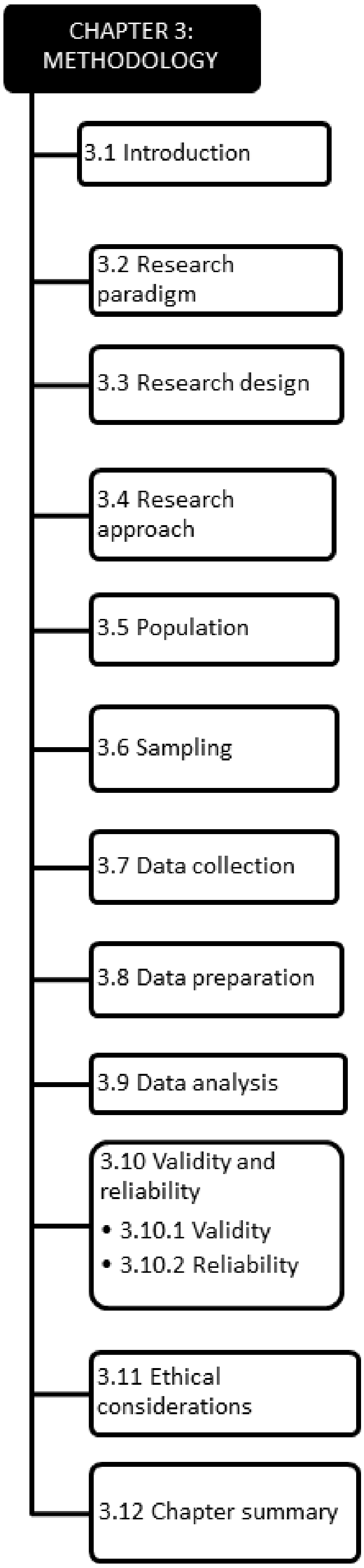
Eskom's correct application of its cost reduction strategies has the potential to prevent an erosion of core and critical skills, improve quality and increase profits even if its sales do not increase in the current financial climate of slow economic growth. All these benefits can be derived from good quality data that will help to ensure that correct decisions are taken when defining and applying its preferred cost saving strategies.

2.6 Chapter summary

The four variables namely transactional data quality: source system, transactional data quality: feedback, costs and monetary savings were unpacked to provide a deeper understanding into transactional data, feedback, costs and monetary savings. The chapter outlined further the relevance of each variable by describing the role that each plays within Eskom and Eskom Distribution. The chapter also noted that an improvement of transactional data quality, by means of feedback, has the potential to enhance monetary savings through a reduction of costs and an improvement in operational efficiency.

The next chapter focuses on the research methodology used in this study.

3 CHAPTER 3: METHODOLOGY



3.1 Introduction

This chapter discusses the methodology that was utilised during this study. The chapter covers the research paradigm, research design, the study population (for technicians and historical data), sampling method, data collection, data analysis, validity and reliability and ethical considerations.

3.2 Research paradigm

A paradigm provides the broader research parameters of a given study and it shapes the philosophy guiding the researcher as they uncover phenomena. Although there is a general lack of consensus on what a paradigm really is, there is a general agreement in mainstream research literature that a combination of ontology, epistemology, and methodology forms the main basis for the paradigm that a researcher can adopt (Johnson & Christensen, 2010; Killam, 2013). Ontology studies the nature of being and includes claims or assertions on the nature of reality. Epistemology is concerned with the study of the nature of knowledge (i.e. whether it is conceived as subjective or objective) and determines the beliefs, foundations, extent and validity of knowledge. This study is founded on developing objective knowledge as it employs quantitative data on transactions in Eskom Distribution to establish if the quality of such data can affect the cost savings of this institution. A Positivist paradigm whose ontology assumes one objective reality and whose epistemology focuses on observable, objective facts or phenomena to gain knowledge (Potter, 2006; Lindgreen, 2008) was thus, adopted in this study.

3.3 Research design

A cross-sectional survey was administered on the key personnel of the Eskom Distribution Free State to determine the impact that transactional data quality from customer calls has on the monetary cost savings. Surveys are applied in quantitative research settings, wherein the population selected is studied in-depth to explore the aspect/s of interest (Marsden & Wright, 2010; Edmonds & Kennedy, 2016).. Data from a survey, which is normally summarised by means of descriptive statistics and insights, is also generalised to the entire population from which the data would have been drawn. A survey has two benefits and these are its low cost and generalisability to larger populations. However, the survey approach has some drawbacks and these include encountering low response rates of between 15% to 25% and respondents not always answering questions truthfully (Marsden & Wright, 2010; Edmonds & Kennedy, 2016). Eskom Distribution has nine operating units with similar standard operating procedures, as a result, a survey of the for Free State Operating Unit (FSOU) was deemed desirable because it is similar to the other 8 Operating Units and closest to the researcher. The survey findings were enriched by analysing a large volume of historical data from the Eskom Distribution Free State customer network centres (CNC). The analysis of all data focussed on the accuracy of customer call transactional data, the financial costs associated with incorrect transactional data and the impact of incorrect transactional data on monetary savings.

3.4 Research approach

A quantitative approach was used to determine the transactional data quality at the source and the receiver as well as to determine the costs of labour hours and kilometres travelled. A quantitative approach was also used to gauge the field technicians' perceptions of the quality of the transactional data and its impact on the unplanned hours and unplanned kilometres travelled.

3.5 Population

A population is the group of subjects (people, elements, objects) that is the focus of the study (Grove, Gray & Burns, 2014). Eskom has 9 Operating Units (OU), one per every South African Province with each containing multiple CNCs. The size and amount of CNCs per OU depends on the amount of customers and size of the electrical network that has to be maintained. The CNCs are located in such a way as to optimise travelling and customer service. For the purpose of this study, the focus was on the FSOU and all the CNCs contained within it. A CNC's staff compliment comprises of a supervisor, technical staff with different skill sets who perform maintenance and repairs on the electrical network, and direct support staff. The researcher focused on the FSOU because he is employed in this operating unit and had easy accessibility to the customer transactional data.

The population consisted of 2 parts, a human population as well as a historical data population. A total of 303 technical staff members, residing at the FSOU CNCs, formed the human population as they interact with transactions from customer call transactional data. The historical data population comprised of the historical archive of records/work orders related to unplanned work undertaken by the CNC technical staff. This historical data consisted of over 800 000 work orders for the period 1 April 2002 to 31 March 2017. One technical staff member is one element (unit) of the human population and a work order as one element of the historical data population.

3.6 Sampling

A sample, which is a subset of the population, is taken from the accessible population (population the researcher has reasonable access to) by utilising a specific sampling method. Each single unit of a sample is termed a subject and the findings from a study of the subjects of a sample are generalised or applied to the larger population (Sekaran & Bougie, 2013; Grove, Gray & Burns, 2014). Samples are desirable in research because it is not always economical, possible or necessary to study an entire population. It is however important to ensure that a sample is representative of the population to enable accurate generalisation (Cargan, 2007). Due to the small staff compliment (303 members) in the Free State Operating Unit CNCs, a census was deemed necessary. Consequently, information was generated from every unit of this population (Cottrell & McKenzie, 2010) that is all 303 members. In addition, access to the historical data allowed the researcher to reference and analyse all records from April 2002 to March 2017. However, the sheer size of the historical data and a higher rate of incomplete records pre 2012 meant that only data from April 2012 to March 2017, totalling 235 945 work orders, was analysed.

3.7 Data collection

A structured questionnaire was used to gather primary data from the human population regarding technical staff's involvement and perceptions on the quality of customer call transactional data in 2017. The self-administered questionnaire comprised ordinal (specifically Likert scale, which is also seen as a rating scale) and ratio scales. Past experience indicated that response rates from field technicians are very low, due to the long time they spend working in the field. However, precise wording ensured that technicians easily and correctly interpret and answer questions related to the concepts. In addition, two Eskom subject matter experts reviewed and verified the wording of the questionnaire so that it could be interpreted correctly. The researchers' supervisor also reviewed the questionnaire with the student for accessibility and intelligibility.

The research delivered the questionnaire via emails to each CNC supervisor who subsequently distributed it to their technical staff. The researcher also made one phone call to each supervisor seeking to explain the purpose of the questionnaire and enlist their cooperation and support. Three response options were offered to respondents and so they were allowed to complete any one of the following:

1. A printed version of the questionnaire answered by hand that the respondent would scan and email to the researcher.
2. An electronic version of the questionnaire on a word processor that the respondent would email to the researcher.
3. Utilise a web browser (either via computer or smartphone) to complete a web based version of the questionnaire.

Numerous follow-ups made to CNC supervisors and technicians sought to solicit a 35% response rate (106 out of 303) from the population. Most of the feedback received was a scanned copy send via email. This data was coded and captured on an Excel spreadsheet and exported to SPSS for detailed data analysis. The historical data was extracted by means of SQL from the company's Maximo application database that houses all the work order records. All records were exported to an Excel spreadsheet for secondary data preparation and initial analysis before it was exported to the SPSS for detailed analysis.

3.8 Data preparation

A normalisation and enrichment of the data was performed to prepare the secondary data for analyses. For example: record descriptions were checked for the wording "Customer Fault" and any meaningful variations that indicated a customer fault. An interpretation of transaction dates allowed for fields to be added, categorise data according to Eskom's financial years, and to determine normal/overtime transactions. Global positioning systems (GPS) coordinates of records were used to calculate distance and to infer the time and cost per transaction. Eskom's minimum wage rate and overtime rules were applied to work orders to calculate the actual overtime hours and costs for each financial year based on the current monetary values. A precompiled report that indicates the mean of the transactional call, and quality per month was referenced to indicate transactional data quality at the source system per secondary data record.

3.9 Data analysis

Data was analysed by applying tools from both descriptive and inferential branches of statistics. A researcher uses descriptive statistics whenever they need to organise and summarise data based on sample demographics (Holcomb, 2016). In addition, inferential statistics are used to draw a conclusion about the population by determining the relationship among variables in a sample and making predictions about the population (McKenzie, 2014). Both branches were relevant in this study's analysis, hence, a descriptive and inferential statistical analysis was conducted using Microsoft Excel and IBM SPSS version 25.

The specific descriptive statistical tools that were employed are frequency distribution, mean and cross tabulation/contingency table analysis, whilst the inferential tools focussed on correlation and regression analysis. Each tool is described in more detail below:

- **Frequency distribution** is a summary of data that is typically presented in tabular form and displays the number of observations per group/category. It assists researchers to gain a deeper understanding of data patterns due to the way data is grouped and summarised. An inherent disadvantage is that these grouped summaries can hide important information (Sharma, 2007). This study's frequency distributions catered for the amount of work orders, data quality of source system, feedback, hour and distance costs and monetary savings.

- **Mean**, which is also known as the average value of a dataset, is part of the central tendency tools. This measure describes a set of data by supplying a value, which is the average of your dataset. The value supplied is calculated by adding all the values and dividing them with the number of these values. A disadvantage of the mean is that outlying values can influence its calculation (Lacrose & Lacrose, 2015). The study's mean was established for the quality of source system data, feedback from technicians, cost of transactional data and monetary savings.

- **Cross tabulation** or contingency table allows a researcher to study the relationship between multiple categorical variables. Data is grouped in a multidimensional table and can indicate how correlations change between different groupings of a variable. It is implemented to gain deeper insights on data by revealing specific trends, patterns or probabilities (Singh, 2007). This researcher performed a cross tabulation on questionnaire data to obtain a greater understanding of the quality of source system data, technician feedback, costs for correct and incorrect transactions during normal and overtime and monetary savings.

- **Correlation analysis** sets out to determine whether a correlation exists between continuous variables such as interval or ratio scale types. Calculated correlation coefficients possess both strength and direction and can range from 1.00 to -1.00. The closer a coefficient value is to 1.00 or -1.00 the stronger the correlation between the variables, whereas the correlation gets weaker as it approaches 0. The direction of correlation can be either positive or negative. A positive coefficient value implies a positive correlation, which means that an increase or decrease in one variable yields a similar effect on the other. A negative correlation indicates that an increase or decrease in one variable results in an opposite effect on the other variable. Statistical significance is also calculated to verify if an identified correlation is significant enough to reject the null hypothesis. It is important to note that correlation does not imply causation (Urdan, 2011). Finally, the correlations established in the study were between data quality of source system and feedback, data quality of source system and costs, data quality of source system and monetary savings, feedback and costs, and costs and monetary savings.

- **Regression analysis**, which is a form of predictive analysis, studies the relationship between one or more independent/predictor variables and a dependent/outcome variable. The focus is on the significance of the impact that a change of an independent variable has on the dependent variable. A regression analysis can assist the researcher to make predictions either within or outside of the dataset (Gordon, 2015). The following variables from the dataset were regressed in this study; transactional data quality on transactional feedback, costs on transactional data quality and transactional feedback and monetary savings on source system data quality, feedback on correct transactions and total cost.

3.10 Validity and reliability

The plausibility of any research is measured by the validity and reliability of the data utilised. According to Kumar (2014:213,215), validity is "the ability of an instrument to measure what it is designed to measure" and a reliable tool is "consistent and stable, hence predictable and accurate". Therefore, a valid and reliable tool will always measure the variables it sets out to measure and produce the same results consistently if conditions are kept the same.

3.10.1 Validity

There are three common types of validity and these are, face and content validity, concurrent and predictive validity, and construct validity. The highly specialised environment and concise requirements of the questionnaire meant that, the researcher had to rely on face and content validity. This type of validity verifies that questions are linked to study objectives and thus measure what they are supposed to measure. It is normally judged by the researcher and experts in the field (Kumar, 2014). The researcher's intimate knowledge of the historical data, environment in which the technicians operate in, and the digital tools used to capture transaction feedback, ensured the face and content validity of the study. Both the face and content validity were done through a careful examination of the questions and subject matter experts in the work environment. The examination ensured that the questions were worded suitably for technicians to understand easily and that each question related to the objectives of this study.

Historical data was extracted by adhering to business rules to ensure only work orders related to this study were extracted. Extracted data was also checked for completeness. Any records that were incomplete and could not be normalised were excluded from analysis as they had the potential to skew the final results. More so, the study leader and statistician appraised the questions for their relevance to study objectives and their completeness for the types of analysis to be conducted.

3.10.2 Reliability

An instrument's reliability can be determined by external and/or internal consistency procedures (Sekaran & Bougie, 2014). The total FSOU field services technician population was approached to respond to the questionnaire, thus ensuring a representative population in order to guarantee statistical significance. However, repeated follow-ups were made to ensure a 35% nominal response rate from the population. The testing conditions were standard in the sense that all respondents were at work and received the same questions. In addition, scoring was constant due to the structured format of the questionnaire. The difficulty of the tests was low as the language used was clear and spoke directly to the technicians' environment.

The concise nature of the questionnaire meant that the Cronbach's alpha could only be to be calculated in instances where multiple questions related to the same concept as per Table 3.1. According to Andrew, Pedersen and McEvoy (2011:202), the purpose of a Cronbach's alpha is to measure "how well a set of variables or items measure a single, unidimensional latent construct". Values can range from 0 to 1, with a desired value between 0.7 and 0.9. One construct scored low at 0.114 and two high at 0.815 and 0.963. The overall internal consistency was generally acceptable but moderate at 0.638. Furthermore, the reliability of the secondary data was ensured through the performance of multiple extractions for the same t period. Record counts and completeness were checked and compared to ascertain that a reliable dataset had been extracted.

Table 3.1: Cronbach's alpha coefficients on questionnaire

Construct	Number of items	Cronbach's alpha	Comment
Transactional data quality - customer side fault Questions: 5, 8	2	0.114	Low internal consistency
Travel distance costs on Correct and Incorrect data Questions: 9, 10	2	0.963	High internal consistency
Work orders performed during Normal and Overtime for Correct and Incorrect data Questions 11, 12, 13, 14	4	0.815	High internal consistency
Overall questionnaire Questions: 4 to 18	15	0.638	Moderate internal consistency

3.11 Ethical considerations

An ethical clearance was received from senior management of Eskom Free State Operating Unit to access and analyse data, with the understanding that the data would be used for academic purposes and kept confidential. Consequently, historical data was aggregated when reporting the results to ensure the confidentiality of customers and financial figures as well as the anonymity of technician information. Data was stored on the Eskom network to ensure it was safe yet accessible to the researcher and statistician. Technicians place a high premium on their anonymity when answering questionnaires. Therefore, anonymity was ensured and the questionnaire questions were kept short and concise r to obtain as many responses as possible. In addition, the respondents were informed of the goals of this project and that their participation was voluntary as well as that the findings from this study would be used for the exclusive purpose of producing the dissertation.

3.12 Chapter summary

The chapter outlined the methodology applicable to this dissertation. It also discussed the research paradigm, research approach, research design and data collection techniques adopted in this study. It is important to note that data from two different population types, namely work orders from a historical dataset, and that from technician feedback extracted via a structured questionnaire was analysed. Within the limitations of the population considered, every effort was made to ensure validity and reliability of the collected and prepared datasets. The chapter also noted that a preliminary analysis was performed in Excel while a more detailed analysis was conducted in SPSS using both descriptive and inferential statistics.

The following chapter presents results of the survey and historical dataset analysis via descriptive and inferential statistics.

4 CHAPTER 4: PRESENTATION OF FINDINGS

CHAPTER 4: PRESENTATION OF FINDINGS

- **4.1 Introduction**
- **4.2 Survey results**
  - *4.2.1 Overview of descriptive data and analysis of survey results*
  - *4.2.2 Inferential statistics of survey results*
  - 4.2.3 Summary of survey results
- **4.3 Results on historical data**
  - *4.3.1 Overview of data summary*
  - *4.3.2 Comparison of transactional data quality: source system with feedback*
  - *4.3.3 Relationship between costs and transactional feedback received and transactional data quality: source system*
  - *4.3.4 Compare cost impact of incorrect transactions to transactional data quality: feedback received and transactional data quality: source system*
  - *4.3.5 Relationship between monetary saving, total cost, transactional data quality: source system and transactional data quality: feedback received*
  - *4.3.6 Summary of historical dataset*
- **4.4 Chapter summary**

4.1 Introduction

In the previous chapter, a breakdown of the methodology applicable to this dissertation was presented and the quantitative nature of this survey was emphasised. The study also described how the data was sourced and which statistical tools were employed to analyse the data. The focus of this chapter is to present results from the statistical analysis applied to the data collected using a questionnaire. The chapter also analyses the historical dataset.

4.2 Survey results

The chapter's sections present the survey results and analysis drawn from 106 responses made by the field technicians. The chapter also presents some descriptive data overview, and descriptive and inferential statistics.

4.2.1 Overview of descriptive data and analysis of survey results

A summary of the survey results organised according to the research questions posed in 1.4.2 is presented in the sections below in conjunction with the analysis performed on appropriate sections.

4.2.1.1 Background variables

The analysed background variables related to demographics and volume of work orders received on a monthly basis. Only two questions linked to demographics were posed to the respondents in order to ensure maximum anonymity. Age, race, experience, seniority and gender can all be used to determine who a person is, especially at smaller CNCs, consequently these questions were avoided to ensure a maximum response rate. The results from an examination of the years of experience at an Eskom CNC, shown in Figure 4.1 and Table 4.1, show that 40.6% of the technicians had 6-10 years, 27.4% had 0-5 years, 21.7% more than 20 years, 8.5% 11-15 years and only 1.9% had 15-20 years. Nonetheless, most of the technicians had more than 5 years of work experience, which implies that the majority of the workforce had sufficient work exposure needed to gain the necessary skills required to fulfil their current responsibilities.

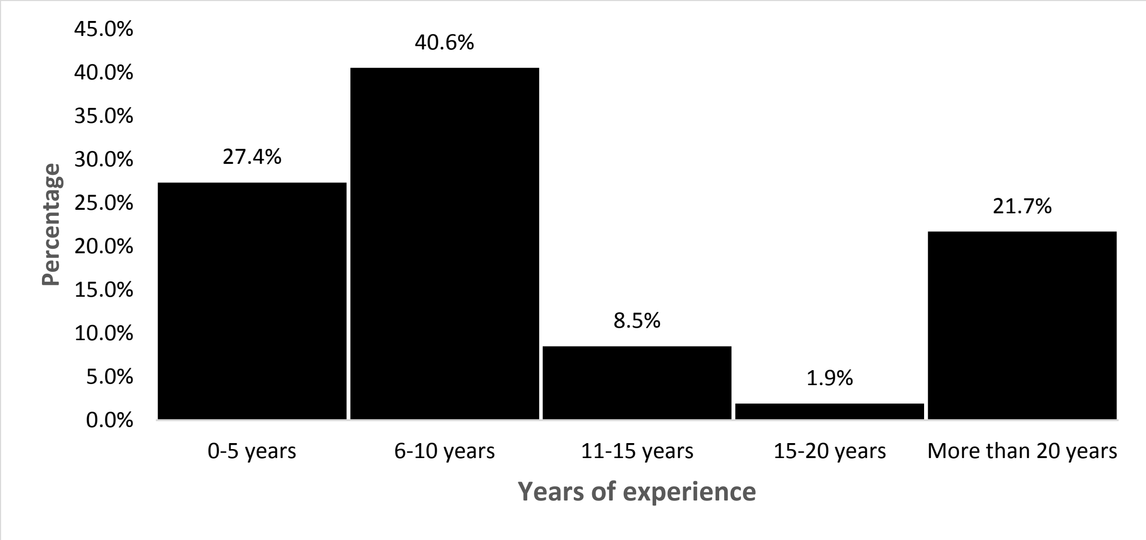


Figure 4.1: Distribution of years of experience at Eskom CNC

Eskom adopted the Enterprise digital assistant (EDA) devices towards the end of the 2011-2012 financial year. Thus, the maximum possible experience a technician has on EDA usage is around 5 years. The distribution analysis of EDA usage, depicted in Figure 4.2 and Table 4.1, reveals that only 8.5% of the technicians had EDA usage experience of 1 year or less, whilst 53.8% had more than 4 years of experience. The EDA usage experience is sufficient for effective utilisation of this tool based on this distribution analysis.

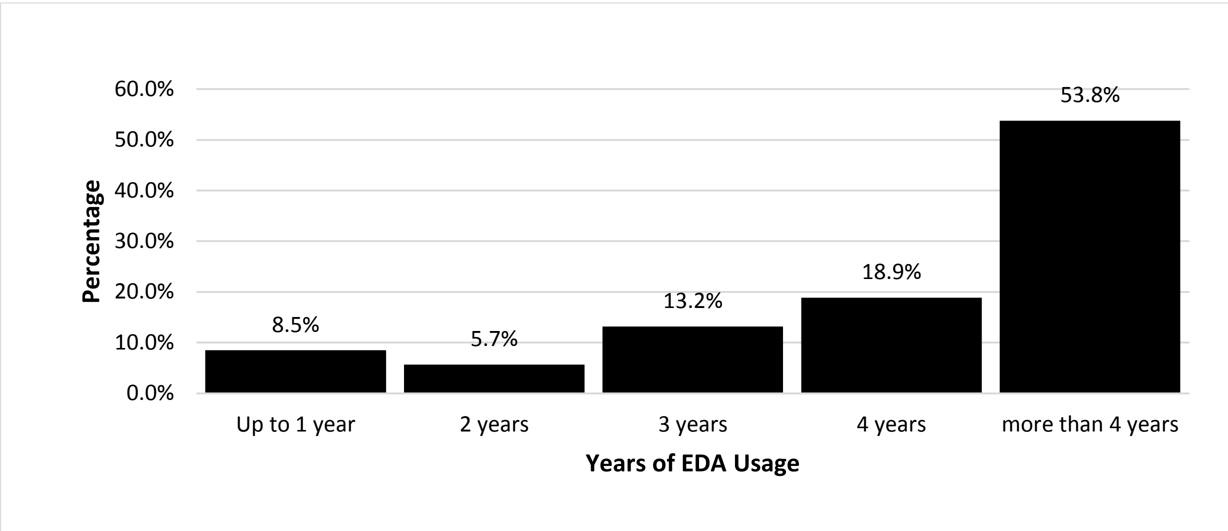


Figure 4.2: Distribution of years of EDA usage

Figure 4.3 and Table 4.1 indicate that the bulk of technicians received less than 20 single customer dispatch work orders per month, 46.2% had 1-10 and 31.1% had

11-20 work orders. Therefore, 77.3% of the technicians received an average of one single customer dispatch work order per working day per month. This implies that technicians were not overburdened with single customer dispatch work orders and could properly evaluate work and provide feedback on the executed work.

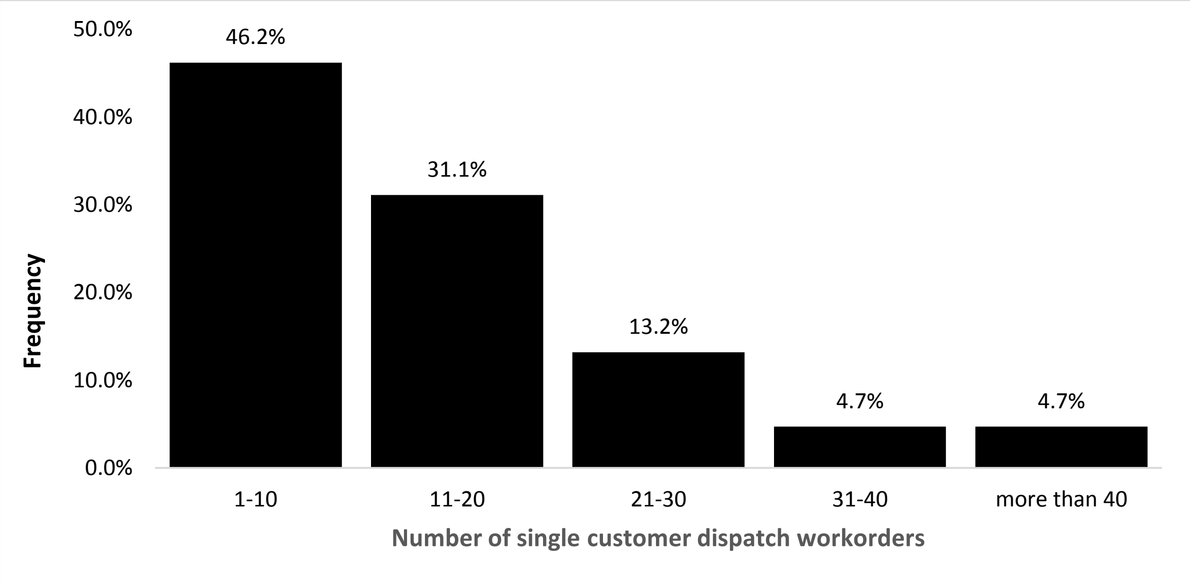


Figure 4.3: Distribution of single customer dispatch work orders received from dispatch per month

Table 4.1: Background variables

Background variables	Category	Frequency	Percent
B1: Please indicate your years of work experience at your Eskom CNC.	0-5 years	29	27.4%
	6-10 years	43	40.6%
	11-15 years	9	8.5%
	15-20 years	2	1.9%
	more than 20 years	23	21.7%
B2: For how many years have you been using an EDA?	Up to 1 year	9	8.5%
	2 years	6	5.7%
	3 years	14	13.2%
	4 years	20	18.9%
	more than 4 years	57	53.8%
B3: How many single customer dispatch work orders do you receive from dispatch per month?	1-10	49	46.2%
	11-20	33	31.1%
	21-30	14	13.2%
	31-40	5	4.7%
	more than 40	5	4.7%

4.2.1.2 Source system data quality

Questions B4, B5 and B8 were, as shown in Table 4.2, used to measure the technician's perceptions of the quality of data received from the source system. The average for B4 was 53.6% and represents overall quality of transactional data from the contact centre. B5 specifically focussed on customer side fault work orders and was posed in an inverse manner. A lower average percentage implies higher data quality percentage from the source system. The average was 31.5 % and this meant that 68.5% of the faults were not related to customer side faults and thus were of good quality. B8 also focussed specifically on customer side faults and 55.7% agreed or strongly agreed that the Contact Centre can improve their ability to identify customer side faults.

Table 4.2: Source system data quality

Source System Data quality		Percentage Categories					
		0-20%	21-40%	41-60%	61-80%	80-100%	Average Percentage
B4: On average, how accurate is the information on the single customer dispatch work orders you receive?	Count	15	21	17	36	17	**53.6%**
	%	14.2%	19.8%	16.0%	34.0%	16.0%	
B5: On average, what percentage of the single customer dispatch work orders that you receive per month, are caused by a fault on the customer side (thus no fault on Eskom side)?	Count	37	37	22	7	3	**31.5%**
	%	34.9%	34.9%	20.8%	6.6%	2.8%	
B8: Please indicate the extent to which you agree with whether the Eskom Contact Centre can do a better job to identify faults that are caused by the customer?		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree+ Strongly agree
	Count	15	6	26	37	22	**55.7%**
	%	14.2%	5.7%	24.5%	34.9%	20.8%	

4.2.1.3 Feedback on transactional data

Questions B6 and B7 focused on the appropriate usage of feedback that indicated a customer side fault as per Table 4.3. The responses to question B6 reveal that feedback indicating a customer side fault was used appropriately for the 52.8% of work orders that were handled. Consequently, in 47.2% of the work orders, customer side faults could not be indicated and thus went unnoticed. Conversely, an analysis of responses to question B7 illustrates that faults on Eskom's side are incorrectly classified as a customer side fault in 25.3% of the cases. The customer side fault feedback indicator is thus not always used correctly. However, the incorrect classification of Eskom faults in B7 tended to balance out the 47.2% of customer side faults which were not indicated as such. The accuracy of the customer side fault feedback indicator is questionable and reveals significant potential for improvement.

Table 4.3: Feedback analysis

Feedback analysis (Based on customer side fault)		Percentage Categories					
		0-20%	**21-40%**	**41-60%**	**61-80%**	**80-100%**	**Average Percentage**
B6: On average, what percentage of the single customer dispatch work orders that are caused by a fault on the customer side, do you mark as "Customer side fault" when putting the work-order in MILE3?	Count	23	26	10	7	40	**52.8%**
	%	21.7%	24.5%	9.4%	6.6%	37.7%	
B7: On average, what percentage of the single customer dispatch work orders that are caused by a fault on Eskom side, do you mark as "Customer side fault" when putting the work order in MILE3?	Count	65	21	8	4	8	**25.3%**
	%	61.3%	19.8%	7.5%	3.8%	7.5%	

4.2.1.4 Cost of transactional data

The main costs for work orders are labour and travel costs. Labour cost is more expensive during overtime when compared to normal work hours due to the higher overtime rates that are applicable. Travel costs comprise of time spent driving that has a higher cost if performed during overtime than normal work hours, and a vehicle cost per kilometre travelled, which is constant and not influenced by overtime or normal work hours. Data transactions are deemed to be correct if a work order is not marked with feedback of a customer side fault. An analysis of the responses to question B11 indicates that 53% of work based on transactions with correct data is performed during normal work hours, whilst those for B13 reveal that 48.9% of correctly captured transaction work is performed during overtime as shown in Table 4.4.

Table 4.4: Correct data transactions during normal work time/overtime

Correct Data transactions (During normal work time/overtime)		Percentage Categories 0-20%	21-40%	41-60%	61-80%	80-100%	Average Percentage
B11: On average, what percentage of the single customer dispatch work orders that are caused by a fault on Eskom side do you respond to during **normal work time**?	Count	17	13	37	15	24	**53.0%**
	%	16.0%	12.3%	34.9%	14.2%	22.6%	
B13: On average, what percentage of the single customer dispatch work orders that are caused by a fault on Eskom side do you respond to during **overtime**?	Count	15	28	27	20	16	**48.9%**
	%	14.2%	26.4%	25.5%	18.9%	15.1%	

Similarly, an analysis of the responses to question B12 shows, in Table 4.5 that, 50% of incorrect data transactions are executed during normal work hours while 45.5% are executed during overtime. Both correct and incorrect transactions show a similar distribution in terms of time executed. Consequently, work time and overtime should not significantly impact cost calculation of one category (read either correct or incorrect transaction) over the other.

Table 4.5: Incorrect data transactions during normal work time/overtime

Incorrect Data transactions (During normal work time/overtime)		Percentage Categories 0-20%	21-40%	41-60%	61-80%	80-100%	Average Percentage
B12: On average, what percentage of the single customer dispatch work orders that are caused by a fault on the customer side do you respond to **during normal work time**?	Count	19	20	31	14	22	**50.0%**
	%	17.9%	18.9%	29.2%	13.2%	20.8%	
B14: On average, what percentage of the single customer dispatch work orders that are caused by a fault on the customer side do you respond to **during overtime**?	Count	18	31	29	13	15	**45.5%**
	%	17.0%	29.2%	27.4%	12.3%	14.2%	

The analysis of travel distance for B9 and B10 (i.e. correct and incorrect transactions) in Table 4.6 indicates that travelling to execute correct transactions is a bit further than incorrect transactions, but the difference is small. Consequently, distance will not have a major impact on the cost calculation of correct versus incorrect transactions. However, the feedback in B15 is interesting. Technicians indicated that they expend further labour cost on 18.9% of incorrect transactions by fixing a customer's fault, even though they are not supposed to.

Table 4.6: Total travel distance and additional costs incurred

Total travel distance for correct or incorect data (km)		Percentage Categories					
		<20 Km	20-40 Km	40-60 Km	60-80 Km	>80 Km	Average Distance
B9: On average, how many kilometres do you travel to a single customer dispatch work order that is caused by a fault on Eskom side?	Count	6	14	19	30	37	**64.7 Km**
	%	5.7%	13.2%	17.9%	28.3%	34.9%	
B10: On average how many kilometres do you travel to a single customer dispatch work order that is caused by a fault on the customer side?	Count	7	17	17	29	36	**63.2 Km**
	%	6.6%	16.0%	16.0%	27.4%	34.0%	
Extra labour costs incurred for customer side faults		Never	Rarely	Sometimes	Frequently	Always	**% Frequently + always**
B15: Please indicate the frequency at which you repair single customer dispatch work orders that are caused by a fault on the customer side (rather than just putting it on MILE3, with customer side fault feedback).	Count	41	19	26	8	12	**18.9%**
	%	38.7%	17.9%	24.5%	7.5%	11.3%	

The results from an analysis of responses to the question on whether f the feedback of customer side fault can be used to determine the cost of such work orders to the business (see statement B16 indicted in Table 4.7) show that, 52.8% of technicians agreed or strongly agreed that it is indeed possible. This figure concurs with the accuracy of the usage of the customer side fault indicator at B6 in Table 4.3, which is also 52.8%. However, only 37.7% felt that Eskom was actually using customer side fault feedback to determine the cost of such work orders (see statement B17). Thus, even though over 50% of the technicians indicated that the customer side fault feedback could be used to determine its cost impact on the business, the confidence levels in the business' ability to utilise feedback in order to gain this insight was significantly lower. This lack of confidence could have an impact on the technicians' motivation to provide appropriate feedback on work orders.

Table 4.7: Usage of feedback to determine cost

Usage of feedback to determine cost		Percentage Categories					
		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree+ Strongly agree
B16: Please indicate the extent to which you agree that it is possible to use the MILE3 feedback of customer side fault to determine how much such work orders cost the business.	Count	7	10	33	31	25	**52.8%**
	%	6.6%	9.4%	31.1%	29.2%	23.6%	
B17: Please indicate the extent to which you perceive Eskom to be using the MILE3 feedback of customer side fault to determine how much such work orders cost the business.		**Never**	**Rarely**	**Sometimes**	**Frequently**	**Always**	**% Frequently + always**
	Count	19	11	36	13	27	**37.7%**
	%	17.9%	10.4%	34.0%	12.3%	25.5%	

4.2.1.5 Monetary savings

The main purpose of statement B18 was to determine if money could be saved by preventing the dispatch of a technician to a customer side fault. The results inTable 4.8 indicate that the majority (67.9%) of technicians agree or strongly agree that money can be saved if they do not have to respond to a customer side fault. By implication, this signifies that an improvement in customer call transactional data quality can result in monetary savings, as it will prevent customer side faults being logged as ESP faults, thereby preventing costs associated with such transactions.

Table 4.8: Monetary savings underlying the prevention of incorrect transactional data quality

Monetary savings embedded within prevention of incorrect transactional data quality		Percentage Categories					
		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Agree+ Strongly agree
B18: Please indicate the extent to which you agree or disagree that it is possible for Eskom to save money if you do not have to go to single customer dispatch work orders that are caused by a fault on the customer side.	Count	11	6	17	30	42	**67.9%**
	%	10.4%	5.7%	16.0%	28.3%	39.6%	

The last question posed to technicians, B19, queried the amount of money in Rand value they perceived could be saved per month if they desist from responding to a customer side fault. Unfortunately, the answers varied widely and were often not in Rand value as requested, but rather statements such as, I do not know, thousands, a lot, plenty, millions etc. Consequently, no average could be computed from the answers supplied. The extent of the answers indicated that most technicians were unsure of the monetary costs involved to service ESP and customer side faults.

However, a cost comparison model could be generated from a combination of data supplied in the questionnaire and used to calculate the cost of valid ESP versus customer side faults. This data includes the average amount of work orders received, percentage of customer side faults versus percentage valid ESP faults, average travel distance, percentage of faults responded to during normal and overtime, average cost per km travelled to faults and the minimum labour rate per fault during normal and overtime. Table 4.9 shows that customer side faults contribute 47.7% to the total cost of work orders during a month, with the portion of 29.1% of the costs contributed during overtime.

Table 4.9: Monetary cost contribution of work orders with correct and incorrect data

<table>
<tr><th>Monetary cost contribution of work orders with correct and incorrect data</th><th>Time category</th><th>Amount of Work orders per category</th><th>Category Total Work orders</th><th>% Cost contribution per category</th><th>% of Total Cost Contribution</th></tr>
<tr><td rowspan="2">Work orders for valid ESP</td><td>Normal work hours</td><td>798</td><td rowspan="2">1535</td><td>20.3%</td><td rowspan="2">52.3%</td></tr>
<tr><td>Overtime</td><td>737</td><td>32.0%</td></tr>
<tr><td rowspan="2">Work orders for customer side faults</td><td>Normal work hours</td><td>730</td><td rowspan="2">1398</td><td>18.6%</td><td rowspan="2">47.7%</td></tr>
<tr><td>Overtime</td><td>668</td><td>29.1%</td></tr>
<tr><td></td><td></td><td colspan="2">2933</td><td colspan="2">100%</td></tr>
</table>

4.2.1.6 Descriptive analysis summary

A descriptive analysis was performed on the questionnaire results. The findings from this descriptive analysis of the questionnaire results are as follows:

- B5 in Table 4.2 indicates that 31.5% of the ESP calls logged transactions captured are related to customer side faults. Consequently, the money spent on 31.5% of the transactions can be avoided if transactional data quality at the source system is improved.
- Feedback accuracy of customer side faults is 52.8% (see B6 in Table 4.3). Subsequently, the impact of customer side faults might be understated. Furthermore, there exists potential to improve feedback accuracy by up to 47.2% (the percentage shortfall) on customer side faults.
- The accuracy of feedback on ESP faults (see B7) can also be improved as 25.3% of these faults are incorrectly tagged as customer side faults. The 25.3% might assist in actual cost calculation through an understating of customer side faults. It is, however, preferable to increase accuracy of feedback.

- 45.5% of customer side fault work orders are executed during overtime as per Table 4.5. This significantly inflates the cost of such work orders due to the labour cost increase experienced during overtime situations.
- The average travel distance in Table 4.6 (see B10) for customer side faults is 63.2km which is close to the travel distance of 64.7km (see B9) for ESP faults.
- In Table 4.8, 67.9% of the technicians agreed or strongly agreed that there are monetary savings embedded within customer side faults.
- Based on the results in Table 4.9, customer side faults contribute 47.7% of the total work order costs. Thus, an improvement of transactional data quality at the source system seeking to eliminate customer side faults can lead to the achievement of a 47.7% potential monetary saving on total work order costs by means of cost avoidance.

4.2.2 Inferential statistics of survey results

An inferential statistical analysis was applied after the descriptive summary and analysis to determine if any statistically significant relationships exist between the constructs under investigation.

4.2.2.1 Relationship between source system data quality and feedback

The relationship between source system data quality and transactional data quality was analysed using a correlation analysis of the questionnaire items that address both variables. The results, s in Table 4.10 below, show that there is no significant relationship between source system data quality and transactional data quality (p-values>0.05).

Table 4.10: Correlations between Source system data quality and feedback

Correlations		B4: On average, how accurate is the information on the single customer dispatch work orders that you receive?	Comment
B5: On average, what percentage of the single customer dispatch work orders that you receive per month, are caused by a fault on the customer side (thus no fault on Eskom side)?	Correlation	-0.101	Not significant
	p-value	0.305	
	N	106	
B8: Please indicate the extent to which you agree or disagree whether the Eskom Contact Centre can do a better job to identify faults that are caused by the customer.	Correlation	-0.097	Not significant
	p-value	0.325	
	N	106	

4.2.2.2 Relationship between source system data quality and costs

A correlational analysis was used to test the relationship between source system data quality and various costs. The tested costs were travel distance and percentage work orders for valid ESP and customer side faults during work hours and overtime. The results are presented in Table 4.11 and Table 4.12 and they show that no significant relationship exists between source system data quality and any of the tested costs.

Table 4.11: Correlation between source system data quality and costs: Travel

Correlations		B4: On average, how accurate is the information on the single customer dispatch work orders that you receive?	Comment
B9: On average, how many kilometres do you travel to a single customer dispatch work order that is caused by a fault on Eskom side?	Correlation	0.112	Not significant
	p-value	0.253	
	N	106	
B10: On average, how many kilometres do you travel to a single customer dispatch work order that is caused by a fault on the customer side?	Correlation	0.099	Not significant
	p-value	0.314	
	N	106	

Table 4.12: Correlation between source system data quality and costs: Percentage work orders

Correlations		B4: On average, how accurate is the information on the single customer dispatch work orders that you receive?	**Comment**
B11: On average, what percentage of the single customer dispatch work orders that are caused by a fault on Eskom's side do you respond to during normal work time?	Correlation	0.152	Not significant
	p-value	0.121	
	N	106	
B12: On average, what percentage of the single customer dispatch work orders that are caused by a fault on the customer side do you respond to during normal work time?	Correlation	0.142	Not significant
	p-value	0.145	
	N	106	
B13: On average, what percentage of the single customer dispatch work orders that are caused by a fault on Eskom's side do you respond to during overtime?	Correlation	0.046	Not significant
	p-value	0.642	
	N	106	
B14: On average, what percentage of the single customer dispatch work orders that are caused by a fault on the customer's side do you respond to during overtime?	Correlation	-0.038	Not significant
	p-value	0.699	
	N	106	

4.2.2.3 Relationship between feedback and costs

The relationship between feedback and costs were tested by means of correlational analysis. The costs analysed were travel distance and percentage work orders related to ESP and customer side faults actioned during normal work time and overtime. No significant relationship exists between feedback and travel distance in Table 4.13.

Table 4.13: Correlation between feedback and travel distance

Correlations		B5: On average, what percentage of the single customer dispatch work orders you receive per month, are caused by a fault on the customer side (that is, no fault on Eskom side)?	**Comment**
B9: On average, how many kilometres do you travel to a single customer dispatch work order that is caused by a fault on Eskom's side?	Correlation	0.150	Not significant
	p-value	0.125	
	N	106	
B10: On average, how many kilometres do you travel to a single customer dispatch work order that is caused by a fault on the customer side?	Correlation	0.177	Not significant
	p-value	0.069	
	N	106	

Table 4.14 shows that a significant positive relationship was established between feedback from technicians and work orders responded to during overtime. Thus, more work orders are handled during overtime if there is a higher occurrence of feedback indicating customer side faults. Nonetheless, the handling of more work orders during overtime will increase the overall transaction costs due to the inherent higher labour costs of overtime transactions.

Table 4.14: Correlation between feedback and ESP, customer side faults during work hours and overtime

Correlations		B5: On average, what percentage of the single customer dispatch work orders that you receive per month, are caused by a fault on the customer side (that is, no fault on Eskom's side)?	Comment
B11: On average, what percentage of the single customer dispatch work orders caused by a fault on Eskom's side do you respond to during normal work time?	Correlation	0.067	Not significant
	p-value	0.496	
	N	106	
B12: On average, what percentage of the single customer dispatch work orders caused by a fault on the customer side do you respond to during normal work time?	Correlation	0.127	Not significant
	p-value	0.195	
	N	106	
B13: On average, what percentage of the single customer dispatch work orders caused by a fault on Eskoms side do you respond to during overtime?	Correlation	0.240*	Significant
	p-value	0.013	
	N	106	
B14: On average, what percentage of the single customer dispatch work orders caused by a fault on the customer side do you respond to during overtime?	Correlation	0.344**	Significant
	p-value	0.000	
	N	106	

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

4.2.2.4 Relationship between costs and monetary savings

Correlational analyses were used to determine the relationships between costs and monetary savings. The costs that were analysed were travel distance and percentage work orders related to ESP and customer side faults actioned during normal work time and overtime. No significant relationships were identified between any of the costs and monetary savings in Table 4.15 and Table 4.16.

Table 4.15: Correlation between costs on kilometres travelled and monetary savings

Correlations		B18: Please indicate the extent to which you agree or disagree that it is possible for Eskom to save money if you do not have to go to a single customer dispatch work order that is caused by a fault on the customer's side.	**Comment**
B9: On average, how many kilometres do you travel to a single customer dispatch work order that is caused by a fault on Eskom's side?	Correlation	0.028	Not significant
	p-value	0.775	
	N	106	
B10: On average, how many kilometres do you travel to a single customer dispatch work order that is caused by a fault on the customer's side?	Correlation	0.018	Not significant
	p-value	0.852	
	N	106	

Table 4.16: Correlation between percentage work order costs and monetary savings

Correlations		B18: Please indicate the extent to which you agree or disagree that it is possible for Eskom to save money if you do not go to a single customer dispatch work order that is caused by a fault on the customer ,side.	**Comment**
B9: On average, how many kilometres do you travel to a single customer dispatch work order that is caused by a fault on Eskom's side?	Correlation	0.028	Not significant
	p-value	0.775	
	N	106	
B10: On average, how many kilometres do you travel to a single customer dispatch work order that is caused by a fault on the customer's side?	Correlation	0.018	Not significant
	p-value	0.852	
	N	106	
B11: On average, what percentage of the single customer dispatch work orders that are caused by a fault on Eskom's side do you respond to during normal work time?	Correlation	-0.022	Not significant
	p-value	0.826	
	N	106	
B12: On average, what percentage of the single customer dispatch work orders that are caused by a fault on the customer's side do you respond to during normal work time?	Correlation	-0.106	Not significant
	p-value	0.277	
	N	106	
B13: On average, what percentage of the single customer dispatch work orders that are caused by a fault on Eskom's side do you respond to during overtime?	Correlation	0.028	Not significant
	p-value	0.777	
	N	106	
B14: On average, what percentage of the single customer dispatch work orders that are caused by a fault on the customer's side do you respond to during overtime?	Correlation	0.163	Not significant
	p-value	0.094	
	N	106	

4.2.2.5 Relationship between quality of source system data and monetary savings

A correlational analysis was performed in order to determine the relationship between the quality of source system data and monetary savings. The results of the analysis, depicted in Table 4.17, indicate that no significant relationship exists between the source system data quality and monetary savings.

Table 4.17: Correlation between source system data quality and monetary savings

Correlations		B4: On average, how accurate is the information on the single customer dispatch work orders that you receive?	Comment
B18: Please indicate the extent to which you agree or disagree that it is possible for Eskom to save money if you do not have to go to a single customer dispatch work order that is caused by a fault on the customer side.	Correlation	0.025	Not significant
	p-value	0.795	
	N	106	

4.2.2.6 Relationship between feedback and monetary savings

A correlational analysis was also performed to determine the relationship between feedback indicating incorrect transactions and monetary savings. The results from the analysis, which are presented in Table 4.18, indicate that no significant relationship exists between feedback indicating incorrect transactions and monetary savings.

Table 4.18: Correlation between feedback and monetary savings

Correlations		B5: On average, what percentage of the single customer dispatch work orders that you receive per month, are caused by a fault on the customer side (thus no fault on Eskom side)?	Comment
B18: Please indicate the extent to which you agree or disagree that it is possible for Eskom to save money if you do not go to single customer dispatch work orders that are caused by a fault on the customer side.	Correlation	-0.024	Not significant
	p-value	0.803	
	N	106	

4.2.2.7 Inferential analysis summary

For all the relationships tested during inferential analysis, one relationship with statistical significance was uncovered between feedback and costs and this was, feedback and work orders responded to during overtime, as reflected in Table 4.14. One can infer that an increase in feedback on customer side faults received can result in increased costs for ESP and customer side faults.

4.2.3 Summary of survey results

The descriptive analysis revealed that, an improvement in customer call transactional data quality at the source can create savings of up to 47.7% for transactions related to customer calls requesting service for an ESP. Inferential analysis revealed one significant relationship between feedback and costs, but no significant relationship between source system transactional data quality and monetary savings. The conciseness of the questionnaire, moderate reliability indicated by Cronbach's alpha test, most technicians' inability to indicate a rand value for expected monetary savings and possible bias in technician's opinion could all play a role in the lack of significant relationships amongst variables tested. Therefore, an analysis of the historical dataset can prove to be valuable as none of the factors influencing the questionnaire will influence the historical dataset analysis results.

4.3 Results on historical data

The historical dataset containing 235 945 records spanned from April 2012 to March 2017. A financial year within Eskom starts on 1 April of a typical year and ends on 31 March of the following year. Therefore, five financial years were analysed and these are, 2012/2013, 2013/2014, 2014/2015, 2015/2016 and 2016/2017. Eskom Distribution Free State comprises a hierarchical structure with three subsequent levels, which are Zones x 2, Sectors x 4 and CNCs x 25.

4.3.1 Overview of data summary

A descriptive overview of the data summary for all financial years combined and disaggregated per hierarchy level, displayed in Table 4.19, contains the following information:

- **Work order:** Total amount of work orders analysed.
- **Percentage work order:** Percentage contributed to the total amount of work orders.
- **Percentage work time:** Percentage of total work orders that were actioned during normal work hours.
- **Percentage over time:** Percentage of total work orders that were actioned during overtime.

- **Source system data quality:** Percentage quality of source system data measured for the work orders.
- **Percentage feedback correct transactions:** Percentage of work orders that were marked as having correct transactional data.
- **Percentage travel distance:** Percentage of total kilometres travelled for the work orders.
- **Percentage time cost:** Percentage contribution to the overall time cost.
- **Percentage total cost:** Percentage contribution to the total monetary cost after the conversion of costs to a monetary value.

On a zone level, Bloemfontein contributes 60.89% to the total work order volume, which is 21.78% more than Bethlehem zone's contribution of 39.11%. The Bloemfontein Sector has the highest work order contribution on a sector level with 31.51% while the Kroonstad Sector has the lowest contribution at 13.58%. At CNC level, Bloemfontein zone has five CNC's with a total work order contribution higher than 5% in comparison to Ladybrand, Selosesha, Hoopstad, Thabong and Welkom towns. The Bethlehem zone only has Bohlokong CNC contributing more than 5%. In addition, the Bloemfontein zone has 13 CNCs compared to Bethlehem Zone's 11 and 4 CNCs more than the Bethlehem Zone thus, contributing greater than 5% to the total work orders. It is therefore, understandable that Bloemfontein has a much higher total work order contribution than the Bethlehem Zone.

The quality of source system data for all hierarchy levels has a very narrow distribution. It consists of a mean of 80%, a maximum deviation from the mean of 0.2% and a minimum deviation of 0%. The feedback indicating correct transactions has a mean of 81.3% with the biggest deviation from the mean being 14% at the Virginia CNC and the closest deviation of 0.1% observed at Welkom Town CNC. The Bloemfontein Zone executed most (54.8%) of its work orders during normal work time, whereas the Bethlehem Zone performed the bulk (52%) of its work during overtime. The reality that labour performed during overtime is more expensive than that is done during normal work hours explains why the Bethlehem Zone's time cost contribution of 42% is higher than its total work order contribution of 39.11%. Similarly, the Bloemfontein Zone's time cost contribution of 58% is less than its total work order contribution of 60.89%, which can be attributed to the fact that most if its work was performed during normal work hours. Travel distance contribution is affected by the travel distance from the CNC to its customers as well as the volume of work orders. CNCs with travel distance contributions higher than their overall work order contribution, indicate that the travel distance from these CNCs to their customers is greater than CNCs where the travel contribution is equal to or less than the total work order contribution.

Table 4.19: Total work order volume distribution

Hierachy	Work order	% Work order	% Work Time	% Over Time	% Source system Data Quality	% Feedback Correct Transaction	% Travel Distance km	% Time cost	% Total cost
Bethlehem Zone	**92280**	**39.11%**	**48.0%**	**52.0%**	**80.0%**	**82.6%**	**39.5%**	**42.0%**	**41.0%**
Bethlehem Sector	***60243***	***25.53%***	***50.1%***	***49.9%***	***80.0%***	***82.0%***	***21.8%***	***26.2%***	***24.3%***
Bethlehem CNC	7996	3.39%	47.4%	52.6%	79.9%	83.8%	4.8%	3.9%	4.2%
Bohlokong CNC	14775	6.26%	49.5%	50.5%	79.8%	80.6%	2.1%	6.1%	4.4%
Harrismith CNC	9956	4.22%	44.9%	55.1%	80.1%	82.1%	3.5%	4.7%	4.2%
Reitz CNC	8185	3.47%	51.1%	48.9%	80.1%	84.9%	2.8%	3.5%	3.2%
Senekal CNC	9204	3.90%	58.4%	41.6%	80.1%	79.1%	4.9%	3.7%	4.2%
Vrede CNC	10127	4.29%	49.7%	50.3%	80.1%	83.2%	3.9%	4.4%	4.2%
Kroonstad Sector	***32037***	***13.58%***	***44.1%***	***55.9%***	***80.1%***	***83.6%***	***17.7%***	***15.8%***	***16.6%***
Bothaville CNC	8206	3.48%	45.7%	54.3%	80.1%	83.5%	4.7%	4.0%	4.3%
Frankfort CNC	7129	3.02%	42.4%	57.6%	80.2%	84.8%	5.4%	3.7%	4.5%
Kroonstad CNC	4260	1.81%	49.4%	50.6%	80.2%	82.4%	2.7%	2.0%	2.3%
Parys CNC	3952	1.67%	44.6%	55.4%	80.0%	79.3%	1.7%	1.9%	1.8%
Vaalpark CNC	8490	3.60%	41.2%	58.8%	80.0%	85.3%	3.1%	4.2%	3.7%
Bloemfontein Zone	**143665**	**60.89%**	**54.8%**	**45.2%**	**80.0%**	**80.5%**	**60.5%**	**58.0%**	**59.0%**
Bloemfontein Sector	***74352***	***31.51%***	***53.3%***	***46.7%***	***80.0%***	***80.1%***	***34.5%***	***31.0%***	***32.5%***
Bloemfontein CNC	9001	3.81%	45.9%	54.1%	80.1%	78.3%	6.7%	4.5%	5.5%
Edenburg CNC	4798	2.03%	49.2%	50.8%	79.9%	76.7%	3.2%	2.3%	2.7%
Jacobsdal CNC	8819	3.74%	46.7%	53.3%	80.1%	80.4%	2.7%	4.0%	3.4%
Ladybrand CNC	13574	5.75%	50.6%	49.4%	79.9%	77.0%	8.4%	6.2%	7.2%
Petrusburg CNC	5471	2.32%	48.3%	51.7%	80.0%	84.1%	2.1%	2.4%	2.3%
Selosesha CNC	23311	9.88%	64.1%	35.9%	80.0%	81.8%	3.6%	6.9%	5.5%
Zastron CNC	9378	3.97%	48.4%	51.6%	80.0%	81.5%	7.7%	4.7%	6.0%
Welkom Sector	***69313***	***29.38%***	***56.5%***	***43.5%***	***80.0%***	***80.8%***	***26.0%***	***27.0%***	***26.6%***
Alma CNC	1808	0.77%	37.5%	62.5%	80.1%	95.0%	0.3%	0.9%	0.7%
Duiker CNC	9515	4.03%	54.4%	45.6%	80.0%	86.5%	6.6%	4.3%	5.3%
Hoopstad CNC	16886	7.16%	48.7%	51.3%	80.1%	79.0%	11.6%	8.1%	9.6%
Thabong CNC	22542	9.55%	58.9%	41.1%	79.9%	77.2%	1.7%	7.5%	5.0%
Virginia CNC	1939	0.82%	38.2%	61.8%	79.8%	95.3%	0.3%	0.9%	0.7%
Welkom Towns CNC	16623	7.05%	66.5%	33.5%	80.0%	81.2%	5.3%	5.2%	5.2%
Grand Total	**235945**	**100.00%**	**52.2%**	**47.8%**	**80.0%**	**81.3%**	**100.0%**	**100.0%**	**100.0%**

4.3.2 Comparison of transactional data quality: source system with feedback

Data quality measured at the source system is based on results from a quality assurance process performed on a 1% sample of the total transactions for a specific month. This quality assurance process consumes a lot of time. Therefore, no more than 1% of the total transactions can be assessed. The quality assurance process evaluates call professionalism when answering calls, whether correct steps were followed to: identify a customer, interpret his fault symptoms, send the fault to the correct department and lastly, if the call was concluded professionally. Call professionalism, which forms part of the score, does not directly associate with transactional data quality, but call content related to transactional data quality comprises 70% of the total score. As a result, most scoring relates to transactional data quality, which is the reason why this score is used as a proxy for transactional data quality at the source system. A comparison of transactional data quality with feedback from technicians reveals whether feedback is impacted by source system data quality or not. Transactional feedback from technicians in the field indicates the correctness of work order/transactional data quality if it describes a valid resolution to an ESP. Feedback stating a customer side fault indicates correctness of the work order/transactional data quality. Consequently, transactional feedback on correct transactions can also serve as a measurement of transactional data quality, but from the data consumer's (field technician's) perspective.

Descriptive analysis was applied in Table 4.20 where each financial year's source system transactional data quality and feedback on correct transactions is compared to its average for all years combined. A value higher than the average is highlighted in green whilst a value lower than the average is highlighted in pink. The colours for source system data quality and feedback indicating correct transactions correspond in the financial years 2014/2015, 2015/2016 and 2016/2017. From this measurement, 60% of the periods analysed showed a corresponding increase/decrease in both source system transactional data quality and feedback indicating correct transactions.

Table 4.20: Source system transactional data quality vs feedback on correct transactions per financial year

Financial Year	Source system transactional data quality	Feedback: % Correct transactions
2012-2013	77.99%	81.37%
2013-2014	80.51%	80.78%
2014-2015	78.26%	80.42%
2015-2016	81.76%	82.50%
2016-2017	81.72%	81.50%
Average	**80.05%**	**81.31%**

Figure 4.4 displays data quality from both source system and transactions with feedback indicating correct data. Transactional data quality from the source system indicates an upward trend, with a noticeable slump during the 2014/2015 financial year. Data quality based on transactional feedback from the data consumer perspective also displays a general upward trend, but it has more declines than source system data quality.

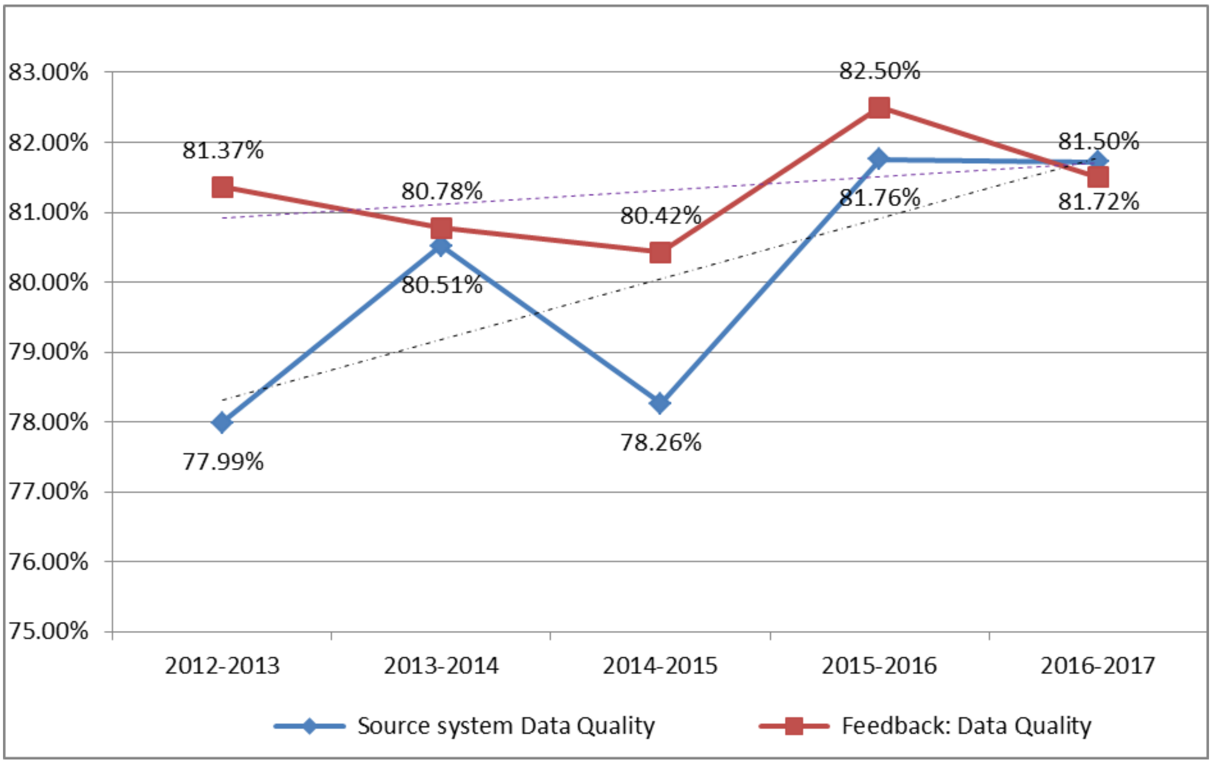


Figure 4.4: Source system and feedback based data quality per financial year

Logistic regression was used to assess inferentially the relationship between transactional data quality from the data creator (contact centre) and transactional feedback from the data consumer (field technician). This tool is preferred as the independent variable, Source system data quality, is a ratio-scale variable measured in percentages and the response variable, transactional feedback, is a binary value where 1=correct and 0=incorrect. The results in Table 4.21 show that transactional feedback is significantly dependent on source system data quality (Wald statistic=18.209, df=1, p-value=0.000).

Table 4.21: Logistic regression of transactional data quality on transactional feedback

Response Variable: Transactional Feedback						
Variables in the Equation	B	S.E.	Wald	df	p-value	Exp(B)
Source system Data Quality	0.008	0.002	18.209	1	0.000	1.008
Constant	0.850	0.145	34.338	1	0.000	2.341

4.3.3 Relationship between costs and transactional feedback received and transactional data quality: source system

In this case, there are two response variables, viz., travel costs and total time cost and two independent variables, viz., transactional feedback and transactional data quality. Travel and time costs were converted to monetary value for analysis purposes. Two regression models, one for each of the response variables were fitted.

4.3.3.1 Regression of travel cost on transactional feedback received and source system data quality

The regression results to assess the effects of transactional data quality based on feedback and source system measurements on travel cost are presented in Table 4.22 below. In this case, the independent variable, transactional feedback, is the binary, hence it will be a dummy variable while source system data quality is continuous. The results indicate that transactional feedback has a significant impact on travel cost (B=3.547, t=6.582, p-value=0.000). Source system data quality also has a significant impact on travel cost (B=0.729, t=10.098, p-value=0.000). Both results indicate that an increase in transaction feedback and source system data quality leads to increased travel costs.

Table 4.22: Regression of travel cost on transactional data quality and transactional feedback

Dependent Variable: Travel Cost	Unstandardised Coefficients		Standardised Coefficients	T	p-value
	B	Std. Error	Beta		
(Constant)	59.720	5.796		10.304	0.000
Transactional feedback	3.547	0.539	0.014	6.582	0.000
Source system data quality	0.729	0.072	0.021	10.098	0.000

4.3.3.2 Regression of time cost on transactional feedback received and: source system data quality

The regression results in Table 4.23 below show that transactional feedback has significant impact on total time cost (B=24.067, t=31.502, p-value=0.000). A value of B=24.067 for transactional feedback means that, total time cost will be higher whenever the feedback is correct. Source system data quality also has significant negative impact on total time cost (B=-1.063, t=-10.382, p-value=0.000). A value of B=-1.063 means the higher the transactional data quality the lower the total time cost, hence, improved transactional data quality results in reduced total time cost.

Table 4.23: Regression of total time cost on transactional feedback and transaction data quality

Dependent Variable: Total time Cost	Unstandardised Coefficients		Standardised Coefficients	T	p-value
	B	Std. Error	Beta		
(Constant)	227.411	8.217		27.677	0.000
Transactional feedback	24.067	0.764	0.065	31.502	0.000
Source system data quality	-1.063	0.102	-0.021	-10.382	0.000

4.3.4 Compare cost impact of incorrect transactions to transactional data quality: feedback received and transactional data quality: source system

Transactions performed on customer side faults are incorrect transactions as they should not have been executed. The mean contribution such transactions had to the total transaction cost is expressed in Figure 4.5. This contribution is then compared to the mean of transactional data quality: feedback received, which is transactional data quality measured by quantifying correct feedback received from technicians, and the mean of transactional data quality: source system, for each financial year. The figure indicates that incorrect transactions had a lower impact on total transaction cost when transactional data quality: feedback received and transactional data quality: source system where higher. In 2015/2016, the mean of transactional data quality: feedback received and transactional data quality: source system were at their highest at 82.50% and 81.76% respectively. However, the cost impact of incorrect transactions was at its lowest at 15.7%.

The 2014/2015 transactional data quality: feedback received, was at its lowest. It had an 80.42% and transactional data quality: source system was at its second lowest point of 78.26%. For the same period, the cost impact of incorrect transactions was at its peak with 18.08%. Therefore, an inverse relationship exists between transactional data quality: feedback and cost impact of incorrect transactions. There is also an inverse relationship between transactional data quality: source system and cost impact of incorrect transactions.

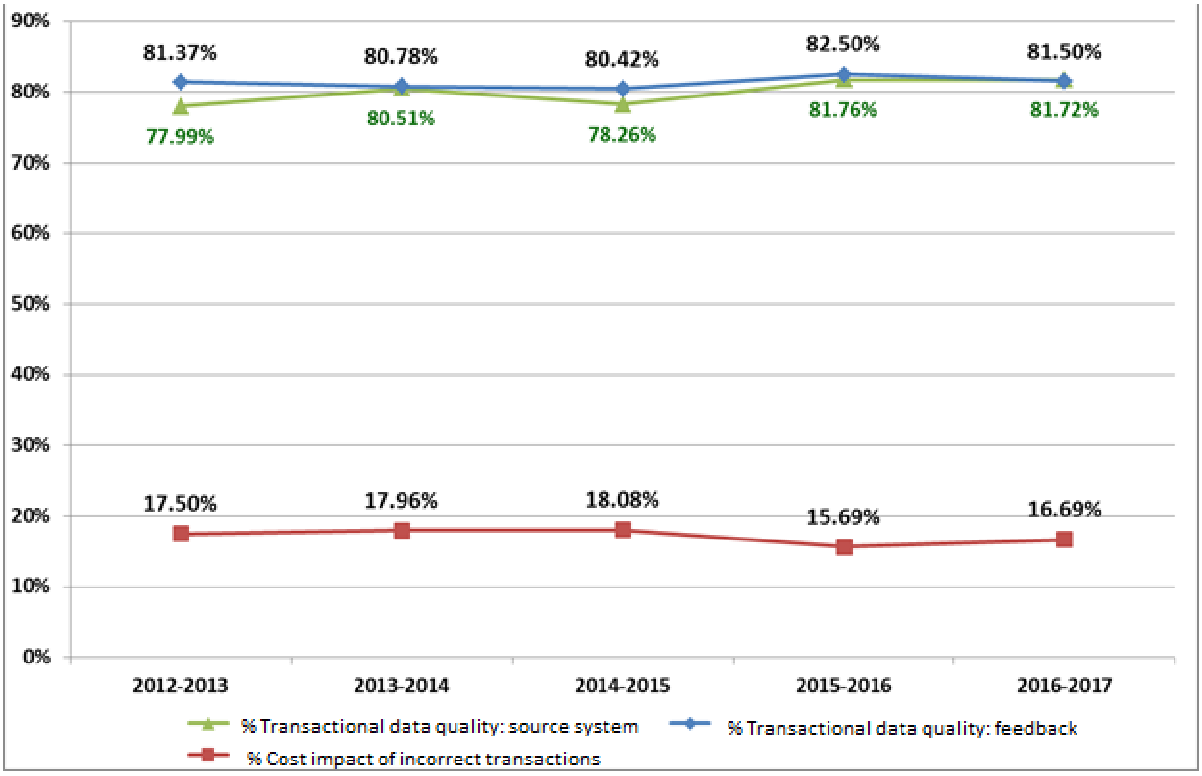


Figure 4.5: Transactional data quality: source system and transactional data quality: feedback received compared to total cost impact of feedback indicating incorrect transactions

In Table 4.24, the same dataset, as in Figure 4.5, is displayed for each financial year with the addition of a mean for each variable over all the financial years combined. For each variable, a value higher than the mean for all financial years combined is highlighted in green, whilst a value lower than the mean is highlighted in pink. In 80% of the time periods analysed, transactional data quality: source system and cost impact of incorrect transactions, confirmed the inverse relationship discovered in Figure 4.5. Similarly, transactional data quality: feedback received and cost impact of incorrect transactions illustrate a corresponding inverse relationship of 80%, but for different time periods. Consequently, an increase in transactional data quality: source system or transactional data quality: feedback received, can result in the lowering of the cost impact of incorrect transactions in 80% of the times.

The observation from the above is that, incorrect transactions should be avoided as far as possible as they contributed an average of 17.2% to the total transaction cost for the analysed financial years. In other words, money can be saved by reducing or avoiding incorrect transactions. As a result, incorrect transaction costs can be used to gauge the savings potential embedded in the improvement of transactional data quality measured from a source system data quality or from feedback on correct transactions perspective.

Table 4.24 : Transactional data quality: source system and transactional data quality: feedback received compared to total cost impact of transactions with feedback indicating incorrect transactions

Fin Year	Transactional data quality: source system (A)	Transactional data quality: feedback received (B)	Cost impact of incorrect transactions (C)	Inverse relationship confirmed: A and C	Inverse relationship confirmed: B and C
2012-2013	77.99%	81.37%	17.50%	Yes	No
2013-2014	80.51%	80.78%	17.96%	No	Yes
2014-2015	78.26%	80.42%	18.08%	Yes	Yes
2015-2016	81.76%	82.50%	15.69%	Yes	Yes
2016-2017	81.72%	81.50%	16.69%	Yes	Yes
Averages	**80.05%**	**81.31%**	**17.18%**	**80%**	**80%**

4.3.5 Relationship between monetary saving, total cost, transactional data quality: source system and transactional data quality: feedback received

As stated previously, the avoidance of costs related to incorrect transactions can translate into monetary savings and will be used as a proxy for monetary savings. Correlation and regression analysis were executed on a summarised version of the work order data analysed thus far. Summarisation was performed per CNC per month on the original dataset. Correlations were tested between monetary savings, transactional data quality: source system, transactional data quality: feedback received and total transaction cost, as shown in Table 4.25.

Table 4.25: Correlation of monetary savings with transactional data quality: source system, transactional data quality: feedback received and total transaction cost

Correlations		Monetary Savings	Comment
Transactional data quality: source system	Correlation	0.040	Not significant
	p-value	0.131	
	N	1442	
Transactional data quality: feedback received	Correlation	0.515	Significant
	p-value	0.000	
	N	1442	
Total Transaction Cost	Correlation	-0.882	Significant
	p-value	0.000	
	N	1442	

The results indicate, no significant relationship between monetary savings and transactional data quality: source system, but a positive significant and strong correlation between monetary savings and transactional data quality: feedback received (correlation = 0.515, p-value = 0.000). There is also a significant strong but negative correlation between monetary savings and total costs (correlation= -0.882, p-value = 0.000). Thus, an increase in the total transaction cost can result in a reduction in monetary savings due to the negative correlation detected. However, an increase in transactional data quality: feedback will yield a corresponding increase in monetary savings as indicated by the positive correlation.

Table 4.26 depicts the regression of monetary savings with transactional data quality: source system, transactional data quality: feedback received and total transaction cost. Monetary saving is significantly affected by both transactional data quality: feedback received (B=27969.543, t=36.479, p-value=0.000) and total transaction cost (B=-0.169, t=-88.065, p-value=0.000). A B=27969.543 for transactional data quality: feedback implies that an improvement in transactional data quality measured via technician feedback will result in increased monetary savings, whereas B=-0.169 for total transaction cost suggests that an increase in total transaction cost will lead to decreased monetary savings.

Table 4.26: Regression of monetary saving with transactional data quality: source system, transactional data quality: feedback received and total cost

Dependent Variable: Monetary Saving	Unstandardised Coefficients		Standardised Coefficients	t	p-value
	B	Std. Error	Beta		
Transactional data quality: source system	3259.311	1730.075	0.017	1.884	0.0598
Transactional data quality: feedback received	27969.543	766.728	0.335	36.479	0.0000
Total Transaction Cost	-0.169	0.002	-0.807	-88.065	0.0000

4.3.6 Summary of historical dataset

The descriptive statistics established from an analysis of the historical dataset indicated the following:

- In Table 4.19 the mean of source system data quality and the mean of feedback for correct transactions differed by only 1.3%. Maximum deviation from the mean for feedback on correct transactions was however significantly higher than source system data quality. Costs are influenced by volume of work orders and distance travelled. Work performed during overtime seems to increase costs more than work order volume alone as per Bethlehem Zone's time cost contribution vs work order volume results.
- For 60% of the cases analysed, transactional data quality at the source and feedback on correct transactions depicted a similar increase/decrease when compared to the mean of each according to Table 4.20.
- The trends for both the quality of source system data and feedback on correct transactions in Figure 4.4 indicate an upward trend, which suggests a steady improvement of transactional data quality.
- A scrutiny of Table 4.24 revealed a possible 17.18% savings potential on total costs for ESP work orders embedded in the improvement of transactional data quality.

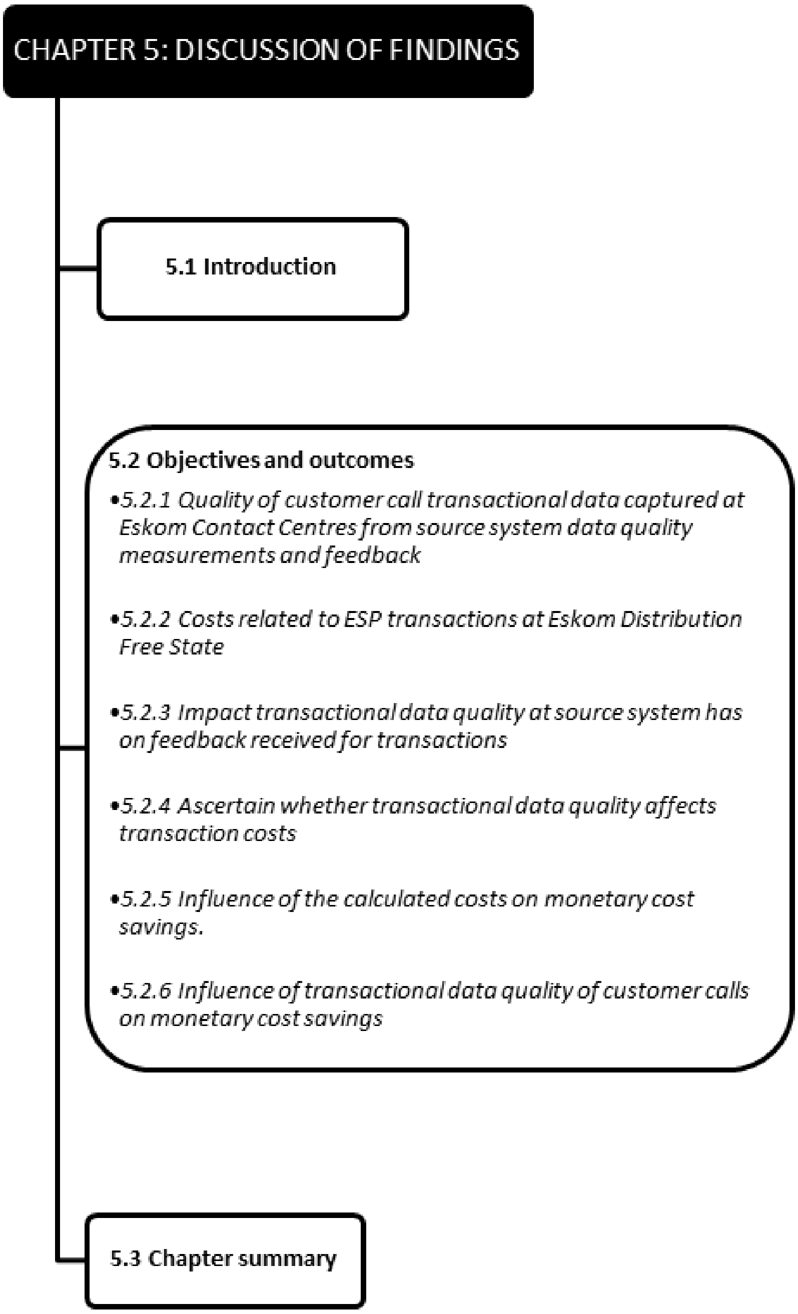
Inferential statistics on the historical dataset revealed that:

- Transactional feedback indicating correct transactions is significantly dependent on the quality of source system data, as shown in Table 4.21.
- According to Table 4.22, travel cost is significantly dependent on transactional data quality and transactional feedback. An increase in transactional data quality or transactional feedback will result in an increase in travel costs.
- Time cost, as shown in Table 4.23, is significantly dependent on transactional data quality and transactional feedback. An increase in transactional data quality results in a decrease in time costs, whereas an increase in transactional feedback will result in increased time costs.
- Correlation and regression analysis in Table 4.25 and Table 4.26 indicates that, on the one hand, an increase in transactional data quality: feedback, leads to a corresponding increase in monetary savings. On the other hand, an increase in total transaction cost will result in decreased monetary savings.

4.4 Chapter summary

This chapter presented the results of descriptive and inferential statistical analyses performed on both survey results gathered from field technicians and the historical dataset comprising ESP logged transactions over 5 financial years. The survey results are based on technician perceptions whereas the historical dataset is based on actual transactions executed and feedback captured against such transactions. The chapter uncovered various results and these are discussed in more depth in the next chapter.

5 CHAPTER 5: DISCUSSION OF FINDINGS



5.1 Introduction

The results from statistical analyses of survey data and that from the historical dataset were presented in the previous chapter. This chapter discusses the findings derived from the analysis in relation with the objectives of this study as stated in Chapter 1.

5.2 Objectives and outcomes

The findings associated with each secondary objective and outcome are discussed before examining the results based on the primary objective.

5.2.1 Quality of customer call transactional data captured at Eskom Contact Centres from source system data quality measurements and feedback

The quality of customer call transactional data captured at Eskom Contact Centres relates to condition 1 [V] and feedback relates to condition 2 [X] of the conceptual framework in section 1.8. The analysis of survey data in Table 4.2 indicates that the overall quality of transactional data from the contact centres was perceived to be 53.6%. Transactional data quality, judging from customer side faults, as an indicator of incorrect data was calculated as 68.5%. The 14.9% difference in Table 4.2 between B4's total transaction quality of 53.6% and B5's transaction quality related to customer side faults of 68.5% suggests the existence of other data quality issues not related to customer side faults, which also emanate from the Contact Centre. The researcher's informal discussions with technicians revealed that the other data quality issues were related to incorrect master data such as customer information and meter data. The majority (55.7%, See Table 4.2) of technicians agreed/strongly agreed that the contact centres can do a better job to identify and prevent customer side faults. The figures revealed that there is potential to increase transactional data quality from the contact centre to a much higher level than what it was during the time of the research. This resonates with Redman's (2013a) suggestion to manage and improve data quality during its creation phase.

The analysis of historical data, as reflected in Table 4.20 and Figure 4.4, revealed that transactional data quality measured at the source system had an average value of 80.05% whereas transactional data quality based on feedback was 81.31%. The mean of both values only differ by 1.26% thus, implying that the metrics used to measure quality from the data creator's perspective closely resemble the data experience from the data consumer. An important difference between the two values is embedded in the volume of data supporting each mean value. Source system data quality only focuses on a small sample of the data whereas feedback data quality is calculated from feedback per transaction.

The difference between data quality measured using the survey and historical data can be ascribed to human perceptions and the granularity of the measurement scale used within the survey. Consequently, historical data can be a more accurate measurement of data quality as perceptions and feelings play a lesser role within this dataset. It is important to note that the customer side fault indicator is not always used correctly. According to Table 4.3, the indicator was only used correctly for 52.8% of the time on actual customer side faults and used incorrectly 25.3% of the time on valid ESP faults. Therefore, the data quality issue is understated on actual occurrences of customer side faults, but overstated when marking a valid ESP as a customer side fault. Feedback from technicians is also data and its importance increases when it is used to improve data quality of transactions. The low percentage of correct usage signifies an issue with the accuracy dimension, due to erroneous interpretation of which feedback option to select. Therefore, technicians need to be educated on the correct usage of feedback to ensure that the root cause of interpretation quality is addressed (Loshen, 2010; DAMA UK Working Group, 2013; Rantala, 2016).

Customer call transactional data quality related to electricity supply problems (ESP) from the historical dataset perspective that is shown in Table 5.1 compares favourably with figures from literature studying international and South African data quality figures. The survey results however compare poorly to the same figures. Historical data reveals that data quality measured at source is on average, 8.77% higher than the average for literature measurements and data quality measured via feedback is 10.03% higher than the literature on average. Thus, the quality of transaction data of customer calls, from a historical data perspective, are better than local and international averages, Conversely, survey results indicate an overall source system data quality that is 17.68% lower than literature average and data quality based on feedback is -2.78% lower. Therefore, the survey results indicate data quality of a lower standard, compared to local and international averages.

Table 5.1: Literature on data quality compared to historical dataset and survey results

Data Type	Quality measured	Source	Data scope	Comparison with Eskom ESP transactional data quality - Historical dataset		Comparison with Eskom ESP transactional data quality - Survey results	
				Source system measurement: 80.05%	Feedback on customer side fault: 81.31%	Overall source system quality: 53.6%	Feedback on customer side fault: 68.5%
Transactional and Master	78.30%	(Röthlin 2004)	International	1.75%	3.01%	-24.70%	-9.80%
Master	78%	(Experian 2013)	International	2.05%	3.31%	-24.40%	-9.50%
Master	73%	(Experian 2017)	International	7.05%	8.31%	-19.40%	-4.50%
Master	50%	(Burrows, 2014)	Local	30.05%	31.31%	3.60%	18.50%
Master	77.10%	(World Economics, 2017)	Local	2.95%	4.21%	-23.50%	-8.60%
Mean	**71.28%**			**8.77%**	**10.03%**	**-17.68%**	**-2.78%**

5.2.2 Costs related to ESP transactions at Eskom Distribution Free State

Costs refer to condition 3 [Y] in the proposed conceptual framework. The quantifiable costs identified are kilometres travelled and hours worked during normal and overtime conditions. These costs can be converted to monetary values for easy comparison and analysis. Overtime exerts a greater impact on the total cost than normal work hours. This occurs because rates for overtime are comparably higher than those for normal time as can be seen when analysing Bethlehem Zone's total work order volume in relation to total cost contribution in Table 4.19. Technicians indicated in Table 4.6 that they performed additional work on 18.9% of customer side fault occurrences, which can potentially increase labour costs on such transactions. However, the additional work can also potentially curb future occurrences of customer side faults, especially if the technicians spent their time to educate customers as a preventative measure. Informal discussions with technicians indicated that customers in general, and especially older customers, could not always interpret questions from the contact centre agents correctly. Due to this misunderstanding, customers could not give correct answers to the agents, thus causing calls to be logged incorrectly. Customers were empowered to identify issues on their side after technicians explained the issue to customers and showed them practically how to determine if the fault was on their side.

5.2.3 Impact transactional data quality at source system has on feedback received for transactions

This section investigates how transactional data quality measured at the source system [V] impacts feedback received on transactions [X]. No significant relationship was discovered within the survey data, but the historical dataset revealed a significant relationship between transactional data quality at the source system and feedback received on correct transactions (see Table 4.20). Consequently, an improvement in data quality during the data creation phase can translate into an improved data consumer experience based on the feedback from field technicians. This finding concurs with calls for improvements in data quality especially the popular data quality dimension of accuracy (Loshen, 2010; DAMA UK Working Group, 2013; Rantala, 2016) as the data creator has to accurately interpret and capture data during the progression of a transaction. From a root cause perspective, the cause of poor data quality is related to interpretation quality. This can be due to the individual subjectivity of the data creator, which leads to an incorrect interpretation and capturing of the cause for the loss of electricity experienced by a customer (McKnight, 2009; Singh & Singh, 2010; Loshen, 2011; Wang et al., 2015). The quality of customer call transactional data can be perceived as being of high value especially because it supports business processes and/or decisions. Hence, it will be worthwhile for Eskom to revisit the acceptable values/range of its call quality measurements to improve its accuracy dimension, if the benefit realised by this are to offset the cost of increases in customer call transactional data accuracy (McGilvray, 2008; Loshen, 2010; DAMA UK Working Group, 2013).

5.2.4 Ascertain whether transactional data quality affects transaction costs

The survey data also sought to explore the influence of transactional data quality: feedback [X] on transaction costs [Y] as well as the impact of transactional data quality: source system [V] on transaction costs [Y]. The findings revealed that a positive statistically significant correlation existed between feedback received and overtime costs incurred (see Table 4.14). This positive correlation implies that an increase in feedback indicating customer side faults from technicians can lead to an increase in overtime costs. The historical data reveals a positive statistically significant correlation between travel cost and feedback received, as well as a positive statistically significant correlation between travel cost and source system data quality (see Table 4.22). Thus, an increase in either feedback indicating correct transactions or source system data quality will result in an increase in overall travel cost.

One would assume an increase in data quality from the source system and increased feedback indicating correct transactions would result in decreased travel cost as it suggests improved data quality. Instead, the higher travel cost resulting from increased feedback could be explained from the premise that the volume of transactions with correct feedback inherently forms part of the total amount of transactions with correct feedback. Therefore, an increase in the total amount of transactions will impact the sum of correct feedback received. An increase in transactions results in rising costs due to an additional cost component per transaction, thus probably explaining the reason behind the rise in travel costs (Vanderbeck & Mitchell, 2015) despite the improved quality of query data capturing. Nonetheless, improved data quality from the source system’s resultant escalation in travel costs is a complex phenomenon to explain.

Table 4.23 shows a positive statistically significant correlation that exists between total time cost and feedback received, whilst time cost has a statistically significant negative correlation with source system data quality. Therefore, increased feedback indicating correct transactions correlates to increased time cost. The expectation can be that an increase in feedback indicating correct transactions will result in a lower total time cost. Instead, the increase can also be influenced by a rise in the total amount of transactions similar to the explanation for Table 4.22 above, resulting in increased total time cost. The negative correlation between time cost and source system data quality can be due to a possible reduction in amount of work orders, as increased source data quality contributes to a reduction in incorrect transactions (Samitsch, 2014; Zhang, 2014).

5.2.5 Influence of the calculated costs on monetary cost savings.

The study also sought to examine the influence of the calculated costs [Y] on monetary cost savings [Z]. No significant relationship between calculated costs and monetary savings could be established from the survey data. A significant negative correlation was established, from the historical dataset, between total costs and monetary savings, as shown in Table 4.25. Table 4.26 also indicates that overall costs have a significant negative impact on monetary savings. These findings point out that any overall costs increase will affect the savings negatively. Based on Metha's (2016) classification, the overall cost consists of direct costs and is calculated from travel and labour costs per transaction. Increased work order volume will increase overall cost as there are costs involved per transaction (Vanderbeck & Mitchell, 2015). Therefore, the highlighted data quality issues indicate that higher costs, which can be linked to more work orders as per Table 4.19, will result in more customer side faults and in that way reduce savings which can be classified as avoidable costs (Holloway, 2016; Mehta, 2016; Willis & Schrieber, 2016).

5.2.6 Influence of transactional data quality of customer calls on monetary cost savings

The influence of the quality of transactional data measured at the source system [V] on monetary savings [Z] as well the impact of the data quality of transactions measured by means of quantifying correct feedback [X] on monetary savings [Z] were examined. The survey results' descriptive statistics shows that, a possible 47.7% savings seems embedded in the improvement of customer call transactional data. The inferential analysis, however, did not indicate any significant relationship between monetary savings and source system data quality in Table 4.17 or feedback indicating incorrect transactions in Table 4.18. As mentioned in section 4.2.3, the lack of significant relationships between variables from the questionnaire could be influenced by the questionnaire's brevity, moderately reliable Cronbach's alpha test, technicians' bias in opinion and their failure to supply a Rand value for anticipated monetary savings.

The historical dataset's descriptive statistics reveal that incorrect transactions had a 17.18% impact on overall transaction costs over the five financial years analysed in Table 4.24. There is also an 80% corresponding increase/decrease between incorrect transaction cost and, either transaction data quality measured at the source, or data quality based on feedback. Inferential statistics revealed no statistically significant relationship between monetary savings and data quality measured at the source. However, feedback indicating correct transactions has a statistically significant positive correlation with monetary savings according to Table 4.25. The fact that only data quality measured from feedback indicating correct transactions has a significant correlation with monetary savings may arise from the reality that data quality measured at the source is compiled from a small sample of data from a specific month and then applied to all transactions of that month.

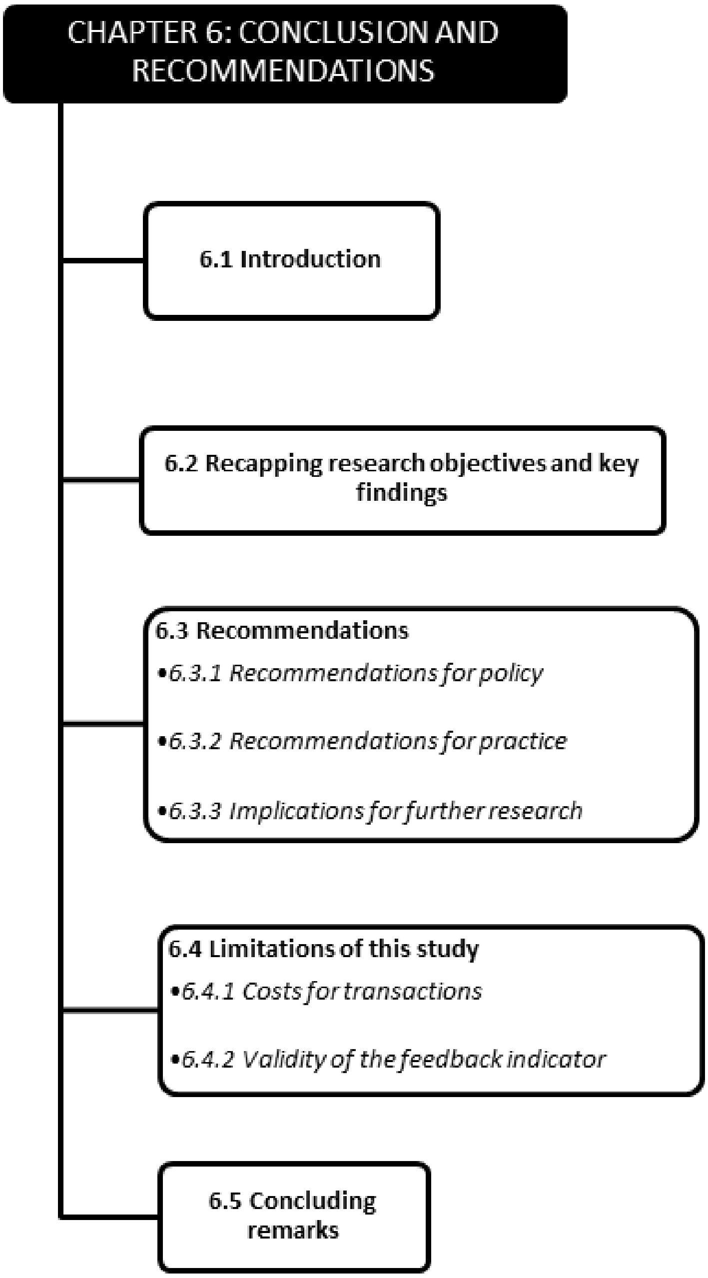
Data quality measured from feedback is, however done per transaction and thus has a much higher accuracy level compared to the small sample considered for source system measurement. Feedback indicating correct transactions in Table 4.26 also significantly impacts monetary savings. Therefore, the 17.18% impact that incorrect transactions have on overall costs, and the 80% relationship it has with feedback indicating correct transactions, is supported by the results from the correlational analysis of the relationship between data quality of transactions measured via feedback and monetary savings. Consequently, a 17.18% savings on ESP transactions exist in terms of avoidable costs, if Eskom increases its customer call transaction data quality by preventing instances of customer side faults from being dispatched. The observed 17.18% savings agrees with Experian's (2015) data quality research, which discloses that organisational profits can increase by 15% on condition that high quality data has been generated. It also resonates with the findings by Wang et al. (2015) that poor data quality has been estimated to cost companies up to 20% of their revenue.

Based on data improvement benefit types defined by Batini and Scannapieco (2016), the 17.18% operational cost savings embedded in transaction data quality improvement forms part of the monetisable benefit category. Two other data improvement benefit categories also exist and these are quantifiable and intangible benefits. Two quantifiable benefits/savings that can be realised from an improvement of customer call transactional data are work hours saved, which could be spend on preventative maintenance, and an improvement in ESP restoration time performance as mandated by NERSA (2002). An important intangible benefit exists in the area of increased customer satisfaction, due to shorter waiting periods for supply to be restored and lower tariff increases, as monetary savings realised reduce operational costs and therefore the need to increase prices. Thus, the overall monetary savings potential could be higher than 17.18%, but as Redman (2013a) points out, there is no reliable method to convert quantifiable and intangible benefits into monetary savings. Ultimately, the decision to expend prevention costs (Batini & Scannapieco, 2016) in order to fix customer call transactional data at the source during its data creation phase, should be based on a cost benefit analysis between savings potential and prevention costs as per Loshen's (2011) recommendation. The fact that data quality measured at the source did not have a statistically significant impact on monetary savings, underscores the importance of creating a feedback mechanism for data consumers. Without the customer side fault feedback indicator, it would have been difficult to identify such transactions, determine the root cause and impact, and improve transaction data quality in order to prevent reoccurrence. (Alexopoulos, Loukis & Charalabidis, 2014).

5.3 Chapter summary

This chapter discussed the results from the previous chapter and linked them to findings from literature. Findings were examined and grouped as per the objective of the study. In general, the survey data's descriptive analysis highlighted some interesting observations. However, it was not supported by subsequent inferential analysis performed. The historical dataset's descriptive analysis was largely in accord with inferential statistical analysis and highlighted that an improvement in transactional data quality can contribute to savings on ESP transactions. Data quality measured at the source system did not offer statistically significant results to support the savings, but the usage of feedback on correct transactions as a data quality measurement strongly supported the identified savings potential.

6 CHAPTER 6: CONCLUSION AN RECOMMENDATIONS



6.1 Introduction

This concluding chapter presents a summary of the key findings discussed in the previous chapter and offers recommendations for Eskom and businesses. The chapter also highlights possible areas for further research and limitations to this study.

6.2 Recapping research objectives and key findings

An analysis of survey and historical transaction datasets using descriptive and inferential statistical tools revealed various insights regarding the main objective and six secondary objectives stated at section 1.4.1. Considered first are the findings regarding the secondary objectives. Thereafter, the focus shifts to the findings related to the main objective.

Secondary objectives one and two sought to establish the quality of customer call transactional data. Objective one determined the data quality of transactions on customer calls from the source system data, whereas objective two verified transaction data quality from feedback provided by the field technicians. The key findings revealed that:

- The ability of data consumers to provide feedback on executed transactions created the opportunity to measure data quality from a data consumer's perspective. Feedback also served as an enabler to quantify costs of incorrect and correct transactions.
- Mean data quality from a data creator perspective measured at the source system and data quality measured via feedback from the data consumer perspective differed by only 1.26%. The implication from a summarised assessment point is that, data quality measurements at the source system closely resembled the experience of data quality by data consumers.
- Eskom's transactional data quality from the historical dataset, measured at the source system was 80.05% and that measured via feedback was 81.31%. These percentages exceeded average data quality measurements found in literature.

- The overall quality of source system data, as observed from the survey results, measured at 53.6% and data quality based on feedback was 68.5%. Both were lower than literature's average data quality measurement values.

Objective three sought to identify the costs related to ESP transactions at Eskom Distribution Free State. The identified costs were labour costs to work and travel during normal time and overtime as well as the running cost per kilometre travelled. These costs were calculated per transaction in monetary value and expressed as a percentage of the total transaction cost.

Objective four sought to determine the impact that transactional data quality at source system had on feedback received from technical field staff/data consumers on transactions. The results exposed that feedback on correct transactions from data consumers was significantly affected by data quality measured at the source system. Therefore, an improvement in data quality at the source system can trigger increased feedback indicating correct transactions, which in turn translate to a better data quality experience by data consumers.

Objective five set out to analyse how transactional data quality measured from the source system and feedback received impact transaction costs. It was found out that:

- Transaction data quality measured from the source system indicated that an increase in data quality resulted in decreased time costs but increased travel costs. An improvement in data quality causing a decrease in time costs makes sense as increased data quality should results in lower overall costs, but the increase in travel costs cannot be explained.
- Transaction data quality measured from feedback resulted in a rise of both travel and time costs. Even though a rise in feedback indicating correct transactions signifies an increase in data quality. It can also be linked to an increase in the total amount of transactions, which can support the overall rise in costs.

Objective six set out to establish the influence of the calculated costs on the monetary savings. Increased overall costs negatively affected monetary savings. An important driver in overall costs is the volume of transactions. The observed current level of data quality of 80.05% at the source system and 81.31% measured via data consumer feedback indicates that an overall increase in transactions will increase the total amount of incorrect transactions, which will in turn reduce savings.

The main objective of this study was to establish if improvements in transactional data quality could translate into financial savings for Eskom Distribution. Hence:

- It was found that granularity of data quality measurements can cause potential analytical challenges. Inferential analysis indicated that only data quality measured from technician feedback had an impact on monetary savings, but not data quality measured at the source from the data creator's perspective. This arises from the fact that data quality measured at the source was based on a small sample of data whereas feedback could be measured over the whole data population, thereby increasing its granularity and accuracy.
- If transactional data quality based on customer call feedback could be increased from 81.31% to 100%, monetary savings of 17.18% can be achieved by means of cost avoidance on electricity supply problems (ESP) transactions. The savings relate to monetisable benefits exclusively, which were calculated from productivity costs. Quantifiable and intangible benefits could be identified, but could not be converted to monetary values; hence, total savings could actually be more than 17.18%.

6.3 Recommendations

This section outlines the recommendations, based on the research outcomes, suggested for Eskom and businesses. In addition, the section explains the implications for future research.

6.3.1 Recommendations for policy

Whenever data quality of significant business value is improved, potential monetary savings and consequent profit increases can be realised because of such improvements. However, businesses tend to focus on quality of its master data and often neglect transactional data quality due to its high volume, volatility and inherent challenges in measuring its quality. Therefore, the recommendation is that Eskom should continually identify and investigate high value transactional data quality as it offers significant savings potential through cost avoidance. In the case of Eskom, a 17.18% cost saving potential on ESP transactions can be achieved, if its customer call transactional data quality could improve from 81.31% to 100%. It is also recommended that Eskom quantify the costs required to improve transactional data quality of its customer calls. The results of such improvements will determine if the savings potential will offset data quality improvement costs and thereby generate net saving outcomes. It should be acknowledged that achieving 100% data quality remains a daunting task. As such, it is better to show management what and where savings potential exists, that they need to determine for the cost involved to improve and verify whether the benefit of the savings will sufficiently justify the expenses required.

Feedback on customer call transactions does not currently form a closed loop as it is supplied but not automatically used to detect and reduce unwanted feedback types. It is therefore, recommended that Eskom harnesses the power of closed feedback loops to leverage feedback received in order to transform data creator behaviour and generate higher quality transactional data. Eskom data capturers are expected to engage in detailed elicitation of customer faults information through probing and proper investigation to generate higher quality transactional data and to reduce the volume of customer side faults. The reduction of such customer side faults can contribute to reduce monetary costs on travel and labour hours, which translate to monetary savings.

6.3.2 Recommendations for practice

The ability to assimilate feedback from data consumers is an important capability any corporate entity needs to establish in order to measure its transactional data quality. Ideally, a closed feedback loop should be created in order to immediately notify a data creator of incorrect transactions so that they can identify and address the root cause. Feedback on transactions enables the measurement of both transactional data quality and the quantification of incorrect transactions. Therefore, large businesses should build the human capital and technological capacity to quantify costs and establish where potential monetary savings resides within specific improvements in transactional data quality.

6.3.3 Implications for further research

The following implications for future research are offered:

- Investigate the disparity between a data consumer's perception of data quality and data quality measured from historical transaction data to uncover:
  - Reasons for this difference.
  - If perceptions or transaction data offer the most reliable depiction of data quality.
- Examine whether an intervention to improve accuracy of feedback from data consumers yield improved feedback and how this will influence the calculation of transactional data quality and costs.

6.4 Limitations of this study

6.4.1 Costs for transactions

Eskom's main financial system, the Systems Applications and Products in data processing (SAP) is not currently configured to differentiate between overtime costs for planned and unplanned work, as the makeup hours are not referenced in these categories. Due to this limitation, a model based on the overtime rules was applied to the extracted transactions to determine costs. As it is against Eskom's human resource policy to reveal each resource's hourly rate, the hourly rate of the lowest graded technical staff that works on ESP work orders was applied to each hour work. The effect of this method will be an understatement rather than overstatement of the costs involved. Travel distance and time was not based on the actual route travelled, but rather on the direct line distance between two sets of GPS coordinates. This will further understate the costs involved as a route travelled will be longer than the direct line distance between two points.

6.4.2 Validity of the feedback indicator

The customer fault feedback indicator was used as a gauge of incorrect transactional data, based on the assumption that technical staff used the indicator correctly under all circumstances. Informal discussions with technical staff revealed instances where there was an underutilisation and over utilisation of the indicator. This was also confirmed by the survey results in Table 4.3.

6.5 Concluding remarks

The study was presented in six chapters. Chapter One set the scene by defining the research problem, supplying a conceptual framework and stating the aim, research objectives and questions. Chapter Two offered insights from literature into the four variables defined in the conceptual framework and elaborated on how each variable plays a role within Eskom's process of servicing customers who experience electricity supply problems (ESP). Thereafter, Chapter Three explained the research methodology in terms of the adopted paradigm and research design. It also described the population under investigation, sampling method, the data collection method, data analysis and how validity and reliability were determined. Chapter Four presented the results from descriptive and inferential statistical analysis. The analysis tools were applied to two datasets, which are survey results and historical data. Chapter Five presented findings on the results and recapped the objectives of the study. Chapter Six concluded the study and offered key findings and recommendations whilst highlighting limitations that need to be considered.

The study realised its objectives and offered insights, which differed substantially when evaluating results from survey data and historical data perspectives. The main research question was answered from a historical data perspective as follows: Eskom Distribution Free State's transactional data quality on customer calls related to ESP can positively impact monetary cost savings on such transactions by up to 17.18%, if data quality can be enhanced from 81.31% to 100%.

7 REFERENCES

Abramovici, A. & Chapsky, J. (2000). *Feedback Control Systems: A Fast-Track Guide for Scientists and Engineers*. New York: Springer Science + Business Media, LLC.

Alexopoulos, C., Loukis, E. & Charalabidis, Y. (2014). A Platform for Closing the Open Data Feedback Loop based on Web2.0 functionality. *JeDem - eJournal of eDemocracy*, 6(1), 62-68.

Alrayes, A.M. (2015). The Impact of System and Data Quality on Organizational Success in the Kingdom of Bahrain. *International Journal of Economics and Management Engineering*, 9(5), 1-7.

Andrew, D.P., Pedersen, P.M. & McEvoy, C.D. (2011). *Research Methods and Design in Sport Management*. Champaign: Human Kinetics.

Arora, R. (ed.) (2016). *Conquering big data with high performance computing*. Austin: Springer.

Bate, N. (2009). *Beat the Recession: A Blueprint for Business Survival*. Connecticut: Easton Studio Press.

Batini, C. & Scannapieco, M. (2016). *Data and Information Quality: Dimensions, Principles and Techniques*. Cham: Springer.

Biehl, R.E. (2016). *Data Warehousing for Biomedical Informatics*. Boca Raton: CRC Press.

Borek, A., Parlikad, A.K., Webb, J. & Wooda, P. (2014). *Total information risk management: Maximizing the value of data and information assets*. New York: Elsevier.

Bovee, M., Srivastava , R.P. & Mak, B. (2003). A Conceptual Framework and Belief-Function Approach. *International journal of intelligent systems*, 18(1), 51-74.

Brodie, M. & Palmer, A. (2016). 'From DevOps to DataOps', in Cutt, S. (ed.) *Getting Data Right. Tackling the Challenges of Big Data Volume and Variety*, Sebastopol: O'Reilly Media, Inc.

Burrows, T. (2014). *ITWeb - SA looks to address data quality*. [Online], Available: https://www.itweb.co.za/content/JOlx4z7kGXY756km [3 January 2018].

Cai, L. & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, 14(2), 1-10.

Cargan, L. (2007). *Doing social research*. Plymouth: Rowman & Littlefield.

Chakraborty, S.K. (2004). *Cost Accounting And Financial Management (For C.A. Course-1)*. New Delhi: New Age International.

Chesmond, C.J. (2014). *Control System Technology*. Brisbane: Elsevier.

Clough, R.H., Sears, G.A., Sears, S.K., Segner, R.O. & Rounds, J.L. (2015). *Construction Contracting: A Practical Guide to Company Management*, 8th Edition. New Jersey: John Wiley & Sons.

Cottrell, R.R. & McKenzie, J.F. (2010). *Health Promotion & Education Research Methods: Using the Five Chapter Thesis/ Dissertation Model*, 2nd Edition. Sudbury: Jones & Bartlett Publishers.

DAMA UK Working Group (2013). *The six primary dimensions for data quality assesment*. [Online], Available: https://www.whitepapers.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQDimensionsWhitePaperR37.pdf [8 April 2016].

DeOreo, P.B. & Wish, J.B. (2015). Medical Director Responsibilities to the ESRD Network. *Clinical Journal of the American Society of Nephrology*, 10(10), 1852–1858.

Dlova, M.R. (2017) Small, Micro and Medium Enterprises Acess to Credit in the Eastern Cape, South Africa, Thesis PhD. University of the Witwatersrand.

Eckerson, W. (2002). *Data quality and the bottom line. Achieving business success through a commitment to high quality data*. Chatsworth: 101communications LLC.

Editors Of Webster's II Dictionaries (ed.) (2005). *Webster's II New College Dictionary*, 3rd Edition. Boston: Houghton Mifflin Harcourt.

Edmonds, W.A. & Kennedy, T.D. (2016). *An Applied Guide to Research Designs: Quantitative, Qualitative, and Mixed Methods*, 2nd Edition. Los Angeles: SAGE Publications.

Entity group (2016). *Why not to put transactional data in an MDM hub*. [Online], Available: http://www.entitygroup.com/wp-content/uploads/2016/05/master_data_management_and_transaction_data_whitepaper.pdf [07 Nov 2018].

Epstein, E. (2012). *Implementing Successful Building Information Modeling*. Norwood: Artech House.

Ernst & Young LLP (2015). *Insiders' Guide to Technology-Assisted Review (TAR)*. Hoboken: John Wiley & Sons.

Eskom Holdings SOC Ltd (2015a). *Environmental, Occupational Health and Safety Incident Management Procedure*, Sunninghill: Eskom Holdings SOC Ltd.

Eskom Holdings SOC Ltd (2015b). *Company information*. [Online], Available: http://www.eskom.co.za/OurCompany/CompanyInformation/Pages/Company_Information.aspx [7 April 2016].

Eskom Holdings SOC Ltd (2016). *Applying the CARAT Principle in Record Keeping and Management*, Sunninghill: Eskom Holdings SOC Ltd.

Eskom Holdings SOC Ltd (2017a). *Eskom Integrated Report*. [Online], Available: http://www.eskom.co.za/IR2017/Documents/Eskom_integrated_report_2017.pdf [3 January 2018].

Eskom Holdings SOC Ltd (2017b). *Design to Cost Bulletin*. 1st Edition, Sunninghill: Eskom Holdings SOC Ltd.

Eskom Holdings SOC Ltd (2018a). *Integrated report*. [Online], Available: http://www.eskom.co.za/IR2018/Documents/Eskom2018IntegratedReport.pdf [29 August 2018].

Eskom Holdings SOC Ltd (2018b). *DTC Change Management Bulletin*. 4th Edition, Sunninghill: Eskom Holding SOC Ltd.

Eskom Holdings SOC Ltd (2018c). *Eskom consults senior executives on Section 189 process*. [Online], Available: http://www.eskom.co.za/news/Pages/2018Nov7.aspx [12 Nov 2018].

Experian Data Quality (2013). *The state of data quality - An Experian Data Quality White Paper*. [Online], Available: https://www.experian.com/assets/decision-analytics/white-papers/the%20state%20of%20data%20quality.pdf [2 April 2016].

Experian data quality (2015). *Global data quality research discussion paper 2015*. [Online], Available: https://www.edq.com/globalassets/uk/papers/global-research-2015_20pp-ext-apr15.pdf [4 April 2016].

Experian Data Quality (2017). *The 2017 global data management benchmark report*. [Online], Available: https://www.edq.com/globalassets/white-papers/2017-global-data-management-benchmark-report.pdf [9 April 2016].

Fürber, C. (2015). *Data Quality Management with Semantic Technologies*. Wiesbaden: Springer.

Gordon, G.A. (2015). *Regression Analysis for the Social Sciences*, 2nd Edition. New York: Routledge.

Grove, S.K., Gray, J.R. & Burns, N. (2014). *Understanding Nursing Research: Building an Evidence-Based Practice*, 6th Edition. Missouri: Elsevier Health Sciences.

Haug, A., Zachariassen, F. & Van Liempd, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management*, 4(2), 168-193.

Holcomb, Z.C. (2016). *Fundamentals of Descriptive Statistics*. New York: Routledge.

Holloway, M.D. (2016). *Spend Analysis and Specification Development Using Failure Interpretation*. Boco Raton: CRC Press.

Hu, Z. (2013). *Electricity Economics: Production Functions with Electricity*. New York: Springer.

Johnson, B. & Christensen, L. (2010). *Educational Research: Quantitative, Qualitative, and Mixed Approaches*. California: Sage.

Keller, W.D. (2015). *Cost and Managerial Accounting II Essentials Volume 2*. New Jersey: Research and Education Association.

Killam, L. (2013). *Research terminology simplified: Paradigms, axiology, ontology, epistemology and methodology*. Sudbury: Laura Killam.

Kubiak, T.M. (2013). *The ASQ pocket guide for the certified six sigma black belt*. Milwaukee: ASQ Quality Press.

Kumar, R. (2014). *Research Methodology*, 4th Edition. California: Sage.

Lacrose, D.T. & Lacrose, C.D. (2015). *Data Mining and Predictive Analytics*, 2nd Edition. New Jersey: John Wiley & Sons.

Law, J. (ed.) (2016). *A Dictionary of Business and Management Oxford quick reference*, 6th Edition. Oxford: Oxford University Press.

Lidwell, W., Holden, K. & Butler, J. (2010). *Universal Principles of Design*, Revised Edition. Beverly: Rockford Publishers.

Lindgreen, A. (2008). *Managing market relationships: Methodological and empirical insights*. Surrey: Gower Publishing.

Loshen, D. (2010). *Master Data Management*. Burlington: Morgan Kaufman Publishers.

Loshen, D. (2011). *The practioner's guide to data quality improvement*. Burlington: Morgan Kaufmann.

Marsden, P.V. & Wright, J.D. (ed.) (2010). *Handbook of Survey Research*. Bingley: Emerald Group Publishing.

Martin, C., Chediak, M. & Wells, K. (2013). *Why the U.S. Power Grid's Days Are Numbered*. [Online], Available: http://www.bloomberg.com/bw/articles/2013-08-22/homegrown-green-energy-is-making-power-utilities-irrelevant [11 Nov 2015].

McCafferty, D. (2016). *How bad data undermines business results*. [Online], Available: https://www.cioinsight.com/it-strategy/big-data/slideshows/how-bad-data-undermines-business-results.html [4 May 2017].

McDonald, D.A. (2009). *Electric capitalism: Recolonising Africa on the power grid*, 1st Edition. Cape Town: HSRC Press.

McGilvray, D. (2008). *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. Burlington: Elsevier.

McKenzie, S. (2014). *Vital statistics - E-Book: An introduction to health science statistics*. New South Wales: Elsevier Health Sciences.

McKnight, W. (2009). Seven sources of poor data quality. *Information Management*, 19(3), 32-33.

Mehta, B.K. (2016). *Cost and Management Accounting*. Agra: SBPD Publications.

Minister of Finance (2015). *Eskom Special Appropriation Bill*. [Online], Available: http://www.gov.za/sites/www.gov.za/files/b_16_-_2015_eskom_special_appropriation_3jun2015_a.pdf [1 June 2016].

Moskwa, L.S. (2015). Is Your Data in Good Order? *Journal Of Pension Benefits: Issues In Administration*, 23(1), 49-50.

Murragapan, P. (2014). *Cost Reduction Strategies*. Subang Jaya: Blue Micro Solutions.

Namu, N.N., Kaimba, G.K., Muriithi, D.K. & Nkari, I.M. (2014). Impact of cost reduction strategies on performance of tea fatories in Embu county, Kenia. *European Journal of Business and Social Sciences*, 3(9), December, 26-48.

National Planning Commission (2012). *National Development Plan 2030*. [Online], Available: www.gov.za/sites/www.gov.za/files/Executive%20Summary-NDP%202030%20-%20Our%20future%20-%20make%20it%20work.pdf [21 September 2015].

NERSA (2002). *NRS 047-1:2002*, Pretoria: South African Bureau of Standards.

O'Shea, S. (2011). *Cambridge business english dictionary*. Cambridge: Cambridge University Press.

Oxford Online Dictionary (2017). *English Oxford Living Dictionaries*. [Online], Available: https://en.oxforddictionaries.com/definition/cost_saving [20 April 2017].

Potter, S. (2006). *Doing postgraduate research*, 2nd Edition. London: Sage.

Protzman, C., Whiton, F., Kerpchar, J., Lewandowski, C., Stenberg, S. & Patrick, G. (2016). *The Lean Practitioner's Field Book: Proven, Practical, Profitable and Powerful Techniques for Making Lean Really Work*. Boca Raton: CRC Press.

Rambe, P. & Modise, D. (2016). Power Distribution at Eskom: Putting Self-Leadership, Locus of Control and Job Performance of Engineers in Context. *African Journal of Business and Economic Research*, 11(1), 45-92.

Rantala, M. (2016) Data quality analysis in industrial maintenance; theory vs. reality, Unpublished master's thesis. Lappeenranta University of Technology.

Redman, T.C. (2013a). *Data quality management past, present and future: Towards a management system for data*. New York: Springer.

Redman, T.C. (2013b). *Data Driven: Profiting from Your Most Important Business Asset*. Boston: Harvard Business Press.

Röthlin, M. (2004) An exploratory study of data quality management. Practices in the ERP software systems context, Proceedings of INFORMATIK 2004, Ulm, 254-258.

Sadgrove, K. (2015). *The complete guide to Business Risk Management*, 3rd Edition. Surrey: Gower publishing limited.

Samitsch, C. (2014). *Data quality and its impacts on decision-making: How managers can benefit from good data*. Wiesbaden: Springer Gabler.

Sekaran, U. & Bougie, R. (2013). *Research methods for business: A skill building approach*, 6th Edition. West Sussex: Wiley.

Sekaran, U. & Bougie, R. (2014). *Research Methods for Business*, 6th Edition. West Sussex: Wiley.

Sharma, J.K. (2007). *Business Statistics*, 2nd Edition. New Delhi: Pearson Education India.

Shim, J.K. (2016). *Accounting and Finance for the NonFinancial Executive: An Integrated Resource Management Guide for the 21st Century*. Washington: CRC Press.

Singh, K. (2007). *Quantitative Social Research Methods*. New Delhi: Sage.

Singh, R. & Singh, K. (2010). A descriptive classification of causes of data quality problems in data warehousing. *IJCSI International Journal of Computer Science*, 7(3), May, 41-50.

Statistics South Africa (2016). *Data Quality Report*. [Online], Available: http://www.statssa.gov.za/publications/03-01-04/03-01-042016.pdf [13 May 2017].

Stern, D.I. (2010). The Role of Energy in Economic Growth. *Crawford School Centre for Climate Economics & Policy Paper No. 3.10*.

Stevenson, A. (ed.) (2010). *Oxford dictionary of English*, 3rd Edition. New York: Oxford University Press.

Topco Media (2014). *Top 500*. [Online], Available: http://www.top500.co.za/company/eskom/ [13 Oct 2016].

Urdan, T.C. (2011). *Statistics in Plain English*, 3rd Edition. New York: Routledge.

Vanderbeck, E.J. & Mitchell, M.R. (2015). *Principles of Cost Accounting*, 17th Edition. Boston: Cengage Learning.

Vermaat, M.E. (2014). *Discovering computers & Microsoft Office 2013: A fundamental combined approach*. Boston: Cengage Learning.

Wang, X.S., Pierce, E.M., Madnick, S., Fischer, R.C. & Zwass, C. (2015). *Information quality*. New York: Routledge.

Wang, R.Y. & Strong, D.M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4), 5-33.

Warren, C.S., Reeve, J.M. & Duchac, J. (2015). *Managerial Accounting*, 13th Edition. Boston: Cengage Learning.

Willis, H.L. & Schrieber, R.R. (2016). *Aging Power Delivery Infrastructures*, 2nd Edition. Boca Raton: CRC Press.

World Economics (2017). *The Data Quality Index (DQI)*. [Online], Available: https://www.worldeconomics.com/Pages/Data-Quality-Index.aspx [3 January 2018].

Zhang, J. (2014) 'Operationalizing data quality through data governance', in *Data governance: Creating value from information assets*, Boca Raton: CRC Press.

8 APPENDICES

8.1 APPENDIX A – Questionnaire

<u>Questionnaire for Eskom CNC technical staff</u>

<u>Part A: Instructions to respondents</u>

This questionnaire aims to ascertain the relationship between transactional data quality of customer calls and monetary cost savings of Eskom Distribution Free State. Please answer all questions accurately and honestly by placing a tick in an appropriate box or by writing your answer in the space provided. The information that you supply will be treated as confidential. Thank you for your willing to complete this questionnaire!

<u>Part B:</u>

1. Please indicate your years of work experience at an Eskom CNC				
1) 0-5 Years	2) 6-10 years	3) 11-15 years	4) 15-20 years	5) More than 20 years
2. For how many years have you been using an EDA ?				
1) up to 1 Year	2) 2 Years	3) 3 Years	4) 4 Years	5) More than 4 years
3. How many single customer dispatch work orders do you receive from dispatch per month?				
1) 1-10	2) 11-20	3) 21-30	4) 31-40	5) More than 40
4. On average, how accurate is the information on the single customer dispatch work orders your receive?				
1) 0-20%	2) 21-40%	3) 41-60%	4) 61-80%	5) 81-100%
5. On average, what percentage of the single customer dispatch work orders you receive per month, are caused by a fault on the customer side (thus no fault on Eskom side)?				
1) 0-20%	2) 21-40%	3) 41-60%	4) 61-80%	5) 81-100%
6. On average, what percentage of the single customer dispatch work orders that are caused by a fault on the customer side, do you mark as "Customer side fault" when putting the work-order in MILE3 ?				
1) 0-20%	2) 21-40%	3) 41-60%	4) 61-80%	5) 81-100%
7. On average, what percentage of the single customer dispatch work orders that are caused by a fault on Eskom side, do you mark as "Customer side fault" when putting the work order in MILE3?				
1) 0-20%	2) 21-40%	3) 41-60%	4) 61-80%	5) 81-100%
8. Please indicate the extent to which you agree or disagree whether the Eskom Contact Centre can do a better job to identify faults that are caused by the customer				
1) Strongly Disagree	2) Disagree	3) Neutral	4) Agree	5) Strongly agree

9. On average how many kilometres do you travel to a single customer dispatch work order that is caused by a fault on Eskom side?				
1) 0-20 km	2) 21-40 km	3) 41-60 km	4) 61-80 km	5) More than 80km
10. On average how many kilometres do you travel to a single customer dispatch work order that is caused by a fault on the customer side?				
1) 0-20 km	2) 21-40 km	3) 41-60 km	4) 61-80 km	5) More than 80km
11. On average what percentage of the single customer dispatch work orders that are caused by a fault on Eskom side do you respond to during normal work time?				
1) 0-20%	2) 21-40%	3) 41-60%	4) 61-80%	5) 81-100%
12. On average what percentage of the single customer dispatch work orders that are caused by a fault on the customer side do you respond to during normal work time?				
1) 0-20%	2) 21-40%	3) 41-60%	4) 61-80%	5) 81-100%
13. On average what percentage of the single customer dispatch work orders that are caused by a fault on Eskom side do you respond to during overtime?				
1) 0-20%	2) 21-40%	3) 41-60%	4) 61-80%	5) 81-100%
14. On average what percentage of the single customer dispatch work orders that are caused by a fault on the customer side do you respond to during overtime?				
1) 0-20%	2) 21-40%	3) 41-60%	4) 61-80%	5) 81-100%
15. Please indicate the frequency at which you repair single customer dispatch work orders that are caused by a fault on the customer side(rather than just putting it on MILE3, with customer side fault feedback).				
1. Never	2. Rarely	3. Sometimes	4. Frequently	5.Always
16. Please indicate the extent to which you agree or disagree that it is possible to use the MILE3 feedback of customer side fault to determine how much such work orders cost the business.				
1) Strongly Disagree	2) Disagree	3) Neutral	4) Agree	5) Strongly agree
17. Please indicate the extent to which you perceive Eskom using the MILE3 feedback of customer side fault to determine how much such work-orders cost the business.				
1. Never	2. Rarely	3. Sometimes	4. Frequently	5.Always
18. Please indicate the extent to which you agree or disagree that it is possible for Eskom to save money if you do not have to go to a single customer dispatch work order that are caused by a fault on the customer side.				
1) Strongly Disagree	2) Disagree	3) Neutral	4) Agree	5) Strongly agree
19. On average, how much money can Eskom save per month if you did not have to go to single customer dispatch work order that are caused by a fault on the customer side				
Estimated amount: R				

8.2 APPENDIX B – Permission to conduct research

Eskom

Date:
16 Jul 2015
Enquiries:
Mr Len Turner
Telephone:
+27 11 800-5184

To: The Registrar
Central University of Technology

ETHICS CLEARANCE: CONFIRMATION OF ESKOM INTELLECTUAL PROPERTY RIGHTS AND SECURITY CLEARANCE FOR MASTERS RESEARCH – MR. CHARL JOHANNES BESTER

This memorandum serves as an ethics clearance; confirmation of Eskom intellectual property rights and security clearance for the continuation of Masters level research and write-up by Mr. CJ Bester. The research topic is "Determine the time and monetary impact of incorrect transactional customer data in Eskom Distribution Free State Field services environment".

Mr. Bester has followed due internal processes in terms of gaining permission for this research.

It must be noted that this general clearance is for a limited period only, which will be for the rest of the financial year 2015 till end 2017, and in no way waives Eskom's Intellectual Property Rights.

Yours sincerely

Len Turner
Senior Consultant
Talent and Skills Management

8.3 APPENDIX C – Extension of permission to conduct research

REQUEST TO EXTEND PERMISSION TO CONDUCT AND COMPLETE RESEARCH AT ESKOM
TOPIC: THE INFLUENCE OF TRANSACTIONAL DATA QUALITY IMPROVEMENTS ON MONETARY SAVINGS OF ESKOM DISTRIBUTION, FREE STATE.

Ms Nozipho Mpanza

Eskom Bloemfontein

Senior Manager – Maintenance and Operations

Email: MpanzaNl@eskom.co.za

Contact no: +27 82 561 4279

Dear Ms Mpanza

Your permission is herewith requested to extend permission for Charl Johannes Bester, a student in the Masters of Business Administration at the Central University of Technology (CUT), to conduct and finalize academic research in your organisation. The thesis is close to completion and currently in its language editing phase.

Background:

Eskom Distribution sponsored my studies towards MTECH Business administration at CUT and initial ethical clearance was granted by Len Turner, Senior Consultant HR, from 2015 to 2017. My research took longer than expected and I consequently require clearance for 2018 and 2019, which will allow me to publish my thesis and mandatory articles. Without publication I will not be allowed to graduate by CUT. Since the process for ethical clearance has changed I humbly request your approval for ethical clearance as senior manager responsible for the data.

Study detail:

Eskom instructed me to perform research on a business challenge. I identified transactional data quality as an area most business do not normally scrutinize as closely as master data for quality issues. I also wanted to verify whether transaction data quality could have an influence on the monetary savings of Eskom Distribution Free State.

The study was based on a quantitative survey design and perspectives of field staff was tested via a questionnaire. Historical workorder data from Maximo was used to support the data from the questionnaire. The mentioned data was analysed by descriptive and inferential statistics to determine the quality of transactional data of customer calls received by field staff in the form of workorders. The impact incorrect data had on operational costs was quantified and then translated into possible monetary savings, if the transactions related to incorrect data could be avoided.

Major findings:

The results of the historical data analysis using mean, frequency distribution, cross tabulation and correlation analysis for survey data and mean distribution, regression analysis and correlation analysis for historical data respectively, revealed potential monetary savings of 17.18% arising from avoidable costs on transactions related to ESP customer calls. These monetary savings were dependent on Eskom's ability to increase its transactional data quality on ESP customer calls from 81.31% to 100%. While it was acknowledged that avoidable costs could only be calculated from quantifiable operational costs, savings would potentially increase if the effects of improved customer service, faster supply restoration times and work hours saved to perform preventative maintenance to reduce overall fault volumes were quantified in monetary terms. If the costs of increasing data quality were lower than the 17.18% monetary savings potential established in the study, such data quality improvement strategies can improve Eskom's financial position.

Please contact my supervisor, Prof Patient Rambe (email address: prambe@cut.ac.za , cellphone number: 073 380 1687) if you have any questions or comments regarding the study. Please sign below to indicate your willingness to extend my ethical clearance for the continuation and completion of my study.

Yours sincerely

CJ Bester

I, Nozipho Mpanza, in my designation as Senior Manager Maintenance and Operations, herewith give my permission for ethical clearance and completion of the study conducted in Eskom Distribution Free State.

Mpanza　　　　　　　　　　20/02/2019

Signature　　　　　　　　　　**Date**

8.4 APPENDIX D – TURNITIN report for plagiarism

ata_Quality_-_9704051_-_10_Mar_2019_-
_Lang_Ed_complete_002.docx

ORIGINALITY REPORT

8%	7%	3%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	uir.unisa.ac.za Internet Source	<1%
2	repository.up.ac.za Internet Source	<1%
3	forddele.dk Internet Source	<1%
4	businessperspectives.org Internet Source	<1%
5	epdf.tips Internet Source	<1%
6	ir.cut.ac.za Internet Source	<1%
7	www.saibw.co.za Internet Source	<1%
8	Carlo Batini, Monica Scannapieco. "Data and Information Quality", Springer Nature, 2016 Publication	<1%

8.5 APPENDIX E – Confirmation of language editing

From : I. Manase (PhD UKZN)

Department of English

University of the Free State

P o Box 339

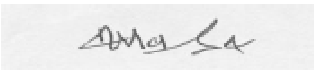
Bloemfontein

Date : 08 March 2019

Confirmation of proofreading and editing of Mr Charl Johannes Bester's MTech dissertation titled: "The influence of Transactional Data Quality Improvements on Monetary Savings of Eskom Distribution, Free State"

This serves to confirm that I have proofread and edited Mr Charl Johannes Bester's above-mentioned dissertation. The suggested sentence and language construction changes have been attended to, and as such, the dissertation is ready for submission.

Sincerely,



Email: irimanase@gmail.com / Manasei@ufs.ac.za