A PARSIMONIOUS DISCRIMINANT ANALYSIS MODEL THROUGH A COMPARISON OF STEPWISE LDA AND FACTOR ANALYSIS-BASED LDA

Ms B S Mheta

23766433

orcid.org/0000-0002-5463-3757

Dissertation submitted in partial fulfilment of requirements for the degree *Magister Curationis* in Statistics at the Mafikeng Campus of the North-West University

Supervisor: Prof T V Montshiwa

Co-supervisor: Dr K N Mpeta

Examination: May 2020

DECLARATION

I, Bonolo S Mheta, student number 23766433, declare that this dissertation is my own unique work. I further declare that I am the sole author of the study and that reproduction and distribution thereof by the North-West University, will not infringe on the rights of any third party. The document has not previously been submitted by me in its entirety or in part, for a degree at this or any other institution. I declare that materials sourced elsewhere other than my original work, have been duly acknowledged in the form of complete references.

……………………………………………. …………………………………………………

Signature Date

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Prof TV Montshiwa and co-supervisor, Dr K Mpeta, for their patience, guidance, motivation, enthusiasm and immense knowledge.

Nobody has been more important to me during my studeis than members of my family. I wish to express my profound gratitude to my sisters, Kelebogile Mnduzulwana and Tshegofatso Mheta, for their support and helping me putting the pieces together. Most importantly, I wish to thank my loving mother, Disekwane Mheta, for providing me with unfailing support and continuous encouragement thought out my years of studies. I will, forever, be grateful for your love.

I wish to thank the North-West University Postgraduate Bursary and the Economic and Management Sciences Faculty Bursary, for providing the funding which allowed me to undertake this research.

Most importantly, I give gratitude to the Lord Almighty, for being my sense of light in darkness.

Thank you!

LIST OF ABBREVIATIONS AND ACRONYMS

CA - Cluster Analysis

CFA - Confirmatory Factor Analysis

DA - Discriminant Analysis

DPLS - Discriminant Partial Least Squares

ELM - Extreme Learning Machine

FA- Factor Analysis

FDA – Fisher's Discriminant Analysis

LDA- Linear Discriminant Analysis

LFDA – Local Fisher Discriminant Analysis

NNC - Nearest Neighbour Classifier

PCA - Principal Component Analysis

PDA - Principal Discriminant Analysis

SDA - Stepwise Discriminant Analysis

ABSTRACT

The aim of this study was to identify the best approach in fitting parsimonious Discriminant Analysis (DA) model(s) with several independent variables, using two comparison models; SDA and CFA-LDA. The objectives of the study were to: compress too many independent variables using Exploratory Factor Analysis (EFA); verify results of Exploratory Factor Analysis (EFA) using Confirmatonary Factor Analysis (CFA); use factor scores from CFA-LDA to fit LDA; use all candidate variables to fit SDA; and use several comparison criteria to compare two models: CFA-LDA and SDA. One thousand three hundred and thirteen (1313) learners were selected to participate in the study, with 23 independent variables. Data was obtained from DataFirst and learners drawn from various schools across the nine provinces of South Africa. Results of Exploratory Factor Analysis (EFA) revealed 16 variables that can be grouped into three factors. CFA models support EFA in selecting the best model fit indices of Standardized Root Mean Squared Residual (SRMR), Root Mean Square Error of Approximation (RMSEA), Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI) and Comparative Fit Index (CFI) with the appropriate cut off criterion within the ranges of adequate fit, found in all the models, with the exception of NFI. The results obtained using Chi-square test of association (<0.000) and Mann Whitney's U test (<0.05) showed significance statistical difference. SDA had high accuracy of classification. The hit ratio was 55.7% while Apparent Error Rate (APER) was 41.3%. Using all 23 independent variables revealed SDA as the efficient and best model selected. This confirms results from the literature and it is concluded that SDA is the best and efficient model to fit a parsimonious model. Thus, the study is relevant and it is recommended that SDA be used in other fields of study.

Keywords: Two-step cluster analysis; Exploratory factor analysis; Confirmatory factor analysis; CFA-LDA; SDA

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER ONE

BACKGROUND AND MOTIVATION

1.1 Introduction

A parsimonious model accomplishes a desired level of explanation or prediction with as few predictor variables as possible. According to Vandekerckhove *et al*. (2015:3), "Parsimony is essential because it helps discriminate the signal from noise, allowing better prediction and generalisation to new data." In addition, parsimony is essential for achieving less complex models (Myung, 2000). Non-parsimonious predictive models are prone to covariate redundancy since they tend to include unnecessary independent variables. Krzanowski (2000) states that redundancy of covariates in discriminant analysis (DA) has been described as a problem that can affect the performance of this classification algorithm. In fact, some authors have argued that there may be a trade-off between the ease of interpretation and the performance of a classifier (Messick, 1995; Bontis, 2001; Lobo, 2008). Krzanowski (2000) further elaborates that when this happens, the researcher then has to make sense of the number of observations and has to try and meet the objectives of the research by selecting an appropriate method of analysis. The field of multivariate analysis is concerned with coming up with solutions to such problems (Krzanowski, 2000). The negative impacts of the redundancy of covariates on the performance of discriminant analysis (DA) motivated the researchers to conduct the current study in order to investigate methods that can enable one to fitting a parsimonious DA model.

With the recent development of high-throughput technologies, for example, in molecular biology, many strategies have been developed to deal with large-dimension data, especially where the number of observations is lower than the number of variables (Mary-Huard and Robin, 2009). In particular, selection and compression of variables have become popular strategies to reduce the dimension of data (Mary-Huard and Robin, 2009). Selection methods allow the discarding of most variables to obtain a classifier based on a few variables only (Mary-Huard and Robin, 2009). However, the selection process can be unstable, an indication that the list of selected variables may depend on the sampling of the original data set into a training sample and a test sample (Mary-Huard and Robin, 2009). Compression methods build combinations of the initial variables in accordance with the label and can be more unstable depending on the sampling, however, the new variables are combinations of all old variables (Mary-Huard and Robin, 2009). It is these limitations, which led to the interest in conducting the current study, in order compare the selection approach to the compression method in an attempt to fitting parsimonious DA models.

As a compression approach, Linear Discriminant Analysis, through principal components analysis (PCA-LDA) or factor analysis (FA-LDA), is proposed and widely used by research in previous studies, in attempt to achieve parsimonious DA models (Pereira *et al.*, 2006; Yang and Yang, 2003; and Zhang *et al*., 2016). In PCA-LDA or FA-LDA, PCA or factor analysis is first used for dimension reduction before fitting a LDA) model. More specifically, PCA or factor analysis is used "to create groups or components to represent variables measuring the same concept while LDA is used to create a model from the principal components or factor scores in order to allocate new subjects" (Maugis *et al*., 2009: 3872). This approach is known to reduce over-fitting of the DA model by projecting the dataset onto a lower-dimensional space that best describes the data (Maugis *et al*., 2009). The purpose of this study was to use this approach to reduce the dimension of the dataset with minimal loss of information and to fit a parsimonious model, using only relevant and theoretically significant covariates. In this study, confirmatory factor analysis (CFA) was used to confirm results of exploratory factor analysis (EFA), prior to fitting LDA and the model founded from a combination of ECA, CFA and LDA, referred to as confirmatory factor analysis LDA (CFA-LDA).

Stepwise Discriminant Analysis (SDA) is an analytic procedure used to reduce the number of variables and a classic feature selection method (Arvanitoyannis *et al*., 1999). SDA is "concerned with selecting the most important variables while retaining the highest discrimination power possible" (Stąpor, 2016: 9). Stepwise discriminant analysis can be formulated as a feature selection problem in the recognition of patterns (Stąpor, 2016), which is the process used to select a subset of relevant features for use in the model to be constructed. As such, this study provides a comparison of SDA, which is a selection method, to a compression approach FA-LDA, with the intention of fitting a parsimonious LDA model.

1.2 Problem statement

When fitting DA models using data with too many variables, it is essential to consider the principle of parsimony. This is because models with too many variables (complex models) are too cumbersome to interpret and may not fit the data well. By taking the principle of parsimony into consideration, one is able to fit DA models that are less complex and simple to interpret. "Goodness of fit seeks a model that can fit the data well with accurate prediction" (Myung, 2000: 3). As such, goodness of fit is a key measure of parsimony. There are several goodness of fit tests such as the Akaike Information Criterion (AIC), classification accuracy rate and the Cramer-von Mises criterion, among others, that one can use to evaluate a DA model for parsimony. However, several studies have used a very limited number of criteria to compare and select parsimonious discriminant analysis models (Hooper *et al*., 2008). For example, Grenoble *et al*. (2007) used only classification models (AIC and BIC) to compare several discriminant analysis models. Other examples include Vandekerckhove (2015) and (Myung, 2000; Wagenmakers & Waldorp, 2006) who used only the AIC and Bayesian Information Criterion (BIC) respectively as their only comparison criterion.

The use of few comparison criteria is problematic since it makes the comparison process prone to selection bias and jeopardises the validity or statistical rigour of results. If not dealt with, model selection bias can lead to the researcher giving misleading recommendations, and what is identified as the best model for fitting data with many independent variables may not really be true. Researchers, who are not experts in the field of Statistics can use the less efficient model, leading to the development of less accurate models with redundant variables. Thus, the need for a study, using several model comparison criteria in the context of DA rather than one or two. This study is significant as it helps confirm the statistical rigour of some suggested DA methods used in fitting models for data with too many covariates and as a way of reducing model selection bias, which can result from using very few model comparison criteria. Hence, the purpose of this study was to address this need.

1.3 Aim and objectives of the study

The aim of this study was to identify the best approach to fitting parsimonious DA model (s) for data with too many independent variables through comparing one commonly used covariate selection approach (SDA) and a commonly used variable compression method (CFA-LDA).

The objectives of the study were to:

1.3.1 Use EFA to compress independent variables;

1.3.2 Use CFA to validate results of EFA;

1.3.3 Fit a Linear Discriminant Analysis (LDA) model, using factor scores from EFA and CFA (CFA-LDA);

1.3.4 Fit SDA, using all candidate variables;

1.3.5 Compare CFA-LDA and SDA, using several model comparison criteria; and

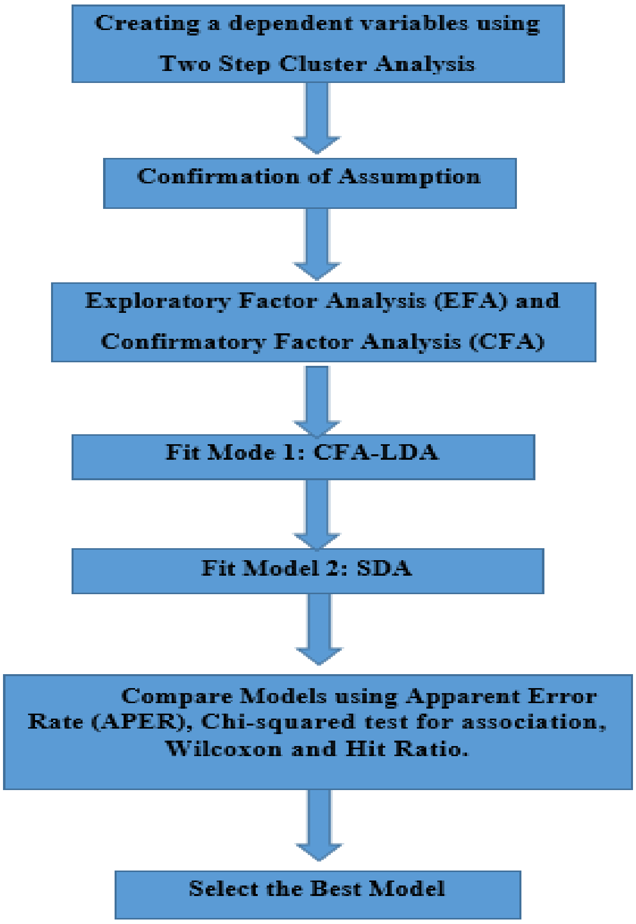
1.3.6 Use the findings of the study to make recommendations.

1.4 Expected contributions of the study

This study is expected to contribute to the literature by extending the scope of previous studies through comparing SDA and CFA-LDA, using as many goodness of fit criteria (model comparison criteria) as possible compared to previous studies which used only a few criteria (one to two). using several selection criteria is expected to ensure that the correct model is selected and recommended thus, minimising model section bias. Such approach will also ensure statistical rigour of findings in the context of modelling data, with many independent variables.

1.5 Significance of the study

Several studies have so far been conducted on CFA-LDA and SDA (Marcialis and Roli, 2002; Maugis *et al*., 2006; Pedagadi *et al*., 2013 and Pereira *et al*., 2006). However, such studies have focused mainly on the efficiency of these models, an indication that more research needs to be conducted on this topic. The findings of this study will assist researchers and statistical analysts with the best method to use when seeking to fit a parsimonious LDA model in instances where too many independent variables have to get into data.



1.6 Figure 1.1: Summary of analytical framework with regard to data

Figure 1.1 provides a summary of the methods used to analyse data in this study. A detailed description of the methods is provided in subsequent sections in this Chapter.

1.7 Ethical considerations

The study did not involve any physical contact with animals or human participations. The postgraduate manual was consulted when conducting this study to ensure that rules and regulations of the North-West University were adhered to. Personal information, such as demographic characteristics of participants used in the study were kept confidential and not be disclosed to anyone. However, such information could be used in academic publications and presentations. Secondary data used in the study was obtained from DataFirst and the researcher(s) had to write a proposal to explain how the data was to be used in this study. Permission to use the secondary data was also requested and granted by the owners of the data. The researcher(s) were granted ethical clearance by the North-West University to enable them to conduct the study. It is worth noting that although the results from the data can be used for academic publications, as a by-product of this research, the focus of the study was actually on the statistical methods and not on the dataset.

1.8 Scope and limitations of the study

The following limitations were encountered during the study: secondary data is considered to achieve the objectives because it is accessible, already cleaned and stored in an electronic format. The methodology adopted in this study focuses on six multivariate methods only which are the Two Step Cluster Analysis, EFA, CFA, CFA-LDA and SDA. The literature reviewed is limited to studies around the multivariate models of interest and the study will focus only on academic sources which are not older than 10 years in order to focus on the latest developments in the literature only. The scope of the study is limited to the methods and not to the data per say.

1.9 Structure of the study

This study is divided as follows:

Chapter 1 is the introduction. It provides the general overview of the study, the statement problem, aim and objectives of the study, significance of the study, analytical framework, ethical considerations, scope and limitations of the study. Chapter 2 focuses on the literature review. The chapter provides insights on empirical studies conducted in the area. Relevant studies are reviewed in order to select the best model in a parsimonious discriminant analysis model, comparing CFA-LDA and SDA. Studies that employed exploratory factor analysis, confirmatory factor analysis, stepwise factor analysis approaches for the same purpose as the current one, were also reviewed in this Chapter. Chapter 3 is the research methodology. In this Chapter, the researcher provides information on how data was collected and the research approach used in conducting the study. Chapter 4 focuses on data analysis and presentation of results. The data analysis and interpretation of results are done to provide a clear picture of the combination of CFA-LDA and SDA, using several a model comparison criteria to select the best model. Chapter 5 provides the conclusion and recommendations for the study. In this Chapter, an overview of the results and conclusions are provided. Recommendations for future studies are indicated and suggestions for an appropriate model are provided.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter focuses on the review of literature related to the study. The studies reviewed focused on Linear Discriminant Analysis (LDA), using Principal Component Analysis (PCA) or Factor Analysis (FA) scores and/or Stepwise Discriminant Analysis (SDA). The Chapter is structured as follows: section 2.2 focuses on the literature review while section 2.3 is a summary of the Chapter.

2.2 Literature review

High dimensionality is defined in PCA and DA as "a large number of principal components or discriminating axes, all of which carry a significant extent of the total differentiation. High dimensionality is desirable in cases where very similar analyses need to be differentiated" (Stewart *et al*., 2014:76). Pandey *et al*. (2011) applied cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and linear discriminant analysis (LDA) in their study to evaluate temporal/spatial variations and to understand a large complex water quality data set of the Fuji River Basin. Data was collected over 8 years (1995–2002) and comprised 12 parameters from 13 different sites and a total of 14 976 observations. The authors used hierarchical cluster analysis to group the 13 sampling sites into three clusters, and referred to them as relatively less polluted (LP), medium polluted (MP) and highly polluted (HP) sites. The classification was based on the similarity of the characteristics of the water quality under study. Subsequently, Pandey *et al*. (2011) applied FA/PCA to the data sets of the three different groups obtained from CA, which yielded five factors for data obtained from the LP cluster, five for data from the MP cluster and three factors for data obtained from the HP cluster.

Pandey *et al*. (2011) further explain that the factors obtained from FA revealed that the parameters responsible for variations in water quality were mostly related to discharge and temperature (natural) and organic pollution in relatively LP areas, organic pollution and nutrients in MP areas, and organic pollution as well as nutrients in HP areas. These results show the benefit of using CA prior to FA when dealing with multidimensional datasets, which could be cumbersome to understand. In addition to CA and FA, the authors used LDA to categorise the water samples based only on the variables obtained from factor analysis, and this also assisted in selecting the most significant variables to retain in the discriminant model.

The final model comprised only six parameters. Pandey *et al*. (2011) argue that LDA assisted in reducing the dimension of the large data set hence, the results of their study confirm the usefulness of multivariate statistical techniques for analysis and interpretation of complex data sets, thus a similar approach was adopted in the current study. Furthermore, the current study is an extention of the one conducted by Pandey *et al*. (2011), as it provides a comparison of the efficiency of a straightforward stepwise LDA model, using all independent variables with a LDA model, taking into consideration,discriminant scores as independent variables.

Qu *et al*. (2003) applied LDA to detect prostate cancer. The authors collected prostate cancer data comprising 248 subjects with 48 538 variables of different protein mass - point (generated using the Protein Chip technology). The authors caution that PCA would have been suitable for data reduction if the sample size was greater than the number of variables. However, the data used by Qu *et al*. (2003) had an over-determined dataset (the number of variables was about 196 times the number of subjects).

The DWT used by Qu *et al*. (2003) reduced 48 538 variables to 1 271 wavelet coefficients, however, only 11 of these coefficients had the highest discrimination as per Mahalanobis distance and these coefficients were included in the LDA model. The results of study showed that "the linear classifier with the 11 wavelet coefficients, detected prostate cancer in a separate test set with a sensitivity of 97% and specificity of 100%" (Qu *et al*., 2003: 143). Although the DWT was efficient in reducing the dimension of the dataset in the study by Qu *et al*. (2003), this method was not considered in this study since it is beyond its scope. In the current study, the researcher(s) used less determined data compared to Qu *et al*. (2003). However, in the current study, the data set was extended by considering a different data reduction method (PCA). Unlike Qu *et al*. (2003), the current study does not only make use of these methods but provides a comparison of the efficiency of a straightforward stepwise LDA model, using all independent variables with a LDA model, taking into consideration, discriminant scores as independent variables.

Jung and Qiao (2014) applied Principal Component Analysis (PCA) as a pre-processing method for various classification methods as follows: LDA; quadratic discriminant analysis (QDA); Support Vector Machines (SVM); Distance-Weighted Discrimination (DWD); and Minimum Distance Empirical Bayes rule (MDEB). The authors conducted repeated experiments to determine the accuracy of classification of these methods. The study by Jung and Qiao (2014) revealed the usefulness of PCA as a dimension reduction method for classification methods. However, in the current study, the researcher(s) was more concerened with comparing the proposed classification methods rather than comparing the efficiency of a straightforward stepwise LDA model, using all independent variables, taking into consideration, LDA model that uses discriminant scores as independent variables (as is the case in the current study). Thus, the current study extends the scope of the one conducted by Jung and Qiao (2014) in this regard. The scope of the current study is, however, limited to LDA, compared to the one by Jung and Qiao (2014).

Ferizal *et al*. (2017) applied Principal Component Analysis (PCA) as a dimension reduction method prior application of Linear Discriminant Analysis (LDA) to explain “the gender recognition system through a human facial image” (Ferizal *et al*., 2017: 1). The authors used several facial attributes from two scenarios: natural face (with no facial mask) and masked face. The methods (PCA and LDA) were then applied to each of the two datasets (with or without face mask). It is important to note that Ferizal *et al*. (2017) compared models built from data collected from a natural face to the one built using data from a masked face, thus differentiating it from the current study. In the current study, a comparison of the efficiency of a straightforward stepwise DA model was done, using all independent variables with LDA model, taking into consideration, discriminant scores as independent variables, thus focuseing on the models and not the data. Simply put, both the current study and the one by Ferizal *et al*. (2017) considered PCA as a dimension reduction method for LDA, however, the former was interested in the models while the latter was interested in the datasets, thus the current study is an extention of the scope of the study conducted by Ferizal *et al*. (2017).

Wu *et al*. (2016) conducted a study to discriminate three kinds of apples, using their proposed method referred to as sorting discriminant analysis (SDA). SDA was compared to LDA, discriminant partial least squares (DPLS) and nearest neighbour classifier (NNC). DPLS was applied using raw data (with the original variables) to reduce the high-dimension of the dataset and to extract the most suitable covariates. Wu *et al*. (2016) argue that DLPS achieves this through the maximisation of the covariance between the predictors and predicts a binary dependent variable. The authors describe SDA as a three-stage method in which PCA reduces the dimension of the original data, and the eigenvectors sorted using an equation which normalises Fisher’s ratio and the eigenvector. The authors further state that the third and final step of SDA, implements LDA to extract discriminant information from the compressed data.

Wu *et al*. (2016) also used a combination of PCA, LDA and NCC to classify apples. To achieve this, the authors reduced the dimension using PCA. The discriminating information from the reduced dimension was extracted using LDA, and the apples classified using NCC based on the information extracted from LDA. The results of the study revealed that LDA and SDA extracted discriminant information was better than PCA and DPLS. However, in the method proposed by Wu *et al*. (2016), SDA was found to best. Although they were found to be useful in extracting discriminant information, SDA, NCC and DPLS are beyond the scope of the current study.

However, PCA and LDA were used in the current study just like in the one by Wu *et al*. (2016), but the current study is an extention of the one by Wu *et al*. (2016) as it provides a comparison of LDA from principal components to LDA from original variables to confirm the efficiency of combining PCA and LDA.

Marcialis and Roli (2006) evaluated face recognition systems based on a fusion of two algorithms referred to as LDA and PCA. The face data bases comprised AT&T datasets with 10 different images of 40 subjects and Yale dataset comprising 11 images per 15 classes. The authors used LDA to factor AT&T original datasets into 5 images per class (200 images) and Yale original datasets with 5 images per class (75 images). Thereafter, the authors tested for the average performance for both datasets using individual face recognition, resulting in fusing between LDA and PCA, which is more accurate.

Marcialis and Roli (2006) state that the combination of PCA and LDA improves the performance of the individual recogniser and exhibits better reliability. The data does not show a strong variation of performance. Marcialis and Roli (2006: 1) maintain there are two main benefits of such fusion as follows: reduction of dependence on environmental conditions with regard to the best individual recogniser; and overall performance improvement over the best individual recogniser. Thus, fusion is investigated under different environmental conditions, namely; "ideal" conditions, characterised by a very limited variability of environmental parameters, and "real" conditions with large variability of lighting and face expressions". Application of face recognition is very critical and it can be concluded that neither LDA nor PCA is better than the other.

Pedagadi *et al*. (2013) applied PCA as a pre-processing step for data reduction and Local Fisher Discriminant Analysis (LFDA) for pedestrian re-identification, using a very high dimension dataset extracted from HSV colour image. The dimension of the data matrix had an initial dimensionality of 11, 253 dimensional vector. Using dimensionality reduction methods, the authors maintain the redundancy in colour space representation, firstly, used PCA with unsupervised data to reduce dimension/ extract data. PCA is used to define dual colour space useful for estimating a reliable embedding space in a subsequent stage. Thereafter, a second supervised data reduction process was conducted with Local Fisher Discriminant Analysis (LFDA), using training data to re-identify problems. To avoid singular matrices, a regularisation step is introduced during LFDA. The regulation assists in achieving a working solution. The last step proposes a characterisation of the Match Characteristic curve, which in essence, reports the percentage of the re-identification problem solved, from 0% to 100%.

Pedagadi *et al*. (2013) conducted an experiment on 3 different datasets in accordance with the steps described above . The first dataset used was the Viper hand generated pedestrian dataset, with 632 variables collected over several months. The second dataset was 3 DPES, with 191 variable images collected during different times of the day. The third and final Cavair dataset consisted of 72 individuals for re-identification images taken from a camera at a shopping centre. The datasets were divided into training and testing. From the datasets obtained, PCA was performed on the mean subtracted data and the results transformed. From the projected data found in PCA, a regularisation step (LFDA) was performed.

Pedagadi *et al*. *(*2013: 3318) conclude that “experiments conducted on three publicly available datasets confirm that the proposed method outperforms the state-of-the-art performance, including all other known metric learning methods”. Furthermore, the method is an effective way to process observations, comprising multiple shots and is non-iterative: the computation times are relatively modest. Marcialis and Roli (2002) applied PCA and LDA to detect face recognition. Fusing two appearance-based (or statistical) approaches to face recognition, PCA representation (“eigenface” approach) and LDA representation (“fisherface” approach), the authors collected AT&T and Yale face dataset comprising 10 different images of 40 subjects. The dataset was subdivided into a training set made up of 5 images per class of 200 images and a test set made up of 5 images per class of 200 images. From both datasets of face images, pre-processing phase such as rotation, re-scaling or normalisation was not needed. Using face recognition on the ATA dataset, Marcialis and Roli (2002) applied PCA transform as a dimension reduction on the original image face space prior application of LDA to explain face recognition system. The authors found that PCA and LDA were not correlated as one could imagine since PCA must be able to reduce high variance and later LDA.

Marcialis and Roli (2002) analysed the differences between these two techniques, but their study did not focus on the possibility of fusing these techniques. It is observed that the apparent and obvious strong correlation of LDA and PCA, especially when frontal views are used and PCA applied prior LDA, discourages the merging of such algorithms. Fusion of LDA and PCA for face recognition and verification is, however, worth theoretical and experimental investigation. Marcialis and Roli (2002) further argue that the combination of PCA and LDA provides improvement with regard to performance, and better identification of reliability and robustness.

Wang *et al*. (2004) applied Fishers Linear Discriminant Analysis (LDA) and Principal Components Analysis (PCA), which are both data / dimension reduction methods, to analyse chemical sensor arrays. A sample size of 15 sensors and 5 coffee beans was used. Wang *et al*. (2004) used two datasets of 2000 and 225 respectively in their study. Both datasets projected PCA against PDA and LDA. To begin with, PDA is used to regularise significance in the covariance of the data. In addition, PDA is used to balance LDA and PCA. On the other hand, LDA is applied to check discrimination in the sub spaces within the projected covariance. Overall, better discrimination results can be found by using both mean and variance information.

Wang *et al*. (2004) used a hybrid model (PCA) that incorporates regularisation parameters of LDA and PCA. Wang *et al*. (2004) conducted a reduction of dimensionality using variance subspace of the PCA dataset. Since the model was biased, LDA was introduced to balance both techniques in accordance with the information provided. Thereafter, using within class variance projection, LDA was applied. LDA has a couple of setbacks/ drawbacks, namely; as a supervised technique, LDA has a tendency to over fit in small-sample-size problems, where dimensionality is higher than the number of vectors in the training stand, and shows more structural problems with LDA when the dataset has more information in the variance than in the mean. Thus, LDA tends to produce poor results. The results revealed that using both techniques (PCA and LDA) is better since PDA provides higher predictive accuracy; and the discriminant and variance projections are located in the appropriate distribution of data.

Kara (2009) applied PCA, LDA and CA to evaluate herbal tea samples in Turkey. Using a sample size of 18 plants, the author used PCA to group the similarities of the whole data set, classified into 5 groups, resulting in 16 elements by extracting values higher than 1. Subsequently, LDA was used to confirm correctly classified predicted groupings/classes. According to Kara (2009), cluster analysis is used to cluster samples into groups that determine the relationships. The classification was based on the similarity of the characteristics of plants under investigation. The study revealed that plants can be grouped into different trace elements.

Yang *et al*. (2008) applied Principal Component Analysis (PCA) as a dimension reduction method prior to application of Linear Discriminant Analysis (LDA) to boost the discriminatory power that explains face recognition systems/tasks. Human face images of data consisting of 400 frontal faces were collected in the study. The data had a high dimension of image space thus, raw image data was used for dimension reduction and creation of sub space for images. Yang *et al*. (2008) concluded that a combination of face tracking and face technology is used to develop a new real time face recognition algorithm. The authors introduced a new face recognition to improve the accuracy of video images using sequences.

Castaño *et al*. (2016) applied LDA, PCA, Extreme learning machine (ELM) and a combination of the three approaches, LDA-PCA-ELM in a study to assess the number of basic functions. The UCI repository dataset used consisted of 15 samples. On a set of problems, LDA approach has a better performance so PCA –ELM algorithm was joined with LDA to perform pruning of hidden nodes hence, the two approaches are combined. In their study, Castaño *et al*. *(*2016*)* introduced an LDA algorithm prior using PCA in pattern recognition and machine learning. LDA is for modelling different classes. A combination of LDA and PCA stands a better chance of class selection and ELM was used for transformation. The LDA-PCA-ELM algorithm is introduced for classification. The method is a competitive proposed approach and improves accuracy. The authors found that classification applied in LDA-PCA-ELM algorithm and ELM algorithm, can be applicable to several problems, such as regression in classification.

Diraman *et al*. (2011) applied Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to evaluate Turkey Virgin Olive oil from Aegean region based on three North Aegean (Ayvalik cultivar), South Aegean (Memecik) and Izmir Peninsula (Erkence cultivar) cultivars of olives in Turkery. The data was collected over 2 years (2001-2002 and 2002-2003) and comprised 187 sample sizes divided into 3 subgroups. The authors used PCA treatment for both functions of high data dimensions to find a pattern, putting them into separate groups. Furthermore, Diraman *et al*. (2011) applied LDA for parameter classification of oil samples. The results revealed that PCA and LDA made good classification approaches.

Sun *et al*. (2011) applied PCA, CA and LDA for mutton samples developing in two different regions of China. The dataset collected from slaughter houses in pastoral and agricultural regions in China, consisted of 99 animals, spanning 7-10 months. Data reduction was performed using PCA while CA was used for clustering similar groups with unsupervised data. The results revealed that using a combination of multi element analysis for PCA, CA and LDA were successful methods for pastoral farmers and agriculturalists in different regions of China.

Yang and Yang (2003) analysed a suggested PCA plus LDA algorithm in building a theoretical foundation framework. The authors performed dimension reduction using PCA before application of LDA, transforming Fisher criterions. The two-linear high dimensional and singular case methods are mostly used in LDA. PCA plus LDA were used to analyse data collected from Oracle Research Lab (ORL). Face image database comprised 10 different images of 40 distinct subjects. The authors found that the traditional LDA had the following limitations:

1) Traditional algorithm cannot be used directly in that the within-class scatter matrix is always singular; and
2) High dimensional image vectors lead to computationally difficulty.

Yang and Yang (2003) concluded that the minimum distance classifier transformed space for feature extracted from fisherfaces and proposed PCA plus LDA algorithm as it is more effective than previous LDA-based methods. This combination does not robust information.

Nwosu *et al*. (2012) used LDA to estimate the three criteria: Fishers criterion; Welchs criterion; and Bayes criterion to determine the best classification in two groups (regions). The dataset comprised 24 observations, equally divided into two groups (regions A and B). LDA, on two groups (regions), was used to analyse a good discriminant function with the apparent error rate (APER) to compare and determine the best criteria among the three criteria, thus, the assumed best criteria are selected by having the least error rate. According to Nwosu *et al*. (2012), APER was compared and the three criteria gave the same results, with the equal apparent error rate as well as with the confusion matrix. In conclusion, the Fishers, Welchs and Bayes criteria performed equally the same.

Chen *et al*. (2000) experimented a new LDA-base face recognition system for small size problems. The authors proposed a new LDA-base technique to solve small sample size problems with a primary testing database comprising 10 different faces (classes) for each of 128 persons (observations). The pixel grouping was done by extracting geometric features and transforming image to normalised size of the face recognition system thereafter, reduction of dimension was performed for feature vectors. LDA proved that vectors derived in the null space of the within-class scatter matrix, provided the best discriminant vector set selected. Additionally, PCAs are equal to the optimal discriminant vector derived in the original space and form the most discriminant vector set. Thus, Chen *et al*. (2000) concluded that there is a significant performance improvement in the new LDA of a face recognition system in accuracy, training efficiency and stability.

Using application to face recognition, Yu and Yang (2001) analysed a proposed LDA algorithm for high dimensional data classification. The data was collected from ORL and consisted of 10 images from 40 individuals and 400 observations. In order to reduce variation, the experiment was done at least 10 times, randomly; choosing 5 images per person for training and another 5 images per person for testing. Data with no dimension reduction results gave an accuracy of 86.6% while with dimension reduction, the accuracy was 90.8%.

Dubuisson *et al*. (2002) performed two proposed statistical-based techniques, PCA and LDA. The authors collected facial expression data comprising 210 subjects with 6 variables of different facial expressions. The test set 1 comprised CMU-Pittsburgh database with 194 different images of 40 subjects and Yale dataset comprising 11 images per 15 classes. PCA was used as a suitable data reduction method, considering a learning set that contains different class samples. In order of importance of facial recognition task eigenvectors,the most discriminant projection was selected on principal components. The forward stepwise selection method was then used to select the most discriminant “optimal” principal component and for stability, the fisher criterion was used to select the optimal set of components. Finally, from LDA, new samples classification was generated for the dimensional discriminant subspace.

The authors proposed a feature selection process for PCA to sort principal components to solve specific recognition tasks. The methods proposed help to improve the accuracy of classification. Dubuisson *et al*. (2002) found that classification results were strongly influenced by the choice of representation and also indicate that a specific representation must be designed for each classifier.

In addition, Zhao *et al*. (2013) used LDA and PCA to analyse wheat grain and soil from China. The sample size collected comprised 61 paired wheat and 22 soil samples. PCA was used to reduce the dimension of the dataset between two kinds of wheat samples, while LDA had two wheat sample groups, split and comprising 41 calibration set samples and 20 test set samples. The authors concluded that out of 22 samples, only 7 where significant. Zhao *et al*. (2013) found that LDA is a favourable subspace on full face recognition system.

Chan *et al*. (2010) conducted a study of facial biometrics to assist with automatic face recognition. The dataset consisted of 400 images. PCA was used to reduce low dimension of original face representations. Chan *et al*. (2010) proposed application of LDA for improving the discriminant. Additionally, the Euclidean distance measurement was used as a benchmark for both PCA and LDA. The results revealed that for both face identification and face verification, LDA gave better results than PCA. The LDA discriminant showed more accuracy on the face biometric study.

On the hand, Pereira *et al*. (2006) conducted a study in Brazil to determine gasoline adulteration by a solvent and identify the solvent added, using infrared absorption spectroscopy and a combination of two multivariate techniques; PCA-LDA. Multivariate techniques were used to analyse the 217datasets collected from October 2003 to May 2004 from a gas station in Minas Gerais, Brazil. Testing for the construction of the model with the confirmation of good quality, the authors used solvent tracer test with 200 samples of adulterated and 17 samples of unadulterated gasoline. The researchers applied PCA for data reduction, resulting in visual principal components information. LDA from the principal components, resulted in 100% validation and classification of modelling sets. The results obtained in this study show that PCA-LDA is a strong technique related to 100% accuracy with FTIR in the quality control of automotive gasoline. The identifying type of solvent added had adulterated gasoline classification of 96% efficiency with sensitivity of 8% and unadulterated gasoline classification of 93% efficiency.

Kher *et al*. (2006) applied Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to analyse High Performance Liquid Chromatography (HPLC) and Infrared Spectroscopy (IR) of blue ballpoint pen ink. The data was collected from 4 wavelengths (254, 279, 370 and 400 nm). In this study, PCA was used to estimate the pen samples to be separated while LDA was applied in the classification of a pen class.

Zhao *et al*. (2010) performed pattern recognition, using PCA and LDA in a study of cracked eggshells from Hangzhou Hennery, China. The experiment had a total sample size of 120 eggs (60 cracked eggs and 60 intact eggs). In order to remove certain information from the dataset, PCA was applied as a data reduction technique. LDA was then used as a grouping procedure.

Calder *et al*. (2001) conducted a PCA and LDA study of facial expression with 50 components. The authors used full image PCA with the first 50 components and three stepwise linear discriminant analyses with the 3 independent variables (facial identity, facial expression and gender). LDA was used for the first 50 components of shape-free PCA, the first 30 components of shape free and the first 20 components of shape information. The authors found that PCA facial expression supports recognition and that PCA of limited number of anatomical feature, points to facial identity and gender. Furthermore, shape components and shape-free components rated the best optimal rate of categorisation. Calder *et al*. (2001) found that LDA and PCA provide a consistent model for psychological models of facial expression.

Skrobot *et al*. (2007) applied PCA and LDA to detect liquid in gasoline at a gas station. The authors collected gasoline, comprising 75 gas samples with 4 solvent / liquid variables of chromatography-mass spectrometry. Skrobot *et al*. (2007:3399) found that “PCA and LDA classification model presented a trend towards classifying mixed samples with a lower amount of solvent in the test set as pure”. Khan and Farooq (2012) applied Principal Component Analysis (PCA) as a dimension reduction method prior application of Linear Discriminant Analysis (LDA) “to maximise discriminatory power” and explain face recognition system. Human face images data, comprising 400 frontal faces, were collected in this study. Raw image data was used for dimension, using PCA prior to creation of a sub space for images. By selecting the most discriminant face projection, a new algorithm was formed. Thus, there is need for a combination of face tracking and face technology to develop a real time face recognition, resulting in improved face recognition.

Furthermore, in their study aimed at forecasting the level of debris flow, Lei *et al*. (2011) combined Principle Component Analysis (PCA) with Linear Discriminant Analysis (LDA) and Data Mining Results (DRS). The authors found that the hazard assessment approach was encouraged since the DRS method was more accurate over PCA and LDA analysis for the flow of debris. Although various methods, such as PCA-LDA algorithms could have been applied, Lei *et al*. (2011) compared Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA). The authors found insignificant results, an indication that both techniques did not meet requirements of threshold variables and advanced data mining techniques (DRS) used in achieving the thresholds, especially in situations where the results were 15% higher with accuracy in performance classifications.

The purpose of the study by Zhang *et al*. (2016:1) was to "investigate the protective effect of total isoflavones from Radix Puerariae (PTIF) diabetic rats". Principal Component Discriminant Analysis (PCA-DA), Partial Least-Squares Discriminant Analysis (PLS-DA), and Orthogonal Partial Least-Squares Discriminant Analysis (OPLS-DA) were used to perform multivariate analysis. In addition, Zhang *et al*. (2016) evaluated oxidative stress and inflammatory cytokines. Eleven potential metabolite biomarkers, mainly related to coagulation, lipid metabolism and amino acid metabolism, were identified by the researchers. PCA-DA scores plots revealed that biochemical changes in diabetic rats were gradually restored to normal after administration of PTIF. Furthermore, the levels of BCAAs, glutamate, arginine and tyrosine increased significantly in diabetic rats. Treatment with PTIF could regulate the disturbed amino acid metabolism. Consequently, PTIF has great therapeutic potential in the treatment of DM, by improving metabolism disorders and inhibiting oxidative damage.

Kıralan *et al*. (2015) used chemometric analysis to determine Turkish oil hazelnut, structured for Triacylglycerol. Discriminant analysis was used in this study to group the samples according to the geological locations. One-way analysis was used to check hazelnut oil cultivars and discriminate analysis to divide samples according to the groups. Fifty (50) samples were selected for the study, comprising 19 hazelnut cultivars, spanning from 2009 – 2010 and collected from 4 provinces in Turkey. The results revealed that hazelnuts with oil and water ranged between 55.01-64.85 and 3.17-4.32. A one-way analysis of variance showed a statistically significant difference between the variables, structured for Triacylglycerol.

Dede *et al*. (2017) used PCA, CA and DA to classify quality of water from nine monitoring stations in Muda River, Malaysia. PCA and CA created two different clusters, using three characteristics of water quality . Discriminant Analysis (DA) was then applied to confirm the clusters. Forecasting was performed for the cluster of membership of new samples created from the discriminant function (DF).

Bansal *et al*. (2012) used PCA and LDA to study face recognition. The authors, firstly, used PCA for dimension reduction of a training data set and image compression. PCA data reduction is used to remove useless and uncorrelated information, referred to as Eigen faces. Secondly, LDA is a technique used to separate data.

Beghdad *et al*. (2011) applied LDA to an intrusion detection method to study the behaviour of students by observing four (4) class sessions at a university. The authors collected real data, comprising four (4) lecture sessions with 13 students per class for three months. Beghdad *et al*. (2011) firstly, used LDA technique to apply an appropriate discrimination between the four lecture sessions. Secondly, "PCA deals with the data in its entirety, for principal components analysis, without paying any particular attention to the underlying class structure" (Beghdad *et al*., 2011). It was concluded that the LDA approach is more suitable than PCA for detection of intrusion.

2.3 Summary of chapter

This chapter has provided a review of literature on the current study. An empirical review was done for studies related to the current study. Most authors in the literature used the traditional SDA approach, however, the study by Wu *et al*. (2016) focused on sorting discriminant analysis (SDA), which is not the traditional stepwise discriminate analysis. SDA was compared to LDA, SDA as a three-stage method, whereby, PCA reduces the dimension of the original data and SDA implements LDA to extract discriminant information from the compressed data. Pandey *et al*. (2011) used CFA since EFA is more effective, however, the results are similar to those of the current study. It is, therefore, concluded that SDA is a more effective technique compared to the combination of CFA-LDA. The current study is different because none of the studies reviewed made use CFA to confirm the results of EFA. Thus, the current study is an extention of the scope of previous studies, considering the fact that the current research makes use of CFA before fitting the CFA-LDA. In Chapter Three, Multivariate assumptions for CFA-LDA and SDA methodologies are discussed in detail, including steps to ensure the model better fits compared to others. The next chapter is the research methodology.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This chapter provides a discussion of the methodology used in conducting the study. The specific objectives of the study as indicated in Chapter One, were to:

- Use EFA to compress independent variables;
- Use CFA to validate results of EFA;
- Fit a Linear Discriminant Analysis (LDA) model using factor scores from EFA and CFA-LDA;
- Fit SDA using all candidate variables;
- Compare CFA-LDA and SDA using several model comparison criteria; and
- Use the findings of the study to make recommendations.

The Chapter is structured as follows: Section 3.1 is the introduction; section 3.2 provides a description of the data and its sources; section 3.3 focuses on the research methodology used in conducting the study; section 3.4 is a description of two-step cluster analysis; section 3.5 provides a discussion on exploratory factor analysis (EFA); section 3.6 focuses on confirmatory factor analysis (CFA); section 3.7 focuses on stepwise discriminant analysis (SDA); 3.8 focuses on CFA-LDA; section 3.9 provides details on SDA; 3.10 provides a comparison of both models (CFA-LDA and SDA); while section 3.11 provides a summary of the chapter.

3.2 Description of data and its sources

Secondary data was used to achieve the aim and objectives of the study. Data used in the study was obtained from Data First, available at: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/627/download/9082. The sample size comprised 1313 learners from different schools across the nine provinces of South Africa (2016). Data was collected using the Innovation Edge to establish the “Early Learning Outcomes Measure 2016”. The Children Observation (ChildObs) ordinal variables adopted in this study were: paid attention; stayed concentrated; careful and diligent; interested and curious; and first used to create the dependent variable, defining how the learner performed during assessment. The two-step cluster analysis, explained later in this chapter, was used to create the dependent variable, using the four variables indicated; the dependent variable was referred to as "child observation". The independent variables consisted of the 23 items used to measure early learning outcomes of the learners under investigation. The outcomes are explained in Table 3.1 and are all measured on a continuous scale.

Table 3.1: Description of independent variables

Name	**Label**
Item 1	Standing on one leg for 10 seconds transformed
Item 2	Catch bean bag both hands transformed
Item 3	Catch bean bag preferred hand transformed
Item 4	Catch bean bag non-preferred hand transformed
Item 5	Cross and square transformed
Item 6	Triangle transformed
Item 7	Draw person transformed
Item 8	String beads transformed
Item 9	Counting transformed
Item 10	Add and sub transformed
Item 11	Sort and class transformed
Item 12	Spatial vocabulary transformed
Item 13	Measurement vocabulary transformed
Item 14	DCCS transformed
Item 15	Pencil tapping transformed
Item 16	Digit span transformed
Item 17	Picture puzzle transformed
Item 18	Empathy transformed
Item 19	Self-awareness transformed
Item 20	Expressive language transformed
Item 21	Expressive vocabulary transformed

Item 22	Oral comprehension transformed
Item 23	Sound discrimination transformed

Statistical Analysis Software SAS® version 9.4 and Statistical Package for the Social Sciences (SPSS) version 25 were used to analyse the data.

3.3 Research methodology

Based on the objectives of the study and the nature of the data collected, a quantitative approach was considered the most appropriate for the study. A quantitative method emphasises on objective measurements and statistical, mathematical, or numerical analysis of data collected through polls, questionnaires, and surveys, or by manipulating pre-existing statistical data using computational techniques (Babbie, 2010 and Muijs, 2010). The advantages of quantitative research are as follows: It is more reliable and objective, it describes assumptions, statistics can be used to generalise findings, it reduces and restructures complex problems in order to reduce the number of variables. It is used to test hypotheses and theories; and subjectivity of researcher in methodology is recognised (Babbie, 2010; and Muijs, 2010).

In this study, the two-step cluster analysis and factor analysis were used as preliminary data analysis methods prior to fitting the CFA-LDA.

3.4 Two-step cluster analysis

The two-step clustering procedure was developed by Chiu *et al*. (2001) to analyse large data. According to Fraley and Raftery (2002) and Hair *et al*. (2010), cluster analysis is a group of multivariate techniques, whose primary purpose is to group objects based on their characteristics. Clustering methods are applied when the intention is to group objects together naturally in various categories (Şchiopu, 2010). According to Moroke (2015), the two-step cluster component of SPSS automatically provides the proper number of clusters if the desired number of clusters is unknown thus, in this study, the two-step cluster analysis was considered and applied.

In step 1, a sequential clustering approach is used. This method scans records one by one and then decides if the current record should merge with previously formed clusters or start a new cluster based on the distance criterion. According to Moroke (2015), by doing this, a new data matrix with fewer cases for the next step is computed. Automatically, SPSS implements this procedure by constructing a modified cluster feature (CF) tree in accordance with MacCallum *et* *al.* (1999). The CF-tree consists of levels of nodes, with each node containing a number of entries and a leaf entry represents the desired sub-cluster (Şchiopu, 2010). According to this procedure, the non-leaf nodes and their entries guide a new record into a correct leaf node (Şchiopu, 2010).

In step 2**,** sub-clusters resulting from the first step are taken as input and grouped into the desired number of clusters (Moroke, 2015). Since the number of sub-clusters is lesser than the number of original records, traditional clustering methods can be used effectively (Moroke, 2015). Similar to agglomerative hierarchical techniques, the pre-clusters are merged using a stepwise procedure (Bacher *et al*., 2004). This procedure is repeated until all clusters are collected in a unit cluster (Bacher *et al*., 2004). Contrary to agglomerative hierarchical techniques, an underlying statistical model is used immediately (Bacher *et al*., 2004).

The two-step cluster analysis approach was used in the current study to test the Children Observation (ChidObs) categorical variables. The approach test if children paid attention, stayed concentrated, were careful and diligent, interested and curious to create one dependent variable defining learners' performance since LDA does not use several dependent variables.

3.5 Assumptions of Exploratory Factor Analysis

In the current study, Exploratory Factor Analysis (EFA) was used to determine the factor structure of the data and to reduce the 23 variables into fewer latent variables (factors) to enable the researcher(s) to use factor scores in CFA-LDA. Factor analysis is an interdependence technique, whose primary purpose is to define the underlying structure among variables analysed (Hair *et al*., 2014). The purpose of data reduction is to reduce the amount of capacity and transformation is done to remove any irregularities that could arise from the data (Hair *et al*., 2014).

3. 5.1 Sufficient sample size

According to Hair *et al*. (2014), when performing factor analysis, the sample size should be 100 or larger to ensure that correlations are reliably estimated. The sample size used in the current study was 1313 learners, which is in accordance with Hair *et al*. (2014). The general rule is to have not less than five times as many observations as the number of variables to be analysed and an observation to variable ratio of, at least, 10:1 is recommended (Hair *et al*., 2014). The observation to variable ratio of the data used in the current study was 55:1, an indication that the sample size was sufficient.

3.5.2 Multicollinearity

Multicollinearity occurs when some or most of the exploratory variables are highly correlated (Ncube and Moroke, 2016). In the current study, multicolinearity between predictor variables was assessed using the Variance Inflation Factor (VIF) and Tolerances (O'brien, 2007). A certain degree of collinearity between the variables is allowed, since it is important to test the assumption of multicollinearity (Ncube and Moroke, 2016) and they should not be perfect multicollinearity between the variables. According to Miles (2014), VIF and tolerance are two closely related statistics for diagnosing collinearity in factor analysis. These statistics are based on the *R*-squared value obtained by factors of a predictor on all of the other predictors in the analysis. Tolerance is the reciprocal of VIF. The rule of thumb is that if the VIF value lies between one and 10, then, there is no serious multicollinearilty and if the VIF is greater than 10, then there is serious multicollinearity (Vu *et al*., 2015). The VIF is given by the following equation, adopted from Vu *et al*. (2015):

$$\mathrm{VIF} = \frac{1}{1-R_j^2}\ , \qquad (1)$$

where, R_j^2 is the coefficient of determination of the independent variable *j* on all other independent variables in the dataset, the axillary regression.

3.5.3 Multivariate normality

Multivariate analysis is using maximum likelihood estimation (MLE) method widely used in factor analysis (Hair *et al*., 2014). To test for multivariate normality, two tests are performed as follows: Mardia's skewness; and Mardia's kurtosis (Loperfido, 2020). According to Sham and Curtis (1995), the p-values can be estimated to any level of accuracy by taking an appropriate number of large samples to test a distribution's departure from normality. If the test statistic compared against chi square is significant, that is, the *p*-value is smaller than 0.05 (Cain *et al*., 2017), the joint distribution of the set of the variables has a significant skewness. Furthermore, if the test statistic compared against the standard normal distribution is significant, the joint distribution has a significant kurtosis. If at least one of these tests is significant, it is inferred that the underlying joint population is not multivariate normal (Cain *et al*., 2017).

Mardia's measure of Multivariate Skewness (MS) by Ncube and Moroke (2016) was adopted in the study as follows:

$$MS = \frac{1}{6n} \sum_{i,j=1}^{n} \left(Y_i^T \, Y_J\right)^3 , \tag{2}$$

where, *n* is the sample size, *Y* is the matrix of random variable of interest, $Y_i \; or \; Y_j$ is the observed data.

Mardia's measure of Multivariate Kurtosis (MK) by Ncube and Moroke (2016) was adopted as follows:

$$MK = \sqrt{\frac{n}{8p(p=2)}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \parallel Y_i \parallel^4 \frac{p(p+2)(n-1)}{n+1} \right\} , \tag{3}$$

where, *n* is the sample size, *p* is the number of parameters to be estimated and Y_i is the observed data.

If the data is not multivariate normal, the current study will screen variables for outliers and remove any multivariate outliers to achieve normality.

3.5.4 Multivariate outliers

Multivariate outliers are the common reason for violating Multivariate Normailty (MVN) assumption. In other words, MVN assumption requires the absence of multivariate outliers. Thus, it is crucial to check if the data has multivariate outliers, before beginning multivariate analysis. "The distance is a metric, which calculates how far each observation is to the centre of joint distribution, which can be thought of as the centroid in multivariate space", (Filzmoser *et al*., 2000: 158). In the current study, multivariate outliers were tested using a scree plot, which suggests that "outliers are thought to be observations coming from one or more different distributions, and extremes are values that are far away from the centre but which belong to the same distribution" (Filzmoser *et al*., 2005: 581). The multivariate outliers, which were the most extreme points in the study, were removed. The deleted points were the outliers identified, and the multivariate threshold corresponded to the distance of the closest outlier, the farthest background individual, or some intermediate distance (Yong, 2013). The Dsquared was the mahalanobis distance equation in accordanance with Penny (1996) as follows:

$$D^2 = (\bar{\chi}_1 - \bar{\chi}_2)' S_{pooled}^{-1} \, (\bar{\chi}_1 - \bar{\chi}_2)$$
$$= a'(\bar{\chi}_1 - \bar{\chi}_2), \tag{4}$$

where, D is the method of detecting outliers in the multivariate data, n is the observations in a p-variable data set, $\bar{x}$ is the sample mean vector and D^2 is the sample covariance matrix.

3.5.5 Reliability

Cronbach's alpha was used to measure consistency and reliability of data in the the study. Cooper and Emory (1995), as cited in Moroke (2015) state that "due to the multiplicity of variables measuring the clusters, Cronbach's alpha is considered the most suitable since it has the most utility of multi-item scales at the interval level of measurement". According to Moroke (2015), a commonly accepted rule of thumb for describing internal consistency, using Cronbach's alpha, is as follows: $\alpha \geq 0.9$ is excellent, $0.8 \leq \alpha < 0.9$ is good, $0.7 \leq \alpha < 0.8$ is acceptable, $0.6 \leq \alpha < 0.7$ is questionable, $0.5 \leq \alpha < 0.6$ is poor and $\alpha < 0.5$ is unacceptable. The Cronbach's αlpha is described by the following equation:

$$\alpha = \frac{kr}{1+(k+1)}, \qquad (5)$$

where, *k* is the average correlation between the variables and *r* is the number of variables (Bonett and Wright, 2015).

3.5.6 Sampling adequacy

MacCallum *et al*. (1999) state that an adequate sample size is partly determined by the nature of the data. According to Moroke (2015), the measurement used to check the adequacy of a sample is the Kaiser-Meyer-Olkin (KMO). The authors further argue that the small values of the *KMO* are a perfect indication that there is a relationship between sets of variables, which cannot be explained by other variables. According to Field (2005) and Moroke (2015), the rule of thumb is that the value of *KMO* closer to 1, is an indication that patterns of correlations are relatively compact and so factor analysis should yield distinct and reliable factors. Moroke (2015) suggests that measures in the ranges of 0.8 or above, are excellent, 0.7 are moderate, 0.6 are mediocre, 0.5 are miserable and below 0.5, are unacceptable. The following formula is used to compute the *KMO* in accordance with Moroke (2015):

$$KMO = \frac{\Sigma\Sigma\, r_{ij}^2}{\Sigma\Sigma\, r_{ij}^2 + \Sigma\Sigma\, a_{ij}^2}, \qquad (6)$$

where, r_{ij} is simple correlation between i-th and j-th variable and a_{ij} is partial correlation between i-th and j-th variable.

3.5.7 Factorability of the correlation matrix

A correlation matrix is not invertible if its determinant is equal to zero (Tabachink and Fidell, 2001) and in sucht case, it is not factorable. The Bartlett's test of sphericity is used to assess the factorability of the correlation matrix. It checks if there is redundancy between variables that can be summarised with some factors (Tabachink and Fidell, 2001). The Bartlett's test for sphericity, tests that the correlation matrix is similar to an identity matrix, which is an indication that each variable correlates only with itself (Sarmento & Costa, 2019). The test statistics can be used to test the following hypotheses:

Ho: Variances are equal for all samples.

Ha: Variance are not equal for one pair or more.

$$\chi^2 = -\left(n-1-\frac{2p+5}{6}\right) \ln |R| \quad , \qquad (7)$$

where, *n* is the sample size, *p* is the number of variables and *R* is the correlation matrix of variables (Sarmento & Costa, 2019).

3.5.8 Communalities

"Factor analysis uses variances to produce communalities between variables" (Yong and Pearce, 2013: 82). The variance is equal to the square of the factor loadings (Child, 2006). Communalities use the squared multiple correlation between each variable and the remaining variables. "Overall, the factor loadings are fairly similar, and one needs to perform rotation regardless of the extraction technique" (Tabachnick and Fidell, 2007). Principal Axis Factor (PAF) was used in Table 7 to produce factors.

3.5.9 Determination of the number of factors to extract

The next step is to determine the number of factors retained, where factors that have an Eigen value equal or larger than 1, are selected as common factors (Suhr, 2006). Factors are rotated for better interpretation since unrotated factors are ambiguous. Orthogonal rotation is when the factors are rotated 90^0 from each other and was applied in this study. The orthogonal technique is a varimax rotation, which minimises the number of variables that have high loadings on each factor and works to make small loadings even smaller. Varimax rotation tends to produce multiple group factors, maintaining orthogonality, thus resulting in increased multicolinearity (loadings of variables on "primary factors" is decreased a bit and loadings on “secondary factors” is raised a bit).

3.5.10 Exploratory Factor Analysis (EFA):

After testing all the assumption and ensuring that they are met, the next step is factor extraction. Factor extraction is a method to reduce data from some variables into fewer factors, in order to describe correlation between observed variables (Yong and Pearce, 2013). Factor extraction in the current study was performed using EFA. The Principal Axis Factor Analysis equation is expressed as follows:

$$L = V'RV, \qquad (8)$$

where, R diagonalisation is accomplished by post-and pre-multiplying it by V matrix and its transpose. V is the eigenvectors and L are eigenvalues (Tabachink and Fidell, 2001).

The next step is computation of factor score coefficients, given as:

$$B = R^{-1}A, \qquad (9)$$

where, R^{-1} is the inverse of the matrix of correlation matrix of correlations among variables and A is the matrix of correlations between factors and variables (Tabachink and Fidell, 2001).

3.5.11 Interpretation of the factor matrix

Factor loadings interpret loadings to determine the strength of the relationship. “Factor Matrix shows the factor loadings prior to rotation whereas the rotated factor matrix shows the rotated factor loadings” (Yong and Pearce, 2013:90) using rotation small suppressing small coefficients to help with the interpretation (Yong and Pearce, 2013). Factors can be identified by the largest loadings, however, it is also important to examine the zero and low loadings in order to confirm identification of factors. The varimax rotated solution provides the simplest interpretation of the structure. Factor loadings with a value ≥ |0.3| may be considered in order to continue with factor analysis, specifically when the sample size is 100 or larger (Hair *et al*., 2010). The factor matrix interpretation contains the final communalities. “If a large number of variables are analysed, verbal summarisies of these results can be used” (Pohlmann, 2004: 13). According MacCallum *et al*. (1999: 274) “larger sample size, higher communalities (low communalities are associated with sampling error due to the presence of unique factors that have nonzero correlations with one another and with the common factors), and high over determination [each factor having at least three or four high loadings and simple structure (few, nonoverlapping factors)], each increase chances of faithfully reproducing the population factor pattern”.

3.5.12 Factor rotation

After factor extraction, factor rotation is performed in attempt to achieve a simple structure. Rotation is done by rotating the axis of the factor, from its centre point to the point to be addressed (Beauducel and Kersting, 2019). Orthogonal rotation was used in this study since the factors were independent. Rotation was done with Qrthogonal equamax rotation, which considers correlated factors. According to Browne (2001:115), “orthogonal and oblique rotation involve the same problem of minimising a complexity criterion, and only the constraints imposed differ”. Previous researches (Treiblmaier and Filzmoser, 2010; and Beauducel and Kersting, 2019) state that in an orthogonal method, when factors are actually correlated, an orthogonal rotation produces an unrealistic solution, while an oblique rotation better reproduces reality. Treiblmaier and Filzmoser (2010) further argue that an oblique rotation does not require factors to be correlated thus, the correlations between the factors will be close to zero if the actual data structure is orthogonal. The stability of results of factors depends on the large samples. When all the steps of EFA are ran, Confirmatory Factor Analysis (CFA) is performed to confirm the results of EFA.

3.6 Confirmatory Factor Analysis (CFA)

CFA was used to verify results of exploratory factor analysis in the study. CFA is a way of testing how well the measured variables represent a smaller number of constructs (Hair *et al*., 2014). Running CFA is to validate the factorial validity of models derived from the results of EFA. CFA is computed using Structural Equation Modelling (SEM). SEM is a hypothesised model that has dependent types of relationships, where each hypothesis represents a specific relationship that must be specified. SEM is achieved through computing the regression coefficients for each path in the structural model as follows:

$$\beta_{y*x} = B_S * \rho_x, \qquad (10)$$

where, β_{y*x} is the observed regression coefficient, B_S is the true structure coefficient and ρ_x is the reliability of the predictor variables (Hair *et al*., 2010). Jackson *et al*. (2009) suggest that SEM model modification should be carried out only when they are theoretically parsimonious.

3.6.1 Goodness of fit of the Structual Equation Modeling (SEM) model

This step aims to avoid under or over-fitting the hypothesised structural model (Hair *et al*., 2010). If need be, the researcher may, either truncate or re-specify the model, delete cases or variables to improve the fit of the SEM model. Measures of overall model fit indicate to which extent a structural equation model corresponds to the empirical data (Hair *et al*., 2010). The basic idea of comparison indices is that the fit of a model of interest is compared to the fit of some baseline model (Hair *et al*., 2010). The indices discussed below are descriptive measures of overall model fit adopted in this study.

3.6.1.1 Standardised Root Mean Square Residual (SRMR)

RMR and SRMR are essentially used to measure how well a model does not fit, since they are based on residuals of the fitted model (Hair *et al*., 2010). SRMR, as proposed by Jöreskog and Sörbom (1993), is an overall badness of fit measure based on fitted residuals. SRMR is an index of the average of standardised residuals between the observed and hypothesised covariance matrices (Chen, 2007). SRMR ranges from 0 to 1 and a good model fit indicated by values under 0.05 or close to zero suggests a good fit (Hair *et al*., 2010) while values smaller than 0.10 can be interpreted as acceptable (Hair *et al*., (2010). Concretely, RMR is defined as the square root of the mean of squared fitted residuals. The equation for SRMR proposed by Cangur and Ercan (2015) is as follows:

$$\text{SRMR} = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}[(s_{ij}-\hat{\sigma}_{ij})/(s_{ii}s_{jj})]^{2}}{p(p+1)/2}} \quad (11)$$

where, s_{ij} indicates a component of *S* sample covariance matrix, $\hat{\sigma}_{ij}$ shows a component of $\sum(\hat{\theta})$hypothesised model while p is the number of observed variables.

3.6.1.2 Adjusted Goodness of Fit Index (AGFI)

"AGIF implies testing how much better the model fits compared to "no model at all" (null model), that is, when all parameters are fixed to zero" (Joreskog and Sorbom, 1993: 123). AGFI adjusts for the model's degrees of freedom relative to the number of observed variables and therefore, rewards less complex models with fewer parameters (Schermelleh-Engel, 2003). AGFI value greater than 0.9 means a satisfactory model fit (Awang, 2012). AGFI was developed by Jöreskog and Sörbom (1993) to adjust bias in GFI, resulting from model complexity. Sarmento and Costa (2019) define the model as follows:

$$\text{AGFI} = 1 - \frac{\chi_t^2/df_t}{\chi_n^2/df_n}, \qquad (12)$$

where, χ_n^2 is the chi square of null model (baseline model), χ_t^2 is the chi square of the target model, df_n is the number of degrees of freedom for the null model and df_t is the number of degrees of freedom for the target model.

3.6.1.3 Root Mean Square Error of Approximation (RMSEA)

RMSEA is related to the residual in the model (Hair *et al*., 2010). RMSEA ranges from 0 to 1 with a smaller RMSEA value indicating better model fit (Hair *et al*., 2010). According to Sarmento and Costa (2019), RMSEA value ≤ 0.05 can be considered a good fit, values between 0.05 and 0.08 as adequate fit, values between 0.08 and 0.10 as mediocre fit while values >0.10 are not acceptable. The following is the RMSEA formula as proposed by Sarmento and Costa (2019):

$$\text{RMSEA} = \sqrt{\frac{1}{n-1}\left(\frac{\chi_m^2 - df_m}{df_m}\right)} \qquad (13)$$

where, χ^2 is Chi square test statistic, *df* is the number of degrees of freedom and *n* is the sample size.

3.6.1.4 Comparative Fit Index (CFI)

CFI is an incremental fit index (Hair *et al*., 2010) and a corrected version of relative non-centrality index (Hooper *et al*., 2008). The extent to which the tested model is superior to the alternative model established with manifest covariance matrix can be evaluated (Chen, 2007). CFI ranges from 0 to 1 with a larger value indicating better model fit. Acceptable model fit is indicated by a

CFI value of 0.90 or greater (Cangur and Ercan, 2015). CFI was developed by Cangur and Ercan (2015) and the equation given as follows:

$$\mathrm{CFI} = 1 - \frac{\max\left[\left(\chi_t^2 - v_t\right), 0\right]}{\max\left[\left(\chi_t^2 - v_t\right), \left(\chi_i^2 - v_i\right), 0\right]} \qquad (14)$$

where, max indicates the maximum value of the values given in brackets, v_t and v_i are the degrees of freedom of the independence model, χ_i^2 and χ_t^2 are test statistics of the independence model and the target model respectively.

3.6.1.5 Normed Fit Index (NFI)

NFI values range from 0 to 1, with higher values indicating better fit (Hair *et al*. (2010). NFI is also referred to as Bentler-Bonett Normed Fit Index (Hooper *et al*., 2008). It is used to analyse the discrepancy between the chi-squared value of the proposed model and the chi-squared value of the null model (Hooper *et al*., 2008). It is considered very good if it is equal to or greater than 0.95, good between 0.9 and 0.95, suffering between 0.8 and 0.9 and bad if it is less than 0.8 (Portela, 2012). The NFI proposed by Ding *et al*. (1995) is defined as:

$$\mathrm{NFI} = \frac{\chi_b^2 - \chi_m^2}{\chi_b^2}, \qquad (15)$$

where, χ_b^2 is the chi- square of the independence model (baseline model) and; χ_m^2 is the chi-square of the target model.

3.7 Assumptions of Stepwise Discriminant Analysis (SDA)

Discriminant analysis is the appropriate statistical technique, when the dependent variable is categorical and the dependent variables are metric (Hair *et al*., 2014). According to Tabachnick and Fidell (2007), a lot of data used in discriminant analysis is first evaluated with the following assumptions:

3.7.1 Multicollinearity

Multicollinearity among the independent variables can markedly reduce the estimated impact of independent variables in the derived discriminant functions, particularly in a stepwise estimation process (Osborne, 2014). Ssee Section 3.5.2 for more information on the same assumptions as EFA).

3.7.2 Multivariate normality

There should be multivariate normality of independent variables. See Section 3.5.3 for additional information on this assumption in EFA.

3.7.3 Absence of outliers

Outliers can have a substantial impact on the classification accuracy of results of any discriminant analysis (Razzak, 2020). To test the stepwise method, Mahalanobis D^2 is considered appropriate and is based on generalised squared Euclidean distance that adjusts for unequal variances (Hair *et al*., 2010). Mahalanobis distance is a statistic used to confirm if there is a possibility of separating the groups and makes use of maximal information available in a stepwise process. See Section 3.5.4 and the equation; they have the same assumptions as EFA.

3.7.4 Stepwise Discriminate Analysis (SDA)

SDAs are equations for testing the significance of a set of discriminant functions. Variance is the set of predictors classified into two sources: variance attributed to differences between groups; and variance attributed to differences within groups. The following equation is the discriminant analysis group in accordance with Tabachnick and Fidell, (2007):

$$S_{total} = S_{bg} + S_{wg} \quad (16)$$

where, S_{total} is the total cross-product matrix, which is partitioned into a cross- product matrix associated with S_{bg}, the difference between groups and S_{wg}, a cross-product matrix of differences within groups.

The next step is the classification equation (j= 1, 2, ……. k) (Tabachnick & Fidell, 2007):

$$C_j = C_{j0} + C_{j1}X_1 + C_{j2}X_2 + \cdots + C_{jP}X_P \ , \quad (17)$$

where, j (C_j) is the score on the classification function for the group; that is, for each X, the predictor is multiplied by the raw score by its associated C_j classification function coefficient summing over all predictors and adding a C_{j0} constant.

After computing the previous step, the last step is to calculate the classification coefficient, calculated using the following equation by Tabachnick and Fidell (2007):

$$C_{j0} = W^{-1}M_j \ , \qquad (18)$$

where, the column matrix of classification coefficients for group j ($C_j| = C_{j1}, C_{j2}, \ldots, C_{jP}$) is found by multiplying the Inverse of W^{-1} , the within- group variance-covariance matrix by a column matrix of means for the group j on the p variables ($M_j = X_{j1}, X_{j2} \ldots, X_{jp}$).

3.7.5 Maximisation of determinant

In E-step, the estimated posterior probability of each data point belongs to a gaussian component (McLahlan, 2004). The calculated rates of correct classification are based on the estimated posterior probabilities for the individual observation discriminant analysis belonging to the branch of classification methods referred to as generative modelling, where one tries to estimate the within-class density of X, given the class label. Combined with the prior probability (unconditioned probability) of classes, the posterior probability of Y can be obtained by the Bayes formula (McLahlan, 2004). This can also be thought of as soft count since one data point can belong to multiple clusters. The equation is given by McLahlan (2004) as:

$$\hat{\Sigma} = \sum_{k=1}^{k} \sum_{gi=k} \left(x^{(i)} - \hat{\mu}_k\right) \left(x^{(i)} - \hat{\mu}_k\right)^{\mathsf{T}} / (N - K) \qquad (19)$$

where, K is within number of classes, $\hat{\mu}_k$ is the number of class k samples, N is the total number points in the data, x is the column vector and N-K unbiased maximum likelihood estimator.

3.7.6 Iteration

Iteration refers to the "continuation of the E-step with the newly estimated parameters and M-step, so on and so forth, until the results converge where the log likelihood starts to flatten out / not increasing significantly anymore" (Yuan *et al*., 2014: 99).

3.8 CFA-LDA

This CFA-LDA algorithm is a combination of the confirmation results (CFA) of EFA and LDA. Firstly, the Principal Axis is used to factor the equamax model to determine the factor scores and save them. Secondly, the created 3 factor scores from the enter method of LDA, which is the enter method (enter all the independents together), is applied instead of SDA. Factors 1, 2 and 3, are retrieved from the factor score used as independent variables for the LDA enter method. The dependent variable used is child observation. The following discriminant model equation was adopted from Tabachnick and Fidell (2007):

$$D_{zi} = d_{i1Z1} + d_{i2Z2} + \cdots + d_{ipZp} \quad (20)$$

where, $D_i$standardised score on the ith discriminant function, which is found by multiplying the standardised score on each predictor (z) by its standardised discriminate function coefficient (d_i) and then adding all predictors.

Box's M is tested and if group sizes are over 30, then MANOVA is robust against violations of homogeneity of variance-covariance matrices assumption. The test has also been criticised for being overly sensitive for large sample sizes.

3.9 SDA

Section 3.4 shows that one dependent variable is created, referred to as child observation and, in SDA, is used as the dependent variable. From Table 1, all the 27 variables are used as independent variables. According to Tabachnick and Fidell (2007), the Wilks' Lambda criteria is one of the minimising stepping methods. This Wilks' Lambda statistic ranges between zero and one, with a value closer to zero (0.000 and 0.004 in the current study), denoting a high level of discriminating power (Davis-Sramek *et al*., 2020). From model 1, which is CFA-LDA and model 2 SD, both model comparisons are checked to select the best model, in accordance with the rule of thumbs.

3.10 Model comparison and selection

The model comparison phase entails comparing the proposed models within a specific sample size and selecting the most efficient model based on the Chi-square test for independence, Apparent Error Rate (APER), Hit Ratio and Mann-Whitney U test as recommended by Hilbe (2014). Parsimony is considered to be important in assessing model fit (Marsh and Hau, 1996) and serves as a criterion for choosing between alternative models.

3.10.1 Chi-square test of association

"The Chi-square test of association is one of the most useful statistics for testing hypotheses when the variables are nominal" (McHugh, 2013:143); the chi-square of association test incorporates both tests (Bolboaca *et al*., 2011). The chi-square test of independence is used to determine whether there is an association or relationship between two or more categorical (i.e., nominal or ordinal) variables (Hahs-Vaughn and Lomax, 2013). "The p-value is a probability that measures the evidence against the null hypothesis; lower probabilities provide stronger evidence against the null hypothesis. Furthermore, the p-value (< 0.05) is used to determine whether to reject or accept the null hypothesis, which states that no independence between two categorical variables exist" (McHugh, 2013: 146).

The hypotheses and formula is given by Marathe and Ryan (2005) as follows:

H_o; there is no difference in association between observed and predicted variables
H_1: there is a difference in association between observed and predicted variables.

$$\chi^2 = \sum \frac{(F_0 - F_e)^2}{F_e}, \tag{21}$$

where, the square of the differences between the F_0 is the observed value and F_e is the expected value in each cell divided by the expected values (F_e), are added to all the cells in the Table.

The chi-square test is used to test the association between the child observation and the predicted child observation variables from the CFA-LDA and SDA models. The p-value is used to determine whether there is association or not between the predicted and observed variables thus, the model is good with a p-value < 0.05

3.10.2 Apparent Error Rate (APER)

APER is an optimistic estimate of Actual Error Rate (AER) (Lawrence *et al*., 2007). APER is easily calculated from the confusion matrix or classification table. Cross-validation is a good technique to test a model on its predictive performance. The confusion matrix shows the actual versus predicted group membership (Lawrence *et al*., 2008). The confusion matrix is represented in Table 2 and in accordance with Helwig (2017).

Table 3.2: Confusion matrix

Actual membership	π_1	π_2	Group sample size
π_1	n_{c1}	n_{m1}	n_1
π_2	n_{m2}	n_{c2}	n_2

An almost unbiased estimate of the expected AER is calculated as follows:

$$\hat{E}(\text{AER}) = \frac{n^*_{m1} + n^*_{m2}}{n_1 + n_2} \times 100\% \qquad (22)$$

$$= \frac{misclassified}{groups\ samples\ sizes} \times 100\%$$

which is the total proportion of misclassified sample observations (Helwig, 2017);

where, n^*_{m1} and : n^*_{m2} are the number of misclassified observations using the above "leave-one-out" procedure. The p-value for a partial F-test is used to determine the significance of the variables added to the classification model. The higher the p-value, the better model.

3.10.3 Hit Ratio or APCR

For the Hit ratio, the correct classification from the confusion matrix is used in the equation and the same group samples from the confusion matrix are applied to get the following equation for APCR or Hit ratio (Helwig, 2017):

$$\text{APCR or Hit Ratio} = \frac{n_{c1} + n_{m1}}{n_1 + n_2} \times 100\%$$

$$= \frac{Correctly\ classified}{groups\ samples\ sizes} \times 100\%, \qquad (23)$$

where $n_{c1} + n_{m1}$ is the number of misclassified observation and $n_1 + n_2$ is the number of misclassified observations. The higher the percentages, the better and efficient the model.

3.10.4 Mann- Whitney U test

This test is a nonparametric alternative to the traditional two-sample t-test (Pallmann, 2016). A popular nonparametric test to compare outcomes between two independent groups is the Mann Whitney U test (Pallmann, 2016). The Mann Whitney U test is used to test whether two samples are likely to derive from the same population (i.e., that the two populations have the same shape)

(Shier, 2004). Some investigators interpret this test as comparing the medians between the two populations (Shier, 2004). The formula in accordance with Pallmann (2016) is defined as follows:

$$U_1 = R_1 - \frac{n_1\,(n_1+1)}{2} - R_1, \quad (24)$$

where, R_1 is the sum of ranks in the sample size and N is the number of items in the sample.
Mann-Whitney U test is used to compare the predicted group membership from the two models (CFA-LDA and SDA) to the observed group membership and to determine whether ($p< 0.05$) predicted group memberships are the same as the observed ones.

3.11 Summary of chapter

The aim of this chapter was to outline and discuss important methodological steps used in this study to compare two models, CFA-LDA and SDA, in analysing multivariate data and discuss tests used in the next chapter. This chapter has also provided a discussion on the dofferent variables used in the study and statistical techniques considered in the study. Furthermore, the two-step analysis described to show how multivariate data can be grouped into appropriate characteristics in the preliminary stage of the analysis. Results of EFA were confirmed using CFA. SDA was used to determine the cost contributor between groups. In addition, a comparison of FA-LDA and SDA model selections criteria were used to select the best model. The results of the techniques provided in this chapter were used to produce and describe the patterns explained in Chapter Four. The two-step cluster analysis, exploratory factor analysis, confirmatory factor analysis, stepwise discriminant analysis, with two methods of comparison (CFA-LDA and SDA) were considered and applied in Chapter Four. The results are reported in Chapter Four in accordance with the objectives stipulated in Chapter One. Thus, the best model will contribute towards narrowing the scope of the study and and filling the gap in the literature. The next chapter is the presentation and interpretation of results.

CHAPTER FOUR

ANALYSIS, PRESENTATION AND AND INTERPRETATION OF RESULTS

4.1 Introduction

The chapter focuses on the presentation and interpretation of the results in accordance with the aim and objectives of the study. The analysis is based on the data analysis steps summarised in Figure 4.1. Section 4.2 focuses on the application of the two-step analysis to create a dependent variable; EFA was tested on 27 independent variables in Section 4.3 and confirmatory factor analysis performed to confirm results of EFA in section 4.4. CFA-LDA and SDA model selection criteria were used to select the best model between the two models. The interpretation is based on best model selected and the underlying framework presented in Figure 1.1 (Chapter One).

4.2. Two-Step Cluster Analysis

The two step cluster analysis is performed using the child observation variables (paid attention, stayed concentrated, careful and diligent and; interested and curious) to create a unified variable to represent “child observation” from the four variables. This section discusses the results of the Two-Step cluster analysis.

Model Summary

Algorithm	TwoStep
Inputs	4
Clusters	3

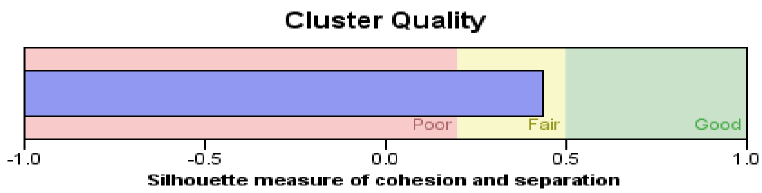


Figure 4.1: Summary of model

Figure 4.1 shows that four variables were clustered and ended up with three clusters. The quality of the cluster shows that three sets of clusters were fair and appropriate to be used for further analysis.

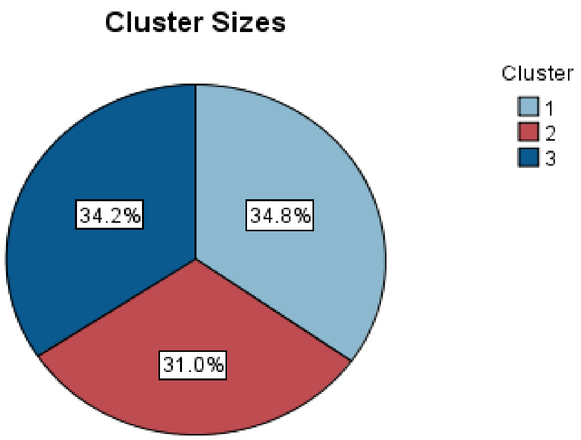


Figure 4.2: Sizes and distribution of clusters

Figure 4.2 shows the size and distribution of clusters within each of the three clusters. The output shows that Cluster 1 comprised 34.8% of learners, 31.0% of learners were allocated to the second cluster while 34.2% were in the third cluster.

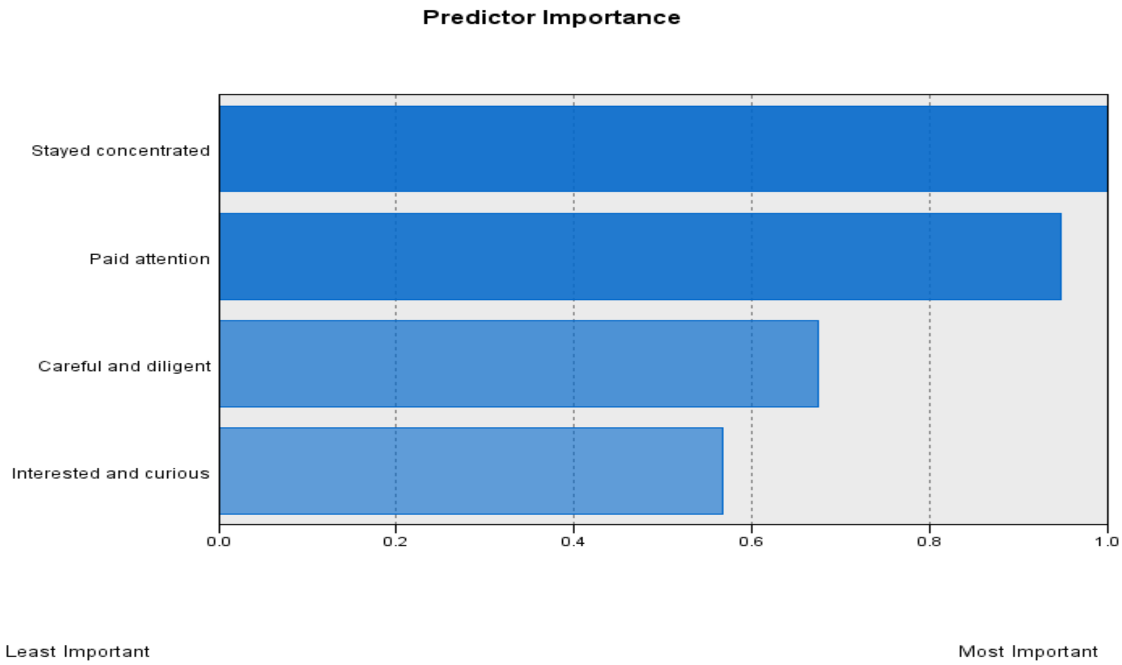


Figure 4.3: Predictor Importance

Figure 4.3 shows the predictor importance of each variable for each continues variable. In Figure 4.3, the importance of all the variables is greater than 0.5 (Tighe and Schatschneider, 2014), thus all variables are important in distinguishing the observation between clusters.

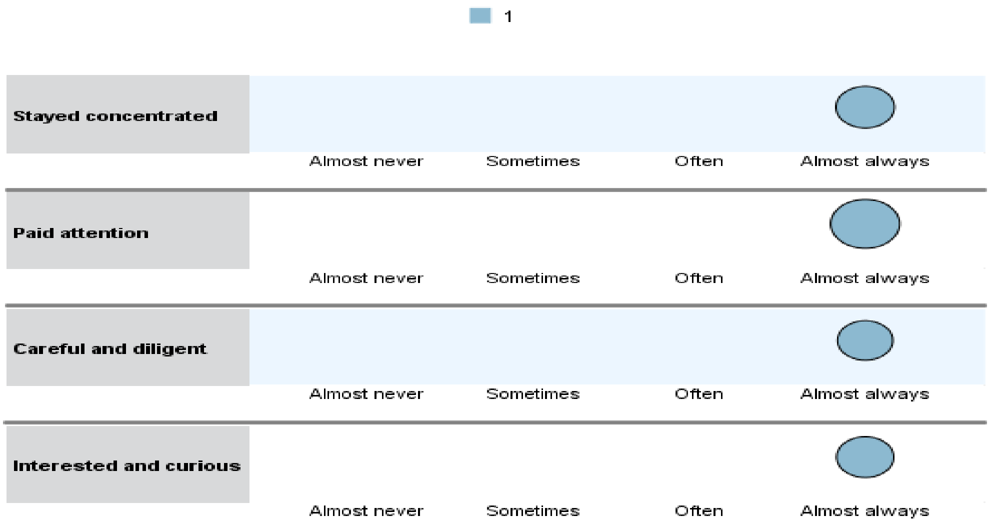


Figure 4.4: Description of Cluster 1

Figure 4.4 shows that learners in cluster 1 almost always stayed concentrated, paid attention, were careful and diligent as well as interested and curious in the learning outcome.

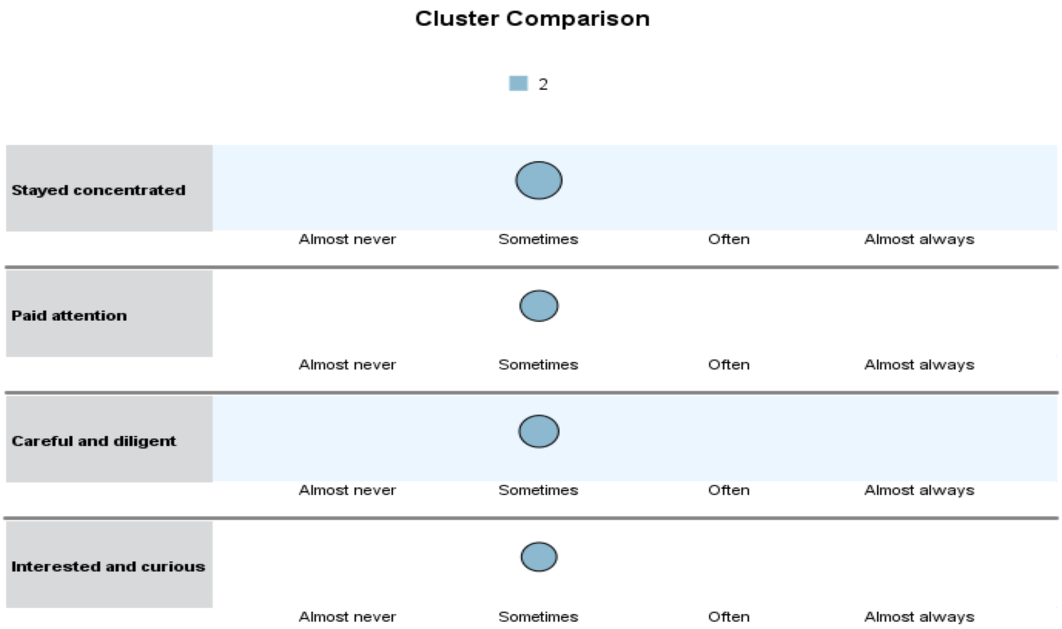


Figure 4.5: Description of Cluster 2

In Figure 4.5, learners in cluster 2 sometimes stayed concentrated, paid attention, were careful and diligent as well as interested and curious in the learning outcome.

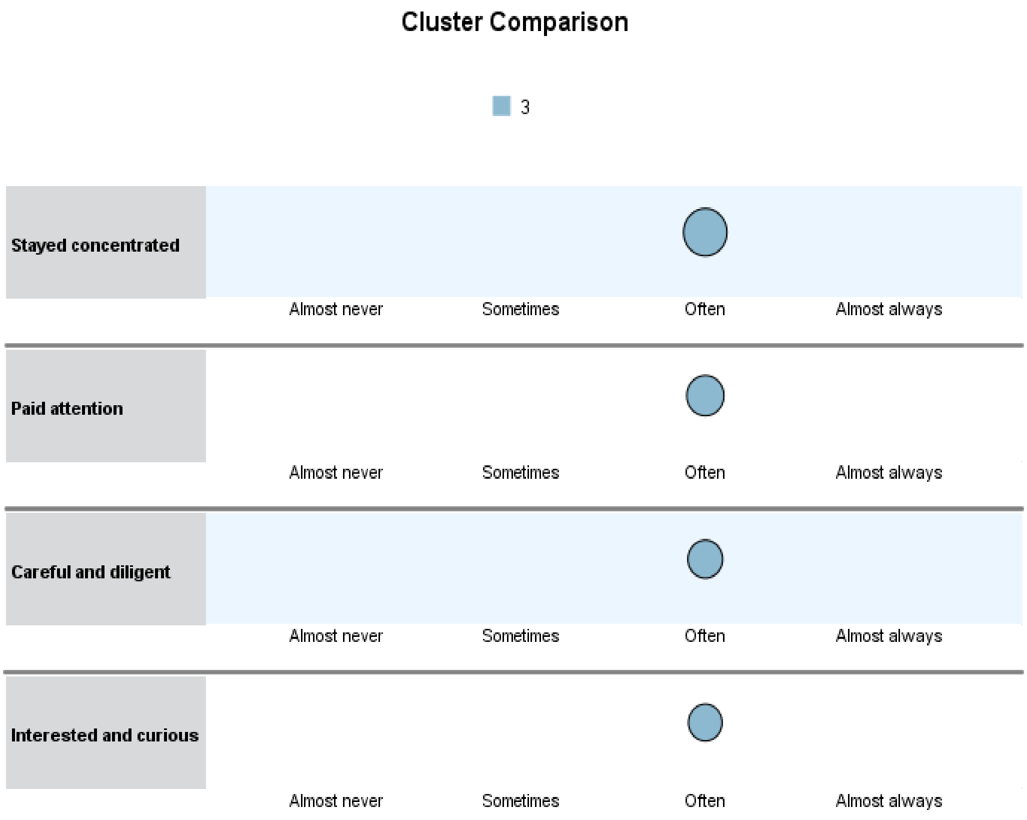


Figure 4.6: Description of Cluster 3

Figure 7 shows that learners in cluster 3 **often** stayed concentrated, paid attention, were careful and diligent as well interested and curious in the learning outcome. The next section focuses on EFA.

4.3 Results of Exploratory Factor Analysis

Table 4.1: Multicollinearity test

Model	Collinearity statistics	
	Tolerance	**VIF**
Standing on one leg for 10 seconds transformed	.927	1.079
Catch bean bag with both hands transformed	.837	1.195
Catch bean bag with preferred hand transformed	.767	1.303
Catch bean bag with non-preferred hand transformed	.797	1.255
Cross and square transformed	.852	1.173
Triangle transformed	.780	1.282
Draw person transformed	.789	1.268
String beads transformed	.816	1.225
Counting transformed	.662	1.510
Add and sub transformed	.734	1.362
Sort and class transformed	.875	1.143
Spatial vocabulary transformed	.927	1.079
Measurement vocabulary transformed	.883	1.133
DCCS transformed	.808	1.238
Pencil tapping transformed	.650	1.539
Digit span transformed	.888	1.126
Picture puzzle transformed	.708	1.412
Empathy transformed	.765	1.307
Self-awareness transformed	.665	1.505
Expressive language transformed	.616	1.624
Expressive vocabulary transformed	.693	1.443
Oral comprehension transformed	.765	1.306

Sound discrimination transformed	.867	1.154

VIF values for all the 23 variables values range between 1.079 and1.624, an indication that the VIF values obtained were less than 10, implying that 23 variables had no serious multicollinearity, which can affect EFA.

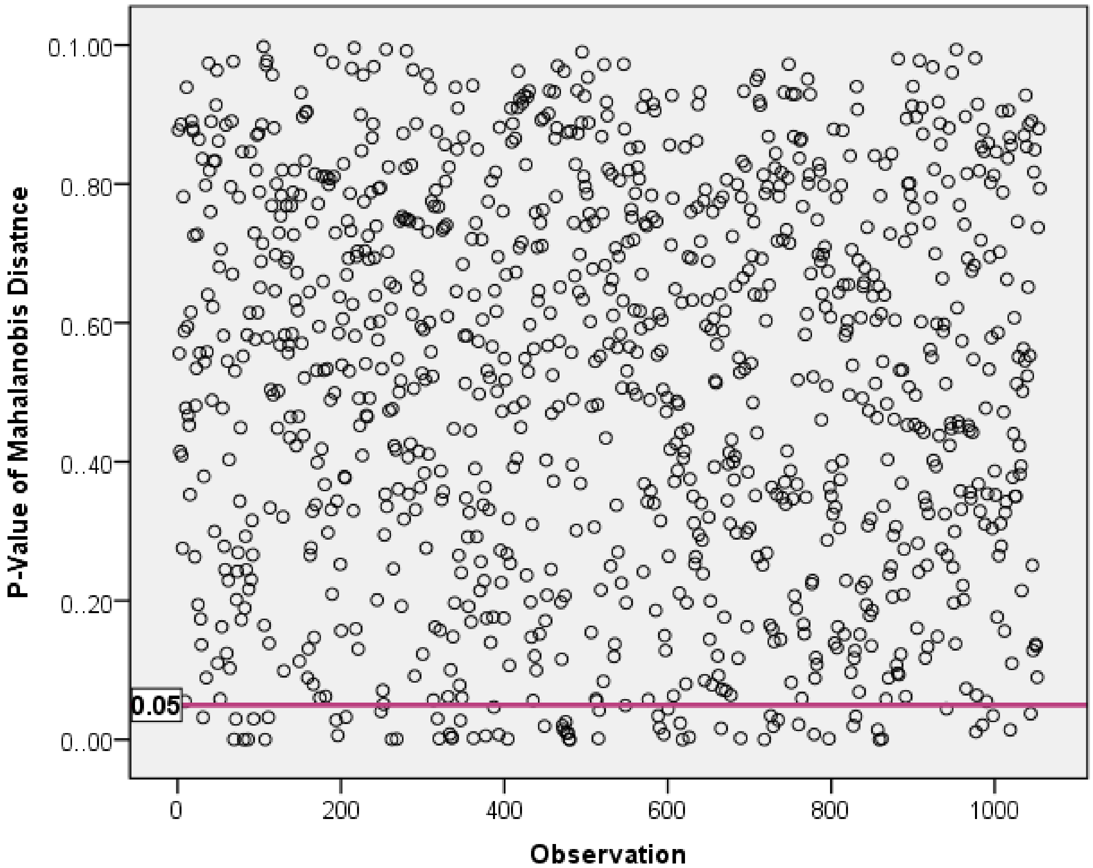


Figure 4.7: Mahalanobis distance of P-Values vs observation number

In the Mahalanobis Distances plot shown in Figure 4.7, some of the p-values of the Mahalanobis distances are under the 0.05 significant level, an indication that they are extreme multivariate outliers. These outliers lead to unrealistic estimations of the correlation and bais of factor analysis. These most extreme points are deleted from the data prior to running EFA.

Table 4.2: Multivariate Normality

Equation	Test statistic	Value	Prob
System	Mardia skewness	4909	<.0001
	Mardia Kurtosis	-5.00	<.0001

Results of the Mardia's test revealed that the p-value was 0.001 for both Mardia skewness and Mardia kurtosis, with p-value<0.05, thus, it is concluded that the joint distribution of variables has a significant skewness and kurtios, an indication that the data was not normally distributed. When the data is large (in this study, there wer 1313 cases as opposed to the recommended minimum of 100 cases), EFA is robust to the violation of this assumption of multivariate normality (Zygmont and Smith, 2014). In addition, if the assumption of multivariate normality is "severely violated", it is recommended that "principal axis factors (PAF)" be used (Castello and Osbourne, 2005). Thus, to address the violation of the multivariate normality assumption, PAF was used instead of EFA.

Table 4.3: Reliability test

Cronbach's Alpha Based on Standardised Items	N of items
.798	23

In Table 4.3, Cronbach's alpha value of 0.798 implies good reliability as described by Cronbach and Shavelson (2004). This is a strong indication of internal consistency among the 23 variables.

Table 4.4: Sampling adequacy and factorability of the correlation matrix

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.844
Bartlett's Test of Sphericity	Approx. Chi-Square	3711.900
	df	253
	Sig.	.000

The KMO (p-value 0.884) in Table 4.4 (Moroke, 2015) suggests that the sample is adequate for EFA and the factors extracted will account for a significant amount of variance. Field (2003) recommends a sample adequacy measure of a value closer to 1, which indicates that the patterns of correlations are relatively compact and so values in excess of 0.80 are meritorious. This indicates sufficient items for each factor.

The Bartlett's test of Sphericity is another indication of the strength of relationship among variables. Since the p-value of the test is less than the significance level of 0.05, it is an indication that factors can be used in factor analysis. The conclusion is that the correlation matrix is significantly different from an identity matrix, in which correlation between variables are all zero. In accordance with Munro's (2005) criterion, it is concluded that, at least, one common factor can be present.

Table 4.5: Communalities for excluded variables

	Extraction
Standing on one leg for 10 seconds transformed	.185
Sort and class transformed	.185
Sound discrimination transformed	.183

The communalities for every value should be higher than 0.3. The communalities represent the relation between the variables and all other variables. Table 4.5 shows three variables extracted, which are "often times variables with low communalities (less than 0.30), eliminated from the analysis since the aim of factor analysis is to try and explain the variance through common factors" (Yong and Pearce, 2013: 83).

Table 4.6: Communalities

	Extraction
Catch bean bag with both hands transformed	.484
Catch bean bag with preferred hand transformed	.633
Catch bean bag with non-preferred hand transformed	.551
Cross and square transformed	.416
Triangle transformed	.481
Draw person transformed	.397
String beads transformed	.335
Counting transformed	.515
Add and sub transformed	.362
Spatial vocabulary transformed	.504
Measurement vocabulary transformed	.531
DCCS transformed	.443
Pencil tapping transformed	.515
Digit span transformed	.334
Picture puzzle transformed	.564
Empathy transformed	.535
Self-awareness transformed	.584
Expressive language transformed	.624
Expressive vocabulary transformed	.556
Oral comprehension transformed	.407

Communalities indicate the common variance shared by factors with given variables and 20 variables are better measurement / correlation of factor analysis communalities. Therefore, 0.3 or greater communalities in accordance with Filieri (2010), is an indication that a larger amount of variance in the variable has been extracted by the factor solution.

Table 4.7: Determination of number of factors

	Eigenvalue	Difference	Proportion	Cumulative
1	3.48	2.43	0.60	0.64
2	1.04	0.130	0.19	0.83
3	1.00	0.40	0.17	1.00

"One criterion that can be used to determine the number of factors to retain is Kaiser's criterion, which is a rule of thumb" (Yong and Pearce, 2013: 85). This criterion suggests retaining all factors with eigenvalues that are above 1 hence, three factors were extracted in this study. In Table 4.7, a proportion of 64% of variance in the three factors is explained by factor 1.

Table 4.8: Rotated Component Factor Matrix

Rotated Factor Pattern				
		Factor1	**Factor2**	**Factor3**
Item 5	Cross and square transformed	0.371		
Item 6	Triangle transformed	0.482		
Item 7	Draw person transformed	0.450		
Item 8	String beads transformed	0.311		
Item 9	Counting transformed	0.593		
Item 10	Add & sub transformed	0.472		
Item 15	Pencil tapping transformed	0.613		
Item 17	Picture puzzle transformed	0.427		
Item 18	Empathy transformed		0.476	
Item 19	Self-awareness transformed		0.651	
Item 20	Expressive language transformed		0.735	
Item 21	Expressive vocabulary transformed		0.582	

Item 22	Oral comprehension transformed		0.404	
Item 2	Catch bean bag both hands transformed			0.488
Item 3	Catch bean bag preferred hand transformed			0.708
Item 4	Catch bean bag non-preferred hand transformed			0.555

The idea of rotation is to reduce the number of factors on which the variables under investigation have high loadings, in accordance with Tabachnick and Fidell (2007). Table 4.8 shows the loadings of the 16 variables on the three factors extracted. The higher the absolute value of the loading, the more the variable contributes to the factor (Hair *et a*l., 2010). The rotated factor pattern shows that all the factor loadings were ≥|0.3| (Hair *et al*., 2010), however, three more variables were excluded from the factor structure due to insignificant loadings, bringing the number of variables to 17. After rotation, eight variables loaded onto Factor 1, five variables loaded onto Factor 2 and three variables loaded onto Factor 3. The criterion of three significant loadings per factor is satisfied (Montshiwa and Moroke, 2014). The results for CFA are presented next.

4.4 Results of Confirmatory Factor Analysis (CFA)

Table 4.9: Path Coefficients

Variable	Predictor	Estimate	Standard Error	t Value
Item 5	**Factor 1**	0.698	0.021	33.618
Item 6	**Factor 1**	0.191	0.008	23.669
Item 7	**Factor 1**	0.462	0.017	27.700
Item 8	**Factor 1**	0.434	0.016	27.073
Item 9	**Factor 1**	0.225	0.009	23.959
Item 10	**Factor 1**	0.235	0.010	24.053
Item 15	**Factor 1**	0.196	0.008	23.711
Item 17	**Factor 1**	0.270	0.011	24.423
Item 14	**Factor 2**	0.477	0.015	32.897
Item 18	**Factor 2**	0.403	0.013	30.766
Item 19	**Factor 2**	0.445	0.014	31.904
Item 20	**Factor 2**	0.693	0.016	43.577

Item 21	**Factor 2**	0.678	0.016	42.637
Item 22	**Factor 2**	0.537	0.015	35.113
Item 2	**Factor 3**	0.728	0.025	29.063
Item 3	**Factor 3**	0.469	0.018	26.055
Item 4	**Factor 3**	0.411	0.017	24.819

All 17 variables allocated to each of the factors were significant at 5% level of significance. According to Moutinho and Hutsheson (2011), parameter estimates are significant at 0.05 level if the t value exceeds 1.96.

Table 4.10: Model fit

Fit Summary	
Standardised RMR (SRMR)	0.0437
Goodness of Fit Index (GFI)	0.851
Adjusted GFI (AGFI)	0.824
RMSEA estimate	0.053
Bentler Comparative Fit Index	0.640
Bentler-Bonett NFI	0.614

A model of the best fit is recognised by looking at individual fit indices for each model during the analysis. The model has an SRMR of 0.0437, which is less than 0.05, indicating an acceptable model fit (Hair *et al*., 2010). In Table 4.10, GFI, AGFI, NFI and CFI are all above the suggested minimum acceptance value of 0.90, indicating a good fit (Hair *et al*., 2010), while the NFI suggests otherwise. RMSEA indicates a very good model fit at 0.053, which is less than 0.08 (Portela, 2012). The CFA fits the models because the comparison of models is recommended to select the appropriate model fit indices in line with the effect of the factors.

4.5 CFA- LDA

Table 4.11: Box's M test statistics (CFA- LDA)

Test Results		
Box's M		61.381
F	Approx.	5.092

	df1	12
	df2	4509054.658
	Sig.	.000

Table 411 presents the Box's M p-value of 0.000, which shows significant results at 0.05 significant level, an indication that not all variances of the population are equal. Box's M is sensitive to large data files, an indication that when there are large number of cases, it can detect even small departures from homogeneity (Zhu, 2018). According to Tavakoli (2012: 261), "if the sample sizes are seriously different in a ratio of 3 or more to 1, this assumption should be tested empirically".

Table 4.12: Eigenvalues and variance proportion

Eigenvalues				
Function	Eigenvalue	% of variance	Cumulative %	Canonical correlation
1	.314	96.6	96.6	.489
2	.011	3.4	100.0	.105

The Wilk's lambda eigenvalue is used to directly measure the function's discriminating power (Cecchi *et al*., 2020). In Table 4.12, the larger the eigenvalue, the better the discriminating power of the function (Bhattachrya and Dutta, 2020). The results show that only two variables were significant.

Table 4.13: Wilks Lambda statistics

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.753	277.019	6	.000
2	.989	10.834	2	.004

Table 4.13 shows that the purpose of Wilks' Lambda is to provide a basis for verifying the statistical significance of the function (Davis-Sramek *et al.*, 2020). This statistic ranges between zero and one, with a value closer to zero (0.000 and 0.004 in the current study) denoting a high level of discriminating power (Tsai and Hsiao, 2010). Wilks Lambda significant value of 0.000 and 0.004 means p<.005, thus, it can be concluded that factors are significantly dependent on child

observation. Wliks lambda is .753 and .989, close to one showing a high degree of separation between two groups.

Table 4.14: Structure matrix

Structure matrix		
	Function	
	1	2
Factor 2	.962	-.031
Factor 1	.783	.530
Factor 3	.171	.767

Table 4.14 shows the means of the discriminant function scores by factors for each function calculated (Bruin, 2010). Calculating the scores of the first function for each case in the dataset, and then looking at the means of the scores by group, revealed that Factor 1 had a mean of 0.962, Factor 2 had a mean of 0.783, and Factor 3 had a mean of 0.171. The more correlated the factors, the more difference between pattern and structure matrix. 0.962 and 0.783 were the largest absolute correlation between each variable and any discriminate.

Table 4.15: Functions at group centroids

Functions at group centroids		
Child observation	Function	
	1	2
Almost always	.587	-.090
Sometimes	.135	.155
Often	-.744	-.049

Table 4.15 shows the means of the discriminant function scores by group for each function calculated (Bruin, 2010). The scores of the first function for each case in the dataset, when calculated, and the means of the scores examined by group, revealed that the almost always group had a mean of 0.587 (-0.090), the sometimes group had a mean of 0.135 (0.155), and the often group had a mean of -0.744 (-0.049).

4.6 SDA

Table 4.16: Box's M tests

Test Results		
Box's M		366.458
F	Approx.	2.733
	df1	132
	df2	2506814.713
	Sig.	.000

Table 4.16 indicates the Box's M p-value of 0.000 showing significant results at 0.05 significance level, an indication that not all the variances were equal

Table 4.17: Eigenvalue and variance proportion

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.526	88.6	88.6	.587
2	.068	11.4	100.0	.252

Table 4.17 shows that the larger the eigenvalue, the better the discriminating power of the function (Bhattacharya, 2020). The only two variables were significant.

Table 4.18: Wilks Lambda statistics

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.614	474.181	22	.000
2	.937	63.518	10	.000

Table 4.18 shows that the purpose of Wilks' Lambda is to provide a basis to verify the statistical significance of the function (Davis-Sramek *et al*, 2020). Wliks lambda statistic of .666 and .961 between zero and one, close to 1, shows a high degree of separation between two groups. See Table 4.13 for the interpretation rule of thumb, which states that there is a statistical significance of the discrimatory power of the discriminant function.

Table 4.19: Structure matrix

Structure matrix		
	Function	
	1	2
Expressive language transformed	.665	.110
Empathy transformed	.536	.077
Self-awareness transformed	.529	.007
Add and sub transformed	.449	.211
String beads transformed	.373	.141
Expressive vocabulary transformed	.359	.089
Sort and class transformed	.332	.054
Oral comprehension transformed	.313	.092
Counting transformed	.278	.179
DCCS transformed	.241	-.233
Sound discrimination transformed	.220	.116
Draw person transformed	.198	.135
Digit span transformed	.165	-.002
Cross and square transformed	.165	.104
Standing on one leg for 10 seconds transformed	.159	.054
Measurement vocabulary transformed	.11	-.059
Spatial vocabulary transformed	.380	-.542
Picture puzzle transformed	.287	.496
Pencil tapping transformed	.368	.475
Catch bean bag both hands transformed	.046	.391
Triangle transformed	.122	.196
Catch bean bag preferred hand transformed	.070	.173
Catch bean bag non-preferred hand transformed	.088	.147

In Table 20, the structure matrix holds the correlations between each variable and each factor same as with orthogonal rotations. 0.665 is the largest absolute correlation between each variable and any discriminate. All 23 variables were retained in the model and no variables excluded.

Table 4.20: Functions at Group Centroids

Functions at group centroids		
	Function	
Child observation	1	2
Almost always	.817	-.194
Sometimes	.069	.385
Often	-.926	-.154

Table 4.20 shows the means of discriminant function scores by group for each function calculated (Bruin, 2010Acording to the Table, if the mean is calculated, taking into consideration scores of the first function for each case in the dataset, an examination of the means of the scores by group shows that the “almost always”, had a mean of 0.695, “sometimes” had a mean of -0.165 while “often” had a mean of -0.887.

4.7 Model comparison

Table 4.21: Chi-square test of association between observed and predicted child observation

	Chi-square tests			
		Value	Df	Asymptotic significance (2-sided)
CFA-LDA	Pearson Chi-square	187.344	4	.000
SDA	Pearson Chi-square	317.216	4	.000

Table 4.21 shows a significant association between the observed child observation variable and the predicted child observation variable from both SDA and CFA-LDA (p-values < 0.000). That is, the predicted group was significantly closer to the predicted thus, both models are efficient according to this test.

Table 4.22: Accuracy of classification

	Child observation	Predicted group membership		
		Almost always	Sometimes	Often
CFA-LDA	Almost always	61.80%	22.80%	15.30%
	Sometimes	40.00%	29.50%	30.50%
	Often	20.40%	17.40%	62.20%
	Child observation			
		Almost always	Sometimes	Often
SDA	Almost always	67.10%	19.90%	13.00%
	Sometimes	32.80%	40.70%	26.60%
	Often	13.40%	19.80%	66.80%

The results in Table 4.22 shows that CFA-LDA correctly classified 61.8% of observations to the "almost always" category, 29.5% were correctly predicted to be in the "sometimes" category and 62.2% were correctly predicted to the "often" category. SDA correctly classified 67.1% of observations to the "almost always" category, 40.7% were correctly predicted to be in the "sometimes" category and 66.8% were correctly predicted to the "often" category. Thus, SDA correctly classified the observations better than CFA-LDA.

Table 4.23: Hit ratio and Apparent Error Rate (APER)

	Child observation	Predicted group membership		
		Almost always	Sometimes	Often
CFA-LDA	Almost always	214	79	53
	Sometimes	122	90	97
	Often	67	57	204
	Child observation			
		Almost always	Sometimes	Often
SDA	Almost always	232	69	45
	Sometimes	100	124	81
	Often	44	65	219

CFA-LDA Hit Ratio: $\frac{214+90+204}{979} \times 100\% = 51.9\%$

CFA-LDA APER= $1 - 0.519 = 48.1\%$

SDA Hit Ratio: $\frac{232+124+219}{979} \times 100\% = 58.7\%$

SDA APER= $1 - 0.587$ = 41.3%

CFA-LDA had 51.9% of original grouped cases correctly classified while SDA had 58.7% of original grouped cases correctly classified. However, the former had a higher APER compared to the latter. Thus, SDA was more efficient than CFA-LDA.

Table 4.24: Comparison of predicted group versus observed group using the Mann-Whitney U test

		Predicted group versus observed group
CFA-LDA	Z	-1.252
	Asymp. Sig. (2-tailed)	0.211
SDA	Z	-0.466
	Asymp. Sig. (2-tailed)	0.641

In Table 4.24, the predicted group memberships from CFA-LDA do not significantly differ from the observed child observation group membership (p-value > 0.05). The same results are shown for the SDA predicted group versus child observation (p-value >0.05). Thus, according to the Mann-Whitney U test, the predicted values for both models are significantly closer to the observed values.

4.8 Summary of chapter

This chapter has provided the analysis and presentation of results. Data reduction was done using EFA and CFA used to confirm results for EFA. CFA-LDA model was fit using factor scores as independent variables for LDA and SDA used all 23 independent variables as independent variables. Results of the model comparison revealed that both models are efficient based on the Chi-square and Mann Whitney's U tests. However, SDA has a higher accuracy of classification compared to CFA-LDA, a high hit ratio and, in turn, a lower APER. Thus, the results show that using all 23 variables retained in SDA, yielded the best model. The next chapter focuses on the discussion and recommendations.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter provides the conclusion to the study. The aim of the research was to determine the best model between SDA and CFA-LDA, using the same comparison criteria. The Chapter is structured as follows: Section 5.1 is the introduction; Section 5.2 provides a summary of each of the objectives of the study and conclusions based on the results presented in Chapter 4; Section 5.3 focuses on the discussion and recommendations for further research; while Section 5.4 is the summary of the study

5.2 Objectives of the study and conclusions

This section focuses on the objectives of the study and a summary of findings discussed in Chapter 4.

Objective 1: Use EFA to compress independent variables

Twenty three (23) independent continuous scale variables were used for EFA, 17 items were left after excluding the items, due to low communalities and factor loadings while three factors were extracted. Thus, this objective was achieved.

Objective 2: Use CFA to validate results of EFA

Results of construct validity showed that the 17 items loaded significantly onto their respective constructs from EFA. Goodness of fit statistics, namely, SRMR, GFI, AFGI, CFI, NFI and RMSEA were examined and it was established that the CFA model was a good fit for the data.

Objective 3: Fit LDA model using factor scores from EFA and CFA (CFA-LDA)

The CFA-LDA model was fit using factor scores saved from the CFA model as independent variables. The LDA was fit using the enter method. The dependent variable was child observation and factor 1, factors 2 and 3 were the independent variables. Thus, the objective was achieved.

Objective 4: Fit SDA using all candidate variables

SDA was fit using all 17 variables as independent variables and child observation as the dependent variable. Thus, this objective was achieved.

Objective 5: Compare CFA-LDA and SDA using several model comparison criteria

The Chi-square test of association, Mann- Whitney U, predictive accuracy, hit ratio and APER were used to compare the results. The Chi-square test of association and Mann-Whitney U yielded inconclusive results as both of these criteria revealed that the predicted values from both models did not differ from the observed ones. However, the predictive accuracy, hit ratio and APER identified SDA to be the most efficient model. Thus, the objective was achieved.

5.3 Discussion and implications for further research

The results revealed that the best and efficient model is SDA. The current study is a contribution to the literature as it extends the scope of the one conducted by Pandey *et a*l. (2011), which focused on a comparison of the efficiency of a straightforward stepwise LDA model using all independent variables and LDA model. The researchers also employed discriminant scores as independent variables examined in the current study, thus, the need to extend the scope of the study. The results of the current study support those of previous studies, which revealed that SDA is more efficient than factor loadings or component-based LDA models (Shrestha and Kazama, 2011 and Wu *et al*., 2016).

Future studies should be conducted to extend the scope of the current study by validating its results using a different dataset. Other exploratory factor analyses other than Principal Axis Factoring can be used prior to using CFA to confirm whether SDA will still outperform CFA-LDA. The same methodology can be used for interval scaled data since the current continuous scaled independent variables were used in the current study. More research can be done on these models, using different sample sizes and different variables to validate the results of the current study. Also, the effect of other data quality issues, such as missing values and non-normality on SDA and CFA-LDA can also be explored by future researchers. Based on the results of this study and previous research, for a parsimonious discriminant analysis, future researchers, especially those who have not specialises in Statistics, are advised to use SDA since it has proved to be more efficient.

5.4 Summary

Multivariate techniques were used in this study to find the most efficient and parsimonious model between CFA-LDA and SDA. The results revealed that SDA had better accuracy with regard to classification, hit ratio and a lower APER. Implications for further research were drawn and all the objectives of the study were successfully achieved.

REFERENCES

Arvanitoyannis, I.S., Katsota, M.N., Psarra, E.P., Soufleros, E.H. and Kallithraka, S. (1999). Application of quality control methods for assessing wine authenticity: Use of multivariate analysis (chemometrics). Trends in Food Science & Technology, 10(10), pp.321-336.

Awang, Z. (2012). Structural equation modeling using AMOS graphic. Penerbit Universiti Teknologi MARA.

Bacher, J., Wenzig, K. and Vogler, M. (2004). SPSS TwoStep Cluster-a first evaluation.

Bansal, A.K. and Chawla, P. (2013). Performance evaluation of face recognition using PCA and N-PCA. International Journal of Computer Applications, 76(8).

Beauducel, A. and Kersting, M. (2019). Identification of facet models by means of factor rotation: A simulation study and data analysis of a test for the Berlin Model of Intelligence Structure.

Beghdad, R., Touati, M. and Berboucha, M.S. (2011). A Data Analysis-Based Approach for Detecting Intruders. Information Security Journal: A Global Perspective, 20(4-5), pp.194-209.

Bhattacharya, A. and Dutta, A. (2020). Predicting a Model for the Financial Risk Tolerance of Retail Investors of Durgapur City on Their Demographic Factors Using Multiple Discriminant Analysis. In Smart Intelligent Computing and Applications (pp. 685-692). Springer, Singapore.

Bolboacă, S.D., Jäntschi, L., Sestraş, A.F., Sestraş, R.E. and Pamfil, D.C. (2011). Pearson-Fisher chi-square statistic revisited. Information, 2(3), pp.528-545.

Bonett, D.G. and Wright, T.A. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. Journal of Organizational Behavior, 36(1), pp.3-15.

Bontis, N. (2001). Assessing knowledge assets: a review of the models used to measure intellectual capital. International journal of management reviews, 3(1), pp.41-60.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis.Multivariate Behavioral Research, 36, 111–150.

Bruin, J. (2010) Newest command to compute new test. Los Angeles (Ca): UCLA: Academic Technology Services, Statistical Consulting Group.

Cain, M.K., Zhang, Z. and Yuan, K.H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. Behavior Research Methods, 49(5), pp.1716-1735.

Calder, A.J., Burton, A.M., Miller, P., Young, A.W. and Akamatsu, S. (2001). A principal component analysis of facial expressions. Vision research, 41(9), pp.1179-1208.

iCangur, S. and Ercan, I. (2015). Comparison of model fit indices used in structural equation modeling under multivariate normality. Journal of Modern Applied Statistical Methods, 14(1), p.14.

Castaño, A., Fernández-Navarro, F., Riccardi, A. and Hervás-Martínez, C. (2016). Enforcement of the principal component analysis–extreme learning machine algorithm by linear discriminant analysis. Neural Computing and Applications, 27(6), pp.1749-1760.

Costello, A.B. and Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical assessment, research, and evaluation, 10(1), p.7.

Cecchi, L., Migliorini, M., Giambanelli, E., Rossetti, A., Cane, A., Mulinacci, N. and Melani, F. (2020). Authentication of the geographical origin of virgin olive oils from the main worldwide producing countries: a new combination of HS-SPME-GC-MS analysis of volatile compounds and chemometrics applied to 1217 samples. Food Control, p.107156.

Chan, L.H., Salleh, S.H. and Ting, C.M. (2010). Face biometrics based on principal component analysis and linear discriminant analysis. Journal of Computer Science, 6(7), p.693.

Chen, L.F., Liao, H.Y.M., Ko, M.T., Lin, J.C. and Yu, G.J. (2000). A new LDA-based face recognition system which can solve the small sample size problem. Pattern recognition, 33(10), pp.1713-1726. .

Child, D. (2006). The essentials of factor analysis. A&C Black.

Chiu, T., Fang, D., Chen, J., Wang, Y. and Jeris, C. (2001). August. A robust and scalable clustering algorithm for mixed type attributes in large database environment. In Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining (pp. 263-268).

Cooper, D.R. and Emory, C.W. (1995). Business Research Methods. Richard D. Irwin. Inc., Chicago.

Davis-Sramek, B., Robinson, J.L., Darby, J. and Thomas, R.W. (2020). Exploring the Differential Roles of Environmental and Social Sustainability in Carrier Selection Decisions. International Journal of Production Economics, p.107660.

Ding, L., Velicer, W.F. and Harlow, L.L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. Structural Equation Modeling: A Multidisciplinary Journal, 2(2), pp.119-143.

Dıraman, H., Saygı, H. and Hışıl, Y. (2011). Classification of three Turkish olive cultivars from Aegean region based on their fatty acid composition. European Food Research and Technology, 233(3), p.403.

Dubuisson, S., Davoine, F. and Masson, M. (2002). A solution for facial expression representation and recognition. Signal Processing: Image Communication, 17(9), pp.657-673.

factors." Quality and Quantity41(5): 673-690.

Ferizal, R., Wibirama, S. and Setiawan, N. (2017). August. Gender recognition using PCA and LDA with improve preprocessing and classification technique. In Engineering Seminar (InAES), 2017 7th International Annual (pp. 1-6). IEEE.

Field, A. (2005). Factor analysis using SPSS. Retrieved March, 17(2019), pp.63-71.

Filieri, R.. (2010). Overcoming knowledge sharing barriers through communities of practice: empirical evidence from a big automotive supplier. Cambridge Scholars Publishing.

Filzmoser, P., Garrett, R.G. and Reimann, C. (2005). Multivariate outlier detection in exploration geochemistry. Computers & geosciences, 31(5), pp.579-587.

Fraley, C. and Raftery, A.E. (2002). Model-based clustering, discriminant analysis, and density estimation. Journal of the American statistical Association, 97(458), pp.611-631.

Grenoble, L and Austin, P.K. (2007). Current trends in language documentation. Language documentation and description, 4, pp.12-25

Hahs-Vaughn, D.L. and Lomax, R.G. (2013). An introduction to statistical concepts. Routledge.

Hair, J. F., Black, W. C., Babin, B. J. & Anderson R. E. (2010). Multivariate Data Analysis: A Global Perspective, Prentice-Hall: Upper Saddle River, New Jersey

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2014). Multivariate data analysis: Pearson new international edition. Essex: Pearson Education Limited.

Helwig, N.E. (2017). Discrimination and classification. Assistant Professor of Psychology and Statistics. University of Minnesota (Twin Cities). Updated 14 March.

Hilbe, J.M. (2014). Modeling count data. Cambridge University Press.

Hooper, D., Coughlan, J. and Mullen, M.R. (2008). Structural equation modelling: Guidelines for determining model fit. Electronic journal of business research methods, 6(1), pp.53-60.

Jöreskog, K.G. and Sörbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. Scientific Software International.

Jung, S. and Qiao, X. (2014). A statistical approach to set classification by feature selection with applications to classification of histopathology images. Biometrics, 70(3), pp.536-545

Kara, D. (2009). Evaluation of trace metal concentrations in some herbs and herbal teas by principal component analysis. Food Chemistry, 114(1), pp.347-354.

Khan, A. and Farooq, H. (2012). Principal component analysis-linear discriminant analysis feature extractor for pattern recognition. arXiv preprint arXiv:1204.1177.

Kher, A., Mulholland, M., Green, E. and Reedy, B. (2006). Forensic classification of ballpoint pen inks using high performance liquid chromatography and infrared spectroscopy with principal components analysis and linear discriminant analysis. Vibrational spectroscopy, 40(2), pp.270-277.

Kıralan, S., Yorulmaz, A., Şimşek, A. and Tekin, A. (2015). Classification of Turkish hazelnut oils based on their triacylglycerol structures by chemometric analysis. European Food Research and Technology, 240(4), pp.679-688.

Krzanowski, W. (2000). Principles of multivariate analysis (Vol. 23). OUP Oxford.

Lawrence, K.D., Pai, D.R. and Kleinman, G., 2009. Bankruptcy prediction in retail industry using logistic regression. Applications of Management Science, 13, pp.61-69.

Lawrence, K.D., Pai, D.R., Klimberg, R. and Lawrence, S.M. (2008). Understanding donor behavior: An empirical study of statistical and non-parametric methods. In Advances in Business and Management Forecasting (pp. 281-291). Emerald Group Publishing Limited.

Lawrence, K.D., Pai, D.R., Klimberg, R.K., Kudyba, S. and Lawrence, S.M. (2007). A classification model for a two-class (New product purchase) discrimination process using multiple-criteria linear programming. In Data Mining Methods and Applications (pp. 317-326). Auerbach Publications.

Lei, T.C., Wan, S., Chou, T.Y. and Pai, H.C. (2011). The knowledge expression on debris flow potential analysis through PCA+ LDA and rough sets theory: a case study of Chen-Yu-Lan watershed, Nantou, Taiwan. Environmental Earth Sciences, 63(5), pp.981-997.

Lobo, J.M., Jiménez-Valverde, A. and Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. Global ecology and Biogeography, 17(2), pp.145-151.

Loperfido, N. (2020). Some remarks on Koziol's kurtosis. Journal of Multivariate Analysis, 175, p.104565.

MacCallum, R.C., Widaman, K.F., Zhang, S. and Hong, S. (1999). Sample size in factor analysis. Psychological methods, 4(1), p.84.

Marathe, R.R. and Ryan, S.M. (2005). On the validity of the geometric Brownian motion assumption. The Engineering Economist, 50(2), pp.159-192.

Marcialis, G.L. and Roli, F. (2002). June. Fusion of LDA and PCA for Face Verification. In International Workshop on Biometric Authentication (pp. 30-37). Springer, Berlin, Heidelberg.

Marcialis, G.L. and Roli, F. (2006). Decision-level fusion of PCA and LDA-based face recognition algorithms. International Journal of Image and graphics, 6(02), pp.293-311.

Mary-Huard, T. and Robin, S. (2009). Tailored aggregation for classification. IEEE transactions on pattern analysis and machine intelligence, 31(11), pp.2098-2105.

Marsh, H.W. and Hau, K.T. (1996). Assessing goodness of fit: Is parsimony always desirable?. The journal of experimental education, 64(4), pp.364-390.

Maugis, C., Celeux, G. and Martin-Magniette, M.L. (2009). Variable selection in model-based clustering: A general variable role modeling. Computational Statistics & Data Analysis, 53(11), pp.3872-3882.

McHugh, M.L. (2013). The chi-square test of independence. Biochemia medica: Biochemia medica, 23(2), pp.143-149.

McLachlan, G.J. (2004). Discriminant analysis and statistical pattern recognition (Vol. 544). John Wiley & Sons.

Messick, S., (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American psychologist, 50(9), p.741.

Miles, J. (2014). Tolerance and variance inflation factor. Wiley StatsRef: Statistics Reference Online.

Montshiwa, .T.V and Moroke, N.D. (2014). Assessment of the reliability and validity of student-lecturer evaluation questionnaire: A case of North West University. Mediterranean Journal of Social Sciences, 5(14), p.352.

Moroke, N.D. (2015). A twostep clustering algorithm as applied to crime data of South Africa.

Moutinho, L. and Hutcheson, G.D. (2011). The SAGE dictionary of quantitative management research. Sage Publications.

Muijs, D. (2010). “Doing Quantitative Research in Education with SPSS”. 2nd edition. London: SAGE Publications.

Myung, I.J. (2000). The importance of complexity in model selection. Journal of mathematical

Ncube, B. and Moroke, N.D. (2016). Structural equation modelling applied to proposed statistics attitudes-outcomes model: a case of a university in South Africa.

Nwosu, F.D., Onuoha, D.O. and Eke, C.N. (2012). Comparison of Three Criteria for Discriminant Analysis Procedure. West African Journal of Industrial and Academic Research, 4(1), pp.18-27.

O’brien, R. (2007). "A caution regarding rules of thumb for variance inflation

Pallmann, P.S. (2016). Sample Size Determination in Clinical Trials with Multiple Endpoints, by T. Sozu, R. Sugimoto, T. Hamasaki, and SR Evans. Biometrical Journal.

Pandey, V.P., Babel, M.S., Shrestha, S. and Kazama, F. (2011). A framework to assess adaptive capacity of the water resources system in Nepalese river basins. Ecological Indicators, 11(2), pp.480-488.

Pedagadi, S., Orwell, J., Velastin, S. and Boghossian, B. (2013). June. Local fisher discriminant analysis for pedestrian re-identification. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on (pp. 3318-3325). IEEE.

Penny, K.I. (1996). Appropriate critical values when testing for a single multivariate outlier by using the Mahalanobis distance. Journal of the Royal Statistical Society: Series C (Applied Statistics), 45(1), pp.73-81.

Pereira, R.C., Skrobot, V.L., Castro, E.V., Fortes, I.C. and Pasa, V.M. (2006). Determination of gasoline adulteration by principal components analysis-linear discriminant analysis applied to FTIR spectra. Energy & Fuels, 20(3), pp.1097-1102.

Pohlmann, J.T. (2004). Use and interpretation of factor analysis in The Journal of Educational Research: 1992-2002. the Journal of Educational research, 98(1), pp.14-23.

Qu Y, Adam BL, Thornquist M, Potter JD, Thompson ML, Yasui Y, Davis J, Schellhammer PF, Cazares L, Clements M, Wright Jr GL. (2003). Data reduction using a discrete wavelet transform in discriminant analysis of very high dimensionality data. Biometrics. Mar;59(1):143-51.

Razzak, I., Saris, R.A., Blumenstein, M. and Xu, G. (2020). Integrating joint feature selection into subspace learning: A formulation of 2DPCA for outliers robust feature selection. Neural Networks, 121, pp.441-451.

Sarmento, R.P. and Costa, V. (2019). Confirmatory Factor Analysis--A Case study. arXiv preprint arXiv:1905.05598.

Schermelleh-Engel, K., Moosbrugger, H. and Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. Methods of psychological research online, 8(2), pp.23-74.

Şchiopu, D. (2010). Applying TwoStep cluster analysis for identifying bank customers' profile. Buletinul, 62, pp.66-75.

Sham, P.C. and Curtis, D (1995). Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. Annals of human genetics, 59(1), pp.97-105.

Shier, R. (2004). Statistics: 2.3 The Mann-Whitney U Test. Mathematics Learning Support Centre. Last accessed, 15, p.2013.

Shrestha, S., Kazama, F. Pandey, V and Babel, M (2011). A framework to assess adaptive capacity of the water resources system in Nepalese river basins. Ecological Indicators, 11(2), pp.480-488.

Skrobot, V.L., Castro, E.V., Pereira, R.C., Pasa, V.M. and Fortes, I.C. (2007). Use of principal component analysis (PCA) and linear discriminant analysis (LDA) in gas chromatographic (GC) data in the investigation of gasoline adulteration. Energy & Fuels, 21(6), pp.3394-3400.

Stąpor, K. (2015). Better alternatives for stepwise discriminant analysis. Acta Universitatis Lodziensis. Folia Oeconomica, 1(311).

Stewart, S., Ivy, M.A. and Anslyn, E.V. (2014). The use of principal component analysis and discriminant analysis in differential sensing routines. Chemical Society Reviews, 43(1), pp.70-84.

Suhr, D.D. (2006). Exploratory or confirmatory factor analysis?.

Sun, S., Guo, B., Wei, Y. and Fan, M. (2011). Multi-element analysis for determining the geographical origin of mutton from different regions of China. Food Chemistry, 124(3), pp.1151-1156.

Tabachnick, B.G. and Fidell, L.S. (2007). Using multivariate statistics (Vol. 5). Boston, MA: Pearson.

Tavakoli, H. (2012). A dictionary of research methodology and statistics in applied linguistics. Rahnama press

Tighe, E.L. and Schatschneider, C. (2014). A dominance analysis approach to determining predictor importance in third, seventh, and tenth grade reading comprehension skills. Reading and writing, 27(1), pp.101-127.

Treiblmaier, H. and Filzmoser, P. (2010). Exploratory factor analysis revisited: How robust methods support the detection of hidden multivariate data structures in IS research. Information & management, 47(4), pp.197-207.

Tsai CF, Hsiao YC. Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. Decision Support Systems. 2010 Dec 1;50(1):258-69.

Vandekerckhove, J., Matzke, D. and Wagenmakers, E.J., (2015). Model comparison and the principle. In the Oxford handbook of computational and mathematical psychology (Vol. 300). Oxford Library of Psychology.

Wagenmakers, E.J. and Waldorp, L. (2006). Special issue on Model selection: Theoretical developments and applications-Editors' introduction.

Wang, M., Perera, A. and Gutierrez-Osuna, R. (2004). October. Principal discriminants analysis for small-sample-size problems: application to chemical sensing. In Sensors, 2004. Proceedings of IEEE (pp. 591-594). IEEE.

Wu, X., Wu, B., Sun, J., Li, M. and Du, H. (2016). Discrimination of apples using near infrared spectroscopy and sorting discriminant analysis. International journal of food properties, 19(5), pp.1016-1028.

Yang, J. and Yang, J.Y. (2003). Why can LDA be performed in PCA transformed space?. Pattern recognition, 36(2), pp.563-566.

Yang, J., Yu, H. and Kunz, W. (2008) December. An efficient LDA algorithm for face recognition. In Proceedings of the International Conference on Automation, Robotics, and Computer Vision (ICARCV 2000) (pp. 34-47).

Yong, A.G. and Pearce, S. (2013). A beginner’s guide to factor analysis: Focusing on exploratory factor analysis. Tutorials in quantitative methods for psychology, 9(2), pp.79-94.

Yu, H. and Yang, J. (2001). A direct LDA algorithm for high-dimensional data—with application to face recognition. Pattern recognition, 34(10), pp.2067-2070.

Yuan, X., Ge, Z. and Song, Z. (2014). Soft sensor model development in multiphase/multimode processes based on Gaussian mixture regression. Chemometrics and Intelligent Laboratory Systems, 138, pp.97-109.

Zhang, Y., Wang, P., Xu, Y. and Meng, X. (2016). Metabolomic analysis of biochemical changes in the plasma of high-fat diet and streptozotocin-induced diabetic rats after treatment with isoflavones extract of radix puerariae. Evidence-Based Complementary and Alternative Medicine, 2016.

Zhao, H., Guo, B., Wei, Y. and Zhang, B. (2013). Multi-element composition of wheat grain and provenance soil and their potentialities as fingerprints of geographical origin. Journal of cereal science, 57(3), pp.391-397.

Zhao, Y., Wang, J., Lu, Q. and Jiang, R. (2010). Pattern recognition of eggshell crack using PCA and LDA. Innovative food science & emerging technologies, 11(3), pp.520-525.

Zhu, A.Y., Von Zedtwitz, M. and Assimakopoulos, D.G. (2018). Responsible Product Innovation. Innovation, Technology, and Knowledge Management.

Zygmont, C. and Smith, M.R. (2014). Robust factor analysis in the presence of normality violations, missing data, and outliers: Empirical questions and possible solutions. The Quantitative Methods for Psychology, 10(1), pp.40-55.