M060070484

Efficiency of the Structural Equation Model and Related Models in Validating the Theory of Planned Behaviour

Kolentino Nyamadzapasi Mpeta

orcid.org/0000-0001-9487-9500

Thesis submitted in fulfilment of the requirements for the degree *Doctor of Philosophy in Statistics* at the North-West University

Promoter: Prof N.D. Moroke

Co-Promoter: Dr L. Gabaitiri

Graduation July 2019

Student number: 25096478

DEDICATION

I dedicate this work to my family and many friends. A special feeling of gratitude to my loving parents whose words of encouragement and prayers sustained me. My beautiful wife, Thelma, Alishah and Aziel - this is for you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

First and foremost, I would like to thank God the Almighty for the gift of life and for enabling me to complete this thesis. Indeed, I can do all things through Him who gives me strength (Philippians 4:13). I am greatly indebted to my promoters, Prof N.D. Moroke and Dr L. Gabaitiri, for their invaluable advice and unwavering support throughout my doctoral program. Prof Moroke, thank you for your patience, concern and encouragement throughout this journey. Without your guidance and help, I would not have attained this point in my academic career. To all the Statistics and Operations Research department members, thank you for your support. Martin, thank you for the R programming and Latex assistance. To members within the Faculty of Economic and Management Sciences, thank you for your valuable comments during the colloquia. To the dean of the faculty, thank you for your personal encouragement and financial support. May the Lord abundantly bless you all. I also want to acknowledge the support from NWU, in the form of the postgraduate bursary as well as skills development funding. These went a long way in enabling me to complete this work. Prof B. Chilisa, without your permission to use the data, this study would not have seen the light of day. My gratitude goes to the school adolescents who contributed to the study as well as the entire UBARP team that was involved in this research.

I am extremely appreciative of my loving wife, Thelma, for her endless support, encouragement and always believing in me. Thelma, you have been an inspiration. To Lili and Azie thank you for your understanding. My siblings and Brian, thank you for checking on the progress. Sabelo and Papula, thank you for your effort in sourcing the library resources that I needed. Finally, it is done!

ABSTRACT

While there are a number of studies that have compared the adequacy of different socio-cognitive models, there appears to be a scarcity of studies that have compared statistical analytical approaches or strategies used to determine the adequacy of these in raising awareness of condom use and therefore mitigating the incidence of HIV and AIDS. This study sought to select an appropriate statistical analysis method for modelling latent variable data among several modelling techniques. Using secondary, cross-sectional data from a randomised control trial involving $n = 794$ Batswana in-school adolescents, the study applied Structural Equation Modelling (SEM), Multiple Regression (MR), Least Absolute Shrinkage and Selection Operator (LASSO) regression and the Generalised Additive Model (GAM) in explicating the influences that explain intention to use condoms among these adolescents and compared the results. Bootstrapping using 1000 samples of size, n = 794, was also carried out for the MR and structural equation models. The predictors of interest were all latent variables derived from the Theory of Planned Behaviour (TPB). Study results revealed that the structural equation model was more adequate for explaining Batswana in-school adolescents intentions to use condoms as it explained 57% of the variance in the model compared to 47.9% in the GAM and 44.7% in both the MR and LASSO regression models respectively. Furthermore, TPB predictors apart from both affective and instrumental attitude in the structural equation model as well as both the bootstrapped models and affective attitude in the remaining models were predictive of condom use intention among Batswana adolescents. Given that identical items were used to measure instrumental attitudes in all models, the study resolved that the difference in significance could be ascribed to the distinct methodological approaches. The structural equation model was more adequate for explaining Batswana in-school adolescents intentions to use condoms than the GAM, MR and LASSO regression models. GAMs would be a good choice in instances

where linearity is not assumed or model is not specified a priori. LASSO regression is very handy in instances where the researcher needs to select a few variables from a myriad variables.

The study recommends the application of SEM when estimating abstract concepts such as attitudes or perceptions towards a certain behaviour since SEM is a more appropriate and adequate approach than MR, LASSO regression and the GAM. This study expanded upon the growing body of HIV/AIDS prevention literature with a new focus on the choice of a relevant statistical analysis methodology. In addition, the TPB is recommended as a framework to establish the prognosticators of condom use intention among Batswana in-school adolescents. The study further recommends that policy makers working on developing HIV education programs or interventions targeted at adolescents should improve the intention to use condoms via promotion of positive instrumental attitudes, subjective norms and perceived behavioural control beliefs of condom use.

Keywords: theory of planned behaviour, generalised additive model, least absolute shrinkage and selection Operator, structural equation modelling, instrumental attitudes

CHAPTER 1. STUDY ORIENTATION

1.1 Introduction to the study

The existence of a variety of statistical techniques that researchers could use to examine relationships among variables necessitates the need for careful method selection.According to Jeon (2015:1634), it is imperative that any meaningful interpretation of statistical calculations must be based on a sound comprehension of appropriate modelling, based on informed methods. When this has been catered for, then the results could be fully tabled to inform decisions made by policy-makers. Such results could also be useful in adding to horizons of knowledge in the research academies.

It is unfortunate that researchers at times apply statistical techniques they saw others apply in the hope of getting results similar to those that others got while lacking a full understanding of whether the techniques applied actually fit the specific needs of their study. The choice of a modelling approach cannot therefore be ignored as it has a bearing on the results obtained and their interpretation. Furthermore, Goodhue et al. (2012) stated that research that compares statistical methods is valuable to researchers as it provides guidance concerning which statistical technique could be more useful and appropriate in a given setting. It is for this reason that this study explored the appropriateness of modelling approaches: Structural Equation Modelling (SEM), Multiple Regression (MR), Least Absolute Shrinkage and Selection Operator (LASSO) regression and Generalised Additive Models (GAMs) in analysing data collected among Batswana in-school adolescents based on the Theory of Planned Behaviour (TPB) while attempting to understand condom use intentions among Batswana adolescents.

The TPB is a well-established socio-cognitive model for predicting a diversity of human behaviours (Ajzen, 2011). The theory postulates that behavioural intention is influenced by attitude, normative beliefs and perceived behavioural control. Although the TPB has been extensively applied in the study of sexual risk behaviour in the western world (Albarracn et al., 2001), its applicability and suitability have however been questioned in non-western and, especially, African settings. This study therefore further validated the applicability of the TPB in predicting condom use intention among Batswana in-school adolescents.

The rest of this chapter is organised as follows: Section 1.2 outlines the background and rationale for the research. The problem statement is defined in section 1.3. Section 1.4 lists the study objectives and research questions are posed in section 1.5. Study hypotheses are stated in section 1.6. The significance of the study is identified in section 1.7 while the organisation of the study is outlined in section 1.8. Lastly, a summary of the whole chapter is given in section 1.9.

1.2 Background of HIV/AIDS in Botswana

Knowledge with respect to adolescents intentions to engage in defensive sexual behaviours is still deficient in numerous nations around the globe, particularly in Sub-Saharan Africa (SSA) where HIV prevalence is the highest (Sacolo et al., 2013). While there is growing evidence that shows that behavioural interventions based on grounded theoretical frameworks and theory-based determinants could reduce HIV risk-associated behaviours such as premarital unprotected sex and having multiple sexual partners, few studies have been conducted in SSA, more especially with Batswana adolescents.

In Botswana, the HIV and AIDS epidemic is largely driven through sexual transmission. The Botswana government therefore recognised behaviour change as the solitary lasting answer to the prevention of the HIV and AIDS epidemic (UNAIDS, 2012). Young people are major casualties in this epidemic hence there is a need to come up with informed, culturally sensitive and effective intervention programmes especially targeted at adolescents aged between ages 15 19 years.

Data from a statistical report published by Statistics Botswana (Botswana. Statistics Botswana. (2013)), revealed high risk sexual behaviour patterns, especially among young men. According to the survey results, about 50% of the young men, aged between 15-19 years, reported more than one sexual partner in the year preceding the report in contrast to their female counterparts at 25.2%.

There was decreased condom use among the general population, for both genders, and across all age groups. For instance, the survey revealed that a decline in condom use was observed in the general population from 90.2% recorded in the 2008 Botswana AIDS Impact Survey (BAIS) III to 81.9% recorded in the 2013 BAIS IV. Furthermore, decreased rates of condom use were evident in all females from 89.5% to 83.14% and all males from 90.4% to 81.2%.

Agyei & Abrefa-Gyan (2016) in their study that sampled 17–24 year old students from the University of Botswana, examined risky sexual patterns and the use of condoms among youth in Botswana. The study indicated that 33% of the sexually active respondents had unprotected sex in the month preceding the survey. The foregoing statistics are indicative of a challenge regarding the use of condoms and point to a need for condom use promotion, especially targeted at young people.

While a number of studies (e.g. Airhihenbuwa & Obregon, 2000; Bryan et al., 2006; Espada et al., 2016; Noar, 2007) have contrasted the adequacy of several socio-cognitive models, to the best of the researchers knowledge, there are no studies that have compared statistical analytical approaches or strategies used to determine the adequacy of these in raising awareness of condom use and therefore mitigating the incidence of HIV and AIDS. It is also interesting to note that most studies about socio-cognitive model applications to condom use intentions focus on young adults and adults. Few studies have fully examined condom use intentions among in-school adolescents. This study therefore attempted to fill the afore-mentioned gap by considering applicable statistical models such as the Structural Equation Model, MR model and related models while simultaneously validating the applicability of the TPB in the Botswana setting.

1.3 Problem Statement

Selection of methodology is an important part of any research study (Davis, 1996; Stevens, 2009). Lowry & Gaskin (2014) concurred and suggested a careful selection of statistical techniques on the basis of the type of data collected. Furthermore, statistical techniques should be carried out in the context of theory using measures derived from theory (Lowry & Gaskin, 2014:123). Whilst MR is one of the most widely used of all statistical methods, it has a number of limitations especially when dealing with latent variables. For instance, regression analysis supposes that independent variables (IVs) are measured without error and this method is not capable of dealing with multiple dependent variables (DVs) simultaneously. The failure to take account of measurement error in parameter estimates is potentially quite severe since findings presented by Principal Component Analysis (PCA)/MR analysis concerning the discriminant and predictive validity may possibly be artifactual (Langdrigde et al., 2007).

In spite of its limitations, MR is often a preferred choice due its relative simplicity (Tabachnick & Fidell, 2014). SEM, on the other hand, is a second generation multivariate method that allows the simultaneous analysis of all variables as opposed to analysing them independently. Besides the capacity to handle both observed and latent variables, SEM can also be applied to assess the reliability and validity of the model measures. Despite the fact that both approaches have been applied to a variety of research problems, adequacy and effectiveness of the methods are rarely taken into consideration. MacCallum & Austin (2000) observed that the constraints of statistical modelling are sometimes not mentioned or disregarded by researchers.

In spite of modern regression techniques such as LASSO regression and GAM expanding modelling technique choices for researchers, they seem to be side-lined. LASSO regression, for instance, provides greater prediction accuracy in addition to increasing model interpretability. GAMs, on the other hand, are a flexible approach that provides excellent fit for both linear and non-liner relationships. They relax the usual parametric assumptions and allow for the uncovering of structure within the relationship between independent and dependent variables (Kapoor et al., 2016). In addition, GAMs can be used to verify results of linear models and are very powerful for prediction and interpolation. This study thus sought to determine and compare the adequacy of the SEM and the MR model as well as LASSO regression and GAMs using data collected on the basis of the TPB.

The TPB, which was the data collection theoretical framework for this study, has been extensively utilised to study condom use intention among different groups, for instance, men who have sex with men (Wolitski & Zhang, 2007), injection drug users (Macalino et al., 2009), female commercial sex workers (Janner et al., 1998), and high school-age adolescents (Bryan et al., 2002; Rannie & Craig, 1997; Wise et al., 2006). The outcomes in a meta-investigation showed that the TPB variables were among the most grounded indicators of condom use (Guo et al., 2014; Sheeran et al., 1999). This theory has likewise demonstrated efficacy in predicting both intention to utilise condoms and actual condom use (Alvarez et al., 2010; Bennet & Bozionelos, 2000).

While the TPB has been utilised as a theoretical framework for predicting condom use in such populations of the European (Carmack & Lewis-Moss, 2009; Mausbach et al., 2009; Munoz-Silva et al., 2009), African (Bryan et al., 2006; Sacolo et al., 2013; Schaalma et al., 2009) and Asian (Cha et al., 2007; Molla et al., 2007) stocks, to the best of the researchers knowledge, the TPB has not been utilised in investigating the influence of attitudes, normative beliefs, and perceived behavioural control with respect to Batswana in-school adolescents condom use intentions.

It is worth noting that theories that could be relevant to certain populations may not be appropriate for other populations as a result of variances in culture, language, history and education. It is for this reason that some authors (e.g. Airhihenbuwa & Obregon, 2000; Campbell & Murray, 2004; Campbell et al., 2007) have intensely quizzed the applicability and suitability of socio-cognitive theories, as well as the TPB, in non-western and, particularly, African settings, advancing cultural and particularly community considerations as more essential. In reaction to this, the TPB has been proven to have good predictive competences outside a Western context where it was first established (Schaalma et al., 2009). Even though the constructs of the TPB are deemed universal, it is recognised that cultural variations have an effect on the dynamics of attitudes, subjective norms, and perceived behavioural control. Since Botswanas culture is unlike cultures in western or Asian countries, it is essential to investigate the Batswana population to ascertain whether the TPB could be a suitable framework to study the elements that motivate condom use intention among Batswana in-school adolescents while applying appropriate statistical methods.

1.4 Aim and Objectives

The aim of this study is to select an appropriate statistical analysis method between SEM, MR, LASSO regression and GAM when dealing with data involving latent variables.
The specific objectives of this study are to:

1. determine the adequacy of the structural equation model, the MR model, LASSO regression and GAM in explaining and predicting condom use intention using Batswana in-school adolescent sample data,

2. apply the TPB to study factors that influence Batswana in-school adolescents' intention to use condoms,

3. formulate suggestions for intervention programs using the findings.

1.5 Research questions

The specific research questions to be answered are:

1. Which model, between the structural equation model, MR, LASSO regression and GAM, is more adequate for explaining and predicting Batswana in-school adolescents intentions to use condoms?

2. Which TPB elements contribute significantly to explaining Batswana in-school adolescents condom use intentions?

3. What suggestions can be made towards intervention program formulation?

1.6 Hypotheses

Research studies conducted in the recent past show that that attitudes towards and beliefs about condoms affect both individuals intentions to use condoms and actual condom use. Such studies have established that adolescents are more likely to use condoms when they identify tangible benefits from such use. Based on the perceived benefits, theses adolescents are likely to cultivate upbeat attitudes toward condoms(Maharaj & Cleland, 2006; Taylor et al., 2014). Thus the first hypothesis tested in this study was:

H_1: Instrumental attitude (Instr_Att) has a positive and significant effect on condom use intention (CdmUse_Intention).

In contrast to H_1, Van Rossem & Meekers (2011), noted that youth are less likely to use condoms when they perceive barriers and develop negative attitudes toward them. Furthermore, young people may neither use nor intend to use condoms when they believe and perceive condoms as unreliable and capable of reducing sexual pleasure (Katikiro & Njau, 2012; Ochieng et al., 2011). This led to the second hypothesis tested in the study:

H_2: Affective attitude (Aff_Att) has a negative and significant effect on condom use intention (CdmUse_Intention).

Bennet & Bozionelos (2000) in their review of 20 studies focusing on the utility of the TPB in predicting condom use established that there was a positive and significant relationship between normative beliefs and condom use intentions in 14 of the studies. Moreover, Ebrahim et al. (2017), in their study of "*psychosocial determinants of intention to use condoms among Somali and Ethiopian immigrants in the U.S.*", hypothesised that higher condom use intentions will be predicted by higher positive attitudes, norms and greater perceived behavioural control. Consequently the following hypotheses were tested in this study:

H_3: Normative beliefs (Norms) have a positive and significant effect on condom use intention (CdmUse_Intention).

H_4: Perceived behavioural control (Perceived control) has a positive and significant effect on condom use intention (CdmUse_ Intention).

1.7 Significance of the study

The findings of this study emphasise the importance of choosing and using appropriate analysis methods when using data with latent variables. Thus statisticians and researchers that apply the recommended approach derived from the results of this study are able to apply appropriate analysis techniques that fit the specific needs of their studies and give robust results. Policy makers on the other hand will benefit from the identification of a more reasonable, data related, analysis method as their policy decisions will be established on more accurate results. Furthermore, researchers and policy makers will be guided by the findings from the TPB validation on the variables that they need to target when designing interventions targeted at condom use among Batswana adolescents. The community will in turn benefit from appropriate theory-guided and culturally sensitive intervention programs that may be formulated on the basis of robust statistical methods that give more reliable results.

1.8 Organisation of the Thesis

The rest of the thesis is organised as follows:

Chapter 2: Literature review

Chapter 2 is a review of existing literature on the subject under consideration in order to identify gaps in the literature and thus buttress the purpose for this research.

Chapter 3: Methodology

Chapter 3 explains in detail the SEM and multiple regression approaches that are applied in this research.

Chapter 4: Results

Chapter 4 is the presentation of data and analysis of the results.

Chapter 5: Summary, Discussion and Conclusions

Chapter 5 provides a summary of the study as well as the discussion of the findings. Conclusions drawn from the findings and implications for practice are also highlighted in this final chapter.

1.9 Summary

In this chapter, an introduction as well as background to the study was given. The importance of selecting appropriate modelling approaches when performing statistical analysis was emphasised.

HIV prevalence statistics as well as condom use rate for Botswana were also presented in this chapter. The statistics indicated a decrease in condom use among the general population, for both genders, and across all age groups. A brief background that focused on TPB as a framework for studying condom use intention was given in the chapter. The chapter then highlighted the aim and objectives of the study. Four hypotheses to be tested in this study were stated. The chapter concluded by considering the significance of the study as well as giving a summary of the organisation of the thesis.

CHAPTER 2. LITERATURE REVIEW

2.1 Introduction

This chapter examines literature and recent studies on condom, especially targeted at adolescents, carried out by other researchers with a view of identifying gaps in this literature such that the contribution of this current study is discernible and justified. To begin with, the researcher highlights the importance of targeting adolescents with regards condom use. A discussion on applications of socio-cognitive theories in behavioural research is then undertaken. The chapter concludes by looking at modelling approaches namely, structural equation modelling, multiple regression alongside its related models and generalised additive models.

2.2 Importance of targeting adolescents

Of the 35 million people living with HIV in 2015 worldwide, a fifth are minors and youth under the age of 25 (UNAIDS, 2014). Adolescents aged 10 to 19 years account for an estimated 2.1 million HIV infections (Idele et al., 2014), and young adults aged 20 to 24 account for an estimated 2.8 million infections (UNAIDS, 2014), implying that almost 5 million young people between the ages of 10 and 24 are living with HIV. Approximately 300,000 new HIV infections occur annually among adolescents aged 15-19 years, based on 2012 estimates (Idele et al., 2014). Worldwide, two-thirds of these infections are among girls, but in some countries more than 80% of new infections are among girls (Idele et al., 2014). The problem of adolescent HIV is concentrated in sub-Saharan Africa, with 82% of the world's HIV-positive adolescents living in this region, mainly in southern Africa (Idele et al., 2014).

Adolescents face critical development tasks such as formation of identity and self-esteem, social and psychological pressures, and the introduction of adult roles and accountabilities which may include income generation and caring for family members (Kapogiannis & Legins, 2014). Yi et al.

(2010) added that adolescents are time and again regarded as being at a life phase of increased experimentation and adventure concomitant with an assortment of risky behaviours, including risky sexual behaviours. Girls and young women face particular environments of risk, including being coerced into marriage or unwanted sexual experiences. All of these facets may place young people in danger of sexual practices which expose them to possibilities of HIV infection, comprising early sexual debut, multiple partners, non-use of condoms, transactional or forced sex, inter-generational sex, and sex under the influence of alcohol or drug use (Kapogiannis & Legins, 2014).

Risky sexual behaviours such as early sexual debut, multiple sexual partners, and non-use of condoms expose and put adolescents at risk of HIV infection (Idele et al., 2014). Adolescence is therefore a critical time to encourage healthy sexual behaviours; healthy practices established during adolescence are likely to be retained through adulthood (Romero et al., 2011). Kapogiannis & Legins (2014) concurred that "adolescence and young adulthood are critical times of life in which attitudes, behaviours, and lifestyles are established which will affect health and well-being throughout the life-course." Jemmott (2012) suggested that young adolescents, before or just after becoming sexually active, are very suitable and important intervention targets due to their high vulnerability and the fact that they are yet to establish habitual sexual behaviour patterns. Available data suggest that a vast number of new infections in many parts of the African continent occur in adolescents, with female adolescents exhibiting a more prominent likelihood to acquire the infection (Okonofua, 2013). The manifestation of new infections among adolescents could be ascribed to young people's engagement in sexual risk behaviours that could lead to unintended health outcomes. In order to navigate the maze around risky sexual behaviours, it is imperative to devise sensitive and effective interventions, taking into cognisance the cultures of the people engaged in the aforesaid risky behaviours, such that, in the penultimate, an acceptable approach is specifically developed to reduce the high risks and manipulative effects for African adolescents.

2.3 Socio-cognitive Theory Applications in Behavioural Research

A whole host of theories whose goal is to appreciate health-related behaviour and offer tools for behaviour modification coexist in health promotion research (Michielsen et al., 2012). Among the most commonly utilised theoretical models in the sphere of HIV/AIDS are the Socio-Cognitive Model (SCM) (Bandura, 1977, 1986, 1994), Information-Motivation-Behavioural (IMB) skills model (Fisher et al., 1996; Fisher et al., 2002; Fisher et al., 2003) and the TPB (Ajzen, 1985, 1991; Fishbein & Ajzen, 1975).

These theories in general indicate that behaviour is mostly influenced by intention(motivation) to practise the behaviour, and that intention is, in turn, determined by an individual's valuation of the consequences of the behaviour (attitude), the behaviour and sentiments of significant others (perceived norms), and personal control over carrying out the behaviour (PBC). Bandura's construct of self-efficacy puts more emphasis on the degree to which a person feels confident that he or she can successfully accomplish the target behaviour. Widespread experiments confirm and support the sufficiency of the TPB, SCM, and IMB model in predicting healthy behaviours generally (Fishbein & Ajzen, 1975; Bandura, 1994; Fisher & Fisher, 1992), especially condom use among adolescents.

2.3.1 The Socio-Cognitive Model

The basis of the SCM is that new behaviours are learnt by either observing the behaviour of others or by direct involvement. According to Bandura (1977), the SCM stresses the important roles played by indirect, representative and self-regulatory processes in psychological functioning and considers human behaviour as a constant collaboration between cognitive behavioural and environmental factors. Central principles of the SCM are:

1. self-efficacy – the belief in one self's capacity to implement the required behaviour

2. situation-outcome anticipation – belief about which outcomes will result with no interfering personal action and

3. action-outcome anticipation - the belief that a specified conduct will not lead to a particular result.

Both outcome expectancies and self-efficacy beliefs play a significant part in embracing fresh health behaviours, eradicating potentially harmful practices and upholding whatever novel behaviours have been attained (Luszczynska & Schwarzer, 2005).

The SCM has been applied in different behaviours for primary prevention such as stop- smoking programmes and problem-solving skills. It has likewise found its part in ancillary prevention programmes such as diabetes education programs and condom use promotion programmes. The social cognitive predictors of condom use have been researched in different populations. Backing for SCT has been established in a research of sexually active college students (Dilorio et al. 2000). Self-efficacy was found to be directly linked to condom use, but positive self-beliefs were also indirectly associated with condom use via the influence of self-efficacy on outcome expectancies.

Consistent with Bandura's theory (Bandura, 1977), self-efficacy predicted emotional state (anxiety), but this state was not linked to health protecting behaviour. Sexually active teenagers who expressed confidence in their skill to wear a condom and proclivity in their ability to decline intercourse with a sexual partner without a condom were more likely to use condoms regularly. Moreover, maintaining positive outcome expectancies, related to condom use, predicted more protective behaviours (Dilorio et al., 2001).

Kanekar & Sharma (2009) conducted a study to determine predictors of safer sex behaviours among sexually active African-American college students using SCT. The study utilised a cross-sectional study design and applied stepwise multiple regression as a modelling technique. Results of the study revealed that self-efficacy toward safer sex ($B = 0.594$, $p < 0.001$) was a significant predictor of safer sex behaviour. Self-efficacy towards safer sex accounted for 14.7% of the variance towards the dependent variable.

2.3.2 The Information-Motivation-Behavioural skills model

The IMB model is one of the few theories that was specially developed to understand HIV risk behaviour (Noar, 2007). The IMB model, proposed by Fisher and Fisher (1992) to explicate HIV-related behaviours, identifies three constructs namely information, motivation, and behavioural skills necessary to participate in a specified health behaviour, as precise and distinct causes of behaviour and behavioural change (Fisher & Fisher, 1992; Norton, 2009). Based on this model, Misovich et al. (2003), defined information as "an initial prerequisite for enacting health behaviour." This consists of not only behaviour-related information but then again myths and heuristics that permit spontaneous or cognitively easy behaviour-related decision-making (Fisher et al., 2003). Motivation comprises of two aspects: personal motivation, which embraces views regarding the intervention result in addition to attitudes toward a specific health behaviour (Fisher et al., 2003), and social motivation, which embraces the perceived social support or social norm for participating in a certain behaviour (Fisher et al., 2003). Behavioural skills, the third factor in the IMB model, are abilities needed for fulfilling specific health behaviour. To enable behavioural modification, behavioural skills in the IMB model stress the development of a person's independent abilities plus boosting professed self-efficacy (Fisher et al., 2003).

Liu et al. (2014) applied the IMB model in their study designed to examine the predictors of regularity of condom use among Chinese college students. Their study followed a cross-sectional research design and applied SEM in the assessment of the IMB model. The final IMB model in the study provided acceptable fit to the data (GFI = 0.992, CFI = 0.992, NFI = 0.989 and RMSEA = 0.028). Additional results from the study showed that preventative behaviour was significantly predicted by behavioural skills ($\beta = 0.754$, $p < 0.001$) while both information and motivation were not significantly associated with preventative behaviour. Information ($\beta = 0.138$, $p < 0.001$) and motivation ($\beta = 0.363$, $p < 0.001$) significantly and positively predicted behavioural skills, which indirectly affected consistent condom use.

The applicability of the IMB model in predicting condom use was likewise tried among approximately 400 sexually active secondary school students in Mbarara, Uganda. According to the results obtained using SEM, the IMB model predicted condom use to some extent (Ybarra et al., 2013). Condom use was precisely predicted by HIV prevention information as well as behavioural skills concerning access to and making use of condoms. Cai et al. (2013) also applied SEM in a cross-sectional, IMB–based study conducted to ascertain predictors of regular condom use among senior high school students in China. The study found that motivation ($\beta = 0.175$, $p < 0.01$) and behavioural skills ($\beta = 0.778$, $p < 0.01$) were significant predictors of consistent condom use. Information was an indirect predictor and was mediated by behavioural skills ($\beta = 0.269$, $p < 0.05$).

2.3.3 The Theory of Planned Behaviour

The TPB has its roots in socio-cognitive theory. The main thrust in TPB is to interrogate the influence of an individual's attitudes, subjective norms and perceived behavioural control on the intentions to carry out a specific behaviour (Ajzen, 1985). The TPB fully explains sexual behaviours in different ethnic adolescent populations insofar as it combines the social and cognitive components in explicating behaviours(Bryan et al., 2002; Cha et al., 2007; Espada et al., 2016; Gebhardt et al., 2003; Jemmott et al., 1998, Kalolo & Kibusi, 2015; Sacolo et al., 2013; Teye-Kwadjo et al., 2017a and 2017b). For example, Jemmott et al. (1998), applied the TPB in a study to elucidate delayed sexual initiation among African American youth while safer sex deliberations and condom procurement in minority inner-city youth were investigated by Bryan et al.(2002) using the same theory. The theory has similarly been expedient in describing the effects of condom use attitudes, norms, and control beliefs on condom use in Hispanic adolescents (Villarruel et al., 2004).

Meta-analytic and review studies offer widespread experiential backing of the TPB in predicting condom use in addition to other health behaviours among diverse populations, for example, adolescents and college students (Albarracín et al., 2001; Armitage & Conner, 2001; Godin & Kok, 1996; McEachan et al., 2011; Sheeran & Taylor, 1999; Webb & Sheeran, 2006). In a meta-analytic analysis of 185 studies, Armitage & Conner (2001), saw that the TPB explained 39 and 27 % of the variance in intention and behaviour, correspondingly. Godin & Kok (1996) obtained comparable results (41% and 34 % variance accounted for by intention and behaviour, respectively) in a review of 56 studies testing the applicability of the TPB to health-related behaviours. Albarracín et al. (2001) meta-analysed 96 studies mainly carried out in Europe and the United States, and resolved that attitude is the leading predictor of condom use intention (r = 0.58; β = 0.47) followed by perceived control (r = 0.45, β = 0.20), and perceived norms (r = 0.39; β = 0.20). Conversely, perceived norms have been identified as being more prognostic of the behaviour in adolescents compared to adults (McEachan et al., 2011). Further confirmation on the impact of the intention's predictors is nonetheless required, particularly in teenage samples.

According to Sheeran & Taylor (1999) and Sutton (1999), a substantial number of empirical studies have applied the TPB to trace the predictors of condom use intentions in heterosexual adolescents based on the TPB. The majority of such studies concur that attitude and subjective norm significantly predicted intention to use condoms. Still more, these studies have confirmed that a combination of attitude and norm contribute significantly to sway adolescents towards use. The significance of perceived control in predicting intention of condom use has similarly been empirically backed, however conclusions have been inconsistent. Ajzen (1991) and Fishbein et al. (1992) indicated that variations regarding the relative importance of predictors are anticipated among diverse population samples.

2.4 Modelling Approaches

A variety of modelling approaches such as logistic regression, General Linear Models (GLMs), Generalised Estimating Equations (GEEs) and Structural Equation Modelling (SEM) are applicable to studies involving the TPB. The various approaches are discussed below with a view of providing justification for the selection of the modelling approaches that were applied in this study.

2.4.1 Logistic Regression

Logistic regression investigates the relationship between a categorical DV and a set of IVs (Keith, 2015; Pituch & Stevens, 2016; Wu & Little, 2011). Similar to standard regression, logistic regression can be applied in a confirmatory model to test the association between explanatory variables and a binary outcome. Wu & Little (2011), further stated that "logistic regression predicts the probability of being a case ($\hat{p}(Y = 1)$) instead of predicting whether someone is a case or not."

A general expression for the logistic regression model is:

$$\ln(odds\ Y = 1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m \tag{2.1}$$

where β_0 represents the predicted value when all the Xs are equal to 0 and $\beta_1, \ldots, \beta_m$ are regression coefficients. β_m is such that for a unit change in the m^{th} IV, the logit changes by β_m units keeping the other IVs constant. Since the logit (ln (odds Y =1)) is not easy to comprehend, researchers frequently choose to explain the influence of the IVs by making use of the odds ratio (OR). For a single unit increase in X_m, the odds ratio increases by a factor of e^{β_m}.

Krugu et al. (2016) applied logistic regression in their study to investigate "Psychosocial correlates of condom use intentions among junior students in the Bolgatang Municipality of Ghana." Results from the study revealed that attitudes toward condom availability, injunctive norms toward condom use, sex experience, perceived susceptibility towards STIs, and perceived behavioural control toward purchasing in addition to making use of condoms set apart people with different levels of intentions to use condoms. Kalolo & Kibusi (2015) performed binary logistic regression to identify factors associated with intention to use and reported use of condoms among adolescents aged between 14 and 19 years in rural Tanzania. The TPB with addition of the empowerment component was used as the theoretical framework for the cross-sectional study. Results obtained in the study revealed that perceived behaviour control predicted intentions to use condoms (AOR = 3.059; 95% CI: 1.324 – 7.065) while a positive attitude (AOR = 3.484; 95% CI: 1.132 – 10.72) and empowerment (OR = 3.694; CI: 1.295 – 10.535) predicted reported condom use.

Multivariate logistic models were applied by Couture et al. (2010) in their study to examine determinants of condom use intentions among female sex workers' clients in Haiti. A cross-sectional survey was carried out with the TPB as a theoretical framework. Their study found that subjective norms (OR = 1.75; 95% confidence interval (CI): 1.06 – 2.88), PBC (OR = 1.34; 95% CI: 1.09 –

1.63) and attitudes (OR = 1.23; 95% CI: 1.04 – 1.44) were predictors of condom use intention, with norms being more significant. On the other hand, Alvarez et al. (2010), conducted multiple logistic regression analyses to evaluate the influences of attitudes, subjective norms, control beliefs (impulse control, condom negotiation, technical skills, condom availability, self-efficacy) and condom use intention predicted occurrence of unprotected sex and proportion of recent protected sex, consistent condom use and condom use at last sex. Compared to adolescents who inconsistently used condoms, adolescents who had better impulse control (OR = 3.92; 95% CI: 1.62 – 9.49) and grander intentions to use condoms (OR = 4.79; 95% CI: 1.0 – 22.8) were three and four times more likely to use condoms consistently. Adolescents who reported greater belief in their condom negotiating skills were less likely to use condoms consistently (OR = 0.32; 95% CI: 0.11 – 0.91). Since condom use intention was not measured as a binary outcome, the logistic regression model was not suitable for this study.

2.4.2 General Linear Models

Traditionally, GLMs have been the predominant analysis tools in the social sciences (Wu & Little, 2011). The General Linear Model (GLM) is a valuable framework for comparing how several variables have an effect on different continuous variables. According to Rutherford (2001:3), the simplest form of the GLM is represented as:

$$Data = Model + Error \tag{2.2}$$

Included under general linear models are techniques such as ANOVA, ANCOVA and MR, which is the most common and adaptable approach (Wu & Little, 2011). Notwithstanding their variations, each of the tests matches the definition in equation (2.2) above. In ANOVA, "data" represents the dependent variable scores, "model" is the experimental conditions and the "error" is the portion of the model that is unexplained by the data. Within regression analysis, "data" represents the dependent variable scores, "model" are the independent predictors and the "error" components are the residuals. ANCOVA, being a combination of ANOVA and regression can also be represented by equation (2.2).

2.4.2.1 Multiple Regression

According to Tabachnick and Fidell (2014) there are three major strategies that can be applied in multiple regression. These are standard (simultaneous) multiple regression, hierarchical (sequential) regression and stepwise (statistical) regression. The three strategies are discussed below with a view of selecting the most suitable approach for this study.

2.4.2.2 Standard Regression

In standard (simultaneous) multiple regression, all IVs or predictor variables are entered into the regression equation in one step. Keith (2015:80) suggested that explanatory research is critically important, specifically when one uses simultaneous regression to establish the extent to which one or more variables exert a demonstrable impact.

Keith (2015) further suggested that standard regression could be useful in instances where a researcher needs to find the extent to which a collection of variables predicts an outcome as well as the relative significance of the various IVs. The ability of standard regression to focus on both the overall effect of all variables and the individual variable effect makes it very useful in explanatory research. Furthermore, when the choice of variables to be included in the regression is theory-based, standard regression gives good effect estimates of the IVs on the DVs. One limitation of standard regression is that, depending on the IVs included in the regression equation regression, coefficients may be unstable.

2.4.2.3 Hierarchical Regression

In hierarchical (at times known as sequential) regression, IVs go into the equation in a sequence indicated by the data analyst or researcher. The order of IV entry may be based on logical or theoretical considerations. At each step, one or more IV is added to the model, and each IV or set of IVs' contribution to the model is assessed. Sequential regression unfortunately tends to overestimate the effects of variables if they are entered into the model too early while underestimating the effects of those variables which are entered later.

2.4.2.4 Stepwise Regression

Stepwise multiple regression, also known as statistical regression, is a way of processing regression in phases. Izenman (2013) suggested two principal types of stepwise procedures: backward elimination, forward elimination along with a fusion method that combines concepts from both main types. Backward elimination, according to Izenman (2013), commences with the complete set of variables and then drops at each step, the variable whose F-ratio,

$$F = \frac{(RSS_0 - RSS_1)/(\nu_0 - \nu_1)}{RSS_1/\nu_1} \tag{2.3}$$

is least, where RSS_0 is the residual sum of squares (with $\nu_0 = n - k$ degrees of freedom) for the reduced model, RSS_1 is the residual sum of squares (with $\nu_1 = n - k - 1$ degrees of freedom) for the larger model and k represents the number of variables in the larger model. Iterations are stopped when all variables retained in the model have F-ratios greater than some predetermined value, $F_{delete} = F_{0.1,\ \ 1,\ n-k-1}$ (Izenman, 2013).

Forward selection, in contrast, initiates with an empty set of variables. The variable with the largest F-ratio is chosen from the variable list at each step, with $\nu_0 - \nu_1 = 1$ and $\nu_1 = n - k - 2$, where k is the number of variables in the smaller model. The chosen variable is included in the regression model and then the enlarged model is refit. The selection of variables for the model is halted when the F value for each variable not currently chosen is lower than a specific fixed value, $F_{enter} = F_{0.25,\ \ 1,\ n-k-1}$ (Izenman, 2013).

Stepwise, forward, and backward methods of regression, have however received more censure than any of the other techniques of multiple regression (Aron & Aron, 1999; Chatterjee & Price, 1991; Cohen, 2001). Frequently, these approaches are disapproved since they yield variable results that are sample restricted and do not precisely or reliably reveal the obtainable relationships in the population. Additionally, stepwise methods have frequently led to erroneous calculations owing to the neglect of correct degrees of freedom, along with incorrect deductions regarding the comparative significance of predictor variables that are statistically reliant on variables previously entered into the investigation (Huberty, 1989; Thompson, 1989). Izenman (2013:146) added that "there is no guarantee that the subsets obtained from either forwards selection or backwards elimination stepwise procedures will contain the same variables or even be the "best" subset." Stepwise regression is consequently used in the exploratory stage of research or for purposes of pure prediction, not theory testing.

Given the foregoing discussion on the different multiple regression strategies and the fact that this study sought to identify the TPB elements that contributed significantly to explaining Batswana in-school adolescents' condom use intentions, standard multiple regression was the most appropriate choice. Standard multiple regression was therefore one of the procedures applied in this study. LASSO regression and GAMs which were also applied in this study are discussed in the sections that ensue below.

2.4.3 LASSO Regression

Lasso regression analysis is a shrinkage and variable selection method for linear regression models that was introduced by Tibshirani (1996). The objective of LASSO regression is to find the subset of predictors that minimises prediction error for a quantifiable dependent variable. It does so by applying a shrinking process where it penalises the coefficients of the regression variables shrinking some of them to zero. The loss function of LASSO can be represented as:

$$L = \sum \left(Y_i - \hat{Y}_i\right)^2 + \lambda \sum |\beta| \tag{2.4}$$

where Y_i represents observed values

$\hat{Y}_i$ represents fitted values

β denotes regression coefficients and

$\lambda \geq 0$ is a tuning parameter.

Variables that end up with a coefficient of zero subsequent to the shrinkage process are excluded from the model. The LASSO regression analysis helps in determining which amongst a set of predictors are most important. Inconsequential variables which are not related to the response variable are excluded thus overfitting is lessened.

2.4.4 Generalised Additive Model

The Generalised Additive Model (GAM) developed by Hastie & Tibshirani (1990), is an extension of the generalised linear model. It is a more versatile approach where each Y_i is linked with X_i by a smoothing function instead of a coefficient β. Its adaptability for non-normally distributed variables is seen as a plus (Tao et al., 2012). The basic additive model can be represented as:

$$E(Y \mid X_1, X_2, \ldots, X_p) = \beta_0 + f_1(X_1) + f_2(X_2) + \ldots + f_p(X_p) \tag{2.5}$$

where $f_i(X_i)$, i = 1,2, ..., p are non-parametric smoothing functions (splines) for explanatory variable X_i. The function f_i is estimated in a flexible way using parametric or non-parametric means thereby affording the potential for better fits than when applying other methods. Introduction of a link function into Additive Models (AMs) results into Generalised Additive Models (GAMs) of the form:

$$g\{E(Y \mid X_1, X_2, \ldots, X_p)\} = \beta_0 + \sum_i f_i(X_i) \tag{2.6}$$

The spline functions in GAMs are penalised splines or smoothing splines aimed at minimising the function:

$$\|Y - X\beta\|^2 + \lambda \int f''(x)^2 dx \tag{2.7}$$

where **Y** is the response vector

X is the data matrix

β is the vector of covariates

λ is a smoothing parameter and

$f''(x)$ is the second derivative of the smoothing function.

GAMs are data-driven rather than model-driven, that is, the resultant fitted values do not come from an a priori model. GAMs are said to be non-parametric (Yee & Mitchell, 1991) or semi-parametric (Guisan et al., 2002), which denotes in this case the absence of a specific functional form of the relationship between the response variable Y and the independent variable X GAMs can handle non-linear, linear and non-monotonic relationships between the dependent variable and independent variables. In contrast to AMs, GAMs are not limited to the normal distribution, but can be applied to any probability distribution from the exponential family (Liew & Forkman, 2015).

2.4.4.1 Smoothing Functions

Equation (2.2) shows that linear models split data into "model + error". Smoothing functions, on the other hand, partition data into "smooth + rough" and attempt to reduce the rough part as much as possible (Hastie & Tibshirani, 1990). While several types of smoothing functions are obtainable, they all rely on the same principles as listed below:

1. A regression model fitted to the surrounding observations predicts each observation in the data set.

2. The curve of the smoothing function is smooth.

3. A smoothing parameter λ controls the smoothness of the curve.

The smoothing parameter λ is frequently determined indirectly through the choice of effective degrees of freedom (edf). The number of effective degrees of freedom is comparable to the number of degrees of freedom of a linear model, which is the number of linear constraints or, in the case of the error, the difference between the number of observations and the number of linear constraints. A high edf value implies a highly non-linear curve. According to Liew & Forkman (2015:44), "The choice of the appropriate level of smoothing, by specifying the edf, is among the most crucial steps in fitting GAMs." An edf value between 3 and 5 is commonly chosen in practice. Cross validation can also be used to automatically choose the number of effective degrees of freedom.

2.4.5 Generalised Estimating Equations

The GEEs methodology, pioneered by Liang and Zeger (1986), is a notable strategy in the analysis of correlated data. GEEs are an expansion of GLMs, which enable regression analyses on dependent variables that are non-normally distributed (McCullagh & Nelder, 1989; Nelder & Wedderburn, 1972). Wang (2014) defined GEEs as "a marginal model popularly applied for longitudinal or clustered- data analysis in clinical trials or biomedical studies." GEEs estimate regression coefficients and standard errors with sampling distributions that are asymptotically normal (Liang & Zeger, 1986). GEE estimates are identical to those produced by ordinary least squares (OLS) regression in the absence of correlation within the response and when the dependent variable follows a normal distribution. The prime aim of GEE is the approximation of the mean model:

$$E(Y_{i,j}|X_{i,j}) = \mu_{i,j} \tag{2.8}$$

where

$$g\left(\mu_{i,j}\right) = \beta_0 + \beta_1 X_{i,j}\left(1\right) + \beta_2 X_{i,j}\left(2\right) + \beta_3 X_{i,j}\left(3\right) + \ldots + \beta_p X_{i,j}\left(p\right) = X_{i,j} \times \beta \tag{2.9}$$

In a GEE model, response variables are represented by $\{Y_{i,1}, Y_{i,2}, ..., Y_{i,n_t}\}$, where $i \in [1, N]$ is the index for clusters or subjects, and $j \in [1, n_t]$ is the index of the measurement within cluster/subject. $\{X_{i,1}, X_{i,2}, ..., X_{i,n_t}\}$ denotes the covariate vector.

In addition, GEEs can be used to analyse main effects and interactions and can be utilised to appraise categorical or continuous independent variables. GEEs are applicable when

1. a generalised linear model regression parameter, β, characterizes systematic variation across covariate levels,
2. the data represents repeated measurements, clustered data or multivariate response, and
3. the correlation structure is a nuisance feature of the data.

Fitting a GEE model needs the researcher "to specify (a) the link function to be utilised, (b) the distribution of the dependent variable and (c) the correlation structure of the dependent variable" (Ballinger, 2004:131). The link function is what "makes generalised linear modelling techniques part of a larger family of log-linear models; nonlinear and distinct from multiple linear regression in the link function and familiar in terms of the string of regression parameters" (Harrison, 2002: 454). The available options for the link function are displayed in Table 2.1 below.

Table 2.1 Distribution choices and link functions available in GEEs

Distribution	Link Function	Description
Normal	Identity link	Fits the same model as the general linear model
Binomial	Logit link Probit link	Fits logistic regression models Fits cumulative probability functions
Poisson	Log link	Fits Poisson regression models

In addition to the link functions displayed in Table 2.1 above, there are power link functions associated with the three distributions listed in the table as well as with the negative binomial and gamma distributions. The power link functions are in the form of any power transformation such as the square root or the square of the variable. There are also reciprocal link functions for the distributions listed in the table in addition to negative binomial and gamma distributions. The reciprocal link functions utilise the reciprocal of the dependent variable($1/\mu$).

McCullagh & Nelder (1989), emphasized the need for the researcher to make every attempt to accurately state the distribution of the response variable, when fitting a GEE, in order that the variance can be efficiently calculated as a function of the mean and regression coefficients can be correctly elucidated. For binary data, researchers must stipulate the binomial distribution. In the case of count data, either the Poisson or negative binomial distribution ought to be indicated, depending on the dispersion of the data (Gardner et al., 1995). Usually, the researcher will have some prior knowledge of the distribution of the response variable.

Specification of the form of correlation of responses within subjects or nested within group in the sample follows specification of the link function and the dependent variable distribution. The specification of the correlation structure will vary depending on the nature of the data collected. Once specified, the working correlation matrix allows GEEs to approximate models that represent the correlation of the responses (Liang & Zeger, 1986). Table 2.2 below gives a summary of commonly used correlation structures.

Table 2.2 Summary of commonly used "working" correlation structures for GEE

Correlation structure	Corr (Y_{ij}, Y_{ik})	Sample matrix
Independent	$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
Exchangeable	$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ \alpha & j \neq k \end{cases}$	$\begin{pmatrix} 1 & \alpha & \alpha \\ \alpha & 1 & \alpha \\ \alpha & \alpha & 1 \end{pmatrix}$
k-dependent	$Corr(Y_{ij}, Y_{i,j+m}) = \begin{cases} 1 & m = 0 \\ \alpha_m & m = 1,2,\dots,k \\ 0 & m > k \end{cases}$	$\begin{pmatrix} 1 & \alpha_1 & 0 \\ \alpha_1 & 1 & \alpha_1 \\ 0 & \alpha_1 & 1 \end{pmatrix}$
Autoregressive AR(1)	$Corr(Y_{ij}, Y_{i,j+m}) = \alpha^m, m = 0,1,2,\dots,n_i - j$	$\begin{pmatrix} 1 & \alpha & \alpha^2 \\ \alpha & 1 & \alpha \\ \alpha^2 & \alpha & 1 \end{pmatrix}$
Toeplitz	$Corr(Y_{ij}, Y_{i,j+m}) = \begin{cases} 1 & m = 0 \\ \alpha_m & m = 1,2,\dots,n_i - j \end{cases}$	$\begin{pmatrix} 1 & \alpha_1 & \alpha_2 \\ \alpha_1 & 1 & \alpha_1 \\ \alpha_2 & \alpha_2 & 1 \end{pmatrix}$
Unstructured	$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ \alpha_{jk} & j \neq k \end{cases}$	$\begin{pmatrix} 1 & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & 1 & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & 1 \end{pmatrix}$

Source: Wang (2014)

For data that are correlated within cluster over time, an autoregressive correlation structure is specified (Wang, 2014). Horton & Lipsitz (1999), recommended the use of an exchangeable correlation matrix when there is no logical ordering for observations within a cluster. GEEs are suitable for analysing longitudinal or clustered data thus they were not an appropriate model in this study as single period data was used.

2.5 Structural Equation Modelling

Ever since Joreskog's (1967) ground-breaking work on the maximum likelihood factor analysis and its subsequent expansions to the estimation of structural equation systems (Joreskog, 1973), SEM has grown into one of the most significant techniques of empirical research. SEM is a technique for determining relationships among unobserved (latent) variables and has been operational since early in the 20th century (Shah and Goldstein, 2006). According to Lei & Wu (2007), SEM refers to a large number of statistical models that are used to evaluate the validity of substantive theories with observed data. Structural equation models (SEMs) allow complex modelling of interrelated multivariate data for assessing interrelationships among observed and latent variables (Lee & Song, 2012). SEM was established to test and improve theoretical models endeavoring to clarify or predict social or behavioural phenomena (Bentler, 1988).

SEMs have been advanced in numerous academic specialties to confirm and test theory (Schumacker & Lomax, 2016). They are frequently used to evaluate unobservable 'latent' constructs. Gefen et al. (2000) indicated that SEMs allow the examination of both the measurement and structural components of a model by analysing the relations among several independent and dependent constructs concurrently. Hair et al. (2010:635) further suggest that "SEM is particularly useful when one dependent variable becomes an independent variable in a later dependence relationship and it gives rise to the interdependent nature of the structural model." Application of SEM is generally acceptable in the social sciences for the reason that SEM has the capacity to assign relationships between latent variables from observable variables (Hancock, 2003). The approach is quite suitable for modelling data in this study.

SEM promotes confirmatory, rather than exploratory, modelling; hence, it is appropriate for theory testing, rather than theory development. It typically begins with a hypothesis, denotes it as a model, operationalises the constructs of interest with a measurement instrument and tests the model. A structural equation model is created on the basis of appropriate and theoretically-sound constructs or the discoveries of a priori investigation, which serve as reasoning for hypothesising causal paths among latent variables and their subsequent indicators (Gallagher et al., 2008). The afore-mentioned steps were applied in the study. Given that this research study involves multivariate dependent-independent relationships, SEM was deemed an appropriate theory-testing model. Byrne (2001) contrasted SEM alongside other multivariate methods and recorded the following four distinctive features of SEM:

1. SEM takes a confirmatory methodology to data analysis by stating the relationship among variables *a priori*.
2. SEM offers clear approximations of error variance parameters.
3. SEM processes integrate both latent and observed variables.
4. SEM has the capacity to model multivariate associations, and approximate direct and

indirect influences of the set of variables under examination.

In the field of SEM, there are two approaches: Covariance-Based SEM (CB-SEM) and component-based SEM (also known as Partial Least Squares (PLS)). Even though CB-SEM and PLS are two different approaches to the same conundrum, they are at variance not only with respect to their basic assumptions and results, but also with respect to their approximation techniques (Hair et al., 2014b; Shook et al., 2004). Reinartz et al. (2009: 334) concur and state that "both start from the same set of theoretical and measurement equations but differ in how they approach the parameter estimation problem." A brief discussion of the two approaches is given in the paragraphs that follow.

2.5.1 Covariance-Based SEM

Tenenhaus (2008) stated that CB-SEM is typically applied with an objective of model authentication and requires a substantial sample (preferably more than 200 subjects). Crisci (2012) concurred that the goal of covariance structure analysis is the reproduction of the covariance matrix of the manifest variables by means of the model parameters and that it is a large sample approach. The covariance structure model comprises of two parts: the structural model and measurement model.

Kaplan (2009) and Kline (2011) concurred and stated that, "SEMs often involve a measurement model that defines latent variables using one or more observed variable, and a structural model that ascribes relationships between latent variables." The general full structural covariance-based model is expressed as:

$$\eta = \mathrm{B}\eta + \Gamma\xi + \zeta \tag{2.10}$$

where η (eta) is vector of endogenous latent variables, B (beta) is matrix of regression coefficients between endogenous latent variables, Γ (gamma) is a matrix of regression coefficients between endogenous and exogenous latent variables, ξ (xi) is a vector of exogenous latent variables and ζ(zeta) is a vector of disturbances.

Additionally, Crisci (2012), stated that for the structural part of the model to be wholly specified the following two, square and symmetric, matrices are required:

$\Phi(phi)$ – a variance-covariance matrix of the exogenous latent variables.

Ψ (psi) – a variance-covariance matrix of the errors.

The measurement model for exogenous variables (X) is expressed as:

$$\boldsymbol{x} = \Lambda_x\xi + \delta \tag{2.11}$$

where, $\boldsymbol{x}$ is a vector of manifest exogenous variables, Λ_x (lambda x) is a of matrix of regression coefficients relating xi to $\boldsymbol{x}$ manifest variables and δ (delta) is a vector of measurement errors in $\boldsymbol{x}$.

The measurement model for endogenous variables (Y) is expressed as:

$$\boldsymbol{y} = \Lambda_y\eta + \varepsilon \tag{2.12}$$

where, $\boldsymbol{y}$ is a vector of manifest endogenous variables, Λ_y (lambda y) is a matrix of regression coefficients relating eta's to y manifest variables and ε (epsilon) is a vector of measurement errors in y.

The measurement model is characteristically regarded as reflective and multivariate normality should be satisfied if approximation is performed by way of the maximum likelihood (ML) method.

2.5.1.1 Component-based SEM

Component-based SEM, also known as partial least squares-path modelling (PLS-PM), in contrast, is primarily utilised for score computation and can be executed on very small samples. PLS-PM aims to account for prediction of the construct relationships (Fornell & Bookstein, 1982; Hair et al., 2010 and 2014b). According to Chin (1998), parameter estimates in the PLS approach are found by basing on the capacity of minimising the residual variances of all dependent variables (both latent and observed). In contrast to CB-SEM, all the latent variables in the PLS-PM approach are expressed in a uniform way without taking into account whether they are endogenous or exogenous variables (Crisci, 2012). The path models in the PLS approach consist of three sets of relations namely, structural model, measurement model and weight relations.

The structural model can be stated as:

$$\xi_m = \mathrm{B}\xi_m + \zeta_m \tag{2.13}$$

where B represents the matrix of all path coefficients in the model and indicates the structural association between the latent variables. ζ_m is the inner residual term.

Just like in the CB-SEM approach, the measurement model in PLS-SEM, identifies the relationships between the constructs and their related indicators. There are however two ways in which the relationships can be represented: reflective way and formative way. In the reflective way, indicator or manifest variables are considered to be reflections of their latent variables thus the direction of causality is regarded to be from the construct to the indicator (Fornell & Bookstein, 1982; Edwards & Bagozzi, 2000; Jarvis et al., 2003). This can be represented in equation form as:

$$x_{pm} = \lambda_{pm}\xi_m + \varepsilon_{pm} \tag{2.14}$$

where x_{pm} is a linear combination of its latent variable ξ_m and λ_{pm} represents the loading coefficient associated with the p-th manifest variable in the m block. ε_{pm} denotes the outer residual term related to the manifest variable.

Indicators, in the formative way, are considered as roots of their latent constructs thus the direction of causality is from the indicator to the construct (Jarvis et al., 2003; Edwards & Bagozzi, 2000).

Formative indicators were originated by Curtis & Jackson (1962) and extended by Blalock (1964). The latent variable ξ_m is assumed to be a linear combination of its manifest variables x_{pm} and expressed as:

$$\xi_m = \Sigma_p \pi_{pm} x_{pm} + \delta_m \tag{2.15}$$

There is a discussion in the scholastic community regarding the value and pertinence of formative measures (Howell et al., 2007; Wilcox et al., 2008; Bagozzi, 2007; Diamantopoulos et al., 2008). For example, Howell et al. (2007), have contended that formative measurement has next to no value and is not an appealing option to the reflective measurement approach. Numerous other authors (e.g., Bollen, 2007; Diamantopoulos et al., 2008) have suggested that formative measures are critical yet are thought little of in underestimated in some fields.

Neither of the SEM approaches is by and large better than the other. For this reason, researchers need to utilise the SEM method that is best suited to their research objective, data characteristics, and model set-up (Fornell and Bookstein 1982; Gefen et al.2011; Reinartz et al. 2009). According to Astrachan et al. (2014), CB-SEM is a confirmatory approach that demands the specification of the full theoretical model prior to data analysis. Since this study was confirmatory, CB-SEM was a suitable approach.

2.6 Application of SEM in studies involving the TPB

SEM has been applied in a few of studies involving the TPB model, including studies in various African settings. For instance, Schaalma et al. (2009) applied SEM in their study aimed at testing the applicability of an expanded form of the TPB on intentions to use condoms among sizeable samples of young people in South Africa and Tanzania. Their study, showed that intentions to use condoms are largely motivated by perceptions of control, perceived social norms and attitudes. This finding was in agreement with researches studies carried out in Europe and the United States of America (Sheeran et al., 1999; Albarracín et al., 2001).

Bryan et al. (2006), also applied SEM to investigate the capability of TPB predictors to explain the proportion of variability in condom use intentions among South African teenagers and ascertain the degree to which, consistent with the TPB, intentions prospectively predict condom use behaviour. Sacolo, et al. (2013) made use of the SEM approach to investigate the TPB for predicting elements related to safer sexual behaviours, including sexual abstinence and condom use, among in-school youths aged 15 to 19 years in Swaziland. Results from the study conducted among Swazi in-school youth, found that perceived control for condom use was the strongest predictor of condom use intention (β = 0.36, p < 0.01) followed by subjective norms (β = 0.27, p < 0.01), and attitudes (β = 0.26, p < 0.01). Espada, et al. (2016) applied the SEM approach in their research that compared the adequacy of the TPB, SCM and the IMB skills model in predicting condom use among Spanish adolescents.

More recently, Teye-Kwadjo et al. (2017b), utilised SEM in their study aimed at determining condom use predictors among heterosexual young people in south-eastern Ghana. The TPB was used as the guiding framework of this study with the results indicating that attitudes toward condom use (b = .38; 95% CI [.14, .62], p < .001) and perceived behavioural control over condom use (b = .47; 95% CI [.31, .63], p < .001) were significantly positively associated with condom use intention. Subjective norms were however not statistically significantly associated with condom use intention (b = .06; 95% CI [-.14, .26], p = .593). Results from the structural model used to examine the direct relationships between the TPB constructs indicated good model fit. The χ^2 test of the model was statistically significant with a value of 241.12 (112, N= 684), p < 0.001, χ^2/df = 2.15, CFI =.967, RMSEA = 0.41; 90% CI [.034, .048]. The cross-sectional structural model explained 61% of the variance in condom use intention.

2.7 Application of MR in studies involving the TPB

Guo et al. (2014) applied multiple regression analysis in their study of TPB condom use predictors of among Chinese college students. The study found that intention to use condoms was statistically significantly predicted by attitudes (β = 0.213), subjective norms (β = 0.259), and perceived behaviour control (PBC) (β = 0.332). All predictors were statistically significant at the 0.001 level.

PBC was the strongest predictor of intention to use condoms. The variables explained 50.4 % of the variance in the model.

Similarly, Jemmott et al. (2007), in a study designed to identify condom use intentions among Xhosa-speaking adolescents in South Africa, established that attitude (B = 0.16, $p < 0.0001$) and PBC (B = 0.46, $p < 0.0001$) were significantly related to condom use intention. In contrast to the findings by Guo et al. (2014), subjective norm was not a significant predictor of intention to use condoms among the Xhosa-speaking adolescents. The overall model was significant at the 0.0001 level and explained 37% of the variance in intention.

Multiple regression was also applied by Alvarez et al. (2010) in a study to identify predictors of condom use among Mexican adolescents. Secondary data collected on the 48-month post-intervention from 157 sexually active adolescents, aged 17 to 21 years were analysed. Consistent with the studies by Guo et al. (2014) and Jemmott et al. (2007), regression analysis revealed that positive attitudes towards condoms ($\beta = 0.67$, $p < 0.001$) were significant predictors of condom use intention. Technical skills ($\beta = 0.13$, p < 0.01) and condom use self –efficacy ($\beta = 0.24$, $p < 0.001$) which were included in this study but were neither part of Guo et al. (2014)'s nor Jemmott et al. (2007)'s study, were also significant predictors of condom use intention.

All in all, the three variables explained 75.5% of the variance in the condom use intention model. Additionally, Alvarez et al. (2010), utilised MR to investigate whether attitudes, subjective norms, control beliefs (impulse control, condom negotiation, technical skills, condom availability, self-efficacy) and condom use intention predicted frequency of unprotected sex and proportion of recent protected sex, consistent condom use and condom use at last sex. Results revealed that intention to use condoms ($\beta = 1.57$, $p = 0.05$), impulse control ($\beta = 1.37$, $p < 0.01$) and condom use negotiation (β = -1.15, $p < 0.05$) were significant predictors of consistent condom use. The remaining three condom use behaviours were not predicted by all the IVs. Since different IVs are investigated across the studies, it is not easy to compare the results effectively.

More recently, Manyaapelo et al. (2017), applied hierarchical linear regression models to establish the distinctive contribution of "*psychosocial correlates of the intention to use condoms among young* *men in KwaZulu-Natal province.*" The study utilised the TPB as the guiding framework and results obtained from the study revealed that subjective norms and perceived behavioural control towards consistent condom use explained 46% of the variance in the intention to use a condom. Subjective norms ($\beta = 0.65$, $p < 0.001$) and perceived behavioural control ($\beta = 0.10$, $p < 0.001$) were significant predictors of intention to use a condom consistently. Attitude, though significantly correlated with intention to use a condom ($r = 0.113$, $p < 0.05$), was nonetheless not a significant predictor of consistent condom use intention.

2.8 Summary

This chapter highlighted condom use studies, especially those directed at adolescents, carried out by other researchers. At the onset, the researcher undertook to highlight the importance of targeting adolescents in relation to condom use. Adolescents are suitable and important intervention targets given the phase of life they are in. As much as possible, the literature focused on TPB studies that involved adolescents and youth. A discussion on applications of socio-cognitive theories in behavioural research which focussed on three theories that are mainly used in the sphere of HIV/AIDS, namely, the socio-cognitive model, the information-motivation behavioural skills model and the theory of planned behaviour was incorporated in the chapter. It is worth mentioning that a variety of modelling techniques can be applied to data analysis involving the TPB as highlighted by studies reviewed in this chapter. So far, logistic regression, multiple regression and structural equation models have been applied in studies based on the TPB. There seems however to be a scarcity of literature on the application of other modelling techniques such as LASSO regression and the GAM in TPB based condom use intention studies. This study thus sought to explore application of these latter techniques besides the currently applied techniques. It is also worth noting that none of the studies in literature compared the efficacy of the modelling techniques.

CHAPTER 3. METHODOLOGY

3.1 Introduction

The basis of all research concerns identifying what forms "valid" research. Valid research is also contingent upon appropriate research method(s) that steer the studies and engagements with epistemic knowledge that is both cutting edge and on the horizons of a specific discipline. A major purpose of nearly all research is to figure out some facet of the world experience in novel ways. Structural equation models offer a suitable environment for sense-making and thus relate philosophy of science principles to theoretical and empirical research (Bagozzi & Yi, 2012). In order to carry out and appraise any research, it is vital to identify what these assumptions are. As stated by Holliday (2002:7), "no matter how extensive the researches, different researchers will always pursue and see very different things in the same setting." The important fact is that the research methodology selected influences the results of the research study. Cooperrider & Srivastva (1987:1) stated that, "through our assumptions and choice of method, we largely create the world we later discover." This chapter offers an explanation and defence of applicable methodology for attaining the research objectives according to the following topics: theoretical framework, research paradigm, the research onion, research philosophy, research approach, research design, research strategy and methods.

3.2 Theoretical Framework

The theoretical framework adopted for this study is the TPB shown in Fig 3.1. The TPB is a well-researched model that predicts behaviour across a variety of settings. It has been described as the most broadly used theory in the investigation of sexual risk conduct in the western world (Albarracín et al., 2001). The TPB, which is an extension of the theory of reasoned action (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975), was developed by Icek Ajzen to predict human behaviour

(Ajzen, 1991). The theory suggests that any behaviour will almost certainly happen when there is a solid goal or strong intention and the capacity to complete the conduct, and when there are no environmental barriers to doing so (Fishbein, 2000; Conner & Armitage, 1998). The TPB endeavours to also predict non-volitional behaviours by integrating perceptions of control over execution of the behaviour as an extra predictor (Ajzen, 1988, 1991).

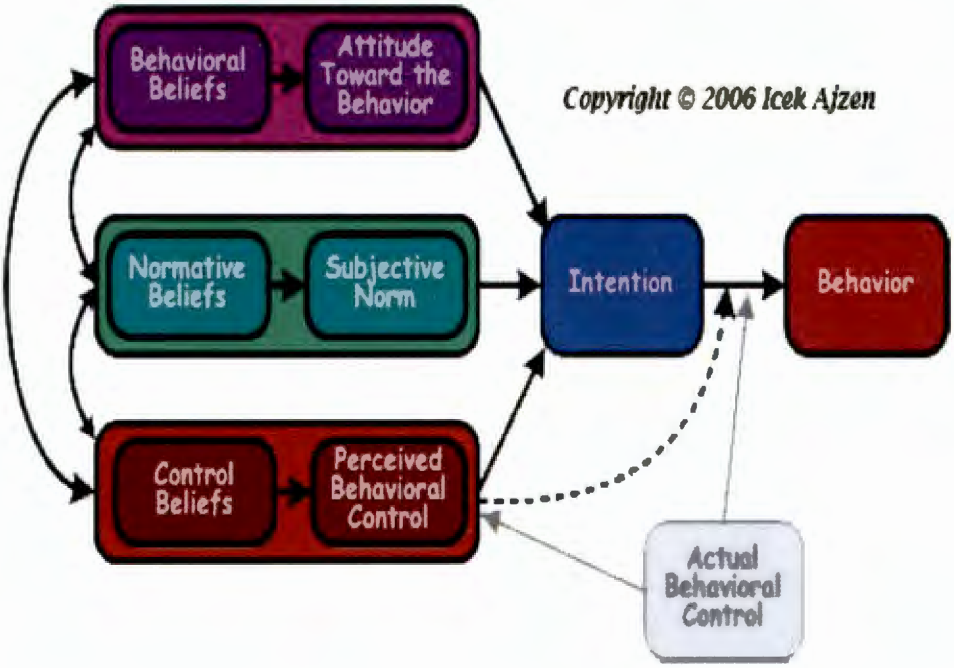


Figure 3.1 Theory of Planned Behaviour (TPB)

Source: http://people.umass.edu/aizen/tpb.diag.html

In the context of TPB, attitude towards the behaviour, subjective norm, and Perceived Behavioural Control (PBC) influence behavioural intention. Attitudes are personal beliefs about the behaviour. They consist of behavioural beliefs and outcome evaluations. Subjective norms are beliefs about what the significant others feel about the behaviour. They are a social pressure to either carry out or not to carry out a given behaviour. Lastly, PBC beliefs are perceptions about one's ability to perform the behaviour. PBC focusses on one's perceptions of how effortless or hard it is for them to carry out the behaviour in question. Mimiaga et al. (2009) suggested that behavioural control is similar to self-efficacy and is determined by the individual's perception of how arduous it is going to be to engage in the behaviour. Bandura (1977) defined self-efficacy as "an individual's belief in their innate ability to achieve goals." According to Ajzen (2012), "PBC is conceptually similar to self-efficacy", a term initiated by Bandura (1977). Both PBC and self-efficacy refer to individuals' beliefs in their ability to perform a given behaviour. The experience of mastery is the most important factor determining a person's self-efficacy. Others, on the other hand, have argued that perceived behavioural control and self-efficacy are conceptually distinct (Trafimow et al., 2002). Since in this study focus was on intention to use condoms as opposed to actual condom use, there was no need to include both PBC and self-efficacy in the models.

When targeted at condom use, the TPB proposes that the intention to use condoms, pooled with PBC will predict the probability of a person's utilisation of condoms. Condom use intention, in turn, is influenced by attitudes toward condom use, subjective norms and perception of control over this behaviour. Generally, the more positive the attitudes and subjective norms, and the greater the perceived behavioural control, the stronger the person's intention to perform the behaviour will be. Given an adequate degree of actual regulation of the behaviour, people are likely to execute their intentions when the opportunity arises. Intention is therefore presumed to be a direct antecedent of behaviour.

3.3 Research Paradigm

All research requires an underpinning for its inquiry, which is provided by worldviews and scientific paradigms. A paradigm comprises of the following: ontology, epistemology, methodology and methods. According to Crotty (1998:10), "Ontology is the study of being." Creswell (2014), Ngulube et al. (2015) and Scotland (2012) concurred as they regarded ontology to be the part of philosophy relating to the examination of the nature of being. Ontological assumptions relate to what forms reality. Researchers have to adopt a standpoint about their views of how things really are and how things in actuality work.

Epistemology has to do with the nature and forms of knowledge (Cohen et al., 2007). Epistemological assumptions relate to how knowledge can be formed, attained and transferred. Crotty (1998:3) defined methodology as "the strategy or plan of action which lies behind the choice and use of particular methods." Consequently, methodology is focussed on answering why, what, from where, when and how data is collected and analysed. Methods, in contrast, are the detailed techniques and procedures that are employed to gather and analyse data (Crotty, 1998). Research methods can be traced back, through methodology and epistemology, to ontological position. Scotland (2012), states that, "It is impossible to engage in any form of research without committing to ontological and epistemological positions."

3.4 The Research Onion: Understanding the Research Process

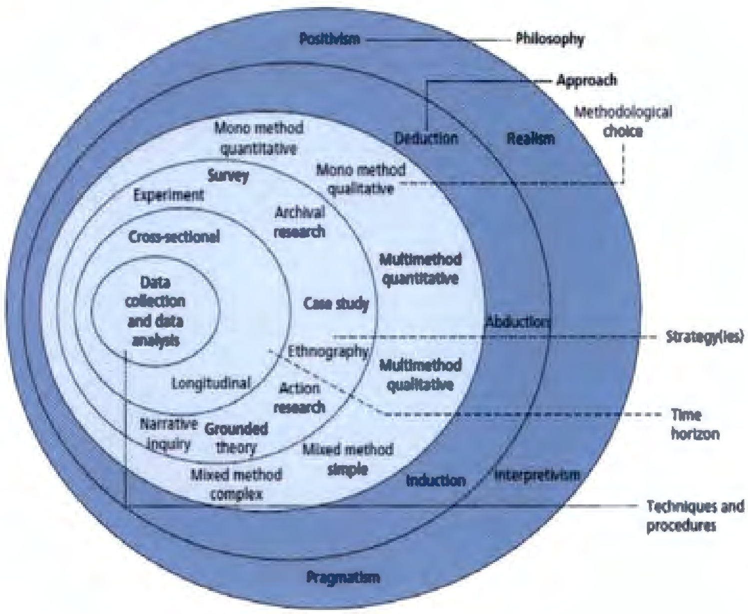


Figure 3.2 The research onion

(Source: Saunders et al., 2012)

As depicted in Fig 3.2 above, every single layer of the onion represents a phase of the research process. According to Bryman (2012), "the research onion's usefulness lies in its adaptability for almost any type of research methodology and its use in a variety of contexts." The different phases of the "onion" are examined and concepts at each phase are explained in the ensuing sections.

3.4.1 Research Philosophy

The research philosophy identification is located within the outer layer of the "research onion". According to Bajpai (2011), "research philosophy deals with the source, nature and development of knowledge." The assumptions generated by a research philosophy offer the rationalisation for how the research will be conducted (Flick, 2011). The adoption of a specific research philosophy has a strong bearing on the approach, methodology as well as the research strategy, data collection methods and statistical techniques to be adopted. This study is based on the positivism philosophy. Carson et al. (2001), stated that statistical and mathematical techniques are central to positivist research. Given that this is a statistical study, the positivism philosophy was fitting.

Positivism as a philosophy sticks to the opinion that no more than factual knowledge acquired through observation (the senses), as well as measurement, is reliable. In positivism research, the task of the researcher is restricted to data collection and interpretation by way of an unbiased method and the research results are generally discernible and can be enumerated. Scotland (2012:10), concurred and stated that in positivism, "the researcher and the researched are independent entities." The ontological view of positivism is one of realism. Realism is the view that objects have existence independent of the knower (Cohen et al., 2007:7). The positivist epistemology is one of objectivism thus positivist researchers venture out into the world neutrally, discovering absolute knowledge about an objective reality.

3.4.2 Research Approach

The second layer of the "research onion" is research approach. Academic researchers use two common approaches: inductive and deductive research. In the inductive approach, the theory is a sequel to data collection and analysis (Saunders et al., 2016). The inductive approach operationalises data collection and subsequent analysis without an apriori theoretical framework.Because of this stance in the execution of research, the research focus can therefore be formed after the data has been collected (Flick, 2011). The deductive approach, on the other hand, foregrounds theory, philosophy and orientation in order to operationalise the research. Based on this stance, the scientific and deductive approach formulates the hypothesis that is generated from pre-existing theory.The hypothesis is tested for its fit or otherwise to pre-existing theory. Testing the hypotheses therefore becomes the de facto reason for all such deductive research (Saunders et al., 2016; Silverman, 2013). Positivism often encompasses the application of existing theory to develop hypotheses or models to be tested in the research process. In this study the TPB, discussed in section 3.2, was applied. The study thus followed a deductive approach.

3.4.3 Research Design

The next three layers of the "onion" focus on the process of research design. The first methodological choice that faces a researcher is choosing whether to pursue a quantitative, qualitative or mixed methods research design (Saunders et al., 2016). Quantitative research is directed at examining and explaining associations between variables, which are measured numerically and analysed by means of a variety of statistical and graphical procedures. While this research design is informed by a positivist philosophy, it can be used to investigate wide-ranging social phenomena, including attitudes and subjective viewpoints. Saunders et al. (2016:166) stated that, "Quantitative research is usually associated with a deductive approach, where the focus is on using data to test theory." Qualitative research, in contrast, is customarily used for exploring the meaning of social phenomena, instead of pursuing causal relationships between confirmed variables (Feilzer, 2010). According to Denzin & Lincoln (2011), qualitative research is often associated with an interpretive philosophy. Included in the methodological choice is the option of using one (mono) research approach, mixed method (usually both a qualitative and quantitative design) or multi-method. The current study used the quantitative research design.

3.4.4 Research Strategy

The research strategy is a plan: it regulates and anticipates what and how the researcher executes the study, always seeking to provide systematic answers to a specific research question. Methodologically, the strategy links philosophy and data collection and analysis methods (Denzin & Lincoln, 2011). Possible research strategies that the researcher can implement include: narrative inquiry, grounded theory, action research, ethnography, case study, archival and documentary research, survey and experiment. The adolescent study, which generated the data used in this study, was an experiment in which participants completed a survey via audio/computer assisted self-interview (ACASI) technology.

3.4.5 Time Horizons

Two forms of time horizons are indicated within the "research onion": the cross sectional and the longitudinal (Bryman, 2012). The cross sectional time horizon is one already established, whereby the data must be collected. A longitudinal time horizon for data collection refers to the collection of data repetitively over protracted period, and is utilised where an important element for the research is investigating change over time (Goddard & Melville, 2004). The school based project, which was a source for the data used in this study, was a longitudinal randomised control trial designed to assess the efficacy of an intervention to promote abstinence and consistent condom use among Batswana in-school adolescents.

Data was collected at baseline (prior to the intervention) and on completion of the intervention. Data for the same participants was also collected at the 3, 6 and 12-month follow up periods. However, in order to avoid repeating the analysis carried out in the project, this study only used the school-based baseline data which had about 800 participants from Gaborone schools and its surrounding areas. According to Teye-Kwadjo et al. (2017a), increasing cross-sectional research indicates that the TPB is robust in predicting intentions to use condoms and condom use behaviour. Thus the cross-sectional time horizon used for this study was suitable.

3.4.6 Data Collection and Analysis

Secondary data from the University of Botswana Adolescent Research Project (UBARP) school based project was used for this study. The project was an NIH (GRANT R24HD056693-05) sponsored collaboration project between the University of Botswana and University of Pennsylvania. It was conducted at the University of Botswana from 2008 to 2012. The UBARP was a 3-in-1 project targeted at school based adolescents, church based adolescents as well as adolescents living with HIV/AIDS (ALWHA) respectively. Secondary targeted outcomes were to limit the number of partners as well as encouraging adolescents to go for HIV testing. The current study also limited the target outcome to condom use intention. Since the data used in this study relates to humans, some ethical considerations had to be undertaken.

Ethics in research are a worldwide set of principles regulating the way in which any research encompassing interaction between the researcher and other human beings or human tissue or data relating to humans is conducted. Ethical considerations in research are essential. For this study, since secondary data was used, permission to use the data was sought in writing from the University of Botswana and was duly granted. Clearance was also sought and given by the ethics committee within the Faculty of Economic and Management Sciences at the North West University.

Variables that were relevant to this study are discussed in section 3.5 below.

3.5 Measures

Manifest variables indicating attitude (affective and instrumental attitude), subjective (injective) norms, perceived controllability and self-efficacy are examined in the condom use intention model. The variables together with their indicators as well as description are displayed in Table 3.1.

Table 3.1 Candidate condom use intention variables for Confirmatory Factor Anaysis (CFA) and SEM analysis

Construct	Item	Mean	Std. Deviation
Norms	NO1	4.19	1.06
	NO2	4.08	1.24
	NO3	4.03	1.26
	NO4	4.03	1.13
Aff_Att	AA1	1.75	1.10
	AA2	2.34	1.25
	AA3	1.65	0.85
	AA4	1.98	1.05
	rAA5[1]	2.51	1.19
	rAA6	2.89	1.13
Instr_Att	IA1	4.34	1.02
	IA2	4.39	0.94
	IA3	4.44	0.86
Perceived_control	PC1	3.97	1.10
	PC2	3.84	1.18
	PC3	4.26	0.90
	PC4	4.27	0.86
	PC5	4.09	1.00
	PC6	4.07	0.96
	PC7	3.53	1.18
	PC8	4.34	0.96
	PC9	4.32	0.84
	PC10	3.35	1.16
	PC11	2.55	1.36
CdmUse_Intention	CUI1	4.40	0.89
	CUI2	4.42	0.85

3.6 Assumptions

Statistical methods are centred on different assumptions that support the methods. It is for the reason that researchers, for example, Gallagher et al. (2008) recommended checking of the data prior to linking it to the Analysis Moment Structure (AMOS) model. They argue that time spent checking the SEM assumptions "will avoid encountering problems at a later stage" (Gallagher et al. 2008: 261). The following assumptions need to be examined:

- Normality of the data
- Outliers
- Multicollinearity
- Heteroscedasticity
- Missing data
- Sample size

Once these assumptions are violated, interpretations and inferences based on the models are unreliable. The foregoing assumptions are discussed in more detail in the sections that follow below.

3.6.1 Multivariate Normality

The CB-SEM ML approach, like many other multivariate statistical techniques, requires data to be multivariate normal (Astrachan, 2014). The normality of the data, which is a fundamental assumption for making justifiable inferences, can be tested by means of several statistical tests or visual inspection (Ramzan et al., 2013). Multivariate non-normality can often be detected through an inspection of outliers. Violating this assumption may result in problems since non-normality affects the accuracy of statistical tests.

In SEM specifically, "non-normality can result in an underestimation of overall model fit, downwardly biased parameter estimates and underestimated standard errors" (Buhi et al., 2007:79).

Since testing whether the assumptions for multivariate normality are met is impractical (Weston & Gore, 2006), scholars suggest examining the distribution of each observed variable for skewness and kurtosis. Zainudin (2012) suggested that the examination of normality in SEM should be done assessing the fit of the measurement model prior to going on to modelling the structural model. Statistical tests and graphical approaches were both applied in the study since their combination enhance judgement on the normality of the data (Ramzan et al., 2013).

3.6.1.1 Mardia's skewness and kurtosis tests

Skewness and kurtosis have for a long time been used in revealing non-normality of univariate data (Pearson, 1930). Non-normality, can generate poor approximations in addition to unacceptably incorrect standard errors and hypothesis test results. As a result, the assumption of normality is much more crucial than in models with non-stochastic exogenous variables. For multivariate normality, Mardia (1970) established two statistics for measuring multivariate skewness and kurtosis. The skewness statistic Mardia's measure of multivariate skewness (MS) is:

$$MS = \frac{1}{6n} \sum_{i,j=1}^{n} (Y_i^T Y_j)^3, \tag{3.1}$$

where n is the sample size, $\mathbf{Y}$ is the matrix of random variable of interest. The kurtosis statistic Mardia's measure of multivariate kurtosis MK is:

$$MK = \sqrt{\frac{n}{8p(p+2)}} \left\{ \frac{1}{n} \sum_{i=1}^{n} ||Y_i||^4 - \frac{p(p+2)(n-1)}{n+1} \right\}, \tag{3.2}$$

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

The null hypothesis of normality is rejected by the skewness test, if MS is excessively large. The test founded on the centralised kurtosis statistic MK rejects the null hypothesis of normality if its absolute value |MK| is too large or when the p-value of the test surpasses the critical value (Shao & Zhou, 2014). Alternatively, Bowman and Shenton (1975) thought about merging skewness and kurtosis into the ensuing statistic:

$$MSK = MS + \|MK\|^2 \tag{3.3}$$

which exhibits an asymptotic distribution in the univariate case. Doornik & Hansen (2008), showed some applied functionality and useful power performance of MSK in the multivariate scenario. Actually, Mardia's test of multivariate kurtosis exhibited good properties for identifying multivariate outliers in a variety of circumstances (Schwager & Margolin, 1982).

AMOS 25.0, used for modelling the data in this study, provides normality checks for data including skewness, kurtosis indexes and Mardia's coefficient which is a test of multivariate normality. Normality assessment results are obtained in AMOS 25.0 once the "test for normality and outliers" output box is selected under the analysis properties. Critical ratios provided by the AMOS output as attached to kurtosis represent Mardia's normalized estimate of multivariate kurtosis. Bentler (2005), suggests that, in practice, values > 5 are indicative of data that are non-normally distributed.

A multivariate kurtosis statistic greater than the critical ratio (c.r.) value indicates that the data is not normally distributed. Additionally, the normality assessment is attained by considering the measure of skewness for each item. An absolute value of skewness equal to 1.0 or lower shows that the data is normally distributed. Skewness values larger than 1.0 in absolute value are however acceptable if the sample size is large (> 200) and the c.r.for skewness does not go above 8 (Zainudin, 2012). Since AMOS was used for analysis in the study, this procedure was used to assess for multivariate normality.

The relation of Mardia's (1970) measure to the Mahalanobis distances is as well beneficial for appreciating the measure. The Mahalanobis distance is calculated using the formula

$$d_i^2 = (x_i - \overline{x})' S^{-1} (x_i - \overline{x}) \tag{3.4}$$

where d_i^2 is the Mahalanobis distance for a given individual, x_i is i^{th} observation and S is the covariance matrix with variances on the diagonal and covariances off the diagonal.

Large values of Mardia's measure, relative to the expected value under multivariate normality, suggest the existence of one or more cases with large Mahalanobis distances. The Mahalanobis distance reveals how distant an individual case is from the centroid of all cases for the predictor variables. When the distance is large, the observation is considered an outlier. If after deletion of cases multivariate normality is still unattainable then there is need to go for Shapiro-Wilk's normality test (Shapiro & Wilk, 1965) discussed in the next section.

3.6.1.2 The Shapiro-Wilk test

The Shapiro-Wilk test for normality is designed to detect all departures from normality. The test rejects the hypothesis of normality when the p-value is less than or equal to 0.05. The Shapiro-Wilk test statistic is considered to be the ratio of two variance estimators, the best linear unbiased estimator (BLUE) and the maximum likelihood estimator (MLE) (Shao & Zhou, 2014). The Shapiro-Wilk test, which was initially designed for testing univariate normality, has the following formula:

$$W_n\left(z_1,\ldots,z_n\right)=\frac{\left\{\sum_{i=1}^{n}a_{n,i}z_{(i)}\right\}^2}{\sum_i^n\left(z_i,\ldots,\bar{z}_n\right)^2} \tag{3.5}$$

where $z_1, \ldots, z_n$ is the observed univariate data, $z, \ldots, z(n)$ are order statistics of $z1, \ldots, zn, \bar{z}_n$ is the sample mean of the univariate variable, and the constants $a_{n,i}$ are $(a_{n,1},\ldots,a_{n,n}) = \left(m^TV^{-1}V^{-1}m\right)^{-\frac{1}{2}}m^TV^{-1}$, with $m=(m_1,\ldots,m_n)^T$ and V the mean covariance of the order statistics of a random standard normal sample of size n, respectively. Browne's (1984) asymptotically distribution-free (ADF) estimation method could be used, if the number of observations is sufficiently large. Boomsma & Hoogland (2001:148) as reported by Schermelleh-Engel et al. (2003: 49) found that for a sample size of n $\leq$ 200, "ADF is a disaster with more than one-third of the solutions being improper."

3.6.1.3 Graphical Approach

Chambers et al. (1983), suggest that "there is no single statistical tool to assess normality that is as powerful as a well-chosen graph." According to Ramzan et al. (2013), a commonly used graphical technique is based on the distribution of ordered Mahalanobis distances of the individual sample points from their mean. Multivariate normality is assessed using a chi-square versus ordered Mahalanobis distance plot (Arifin, 2015). The technique follows the following three steps:

- Mahalanobis distances are sorted in ascending order.
- Quantiles associated with the upper percentiles of the chi-square distribution are calculated.
- Pairs of the quantiles and Mahalanobis distances are then plotted to obtain a scatter plot.

A multivariate normal distribution is shown by the points forming a straight line (Burdenski, 2000). This graphical approach was applied as one of the methods for assessing multivariate normality in this study.

In multivariate analyses, if variable distributions are non-normal it is often necessary to transform the variable to form a new variable with a normal distribution. Transformation of variables could however lead to complications in interpretation. In SEM if variables are non-normally distributed it is completely logical and possibly even better to decide on an estimation method that addresses the non-normality as an alternative to transforming the variable. Non-normal items from the measurement model may be deleted before continuing with the analysis. Alternatively, the farthest observation from the centre, based on the Mahalanobis distance, may be removed. According to Zainudin (2012:73), "the most popular method lately is to continue with the analysis using the ML approach and reconfirm the result of analysis through bootstrapping." The suggestion by Zainudin (2012) was carried out in this study.

3.6.2 Outliers

SEM assumes that the data should be free of outliers. Byrne (2010: 105) defines outliers as "cases whose scores are substantially different from all others in a particular set of data." Outliers affect the model significance (Garson, 2015). Multivariate outliers can be examined and detected using the squared value of Mahalanobis distance as indicated in section 3.6.1.1 above. AMOS 25.0 calculates the squared values of Mahalanobis distance and also provides information related to possible outliers (Byrne, 2010). Usually, an outlying case will have a squared Mahalanobis distance value that stands apart from all other squared values. Gallagher et al. (2008) advise that the decision of whether to delete or retain outliers should be given careful consideration as important information may be lost when excluding them.

3.6.3 Multicollinearity

Multicollinearity refers to situations where measured variables are so highly correlated such that they are in essence redundant (Weston & Gore, 2006). Since related measures are used as indicators of constructs, they suggest that there is a possibility that the measures may be too highly related for certain statistical operations to function properly. Multicollinearity could as a result lead to erroneous parameter estimates and even cause statistical non-significance of parameter estimates (Grewal et al., 2004), thus leading to wrong interpretation or removal of significant predictors from the model. A rough guideline for checking multicollinearity is to screen bivariate correlations. According to Kline (2011), bivariate correlations greater than r = 0.85 can be indicative of potential problems. Most regression software packages have a "tolerance" parameter as part of the analysis output. Low tolerance values point to stronger relationships between IVs. Myers (1990) and Stevens (2009) suggest that tolerance values in the range of 0.1 are problematic. Related to tolerance is a statistic known as Variance Inflation Factor (VIF), which is calculated using the formula:

$$VIF = \frac{1}{Tolerance} \tag{3.6}$$

According to Cohen et al. (2003), Myers (1990) and Stevens (2009), a VIF value of 10 is indicative of high correlation that may be problematic.

3.6.4 Missing Data

In the real world, missing values occur in countless data sets, in spite of the best efforts at preventing such occurrences. Missing data are ever-present throughout the social, behavioural, and medical sciences (Enders, 2010). Missing data arise due to many causes, including hardware failure, software bugs, missed appointments, and case attrition thus causing a problem for researchers using SEM techniques for data analyses. A few missing values, such as less than 5% on a solitary variable, in a huge sample might be of little concern. Enders (2010) and Zhang & Little (2009) also echoed that missing observations can be problematic in SEM and suggested that the problem should be dealt with prior to data analysis. Hair et al. (2014a) stated that missing data causes two main complications: (1) it reduces the capability of statistical tests to indicate a relationship in the data set, and (2) it produces biased parameter approximations. Missing data may therefore create problems for model estimation and hypothesis testing. The possible effects of missing data hinge on the rate of occurrence, the pattern of missing observations, and the reasons for the missing value (Tabachnick & Fidell, 2014).

The extent of missing data is directly related to the quality of statistical interpretations. There is, however, no recognised limit from the literature concerning an acceptable proportion of missing data in a data set for one to make justifiable statistical inferences. For example, Schafer (1999) states that a missing rate of 5% or less is insignificant. Savalei & Bentler (2006: 355) concur and suggest that, "missing data below 5% might be ignored by use of listwise deletion when computing correlation or covariance matrices."

Bennett (2001), on the other hand, maintains that "statistical analysis is likely to be biased when more than 10% of data are missing." Savalei (2008), showed that the maximum likelihood goodness of fit index, chi-square, declines more quickly than usually stated in the literature when the data are highly non-normal and with a high proportion of missingness. Moreover, the extent of missing data is not the solitary measure by which a researcher evaluates the missing data problem. Tabachnick & Fidell (2014) suggest that the missing data mechanisms and the missing data patterns have more bearing on research results than does the extent of missing data.

3.6.5 Missing Data Mechanisms

In order to develop more understanding in the missing data problem, the reasons for missing data are singled out in a number of missing data mechanisms. These mechanisms define the fundamental reason of missing data and were first defined by Rubin (1976). Missing data mechanisms relate to the statistical relationship between variables and the probability of missing data. Rubin (1976) and Little & Rubin (2002) differentiated three missing data mechanisms: Missing Not At Random (MNAR), Missing At Random (MAR) and Missing Completely At Random (MCAR). These missing data mechanisms are significant since they are assumptions for the missing data treatment methods. It is worth noting that the missing data mechanisms are not features of a complete data set, but the mechanisms are just assumptions which have some bearing on the performance of different missing data methods.

Data is regarded as MNAR when the existence of missing data for a specified variable is related to that variable itself even after controlling for other variables in the dataset. According to Enders (2010), the probability distribution of MNAR can be written as

$$P(R|Y_{obs}, Y_{mis}, q) \tag{3.7}$$

where Y is a matrix of the whole dataset

Y_{obs} represents the observed part of the data

Y_{mis} represents the missing part of the data

R is a missingness matrix

q is a vector of parameters describing the relationship between missingness, R and the dataset, Y.

Equation (3.6) indicates that "the probability of whether a position in R is 0 or 1 depends on both Y_{obs} and Y_{mis} and this relationship is governed by q" (Nakagawa, 2015:84). Notably, q is known as the mechanism of missing data and provides the foundation for differentiating between MNAR, MAR and MCAR.

The MAR mechanism is at work when the probability of missing data in a variable is related to some other variable(s) in the dataset. The probability distribution for MAR is expressed as:

$$P(R|Y_{obs}, q) \tag{3.8}$$

This implies that missingness depends only on Y_{obs} and the relationship is governed by q.

Lastly, when missingness does not depend on either Y_{obs} or Y_{mis} values in the dataset, the missingness is considered to be missing completely at random (MCAR). The probability distribution for

MCAR is expressed as:

$$P(R|q) \tag{3.9}$$

This implies that the probability of missingness is independent of the data. However, whether positions in R take 0 or 1 is still governed by q.

Hair et al. (2014a) suggested that if the pattern of missing data is systematic (i.e. non-ignorable or is not missing at random), any method used to treat this missing data could probably produce biased results, whereas, if the missing data is distributed in a random manner with no clear pattern (i.e. missing completely at random = MCAR), any remedy to treat this problem is expected to produce acceptable results. The different possible techniques for treating missing data are discussed below and the strategy that was applied in this study is indicated.

3.6.5.1 Listwise Deletion

Listwise deletion is a frequently used and easily executed deletion strategy (Pituch & Stevens, 2016). In this method, cases that have any missing data are deleted from the analysis. Kang (2013) and Pituch & Stevens (2016) concur that listwise deletion is the default option for handling missing data in most software programs. A primary limitation of listwise deletion is that it usually requires that the data are MCAR and may severely decrease statistical power if many cases have missing data on one or more variables. With a proportion of 5% or less missing data, however, Roth (1994) states that "listwise data deletion is a defensible strategy for handling the incomplete data problem." When the sample size is greater than 250 and the proportion of missing data for the analysed variables is less than 10%, Gallagher et al. (2008) recommend the implementation of listwise deletion. Kang (2013) suggests that, "if there is a large enough sample, where power is not an issue, and the assumption of MCAR is satisfied, the listwise deletion may be a reasonable strategy." This study applied steps suggested by Byrne (2001) for dealing with missing data, which are: (1)Investigation of the total amount of missing data, (2) Investigation of the pattern of missing data, and (3) applying an appropriate technique to deal with missing data.

3.6.6 Sample Size

Sample size considerations are very important before running a SEM analysis. While large samples result in less sampling error compared to small samples, some difference of opinion exists with regard to recommended sample sizes for SEMs (Bagozzi & Yi, 2012). While Loehlin (2004) recommends that a model with 2-4 latent factors needs at least 100 responses, Bentler & Chou (1987) recommended that there should be 5 responses per estimated parameter. Kline (2011) suggested that a larger sample size of more than 200 is more appropriate for SEM. According to Gallagher et al. (2008) "prevailing agreement has long been that SEM requires a large sample size."

According to Meyers et al. (2013), one rule of thumb is that "sample size should be at least 50 more than 8 times the number of variables in the model." Stevens (2009) on the other hand suggested that there should be at least 15 cases per measured variable or indicator.

Hair et al. (2014a) took into consideration influencing factors such as the number of latent variables, lowest number of indicators in a latent variable and communalities in determining the appropriate sample size. Ullman (2014), however, suggested that models with strong expected parameter estimates, reliable measured variables and well-defined constructs may require less data. The required sample size is dependent on the data quality, complexity of the model and the estimation method that is applied. Generally, larger samples are required for non-normal data.

This study used a sample of 794 participants. With 20 observed variables considered as possible model candidates, the resultant cases-to-observed variables ratio, found by dividing 794 by 20, was approximately 40:1. This ratio is much higher than Stevens' (2009) suggestion of at least 15 cases per observed variable or indicator. Hair et al. (2014a: 574) recommended a minimum sample size of 300 for models containing seven or fewer constructs with lower communalities (<0.45). The sample size used in this research was therefore adequate.

3.7 Steps of Structural Equation Modelling

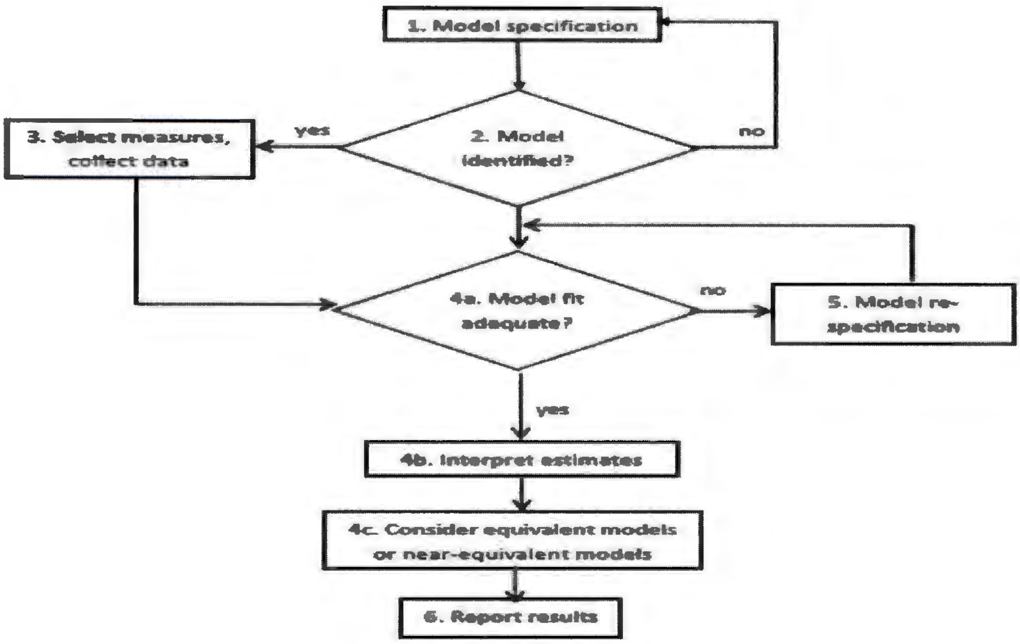


Figure 3.3 Steps of SEM

(Source: Kline, 2011:92)

Fig 3.3 displays the steps that are followed in conducting structural equation modelling. As depicted in Fig 3.3, the application of SEM involves an iterative process consisting of five consecutive steps: model specification, model identification, model estimation, model fitting, and model modification (Bollen & Long, 1993; Schumacker & Lomax, 2016). The steps are discussed in detail in the sections that follow.

3.7.1 Model Specification

Specification of a model is the first and most important step in SEM analysis. A combination of theory and empirical research results usually guides the model specification (Hox & Bechger, 1998). In this step, the model and hypotheses to be tested are presented most often through a diagram.

Decisions pertaining to the variables to be included in the model along with how these variables are related are undertaken in this step. If a variable or parameter is erroneously omitted or included in a model, a mis-specified model can result. Mis-specified models in turn result in biased parameter estimates and may not fit the data according to global fit indices. The proposed models for this study are specified in sections 3.7.2 and 3.7.3.

There is no consensus among researchers on how SEMs should be represented. Bentler(2010), for instance,would rather use equations and simple diagrams while others such as Iacobucci (2009), would rather use matrix algebra and diagrams plus Greek letters. The preference is essentially a matter of taste and a researcher's earlier experience with one method or another. Although the first is less daunting than the latter, it is recommended that one becomes acquainted with both the utilisation of equations and matrices and Greek letters. SEMs are most often represented graphically. Path diagrams generally are integral to model specification because they allow the researcher to clearly show the hypothesised set of relationships in the model. These diagrams clarify a researcher's ideas about the relationships among variables and are directly translatable into the equations deployed consequently for the analysis.

Whilst a number of conventions are utilised in creating SEM path diagrams, observed or manifest variables are commonly represented using rectangular shapes while latent variables or constructs are signified by circular or elliptical shapes. Residuals are also represented by circular or elliptical shapes. Arrows are used to depict hypothesised relationships between variables. Straight arrows represent direct effects between two variables. Curved arrows or two-headed arrows represent correlation between variables. A complete SEM model comprises of measurement and structural models. A model ought to be developed centred on certain fundamental hypothesis. In this study, model specification was done graphically (**See Fig 3.4 and Fig 3.5**) as well as using equations (**See Equations 3.10 – 3.35**)

3.7.2 Measurement Model

According to Khine et al. (2013), the measurement model links observed responses or 'indicators' to latent variables and in some cases to observed covariates (i.e., the Confirmatory Factor Analysis (CFA) model). Garson (2015) concurred that the measurement model step is called CFA model since its purpose is to validate (confirm) the way the researcher has measured the latent variables in the model. Using the graphical convention discussed in section 3.6, the proposed measurement model for this study is shown in Fig 3.4.

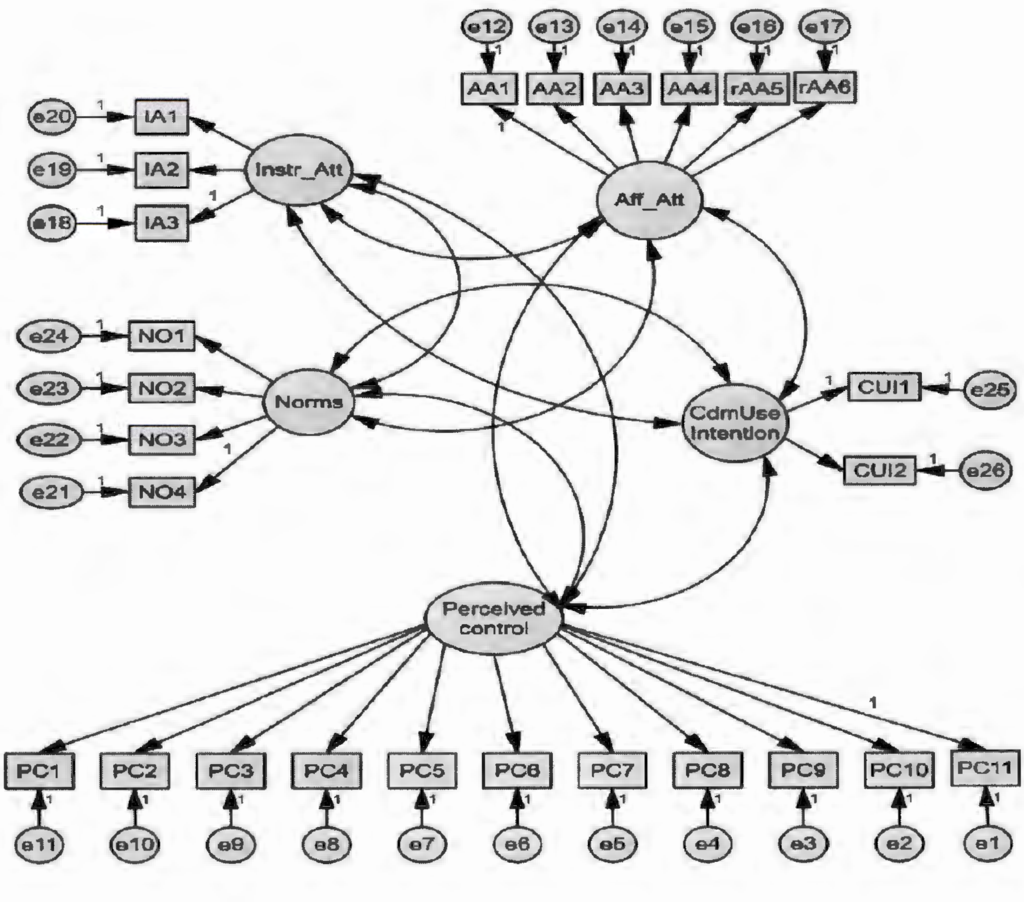


Figure 3.4 Hypothesised measurement model

Letting ξ_1= Perceived_control, ξ_2= Aff_Att, ξ_3=Instr_Att, ξ_4= Norms and η = CdmUse_Intention and following equations (2.7) and (2.8) given in section 2.5.1, the hypothesised measurement model can be expressed as Equations (3.10) – (3.35):

$$PC11 = \xi_1 + \delta_1 \tag{3.10}$$

$$PC10 = \lambda_2\xi_1 + \delta_2 \tag{3.11}$$

$$PC9 = \lambda_3\xi_1 + \delta_3 \tag{3.12}$$

$$PC8 = \lambda_4\xi_1 + \delta_4 \tag{3.13}$$

$$PC7 = \lambda_5\xi_1 + \delta_5 \tag{3.14}$$

$$PC6 = \lambda_6\xi_1 + \delta_6 \tag{3.15}$$

$$PC5 = \lambda_7\xi_1 + \delta_7 \tag{3.16}$$

$$PC4 = \lambda_8\xi_1 + \delta_8 \tag{3.17}$$

$$PC3 = \lambda_9\xi_1 + \delta_9 \tag{3.18}$$

$$PC2 = \lambda_{10}\xi_1 + \delta_{10} \tag{3.19}$$

$$PC1 = \lambda_{11}\xi_1 + \delta_{11} \tag{3.20}$$

$$AA1 = \xi_2 + \delta_{12} \tag{3.21}$$

$$AA2 = \lambda_{13}\xi_2 + \delta_{13} \tag{3.22}$$

$$AA3 = \lambda_{14}\xi_2 + \delta_{14} \tag{3.23}$$

$$AA4 = \lambda_{15}\xi_2 + \delta_{15} \tag{3.24}$$

$$rAA5 = \lambda_{16}\xi_2 + \delta_{16} \tag{3.25}$$

$$rAA6 = \lambda_{17}\xi_2 + \delta_{17} \tag{3.26}$$

$$IA3 = \xi_3 + \delta_{18} \tag{3.27}$$

$$IA2 = \lambda_{19}\xi_3 + \delta_{19} \tag{3.28}$$

$$IA1 = \lambda_{20}\xi_3 + \delta_{20} \tag{3.29}$$

$$NO4 = \xi_4 + \delta_{21} \tag{3.30}$$

$$NO3 = \lambda_{22}\xi_4 + \delta_{22} \tag{3.31}$$

$$NO2 = \lambda_{23}\xi_4 + \delta_{23} \tag{3.32}$$

$$NO1 = \lambda_{24}\xi_4 + \delta_{24} \tag{3.33}$$

$$CUI1 = \eta + \varepsilon_1 \tag{3.34}$$

$$CUI2 = \lambda_{26}\eta + \varepsilon_1 \tag{3.35}$$

where λ_i is the loading of the observed variable on the common factor ξ_j for j = 1,2,3,4
δ_i is the residual error
The CFA (measurement) model permits for inferential testing of two classes: the significance of each of the factor loadings, and global fit of the model (Iacobucci, 2009). Garson (2015) further stated that CFA establishes convergent and divergent validity in the proposed model. A good SEM analysis fit demonstrates that the indicator variables reflect the latent variables which they represent and that the latent variables differ from each other.
CFA was performed by co-varying all constructs contained in the model. The validity of each construct was then assessed through assessment of model fit and construct validity and reliability. To ensure adequacy of a measurement model, Carmack & Lewis-Moss (2009) and Hair et al. (2014a), recommended inclusion of items in the model that are consistent with theory and having factor loadings of 0.6. Field (2013), then again, advocated the recommendation of Guadagnoli & Velicer (1988) to consider a factor as dependable if it has at least four loadings of at least 0.6 regardless of sample size. Stevens (2009) suggested utilising a cut-off of 0.4, notwithstanding the sample size, for explanatory objectives. In instances where items do not have the same frequency distributions Tabachnick & Fidell (2014) followed Comrey & Lee (1992) in proposing more rigorous cut-offs ranging from 0.32 (*poor*), 0.45 (*fair*), 0.55 (*good*), 0.63 (*very good*) to 0.71 (*excellent*). The latter proposal is adapted for this study. A cut-off of 0.5 is therefore applied in this study. The proposal is in line with the unidimensionality procedure (Zainudin, 2012) which could be applied to the measurement model in order to remove indicator items that have lower standardised factor loadings (< 0.5). As indicated by Zainudin (2012), unidimensionality is attained when the factor

loading of items are at least 0.50 for newly developed scales and 0.60 for proven scales. Once the CFA upholds the measurement model, testing of the structural model can then be carried out.

3.7.3 Structural Model

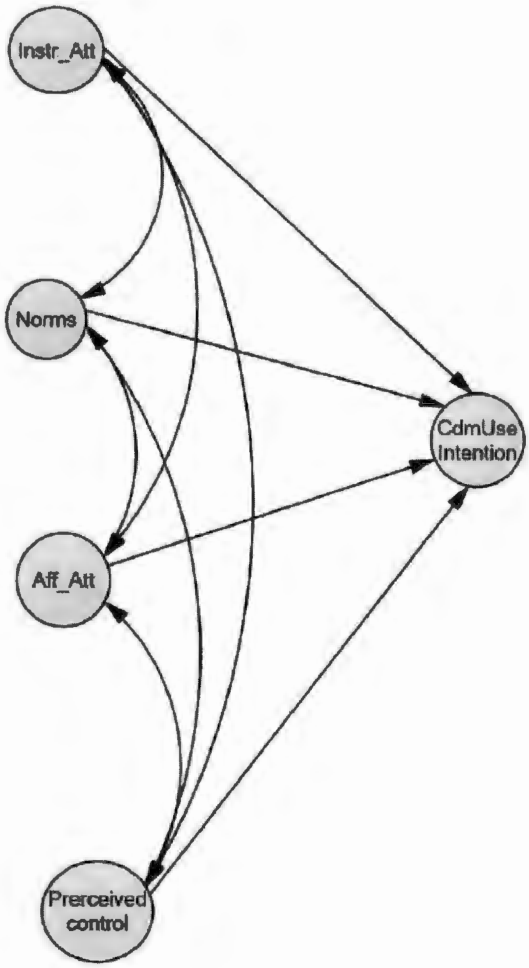


Figure 3.5 Hypothesised structural model

Fig 3.5 shows the proposed structural model for this study. This structural model involves stating structural relationships between latent constructs which can be linked to measured variables with a reliance relationship. Structural models contrast from measurement models in that the prominence changes from the association between latent constructs and their measured variables to the type and extent of the association between constructs (Hair et al., 2014a). Two kinds of connections are conceivable among constructs. The first is a reliance relationship, which is constantly represented by a straight arrow and used between an exogenous construct and an endogenous construct. The second is a correlation relationship, which is delineated by a two-headed arrow connection, which can only be shared between exogenous constructs.

Following Equation (2.6) from section 2.5.1, the hypothesised structural model depicted in Fig 3.5 can be expressed in equation form as:

$$\eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_3 + \gamma_4\xi_4 + \zeta_1 \tag{3.36}$$

3.7.4 Model Identification

According to Kelloway (1998) and Schumacker & Lomax (2016), a model may be considered identified, if it is theoretically possible to establish a unique estimate for each parameter. Every potential parameter in SEM is either free (unknown and hence it must be estimated), fixed (not free, typically fixed at either 0 or 1) or constrained (unknown but constrained to some other free parameter). Model identification is carried out after the model has been specified. According to Schumacker & Lomax (2016), the question of interest under model identification is:

"On the basis of the sample data obtained in the sample variance-covariance matrix S and the theoretical model implied by the population variance-covariance matrix Σ, can a unique set of parameters be found?"

For a model to be identified, every parameter must be identified. The first step in the identification process involves counting the number of data points and the number of parameters that are to be estimated (Tabachnick & Fidell, 2014). The number of data points is the number of sample variances and covariances found by applying the formula:

$$Number\ \ of\ \ data\ \ points = \frac{p(p+1)}{2} \tag{3.37}$$

where p is the number of measured variables.

There are three basic levels of identification and these can be assessed by considering the models' degrees of freedom (df). The degrees of freedom are calculated using the equation:

$$\nu = \frac{p(p+1)}{2} - k \tag{3.38}$$

where k is the number of parameters to be estimated. Models are classified as under-identified, just-identified or over-identified. An under-identified model is also referred to as not identified model. For this model, there is not enough information (i.e. there are fewer data points than parameters to be estimated) to uniquely estimate every parameter in the model; $df < 0$. The number of parameters may be reduced by fixing, constraining, or deleting some of them. A just-identified model on the other hand provides just enough information in the data (i.e. the same number of data points as parameters to be estimated) to estimate uniquely every parameter; df = 0. Hypotheses about the model adequacy cannot be tested while hypotheses about specific paths in the model can be tested. The model cannot however be modified. Lastly, an over-identified model has more than enough information (i.e. more data points than parameters to be estimated) to uniquely estimate every parameter (df > 0) and this is the preferred model. This type of model allows for modification.

3.7.5 Model Estimation

Once a model has been specified, factor loadings and covariances can be estimated. Model estimation consists of approximating the parameters of the hypothetical model with the end goal that the theoretical parameter estimates produce a covariance matrix that is as close as conceivable to the observed covariance matrix. In the estimation process, an estimation technique is utilised to find approximations of the parameters in θ to minimise the difference between S, the observed covariance matrix, and $\Sigma(\theta)$. SEM differs from other multivariate methods in that it is a covariance structure analysis method rather than a variance analysis method (Hair et al., 2010). Apart from the Maximum Likelihood Estimation (MLE), the typically referred SEM technique (Hair et al., 2010), other estimation techniques that are applicable in SEM are described in literature, including unweighted least squares (ULS), weighted least squares (WLS), generalised least squares (GLS) and asymptotic distribution free (ADF) methods. The two estimation methods utilised in this study are discussed below.

3.7.5.1 Maximum Likelihood Estimation

The MLE is a common method for estimating model parameters. Hair et al (2010: 663) indicated that "MLE is preferred over alternative methods of estimation such as generalised least squares (GLS), unweighted least squares (ULS), ADF, etc." MLE uses derivatives to minimise the fit function:

$$F_{ML} = \log|\Sigma(\theta)| \; + tr\left(S\Sigma^{-1}(\theta)\right) - \log|S| \; - p \tag{3.39}$$

where F_{ML} is the ML fit function.

The probability of observing S as a sample of the population with Σ is given by Wishart distribution:

$$p\left(S \mid \theta\right) = W\left(S, \Sigma\left(\theta\right), n\right) = \frac{e^{-\frac{n}{2}tr(S\Sigma(\theta)^{-1})}}{|\Sigma(\theta)|^{n/2}}C \tag{3.40}$$

The probability of observing perfect fit, i.e. $S = \Sigma$ is given by

$$p\left(S \mid \right) = W\left(S, S, n\right) = \frac{e^{-\frac{n}{2}tr(SS^{-1})}}{|S|^{n/2}}C \tag{3.41}$$

The Likelihood Ratio (LR) is

$$LR = \frac{W\left(S, \Sigma\left(\theta\right), n\right)}{W\left(S, S, n\right)} \tag{3.42}$$

$$= \frac{e^{-\frac{n}{2}tr(S\Sigma(\theta)^{-1})}}{|\Sigma(\theta)|^{n/2}}C.\frac{|S|^{n/2}}{e^{-\frac{n}{2}tr(SS^{-1})}C} \tag{3.43}$$

$$= \; e^{-\frac{n}{2}tr(S\Sigma(\theta)^{-1})}|\Sigma\left(\theta\right)|^{-n/2}e^{\frac{n}{2}}tr\left(SS^{-1}\right)|S|^{-n/2} \tag{3.44}$$

Taking the log gives the Log-Likelihood Ratio (LLR):

$$LLR = -\frac{n}{2}tr\left(S\Sigma(\theta)^{-1}\right) - \frac{n}{2}log\left|\Sigma(\theta)\right| + \frac{n}{2}tr\left(SS^{-1}\right) + \frac{n}{2}log\left|S\right| \tag{3.45}$$

Since $tr\left(SS^{-1}\right) = tr\left(I\right) = p$, the number of observed variables, we can write

$$LLR = -\frac{n}{2}tr\left(S\Sigma(\theta)^{-1}\right) - \frac{n}{2}log\left|\Sigma(\theta)\right| + \frac{n}{2}p + \frac{n}{2}log\left|S\right| \tag{3.46}$$

Multiplying by $-\frac{2}{n}$ gives the ML fit function:

$$F_{ML} = -\frac{2}{n}LLR = tr\left(S\Sigma(\theta)^{-1}\right) + log\left|\Sigma(\theta)\right| - log\left|S\right| - p \tag{3.47}$$

The ML estimator assumes that the variables in the model are conditionally multivariate normal. A shortcoming of MLE is the strong assumption of multivariate normality, since contraventions of distributional assumptions are widespread and often inevitable in practice and can possibly give rise to extremely inaccurate outcomes (Schermelleh-Engel et al., 2003). On the other hand, MLE appears to be essentially robust against violation of the normality assumption (Boomsma & Hoogland, 2001; Chou & Bentler, 1995; Curran, West & Finch, 1996; Muthen & Muthen, 2002). Furthermore, Schermelleh-Engel et al. (2003:26) stated that

"simulation studies suggest that under conditions of severe non-normality, ML parameter estimates are still consistent but not necessarily efficient."

3.7.5.2 Unweighted Least Squares (ULS)

The ULS fitting function is comparable to ordinary least squares (OLS) estimation in regression. The function is represented by the formula:

$$F_{ULS} = \left(\frac{1}{2}\right) tr\left[\left(S - \Sigma(\theta)\right)^2\right] \tag{3.48}$$

The ULS fitting function differs from the other functions in that it is not built on an assumption of multivariate normality in the data. Bollen (1989a: 112) concurred that "ULS is a consistent estimator, and it makes no distributional assumption about the observed variables." In the event that the normality assumption is not met, the ULS approach will be applied. Results obtained using this approach will be compared with those obtained from application of the MLE approach. Choice of the estimation method to be used is dependent on whether the data are normally distributed or not. For instance, the ULS estimates have no distributional suppositions but require that the scale of all the observed variables be the same to guarantee consistent estimates. The ML and GLS methods, on the other hand, presuppose multivariate normality while they are not scale dependent. In instances where the normality supposition is violated, Yuan Bentler (1998) recommend the utilisation of an ADF method such as the WLS estimator or the ULS approach which does not presuppose normality. Research has demonstrated that MLE produces efficient and reliable estimates hence MLE was applied as the initial estimation method in this research study.

3.7.6 Model Evaluation

After confirming the fit of the measurement model, the structural model can then be analysed to decide the degree to which the model is upheld by the sample data. Thus model evaluation is about assessing how well the hypothesised model fit the data and is performed at two levels: global and parameter levels. Global fit indices are used to assess fit of an entire model. The global fit indices evaluate the degree of discrepancy between the theoretical covariance matrix Σ and the sample covariance matrix S. There is no single best fit statistic thus it is prudent to report multiple fit indices.

It is worth noting that there is extensive disagreement concerning fit indices. Some researchers (e.g. Barrett, 2007) do not think that indices add value to the analysis and prefer that only the chi square should be interpreted. Others (e.g. Hayduk et al., 2007) contended that cut-offs for a fit index can be deceptive and subject to abuse. Most analysts, however, believe in the importance of fit indices, but warn against strict dependence on cut-offs. Table 3.1 shows some commonly used fit indices and cut-offs.

Table 3. 2 Commonly Used Model Fit Indices in SEM

Category	**Index name**	**Fit Criteria**	**Literature**
Factor Loading	Standardized Regression Weight	weight 0.5	Hair *et al.* (2010)
Absolute Fit	Chi-square (χ^2)	p > 0.05	Wheaton *et al* (1977); Bollen (1989b); Kaplan (2009); Kline (2011)
	Root mean square error of approximation (RMSEA)	Close fit $\leq$ 0.05 Reasonable fit : 0.05-0.08 Poor fit: $\geq$ 0.10	Browne & Cudeck (1993); Steiger & Lind (1980); Steiger (1990)
	Goodness-of-fit-index (GFI)	$\geq$ 0.9	Jöreskog & Sorbom (1984); Hoyle & Panter (1995)
Incremental Fit	Comparative fit index (CFI)	Great > 0.95 Traditional: 0.9 Sometimes acceptable: 0.8	Bentler (1990); Hu & Bentler (1999)
	Tucker-Lewis index (TLI)	> 0.9	Bentler & Bonnet (1980); Hu & Bentler (1999)
	Normed fit index (NFI)	> 0.9	Bentler (1990)
Parsimonious fit	Chi-square/df	< 5.0	Marsh & Hocevar (1985)
	Adjusted GFI (AGFI)	> 0.9	Jöreskog & Sorbom (1984); Tanaka & Huba (1985)

Sources: Bowen & Guo (2011); Hair et al. (2010); Kline (2011); Pituch & Stevens (2016)

Over the past 25 years, at least 24 fit indices have been proposed (Klem, 2000). In addition to the controversy about fit indices referred to in the foregoing paragraph, there is lack of agreement among SEM researchers regarding classification of these indices. There is also disagreement about which individual fit measures might best be classified together.

Arbuckle (2010) devised an eight-category scheme (parsimony, sample discrepancy, population discrepancy, information theoretic, baseline model, parsimony adjusted, goodness of fit and miscellaneous), whereas Tabachnik & Fidell (2014) suggested a five-category system (comparative, absolute, proportion of variance, parsimony and residual based). Hair et al. (2014a) and Jaccard & Wan (1996) on the other hand suggested a triple group scheme (absolute, relative and parsimonious). The most-cited organisation system appears to be the triple classification scheme of absolute, relative and parsimonious fit indices (Meyers et al., 2013). They added model comparison indices AIC (Akaike Information Criterion), BCC (Browne-Cudeck Criterion), BIC (Bayesian Information Criterion) and ECVI (Expected cross-validation index) to the scheme. Absolute, relative and parsimonious fit indices together with model comparison indices were considered and are reported for this study. Specifically, χ^2, χ^2 /df, GFI, RMSEA, NFI, CFI and AGFI are reported as fit indices in this study.

Absolute fit indices address the question: Is the residual or unexplained variance after model fitting appreciable? The most basic assessment of how well a researcher's theory fits the sample data is thus provided by absolute fit indices. Included under absolute fit indices are the χ^2, GFI, SRMR and RMSEA. Relative fit indices (McDonald & Ho, 2002) or comparative indices (Miles & Shevlin, 2007) on the other hand address the question: How well does a particular model explain a set of observed data compared to other possible models? Under relative or comparative fit indices, the NFI and CFI are reported in this study. Parsimony fit indices capture the goodness of fit of a proposed model while adjusting the number of parameters to be estimated. The AGFI was considered under the parsimony fit indices. Lastly, model comparison indices are used to compare different models and are applicable when MLE is used (Anderson, Burnham & White, 1998). The

AIC, BCC and ECVI will be considered under the model comparison indices. The various indices that were utilised for model evaluation in this study, are discussed in more detail below.

3.7.6.1 Chi-square

The χ^2 statistic is the most important absolute fit index. Historically, the χ^2 was used to evaluate model fit with a non-significant χ^2 value preferred. The chi-square test tests the hypothesis

$H_0 : \Sigma = \widehat{\Sigma}$ versus

$H_1 : \Sigma \neq \widehat{\Sigma}$

with the test statistic $X^2 = (N-1)\left\{\ln|\Sigma| + tr\left(S\Sigma^{-1}\right) - \ln|S| - p\right\} \sim \chi^2_{\nu}$ where S is the sample covariance matrix and Σ is the predicted covariance matrix. Low ^{2} values relative to the degrees of freedom with an insignificant p value ($p > 0.05$) would support the model as representative of the data. For models with large samples (which is typically the case in SEM), χ^2 will usually be statistically significant. Due to its sensitivity to sample size, the χ^2 statistic nearly always rejects the model when large samples are used (Bentler & Bonnet, 1980; Joreskog and Sorbom, 1993) thus Wheaton et al.'s (1977) relative chi-square (χ^2/ν) is sometimes preferred. Wheaton et al (1977) and Tabachnick & Fidell (2014) recommend that $2 \leq \frac{\chi^2_{\nu}}{\nu} \leq 5$. Joreskog & Sorbom (1996) and Bentler (1990), advised against the sole use of the χ^2 value in assessing the model's overall fit due to its sensitivity to sample size.

3.7.6.2 Goodness of Fit Index

Kline (2011:207) defined the GFI as "an absolute fit index that estimates the proportion of covariance in the sample data matrix explained by the model". The GFI thus estimates how much better the researcher's model fits compared with no model at all (Joreskog, 2004). Analogous to R^2 in regression models, the GFI calculates the proportion of variance accounted for by the estimated population covariance (Tabachnick & Fidell, 2014). It is calculated using the formula:

$$GFI = 1 - \frac{tr\left[(\Sigma^{-1}S - I)^2\right]}{tr\left[(\Sigma^{-1}S)^2\right]} \tag{3.49}$$

According to Garson (2015), there are some problems associated with the GFI. While its values are expected to lie between 0 and 1, the GFI can at times yield meaningless negative values. Additionally the GFI can be large even for poorly specified models and is pushed up by large samples.

3.7.6.3 Root Mean Square Error of Approximation

The RMSEA indicates how well a model with unknown but optimally chosen parameter estimates would fit the population covariance matrix (Byrne, 1998). The following formula is used to calculate RMSEA:

$$RMSEA = \sqrt{\frac{|\chi^2_\nu - \nu|}{N-1}} \tag{3.50}$$

Kenny (2014) indicated that currently, the RMSEA is the most popular measure of model fit and is reported in almost all papers that use CFA or SEM. MacCallum, Bowne & Sugawara (1996) used 0.01, 0.05 and 0.08 to specify excellent, good and poor fit, respectively. However, others have suggested 0.10 as the cut-off for poor fitting models. Use of confidence intervals and tests of PCLOSE can assist in comprehending the sampling error in the RMSEA. The lower limit of the 90% confidence interval should be preferably be very close to zero while the upper limit should be less than 0.08 (Kenny, 2014).The PCLOSE on the other hand gives a one-sided test of the null hypothesis that the RMSEA equals 0.05.

3.7.6.4 Normed Fit Index

In SPSS AMOS, the normed fit index is labelled NFI. "Normed" means that the index varies from 0 to 1, with 1 indicating perfect fit. The NFI was proposed by Bentler and Bonnett (1980) and assesses the model by comparing the χ^2 value of the proposed model to the χ^2 value of the null model. According to Garson (2015), the NFI reflects the proportion by which the researcher's model improves fit compared to the null model. It is calculated using the formula:

$$NFI = \frac{\chi_0^2 - \chi_\nu^2}{\chi_0^2} \tag{3.51}$$

where χ_0^2 is the χ^2 value for the null model

χ_ν^2 is the χ^2 value for the proposed model

ν is the degrees of freedom

NFI values above .95 are good (Schumacker & Lomax, 2016), between .90 and .95 adequate and below .90 point to a need to re-specify the model. The NFI may underestimate fit for small samples and does not penalise models for lack of parsimony.

3.7.6.5 Tucker-Lewis Index

The Tucker-Lewis index (TLI, also known as the Non-Normed Fit Index (NNFI)) compares the fit of the existing model with that of the null or independence model. It provides a small correction for parsimony and is relatively independent of sample size (Tanaka, 1993). The TLI is calculated using the formula:

$$TLI = \frac{\chi_0^2/\nu_0 - \chi_\nu^2/\nu}{\frac{\chi_0^2}{\nu_0} - 1} \tag{3.52}$$

Values of the TLI close to 1.0 indicate a better fit while values over .95 and .90 represent a good and an adequate fit of the model data respectively (Hu & Bentler, 1999). According to Gallagher et al. (2008), the TLI has superseded the NFI as the latter could not deal well with small samples.

3.7.6.6 Comparative Fit Index

The Comparative Fit Index (Bentler, 1990) is a revised form of the NFI which takes into account sample size (Byrne, 1998) and performs well even when sample size is small (Tabachnick and Fidell, 2014). The CFI is calculated using the formula:

$$CFI = 1 - \frac{\chi_\nu^2 - \nu}{\chi_0^2 - \nu_0} \tag{3.53}$$

Values for this statistic range between 0 and 1with values close to 1.0 considered as indicating good fit. A value of CFI ≥ 0.95 is recognised as indicative of good fit (Hu & Bentler, 1999). According to Hooper, Coughlan & Mullen (2008), the CFI is included in all SEM programmes and is one of the most popularly reported fit indices.

3.7.6.7 Adjusted Goodness of Fit Index

The AGFI adjusts the GFI based on the number of parameters in the model, with more saturated models reducing fit (Tabachnick & Fidell, 2014). More parsimonious models are favoured while complicated models are penalised. AGFI tends to increase with sample size. It is calculated using the formula:

$$AGFI = 1 - \frac{tr\left[(\Sigma^{-1}S - I)^2/\nu\right]}{tr[2(\Sigma^{-1}S)^2/p(p+1)]} \tag{3.54}$$

3.7.6.8 Akaike Information Criterion

The AIC is a comparative measure of fit thus it can only be considered when two different or competing models are estimated. It is a useful cross-validation index because it tends to select models that would be selected if results were cross-validated to a new sample (Loehlin, 2004). Similar to the chi-square index, the AIC reveals the degree to which the observed and predicted covariance matrices differ from each other. However, in contrast to the chi-square index, the AIC penalises models that are too complex. Small AIC values indicate a better fit hence the model with the least AIC is regarded the best fitting model. There are somewhat different formulae for the AIC found in the literature:

$$AIC = \chi^2 + m\,(m+1) - 2\nu \tag{3.55}$$

where m is the number of variables in the model and ν is the degrees of freedom. When means are included in the model, the formula becomes

$$AIC = \chi^2 + 2p \tag{3.56}$$

where p is the number of free parameters (i.e. parameters to be estimated) in the model.

3.7.6.9 Browne-Cudeck Criterion

The Browne-Cudeck Criterion (BCC; Browne & Cudeck, 1989) is similar to the AIC in that it measures the balance between model complexity and model fit. The BCC however imparts a higher penalty than the AIC when the model is complex. The BCC is calculated using the formula:

$$BCC = \frac{\chi^2}{n} + \frac{2k}{n-v-2} \tag{3.57}$$

3.7.6.10 Expected Cross Validation Index

ECVI is recommended as a way of evaluating, in a single sample, the likelihood that the model cross-validates similar-size samples from the same population. It measures the discrepancy between the fitted covariance matrix in the analysed sample, and the expected covariance matrix that would be obtained in another sample of equivalent size (Kaplan, 2009). ECVI is calculated using the formula:

$$ECVI = \frac{\chi^2}{n-1} + \frac{2k}{N-1} \tag{3.58}$$

Application of the ECVI assumes a comparison of models whereby an ECVI index is calculated for each model and then all ECVI values placed in rank order. The model having the smallest ECVI value shows the greatest potential for replication. It is possible to take the precision of the estimated ECVI value into account through the formulation of confidence intervals. By reporting an ECVI value within the bounds of a 95% confidence interval, one can argue that over all possible randomly sampled ECVI, 95% of them will fall within the upper and lower limits of the interval constructed (Byrne, 2001.).

3.7.6.11 Hoelter Index

Hoelter's critical N, also known as the Hoelter index, is used to judge whether the sample size is adequate. The index is calculated using the formula:

$$N = \frac{(n-1)\chi^2(crit)}{\chi^2} + 1 \tag{3.59}$$

where n is the sample size.

Alternatively, if $\chi^2(crit)$ is unknown, the Hoelter index is calculated using the formula:

$$N = \frac{\left[1.645 + \sqrt{2\nu - 1}\right]^2 + 1}{\frac{2\chi^2}{n-1} + 1} \tag{3.60}$$

By convention, sample size is adequate if N>200 and χ^2 is statistically significant. Hu and Bentler (1998), however, do not recommend this measure.

There are occasions when model-fit indices are not acceptable because the sample data do not support the hypothesised model. In such an instance, there is need for the re-specification of the theoretical model. Schumacker & Lomax (2016) recommended that researchers ought to consider whether each parameter is statistically significant from zero, runs in the expected direction, and could be meaningfully interpreted if they were determined to establish the fit of individual parameters. Finally, the parameter estimates should remain within expected values (i.e., correlations should not exceed 1, variances should have positive values). This recommendation was adopted in this study.

3.7.7 Model Modification

The final step of SEM involves model modification or re-specification. In this step, model modification methods are applied in order to find a model that best fits the data. The step is iterative and may involve more than a single iteration before an acceptable model fit is attained. Modification indices from the AMOS text output can also be used to inform the addition of covariance arrows to the error terms having the highest modification index value but belonging to factors loading onto the same latent variable if necessary. This is repeated for a single pair of error terms at a time until an acceptable model fit is achieved for all fit indices. Once the modification iterations result in acceptable model fit, obtained parameter estimates are then be used to confirm the convergent and discriminant validity for the latent variables. In this study, model re-specification (modification) began by deletion of factors with factor loadings below the cut-off. Acceptable model fit was attained following this initial step hence use of modification indices was not necessary.

3.7.8 Validity

Validity is the ability of an instrument to measure what it is supposed to measure for a latent variable. Two types of validity are required for each measurement model and they are described below:

3.7.8.1 Convergent Validity

Convergent validity refers to the degree to which indicators of a specific construct share a high proportion of variance in common. Results from the CFA were used to analyse the convergent and discriminant validity of the model. Convergent validity in the study was assessed using the average variance extracted (AVE) which is found by averaging the squared factor loadings for each construct as shown in the formula below:

$$AVE = \frac{\sum \lambda_{ij}^2}{n} \tag{3.61}$$

where λ_{ij} represents the factor loading i on factor j and n is the number of indicators for each construct.

According to Hair et al (2010), an AVE of 5.0 or higher is an adequate indication of convergence.

3.7.8.2 Discriminant Validity

Discriminant validity, on the other hand, refers to the extent to which a construct is fully distinct from another construct both in terms of how much it correlates with other constructs and how distinctly measured variables represent only this single construct. Discriminant validity was measured by comparing the AVE estimates for each factor with the squared inter-construct correlation (SIC) for that factor. Evidence of discriminant validity is shown by AVE > SIC (Campbell & Fiske, 1959; Heeler & Ray, 1972; Thompson, 2003; Hair et al. 2014a). The average variance extracted (AVE) as well as the inter-correlation matrix was used to assess discriminant validity in this study.

3.7.9 Reliability

In survey research, the term *reliability* reflects the internal consistency and how repeatable the method of a research is (Rubin et al, 2010). Reliability is indicative of the extent to which the indicators all measure the same thing. Reliability analysis allows the study of measurement scale properties and the individual items in the scale. According to Bryman & Cramer (2005), internal reliability is particularly important when there are multiple measurement items for each construct. Two criteria for assessing reliability are available and they are described below:

3.7.9.1 Internal Reliability

Internal reliability assesses the consistency of the results for different items for the same construct within a given measure. In this study, the Cronbach's coefficient alpha (Cronbach, 1951) was used to assess the internal consistency of the items under the main constructs. Cronbach's alpha "measures the extent to which item responses obtained at the same time correlate with each other" (Gallagher et al. 2008:268). The formula for calculating Cronbach's alpha is:

$$\alpha = \frac{N\overline{c}}{\overline{v} + (N-1)\overline{c}} \tag{3.62}$$

where N = the number of items forming the construct

$\overline{c}$ = average covariance between item-pairs and

$\overline{v}$ = average variance

If the score is high or within acceptable limits, validation can proceed. An alpha value of 0.6 or greater was considered adequate for this study.

3.7.9.2 Composite Reliability

According to Hair et al. (2010), composite reliability is an estimate of the extent to which a set of latent construct indicators share in their measurement of a construct. Composite reliability is calculated using the formula:

$$CR = \frac{(\sum \lambda_{ij})^2}{(\sum \lambda_{ij})^2 + \sum \varepsilon_{ii}} \tag{3.63}$$

where ε_{ii} is the variance of the error term corresponding to indicator i.
It is worth noting that "there are no universally accepted standards" pertaining "minimally acceptable indicator and composite reliabilities" (Bagozzi & Yi, 2012:17). In this study composite reliability values ≥ 0.6 (Hair et al, 2010; Zainudin, 2012) were considered as acceptable.

3.8 Bootstrapping

Efron (1979) developed the bootstrap as a general technique for evaluating the statistical precision of an estimator. Hinkley (1988) stated that the principle of bootstrap methods is the simulation of pertinent attributes of a statistical technique with minimal model assumptions, Bootstrapping functions as a resampling technique by which the original sample is deemed to represent the population (Byrne, 2010). Multiple subsamples of the same size as the parent sample are then drawn randomly, with replacement, from this population and provide the data for empirical investigation of the variability of parameter estimates and indices of fit. The bootstrap procedure thus allows the researcher to assess the stability of parameter estimates and report their values with greater degree of accuracy. According to Yung & Bentler (1996), the bootstrap procedure affords a tool for tackling situations where assumptions of large sample size and/or multivariate normality may not hold.
Bootstrapping has proved a robust resilience as a useful technique for the following reasons:

- It is easy to perform

- It assesses uncertainty about a parameter where more conventional classical statistics methods are not available
- It makes use of the increased power and ease of use of computers and
- It matches conventional statistics methods where they are available, especially when large data sets are involved.

3.8.1 Bootstrap Approaches

There are two approaches to bootstrapping namely: parametric bootstrap and non-parametric bootstrap. The parametric bootstrap is applicable when parameters of a population or probability distribution are estimated from a known distribution. The non-parametric bootstrap, on the other hand, is used to estimate parameters of a population or probability distribution in situations where the distributional form is unknown. The two approaches use the same procedure but differ in the way they simulate data. The parametric bootstrap generates simulated values from a fitted model while the non-parametric bootstrap resamples original data. Non-parametric bootstrapping was applied in this study.

3.8.2 Non-parametric Bootstrap Procedure

The principal steps in the non-parametric bootstrap procedure which was applied in this study are:

Step 1. An empirical probability distribution, F_n, is constructed from the sample by assigning a probability of $1/n$ at each point, $x_1, x_2, \ldots, x_n$ of the sample. This represents the empirical distribution function of the sample.

Step 2. From the empirical distribution, F_n, a random sample of size n is drawn with replacement.

Step 3. The required statistic, $\hat{\theta}$, is calculated for the sample generated in Step 2 above.

Step 4. Steps 2 and 3 are then repeated B times in order to create B resamples. The value of B is usually at least equal to 1000 when an estimate of the confidence interval around $\hat{\theta}$ is required.

The distribution of these B estimates of represents the bootstrap estimate of uncertainty about the true value of .

3.8.3 Bias and Standard Error Estimation

Bias can be viewed as the tendency of a sample statistic to systematically under- or over-estimate a population parameter. If $\hat{\theta}$ is an estimate of the parameter θ then the bias for the estimator $\hat{\theta}$ of θ is given by $E\left(\hat{\theta}\right) - \theta$. Moreover, an estimate of the bootstrap standard error is calculated using the formula:

$$SE\left(\hat{\theta}\right) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_b - \overline{\theta}\right)^2} \tag{3.64}$$

where $\hat{\theta}_b$, $b = 1, 2, \ . \ . \ ., B$ denotes the B estimates of θ and $\overline{\theta} = (1/B)\sum_{b=1}^{B}\hat{\theta}_b$ denotes the mean of the estimates across the B bootstrap samples.

3.8.4 Bias-Corrected (BC) Bootstrap Confidence Intervals

Singh & Xie (2008) referred to a confidence interval for a given population parameter θ as a sample based range $[\hat{\theta}_1, \hat{\theta}_2]$ for the unknown quantity θ such that it would lie within the range with a high specified probability. The two frequently used confidence levels are 95% and 99%.Unlike the usual confidence interval, the BC bootstrap confidence interval adjusts for bias in the bootstrap distribution. Puth et al. (2015) stated that this method is founded on the assumption that there is a monotonic increasing transformation f () of the estimator $\hat{\theta}$ such that the transformed values $f(\hat{\theta})$ are normally distributed with mean $f(\theta) - z_0$ and standard deviation one.

3.8.4.1 Calculation of 95 per cent interval

Carpenter & Bithell (2000) suggested the following steps for calculating a 95% bootstrap confidence interval:

Let

$$\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_{1000}^* \tag{3.65}$$

represent the ordered bootstrap sample set such that $\hat{\theta}_i^* < \hat{\theta}_j^*$, for $1 \leq i \leq j \leq 1000$.

Step 1. Count the number of members of equation (3.65) that are less than $\hat{\theta}$ (calculated from the original data). Call this number p and set $b = \Phi^{-1}(p/B)$.

Step 2. Calculate $Q = (B+1)\Phi(2b - z_{0.05})$, where $z_{0.05} = -1.64$. Q is the percentile of the bootstrap distribution required for the upper endpoint of the bias corrected confidence interval.

Step 3. Estimate the endpoint of the interval by $\widehat{\theta}^*_{int(Q)}$, where int(Q) implies 'take the integer part'. If a more accurate estimate is required, interpolation can be used between the members of equation (3.65) by letting the nearest integers to Q to be a, b, such that $a < Q < b$ and $b = a + 1$. The Qth percentile is then estimated by

$$\hat{\theta}_Q^* = \hat{\theta}_a^* + \frac{\Phi^{-1}\left(\frac{Q}{B+1}\right) - \Phi^{-1}\left(\frac{a}{B+1}\right)}{\Phi^{-1}\left(\frac{b}{B+1}\right) - \Phi^{-1}\left(\frac{a}{B+1}\right)}\left(\hat{\theta}_b^* - \hat{\theta}_a^*\right) \tag{3.66}$$

The one-sided bias-corrected interval is $\left(-\infty, \hat{\theta}_Q^*\right)$.

Two-sided 95% confidence intervals are formed as the intersection of two one-sided 97.5% intervals (Carpenter & Bithell, 2000).

3.8.5 Bias-Corrected and Accelerated (BCa) Bootstrap Confidence Intervals

BCa intervals use percentiles of the bootstrap distribution without necessarily using the 100α-th and 100(1- α)-th percentiles. The main advantage to the BCa interval is that it adjusts for bias and skewness in the bootstrap estimates distribution. The BCa interval involves the estimation of two parameters, the bias-correction parameter, z_0 and the acceleration parameter, a. According to Efron & Tibshirani (1993), BCa intervals have the form:

$$\left[\hat{\theta}_{lo}, \hat{\theta}_{up}\right] = \left[\hat{\theta}^*_{(\alpha_1)}, \hat{\theta}^*_{(\alpha_2)}\right], \tag{3.67}$$

where $\hat{\theta}_{lo}$ is the lower confidence limit value

$\hat{\theta}_{up}$ is the upper confidence limit value

$\hat{\theta}^*$ are bootstrap parameter estimates

$$\alpha_1 = \Phi\left(z_0 + \frac{z_0 + z_\alpha}{1 - a(z_0 + z_\alpha)}\right) \tag{3.68}$$

and

$$\alpha_2 = \Phi\left(z_0 + \frac{z_0 + z_{1-\alpha}}{1 - a(z_0 + z_{1-\alpha})}\right) \tag{3.69}$$

$\Phi(.)$ is the cumulative distribution function (cdf) of the standard normal

z_α is the 100α-th percentile of the standard normal.

It is worth noting that when $a = z_0 = 0$, then

$$\alpha_1 = \Phi(z_\alpha) = \alpha \quad \text{and} \quad \alpha_2 = \Phi(z_{1-\alpha}) = 1 - \alpha \tag{3.70}$$

Efron & Tibshirani (1993) added that z_0 is found directly from the proportion of bootstrap replications less than the original estimate $\hat{\theta}$,using the equation

$$z_0 = \Phi^{-1}\left(\frac{\#\{\hat{\theta}^*(b) < \hat{\theta}\}}{B}\right) \tag{3.71}$$

where $\Phi^{-1}(.)$ is the inverse function of the standard normal cdf.

The acceleration, a, is obtained using the formula:

$$a = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(.)} - \hat{\theta}_{(i)}\right)^3}{6\left\{\sum_{i=1}^{n}\left(\hat{\theta}_{(.)} - \hat{\theta}_{(i)}\right)^2\right\}^{3/2}} \tag{3.72}$$

where $\hat{\theta}_{(.)} = \sum_{i=1}^{n} \hat{\theta}_{(i)}/n$

3.9 Multiple Regression Modelling

The general regression model expresses the DV as a linear combination of the IVs. It can therefore be represented as an equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + e \tag{3.73}$$

where Y is the DV; X_1, X_2, ..., X_p are IVs; β_1, β_2,..., β_p are regression weights; β_0 is the regression constant or intercept and e is the residual.

In this study, the model is represented as:

$$CdmUse\_Intention = \beta_0 + \beta_1 Aff\_Att + \beta_2 Instr\_Att + \beta_3 Norms + \beta_4 Perceived\_control + e \tag{3.74}$$

3.9.1 Computation of composite variables

Composite variables that were to be included in the MR model were computed by averaging the item scores for the construct items confirmed by the SEM measurement model. After computing composite variables, the regression model was fitted following the steps shown in Fig 3.6 below:

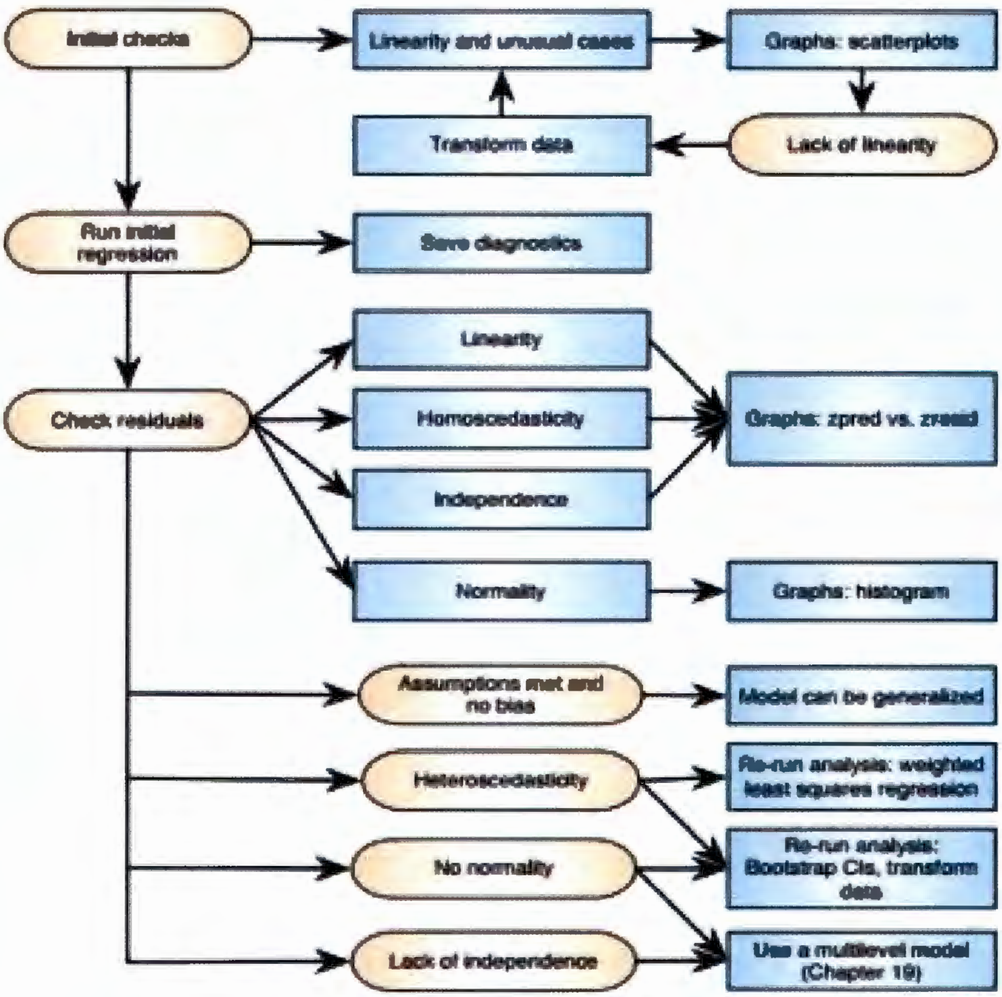


Figure 3.6 Regression Flow Diagram

(Source: Field, 2013:316)

3.9.2 Exploratory Data Analysis

Before running the multiple linear and LASSO regression model, the descriptive statistics of attitudes toward condom use, subjective norms, PBC, and intention to use condoms were computed. The descriptive statistics help in the identification of outliers. Tabachnik & Fidell (2014) suggested the removal of values that fall outside the range "±3.29 standard deviations" away from the mean. Skewness and Kurtosis indices were also used to check for influential cases or outliers in the data. A detailed discussion on skewness and kurtosis was covered in section 3.6.1.1.

3.9.2.1 Linearity

Multiple regression assumes that the relationship between the dependent and independent variables is linear. In the event that the relationship between the DV and IVs is non-linear, the results of the regression analysis will under-estimate the true relationship (Osborne & Waters, 2002). Consequently, Osborne & Waters (2002:1) further explained that "this under-estimation carries two risks: increased chance of a Type II error for that IV, and in the case of multiple regression, an increased risk of Type I errors (over-estimation) for other IVs that share variance with that IV." Casson & Farmer (2014) propose that the linearity assumption can be assessed by examining the relationship between the dependent and independent variables on scatter plots. Bivariate scatterplots of the DV (*CdmUse_Intention*) and the IVs (*Aff_Att*, *Instr_Att*, *Norms* and *Perceived-control*), as advocated by the above-mentioned authors, were used to visually examine conformance of the study variables to the linearity assumption. In addition, Pearson correlation coefficients were calculated to determine the strength of the relationships between variable pairs.

3.9.2.2 Diagnostics

Once the preliminary analyses were concluded, the regression model was then run. The model was checked for multicollinearity by examining the collinearity diagnostics. Residual plots were then be used to check error variance assumptions (linearity, normality and homogeneity of variance). The significance of coefficient estimates was examined and used to trim the model.

3.9.2.3 Linearity of Residuals

Linearity of the relationship between residuals of prediction and predicted DV scores is assumed. Violation of this assumption may possibly bias the regression coefficients, standard errors and test of significance (Keith, 2015; Tabachnik & Fidell, 2014). According to Keith (2015), assessment of this assumption can be done by inspecting a scatterplot of standardised residuals against standardised predicted values. If nonlinearity is existent the general shape of the scatterplot is curved instead of rectangular (Tabachnik & Fidell, 2014). Alternatively, one can fit a locally estimated scatterplot smoothing (loess) line to the scatterplot. Conformity to the linearity assumption would be shown by the loess line coming close to a horizontal regression line that passes through zero. Cohen et al. (2003:111) remark that the loess line should resemble "a young child's freehand drawing of a straight line." Significant departure from linearity would be indicated by a curved loess line.

3.9.2.4 Normality of Residuals

The normality of residuals assumption implies that the residuals (errors) are normally distributed about the predicted DV scores (Casson & Farmer, 2014; Tabachnik & Fidell, 2014). Normality of residuals can be assessed via histograms or scatterplots of residuals versus predicted values. Another, more precise method, is the P – P plot (or alternatively, Q – Q plot). A P – P plot of the residuals displays the value of the expected cumulative probability on one axis and the observed cumulative probability on the other. When the normality of residuals assumption is met, a superimposed normal curve on the histogram suggests that residuals are normal (Keith, 2015) while the residuals scatterplot would show a rectangular distribution with a concentration of scores along the centre. In the case of the P – P plot, the thick line will come close to the diagonal straight line when the residuals are normally distributed. A Q – Q plot of residuals on the other hand displays the value of observed residuals on one axis and the expected normal value of the residuals on the other. Similar to the P – P plot, if the residuals on the Q – Q plot are normally distributed, the thick line will come close to the diagonal straight line. This study made use of the scatterplots, histogram and the P – P plot to test the normality of residuals assumption.

3.9.2.5 Homoscedasticity and Heteroscedasticity of Residuals

Homoscedasticity signifies that the variance of errors is approximately constant for all predicted DV scores (Tabachnick & Fidell, 2014). In instances where the variance of errors changes for different values of the IV, heteroscedasticity is shown (Osborne & Waters, 2002). According to Berry & Feldman (1985) and Tabachnick & Fidell (2014), slight heteroscedasticity has inconsequential effect on significance tests; on the other hand, when heteroscedasticity is evident it can cause serious misrepresentation of conclusions as well as extremely weaken the analysis thereby increasing the possibility of a type I error. This assumption can be confirmed by visual examination of a scatterplot of the regression standardised residuals alongside the regression standardised predicted value.

According to Kline (2011), the homoscedasticity and normality of residuals assumptions are less critical for the reason that regression is fairly robust to their violation. Keith (2015) added that the violation of the normality of residuals assumption is only worrying in the case of small samples. Moreover, even if the errors are not from a normal distribution, the central limit theorem could allow for the satisfying of the normality assumption since large samples ($n > 50$) are robust to the violation of the normality assumption (Lumley et al., 2002).

It is worth noting that since LASSO regression is applied to facilitate variable selection, its assumptions are those of the type of model it is applied to. Thus in the study, assumptions for ordinary least squares (OLS) regression apply to the LASSO regression.

3.9.3 Steps in Generalised Additive Modelling

The sections below outline the steps that are followed when GAMs are applied.

3.9.3.1 Model Specification in GAMs

If $g(\mu)$ is in general an increasing or decreasing function in x, that is not a perfect straight line, but is a bending curve, it can be modelled as a combination of a straight line and a smoothing function. A model with a single predictor variable is specified as:

$$g(\mu) = \beta_0 + \beta_1 x + f(x) \tag{3.75}$$

In this case, the effect of the smoothing function $f(x)$ is separated from the parametric linear effect $\beta_1 x$. According to Liew & Forkman (2015), if $\beta_1 x$ was excluded from Eqn (3.75), the smoothing function $f(x)$ would express the totality of all linear and non-linear effects of x.

3.9.4 Estimation in GAMs

Estimation in GAMs is carried out through a combination of backfitting and iteratively reweighted least squares. The functions f_i and β_k are determined empirically according to the data and the assumed model. The deviance is calculated from the model. The PROC GAM procedure of the SAS System and the *gam* function in R fit LOESS and cubic spline smoothing functions, using the backfitting algorithm (Hastie & Tibshirani, 1986; 1990). The algorithm is presented below.

3.9.4.1 Backfitting Algorithm

The basic idea behind backfitting is estimation of each smooth component of an additive model by iteratively smoothing partial residuals from the additive model with respect to the covariates that the smooth relates to. The backfitting algorithm is outlined below:
Let

$$Y = \beta_0 + f_1(X_{1)} + f_2(X_2) + \ldots + f_p(X_p) + e \tag{3.76}$$

Step 1: Set $\beta_0 = mean(Y)$

Step 2: Initialise $f_j(X_j) = f_j(X_j)^0$

Step 3: Iterate and cycle over the p variables for

$$f_j = S_j(Y - \beta_0 - \sum_{p \neq j} f_p | X_j) \tag{3.77}$$

until the f_i do not change

3.9.5 Assessing Model Fit in GAMs

Model fit in GAMs is described graphically. Fitted values are plotted against values of the predictor variable. In order to show how well the curve fits the observed values, the graph may include the observed values. Partial residuals are valuable for the examination of predictor variables (Guisan et al., 2002; Wood, 2006). In partial residual plots, partial residuals are plotted versus predictor variable values. A good fit is shown by partial residuals that are randomly distributed around the curve for the smoothing function. In addition, it is possible to approximately establish whether a certain predictor variable is significant or not, by comparing deviances (Hastie & Tibshirani, 1990). Deviance is calculated using the formula:

$$D = 2\left(\log L\left(saturated\ model\right) - \ \log L\ \left(fitted\ model\right)\right) \tag{3.78}$$

According to Agresti (2007), the deviance is the likelihood-ratio statistic for comparing the saturated model to the fitted model. It is a test statistic for the hypothesis that all parameters included in the saturated model but not in the fitted model equal zero. For large samples, the statistic has an approximate chi-square distribution with df equal to the difference between the residual values for the separate models. In the case of two models, designated by M_1 and M_2, where M_1 is nested in M_2, large test statistics and small p-values suggest that model M_1 fits more poorly than M_2. Furthermore, Hastie & Tibshirani (1984) stated that in the case of Gaussian errors the deviance is identical to the residual sum of squares (RSS).

3.9.6 Summary

This chapter highlighted the procedures undertaken in this study in order to fit data to the SEMs, GAMs, MR as well as application of LASSO regression. The chapter began by discussing the theoretical framework that was utilised in the data collection and the philosophical underpinnings associated with the research process. The variables or measures (both indicators and latent variables) used in this study were also delineated. Moreover, the chapter highlighted and discussed the test statistics and fit indices to be employed in the ensuing chapter as well as the steps that are pursued when applying SEM, GAMs, MR and LASSO regression. Absolute, relative and parsimonious fit indices together with model comparison indices were considered and are reported in Chapter 4. It was noted that a number of estimation methods were obtainable when using SEM, with the MLE being the most preferred estimation approach. A discussion of the bootstrapping approach was carried out. In the case of MR, an outline for computing the composite variables was given. The chapter also reflected on diagnostic considerations related to the MR model before concluding with steps that are followed in generalised additive modelling.

CHAPTER 4. DATA ANALYSIS AND RESULTS

4.1 Introduction

Chapter 4 presents the results and engages with the analysis of data. Anderson & Gerbig's (1988) two-step approach was applied to the TPB based condom use intention structural equation model discussed in Chapter 3. The approach involves separate estimation of the measurement model before the estimation of the structural model. Prior to the estimation of parameters, data screening was carried out. Descriptive statistics for the variables of interest were also computed. Parameters in the structural equation model were initially estimated using a maximum likelihood estimation (MLE) procedure in AMOS. Due to violation of the multivariate normality, parameter estimation in the structural was also carried out using the unweighted least squares (ULS) method. The SPSS linear regression procedure was used to estimate parameters in the multiple regression model. The Glmnet procedure for LASSO regression in R and the generalised regression option in JMP Pro 13 were used for variable selection as well as fitting the LASSO regression model. GAM regression was done using the GAM procedure in SAS as well as the mgcv and gam procedures in R.

4.2 Data Screening results

Both case screening and variable screening were performed on the data. Case screening was performed in order to investigate data completeness as well as unengaged responses. Variable screening, on the other hand was carried out to explore characteristics of the dataset focusing on outliers, skewness and kurtosis. Results are presented and discussed in section 4.16.

4.3 Missing Data

The Missing Value Analysis (MVA) procedure in IBM SPSS version 25 was used to analyse cases for missing values. An overall summary of missing values is shown in Fig 4.1.

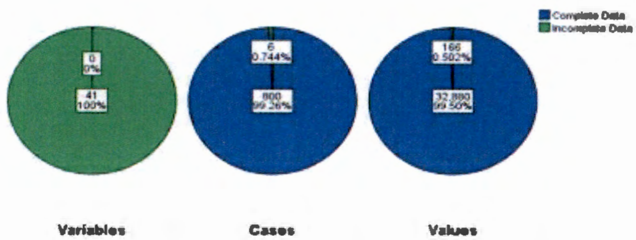


Figure 4.1 Overall summary of missing values

As shown in Figure 4.1, all the variables of interest had some missing values. The proportion of missing data within the variables of interest was very low, with an overall proportion of this missing data at 0.5%. Only 6 cases out of 806 (0.74%) had missing data. Olinsky et al. (2003) suggest that a few missing values, such as less than 5% on a single variable, in a large sample may be of little concern. Since the proportion of missing data was very low, listwise deletion (Kang, 2013; Roth 1994) was applied following the suggestions of Kang (2013) to deal with missing data and mitigate the chances of inaccuracy.

4.4 Unengaged responses

Unengaged responses were assessed by calculating the standard deviation for all possible model items. Cases with standard deviation values less than 0.5 (Gaskin, 2012) were deleted. Six such cases were present thus resulting in the sample size decreasing by a further 6 cases. Variables were then assessed for outliers, skewness and kurtosis.

4.5 Sample Statistics

Of the 794 participants, 53.7% were female while 46.3% were male. The participants' ages ranged from 13 to 18 years with a mean of 14.7 and standard deviation of 0.93 years. The majority (87.2%) of the participants had never had sexual intercourse while the remaining 12.8% indicated that they had had sex before. Participants' knowledge regarding STIs and safer sex practices was assessed by means of 24 yes – no questions. In order for participants not to guess, the “I don’t know” response option was available. A correct answer was given a single point while an incorrect or an “I don’t know” answer was not awarded any point. Thus the total possible knowledge score ranged from 0 to 24 points. Knowledge concerning STIs and knowledge of HIV prevention differed extensively within the sample. The total knowledge score for this sample ranged from 5 to 24 with a mean score of 16.55 and a standard deviation of 3.55.

Included in the knowledge questions were 5 questions associated with condom use, storage and technical skills about use of condoms. Table 4.1 shows the frequency distribution of the number of adolescents who responded correctly and incorrectly to the questions.

Table 4.1 Condom use knowledge frequency distribution (n =793)

Question/Statement	Correct	Incorrect
Can people reduce their chances of getting HIV/AIDS by using a condom correctly every time they have sex?	675(85.1)	118(14.9)
When a condom is placed on the penis, space should be left at the tip of the condom.	457(57.6	336(42.4)
Storing or carrying condoms in a hot or warm place can destroy their effectiveness.	554(69.9)	239(30.1)
The penis should be hard when the condom is put on it.	502(63.3)	291(36.7)
If you place a condom on the penis the wrong way, you should start over with a new condom.	539(68.0)	254(32.0)

As shown in Table 4.1, generally the majority of adolescents in the sample were familiar with condom use knowledge. An overwhelming majority (85.1%) of the adolescents in the sample knew that the risks of getting HIV/AIDS could be reduced by using a condom correctly every time people have sex. Approximately two-thirds of the adolescents in the sample were aware that condom effectiveness could be weakened by heat, a condom had to be worn on a stiff penis and placing a condom on the penis the wrong way required starting over with a new condom respectively. About 42% of the adolescents were oblivious to the need to leave space at the tip of the condom when placing the condom on the penis.

4.6 Variables used in the data analysis and their descriptive statistics

Table 4.2 Descriptive statistics for items used in the validation of models

Construct	Item	Mean	Std. Deviation
Norms	NO1	4.19	1.06
	NO2	4.08	1.24
	NO3	4.03	1.26
	NO4	4.03	1.13
Aff_Att	AA1	1.75	1.10
	AA2	2.34	1.25
	AA3	1.65	0.85
	AA4	1.98	1.05
	rAA5[1]	2.51	1.19
	rAA6	2.89	1.13
Instr_Att	IA1	4.34	1.02
	IA2	4.39	0.94
	IA3	4.44	0.86
Perceived_control	PC1	3.97	1.10
	PC2	3.84	1.18
	PC3	4.26	0.90
	PC4	4.27	0.86
	PC5	4.09	1.00
	PC6	4.07	0.96
	PC7	3.53	1.18
	PC8	4.34	0.96
	PC9	4.32	0.84
	PC10	3.35	1.16
	PC11	2.55	1.36
CdmUse_Intention	CUI1	4.40	0.89
	CUI2	4.42	0.85

Table 4.2 shows the variables used in the data analysis and their descriptive statistics. For most of the constructs, item mean scores were above the midpoint of 3.00, indicating more positive dispositions. However, all items for the *Aff_Att* construct had item means below the midpoint of 3.00, with a range of 1.65 to 2.89. The values were consistent with the way the statements were phrased. Subjects with positive affective attitudes were expected to either disagree or disagree strongly with the statements hence giving overall low scores for the scale.

4.7 Test of the TPB measurement model

The measurement model was used to test the hypothesis that the variables CUI1 and CUI2 are indicators of condom use intention (**CdmUse_Intention**); NO1, NO2, NO3 and NO4 are indicators of normative beliefs (**Norms**); AA1, AA2, AA3, AA4, rAA5 and rAA6 are indicators of condom use affective attitude (**Aff_Att**); IA1, IA2 and IA3 are indicators of condom use instrumental attitude (**Instr_Att**) and PC1, PC2, PC3, PC4, PC5, PC6, PC7, PC8, PC9, PC10 and PC11 are indicators of condom use perceived control beliefs (**Perceived_ control**).

4.8 SEM Model Identification results

As mentioned in Chapter 3, the identification of a model refers to the presence of sufficient information (an adequate number of observed variances and covariances) to allow estimation of all the model parameters. The model parameters to be estimated are the covariances of the latent variables, the number of factor loadings and the measurement error variances. An over-identified model (i.e a model having a positive number of degrees of freedom) is preferred since it contains more than enough information to allow every parameter to be distinctively estimated.

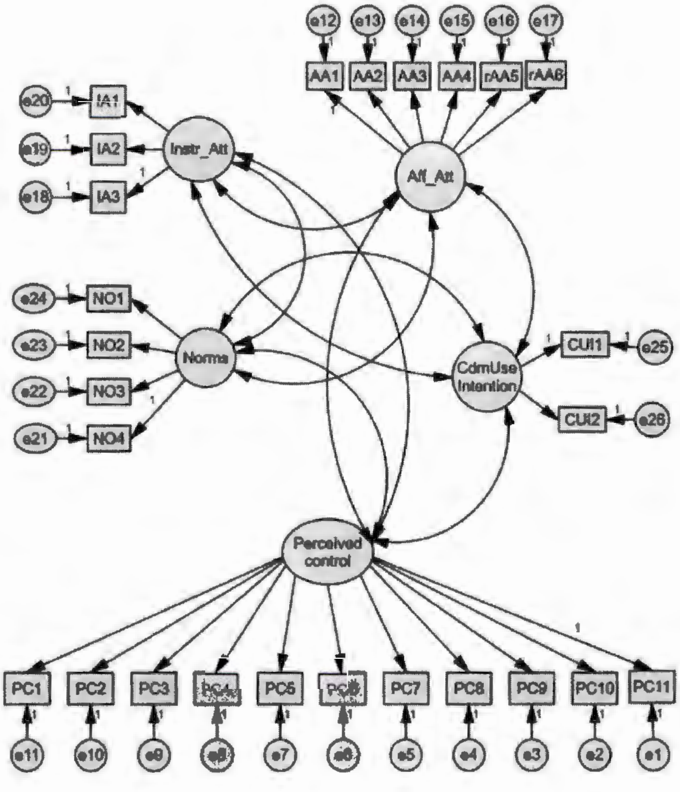


Figure 4.2 Condom use intention measurement model

As shown in Fig 4.2, with p = 26 observed variables in the measurement model, the number of variances and covariances is 26 (26+1)/2 = 351. The model parameters to be estimated are 10 covariances of latent variables, 26 factor loadings and 26 measurement error variances (giving k = 62). The number of degrees of freedom using is 289 (351 – 62) df. Since the df > 0, the model is over-identified and thus can be tested.

4.9 SEM Model Estimation

Anderson & Gerbig's (1988) two-step SEM strategy using IBM SPSS AMOS 25.0 was used to perform the SEM. The strategy involves separate estimation of the measurement model before the estimation of the structural model. Parameters were estimated using a maximum likelihood estimation (MLE) procedure. Fig 4.3 shows the CFA standardised estimates in the form of a path diagram.

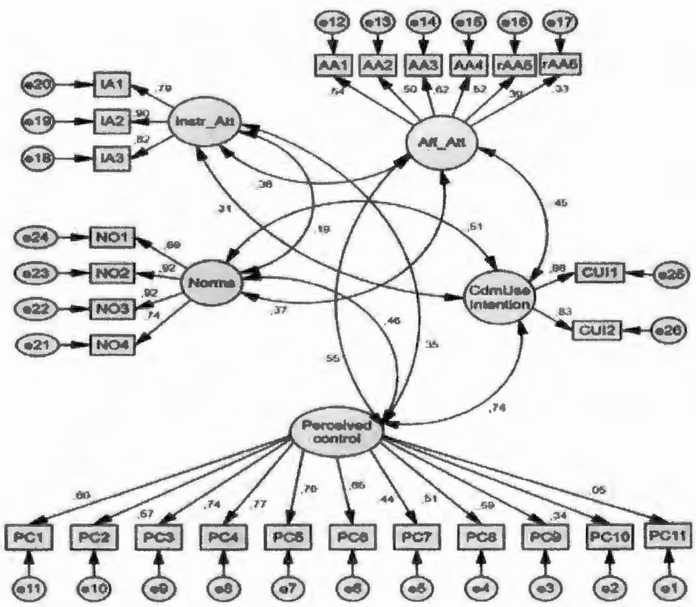


Figure 4.3 Path diagram showing CFA standardised estimates

Fig 4.3 shows factor loadings as well as correlation coefficients between the latent variables. Values on doubled headed arrows represent correlations (e.g., the correlation coefficient between Instr_Att and Normative beliefs is r = 0.19). The correlations among the factors ranged from -0.55 to 0.74, indicating that there is sufficient discriminant validity among the latent constructs (Bollen, 1989a; Kline, 2011). An inspection of the factor loadings shows that the values ranged from 0.05 to 0.92 with a few of the values falling below the proposed 0.5 cut-off. Factors with loadings below 0.5 are candidates for deletion from the model if the model fit is not adequate.

4.10 Original SEM Model Evaluation results

As indicated earlier in section 3.7.6, SEM has no single statistical test that best describes the strength of model prediction. A combination of different types of indices is therefore used in evaluating model fit. The following tables highlight the indices of interest for the original (proposed) model for this study.

Table 4.3 CMIN and CMIN/DF

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	62	**1112.738**	289	0.000	**3.850**
Saturated model	351	.000	0		
Independence model	26	8545.310	325	0.000	26.293

Information about our model is contained in the row labelled **Default model** in Table 4.3. CMIN stands for the minimum discrepancy between the model and the data. CMIN is equivalent to the χ^2 value. As noted in section 3.7.4.1, low $\chi 2$ values relative to the degrees of freedom with an insignificant p value ($p > 0.05$) would support the model as representative of the data. The $\chi 2$ test of the model was statistically significant with a value of 1112.738 (289, $N = 794$), $p < 0.001$. This however is inconclusive as the $\chi 2$ statistic almost continually rejects the model when samples of large size are used. It is for this reason that the relative chi-square (χ^2/ν) is preferred. The relative chi-square is represented in the table by CMIN/DF and falls within the acceptable range of between 2 and 5.

Table 4.4 shows the Goodness of Fit Index (GFI) and the Adjusted Goodness of Fit Index (AGFI).

Table 4.4 GFI and AGFI

Model	RMR	GFI	AGFI	PGFI
Default model	0.066	**0.895**	**0.873**	0.737
Saturated model	0.000	1.000		
Independence model	0.278	0.351	0.299	0.325

As shown in Table 4.4 both indices were below the recommended 0.9 cut-off value listed in Table 3.3. There was therefore a need to re-specify the model in order to improve the model fit.

Table 4.5 below shows values for the normed fit index and the comparative fit index.

Table 4.5 CMIN and CMIN/DF

Model	NFIDelta1	RFIrho1	IFIDelta2	TLIrho2	CFI
Default model	**.70**	.854	.900	.887	**.900**
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

The CFI value (0.9) suggested an acceptable fit while that of NFI (0.7) was below the recommended 0.9 cut-off value. NFI values less than 0.9 can usually be improved considerably. There was therefore a need to re-specify the model in order to improve the model fit.

Table 4.6 displays values for the RMSEA.

Table 4.6 CMIN and CMIN/DF

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	**.060**	**.056**	**.064**	.000
Independence model	.179	.175	.182	.000

The RMSEA value obtained for our model was 0.06 with a 90% confidence interval of 0.056 to 0.064. Thus the RMSEA indicated adequate fit since the value lies within the 0.05 – 0.08 interval. In summary, the model produced acceptable fit indices for the relative chi-square (CMIN/DF), CFI and RMSEA; however, the GFI, NFI and AGFI were inadequate. The model thus needed to be modified.

4.11 Modification of original SEM model

The first step in modifying the model involved removing factors that had factor loadings less than 0.5 from the model. The resultant modified model is shown in Fig 4.4 below.

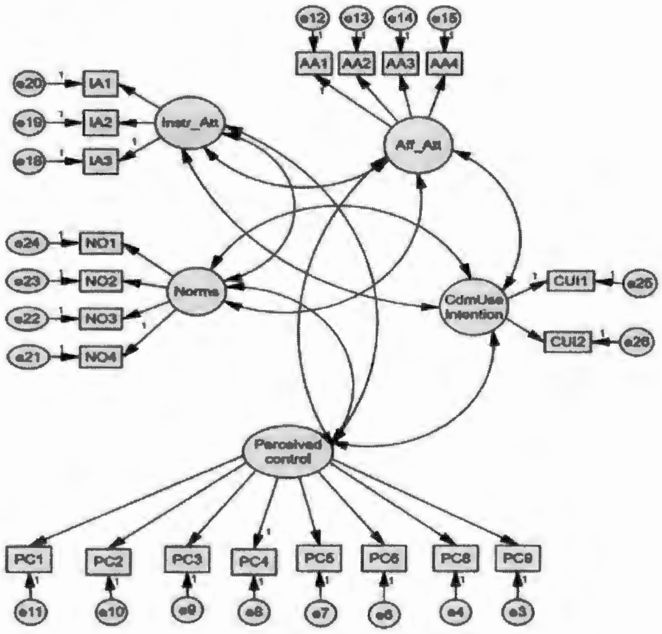


Figure 4.4 Re-specified model 1

Fig 4.4 shows the modified model after PC7, PC10 and PC11 from the Perceived control latent variable as well as rAA5 and rAA6 from the Aff_Att latent variable were deleted from the original model given in Fig 4.2. Parameter estimates and model fit indices for this respecified model were computed and are shown in Appendix 2.

Table 4.7 shows the model fit comparison for the two models.

Table 4.7 Model Fit Comparison

Fit Index	Original Model	Respecified Model 1	Recommended Thresholds
χ^2(Cmin)	1112.738	671.467	$p > 0.05$
χ^2/df	3.850	3.751	< 5
GFI	0.895	0.921	≥ 0.90
RMSEA	0.060 (0.056, 0.064)	0.059 (0.054, 0.064)	≤ 0.08
CFI	0.900	0.935	> 0.90
NFI	0.870	0.914	> 0.90
AGFI	0.873	0.899	> 0.90
AIC	1236.738	775.467	No threshold but compares values in alternative models. Lower values are better
BCC	1241.109	778.434	
ECVI	1.560 (1.435, 1.693)	0.978 (0.883, 1.083)	

Deletion of indicators having factor loadings < 0.5 resulted in the improvement of all fit indices as shown in Table 4.7. The relative chi-square ($\chi 2/df$), CFI and RMSEA were still adequate and showed better values. The GFI which was below the recommended threshold in the proposed model had an acceptable value of 0.921 in the re-specified model. The NFI value also improved to a value above the recommended cut-off value after the re-specification. Although the AGFI showed an improvement, its value was still below the recommended threshhold. There was therefore room to improve the model further. Overall, the re-specified model was better as evidenced by the lower AIC, BCC and ECVI values. A look at the factor loading estimates revealed that the factor loading estimate for PC8 was 0.496 (see standard regression weights in Appendix 2) and hence fell below the 0.5 cut-off value. PC8 was deleted as the next step in the modification process to give the second modified model shown in Fig 4.5 below.

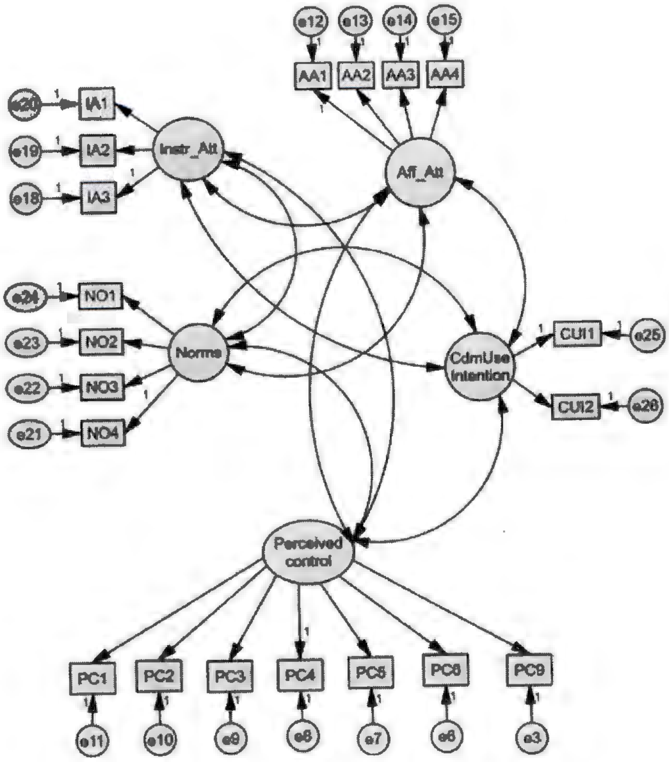


Figure 4.5 Re-specified model 2

The second re-specification resulted in the improvement of all fit indices to levels above the recommended threshold values. Re-specified model 2 thus became the final measurement model.

The model fit indices for the three models are displayed in Table 4.8.

Table 4.8 Model Fit Comparison of the three SEM Models

Fit Index	Original Model	Respecified Model 1	Final Model	Recommended Thresholds
χ^2 (Cmin)	1112.78	671.467	616.839	$p > 0.05$
χ^2/df	3.850	3.751	3.855	< 5
GFI	0.895	0.921	0.925	≥ 0.90
RMSEA	0.060 (0.056, 0.064)	0.059 (0.054, 0.064)	0.060 (0.055, 0.065)	≤ 0.08
CFI **NFI**	0.900 0.870	0.935 0.914	0.938 0.918	> 0.90 > 0.90
AGFI	0.873	0.899	0.901	> 0.90
AIC	1236.738	775.467	716.839	No threshold but compares values in alternative models. Lower values are better
BCC	1241.109	778.434	719.559	
ECVI	1.560 (1.435, 1.693)	0.978 (0.883, 1.083)	0.904 (0.813, 1.005)	

Table 4.8 shows that in the original model three indices (GFI, NFI and AGFI) had values below the recommended cut-offs. After the first re-specification, there was an improvement in all three indices. GFI and NFI values were now acceptable but AGFI was still below the recommended cut-off thus there was need for another re-specification. A second re-specification resulted in all three indices having values above the 0.9 cut-off level. Constructs within the model could then be assessed for validity.

4.12 Convergent Validity

To ensure convergent validity, the researcher checked if items loaded on their respective constructs with standardised loadings greater than 0.5, AVE > 0.5, CR > 0.6 and item-total correlation > 0

.6 (Hair et al. 2010). The results obtained for the measurement model, using excel, are given in Table 4.9.

Table 4.9 Measurement Model Convergent Validity Results

Constructs		Cronbach's test				Measurement item loadings ()
		Item-total correlation	value	Composite Reliability	AVE	
Affective Attitude (Aff_Att)	AA1	**0.427**	**0.651**	0.659	**0.327**	0.561
	AA2	**0.430**				0.528
	AA3	**0.451**				0.628
	AA4	**0.444**				0.566
Instrumental Attitude (Instr_Att)	IA1	0.729	0.873	0.878	0.707	0.792
	IA2	0.808				0.903
	IA3	0.743				0.823
Perceived control	PC1	**0.559**	0.842	0.846	**0.443**	0.604
	PC2	**0.522**				0.570
	PC3	0.692				0.764
	PC4	0.700				0.784
	PC5	0.631				0.694
	PC6	**0.565**				0.630
	PC9	**0.489**				0.580
Normative Beliefs (Norms)	NO1	0.672	0.890	0.891	0.676	0.685
	NO2	0.821				0.916
	NO3	0.827				0.921
	NO4	0.726				0.739
Condom use intention (CdmUse_Intention)	CUI1	0.735	0.847	0.848	0.736	0.884
	CUI2	0.735				0.831

As can be seen in Table 4.9, all items had factor loadings greater than 0.5. The composite reliability (CR) for each latent variable exceeded 0.6 as required. The Cronbach alpha values for all variables except Aff_Att were high (> 0.8). While the Cronbach alpha value for Aff_Att fell below the target value of 0.7, it was however marginally acceptable. Despite the shortcomings indicated for the affective attitude and perceived control constructs, all other requirements were met satisfactorily. The measured variables were therefore all good indicators of their respective factors (latent variables). The final measurement model is shown in Fig 4.6.

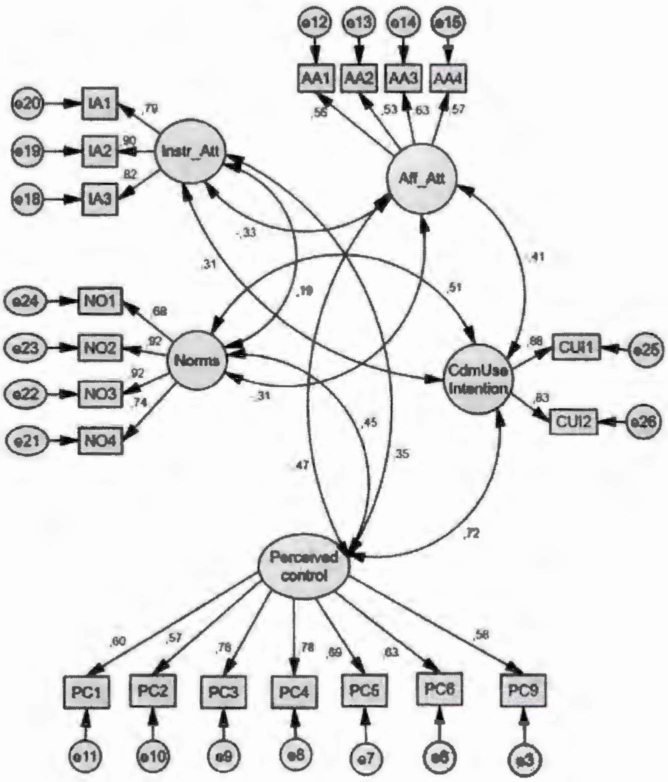


Figure 4.6 Final Measurement Model estimates

Figure 4.6 shows the standardised parameters (factor loadings and correlations) for the final measurement model. The measurement item loadings were all statistically significant ($p < 0.001$) and ranged from 0.53 (indicative of moderate strength) to 0.92 (indicative of high strength).

4.13 Discriminant validity

Discriminant validity is the extent to which a construct. This discriminant validity factors how much the construct correlates with others. Essentially the discriminant component lies in the distinctness that the variables only represent this single construct. The inter-correlation matrix as well as AVE were used to assess discriminant validity. Values are displayed in Table 4.10.

Table 4.10 Discriminant validity for the measurement model

	1	2	3	4	5
1. Aff_Att	**0.572**				
2. Instr_Att	-0.333	**0.841**			
3. Norms	-0.311	0.189	**0.822**		
4. Preceived_control	-0.473	0.347	0.451	**0.666**	
5. CdmUse_Intention	-0.413	0.306	0.513	0.724	**0.858**

Off diagonal values in Table 4.10 show correlations between the pairs of latent variables. As shown in the table correlations range from 0.189 to 0.724, suggesting that collinearity is not an issue in this model since all the correlations are below 0.85. All correlations were statistically significant ($p < 0.001$). Diagonal elements (in bold) show the square root of the AVE for a given construct. Discriminant validity was measured by comparing these values with the inter-construct correlations. If the off-diagonal elements are less than the square-root of the AVE in the corresponding rows and columns then discriminant validity is achieved (Fornell, Tellis & Zinkhan, 1982). As can be seen in the above-mentioned table, all inter-construct correlations are less than the square root of the AVE. Overall, these results confirm the existence of discriminant validity of the measurement used in this study.

Based on the evidence of the model fit indices and confirmation of convergent and discriminant validity, it was therefore concluded that the re-specified hypothesised measurement model was acceptable. The results from the confirmatory factor analysis supported further use of the measurement model as part of the structural model. Before proceeding to the structural model, data was assessed for normality and the results are presented in subsequent section.

4.14 Assessment of normality of the data

Fig 4.7 shows a plot of the chi-square percentiles against the ordered mahalanobis distances. The points in the plot do not follow a straight line thus the data does not follow a multivariate normal

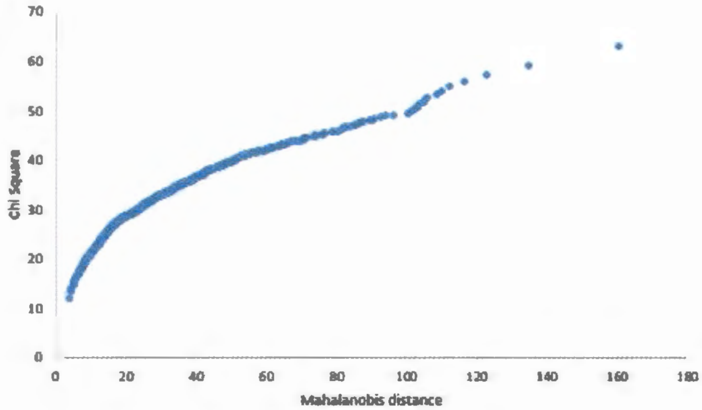


Figure 4.7 Chi-square probability plot

distribution. Additional valuable information which is evident from this plot is the presence of a multivariate outlier. Besides the graphical approach, normality assessment was also done using AMOS 25.0. Table 4.11 shows the normality assessment output obtained from the "test for normality and outliers" procedure in AMOS 25.0.

Table 4.11 Assessment of normality results

Variable	min	max	skewness	c.r.	kurtosis	c.r.
NO1	1,000	5,000	-1,451	-16,688	1,583	9,103
NO2	1,000	5,000	-1,326	-15,256	0,658	3,782
NO3	1,000	5,000	-1,242	-14,285	0,443	2,546
NO4	1,000	5,000	-1,202	-13,831	0,673	3,872
IA1	1,000	5,000	-1,948	-22,414	3,516	20,224
IA2	1,000	5,000	-2,020	-23,238	4,189	24,093
IA3	1,000	5,000	-2,115	-24,326	5,152	29,633
CUI2	1,000	5,000	-1,872	-21,531	3,992	22,963
CUI1	1,000	5,000	-1,876	-21,582	3,827	22,014
AA4	1,000	5,000	0,905	10,414	0,143	0,824
AA3	1,000	5,000	1,246	14,329	1,240	7,130
AA2	1,000	5,000	0,501	5,767	-0,807	-4,642
AA1	1,000	5,000	1,517	17,452	1,527	8,785
PC1	1,000	5,000	-1,195	-13,746	0,881	5,066
PC2	1,000	5,000	-0,983	-11,313	0,197	1,130
PC3	1,000	5,000	-1,646	-18,934	3,228	18,566
PC4	1,000	5,000	-1,544	-17,760	3,044	17,506
PC5	1,000	5,000	-1,229	-14,143	1,284	7,384
PC6	1,000	5,000	-0,985	-11,330	0,585	3,365
PC9	1,000	5,000	-1,436	-16,520	2,337	13,442
Multivariate					310,268	147,359

The univariate normality assessment was done by evaluating the measure of skewness and kurtosis for each item. The skewness and kurtosis indices showed acceptable ranges based on Kline's (2011) recommendations that the skewness and kurtosis indices should not exceed |3| and |10| respectively to ensure normality of the data. The data in this study thus satisfied the univariate normality assumption. However, since the multivariate kurtosis statistic exceeded c.r. (147,359), the assumption of multivariate normality was not met. This is in agreement with the results from the graphical analysis in the preceding section.

4.15 Structural Model

Given that the multivariate normality assumption was not met, both the ML (Fig 4.8) and ULS estimation methods were applied in fitting the structural equation model.

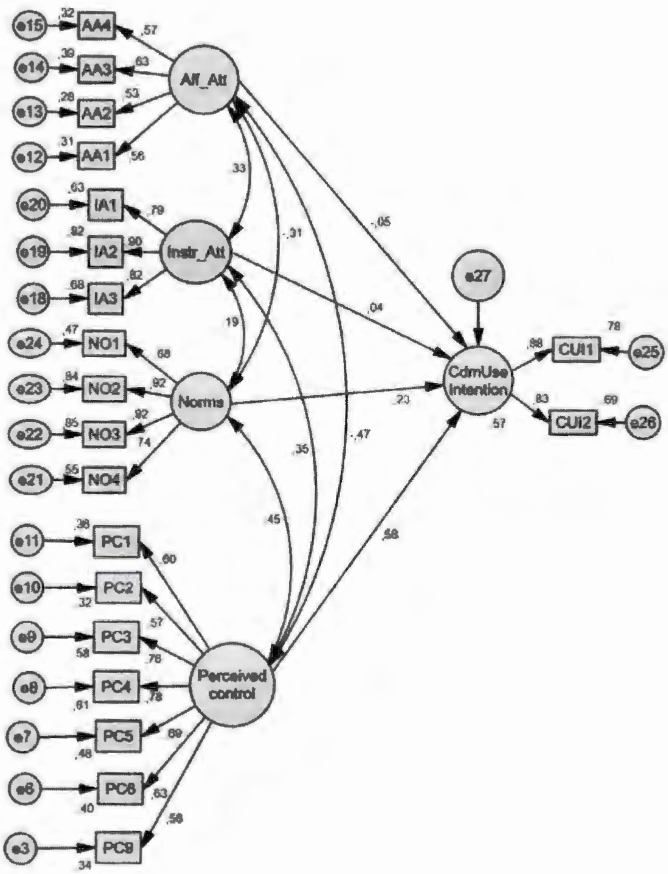


Figure 4.8 The Structural Model using MLE method

Fig 4.8 presents all the standardised coefficients and correlations between the *exogenous variables perceived control, norms, instr_att and aff_att and the endogenous variable, cdmuse intention* in the full structural model. The output in Fig 4.8 indicates that 57% of condom use intention coud be estimated by using the foregoing four exogenous constructs in the model. Table 4.12 shows the unstandardised, standardised and significance levels for the model in Fig 4.8.

Table 4.12 Unstandardised, Standardised and Significance Levels for Structural Model using MLE method (Standard errors in brackets; N = 794

Parameter Estimate	Unstandardised	Standardised	P
CdmUse_Intention <— Aff_Att	-0.068 (0.057)	-0.053	0.229
CdmUse_Intention <— Instr_Att	0.049 (0.038)	0.044	0.201
CdmUse_Intention <— Norms	0.211 (0.033)	0.226	***
CdmUse_Intention <— Perceived_control	0.676 (0.053)	0.582	***

*** $p < .001$

Table 4.12 shows that *Norms* (β = 0.226) and *Perceived control* (β = 0.582) had a significant influence on intention to use condoms. Both preceding predictors were significant at the 0.001 level ($p < 0.001$). *Aff_Att* and *Instr_Att* did not have a significant influence on condom use intention. Loehlin (2004) advises against prescriptively leaving out all paths that are non-significant. Goodboy & Kline (2017:73), concurred and added that "if a path was theoretically justified in the first place, then it should be retained regardless of whether its coefficient is significant or not significant." Thus the structural model depicted in Fig 4.8 was not modified in any way. Substituting the values of the unstandardised coefficients in equation (3.36) gives

$$\eta = -0.068Aff\_Att + 0.049Instr\_Att + 0.211Norms + 0.676Perceived\ control \tag{4.1}$$

4.16 Structural equation model using ULS method

Since the multivariate normality assumption was not met as indicated in section 4.14, the recommendation of Yuan and Bentler (1998) was taken into consideration. The ULS estimation method which does not assume normality was applied and then the results obtained were compared with those found using the ML estimation method. The results obtained from the ULS estimation are shown in Fig 4.9

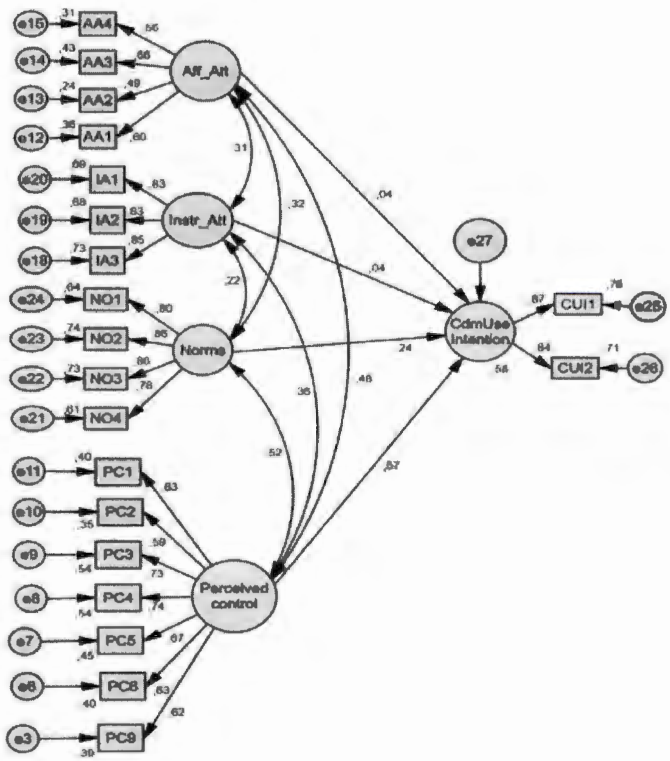


Figure 4.9 The Structural Model using ULS method

Fig 4.9 presents all the standardised coefficients and correlations between the exogenous variables *perceived control*, *norms*, *instr_att* and *aff_att* and the endogenous variable, cdmuse intention in the full structural model. The output in Fig 4.9 indicates that 58% of condom use intention coud be estimated by using the foregoing four exogenous constructs into the model. Table 4.13 shows the unstandardised and standardised parameter estimates for the model in Fig 4.9.

Table 4.13 Unstandardised, Standardised and Significance Levels for Structural Model using ULS method (Standard errors in brackets; N = 794

Parameter Estimate	Unstandardised	Standardised
CdmUse_Intention <—Aff_Att	-0.048	-0.041
CdmUse_Intention <—Instr_Att	0.046	0.043
CdmUse_Intention <—Norms	0.208	0.239
CdmUse_Intention <— Perceived_control	0.693	0.569

It is worth noting that as shown in Table 4.13, when the ULS estimation method is applied, neither the standard errors nor p-values are calculated in AMOS 25.0. Similary, PROC CALIS provides neither standard errors nor test statistics when the ULS approach is used. A comparison of the standardised parameter estimates shown in Tables 4.12 and 4.13 revealed minor variations in the values obtained. Estimates obtained using the MLE method, except for the *Norms* coefficient (= 0.226) were generally higher than those from the ULS approach. The total variation explained by the model using the ULS estimation method was however slightly higher (58%) than the one explained in the model using the MLE (57%).

4.17 SEM Bootstrapping Results

As indicated in section 3.6.1.3, reconfirmation of the structural equation model results was done through bootstrapping. Bootstrapping was carried out using 1000 samples and the results are displayed in Table 4.14.

Table 4.14 Bootstrap Results

(n = 794, B = 1000)

Parameter Estimate	Mean of estimates	S.E.	Bias	*P*	95% BC Confidence Interval
CdmUse_Intention <— Aff_Att	-.068	.068	.000	.256	(-.219, .058)
CdmUse_Intention <— Instr_Att	.050	.040	.002	.191	(-.027, .134)
CdmUse_Intention <— Norms	.209	.045	-.002	.001	(.133, .314)
CdmUse_Intention <— Perceived_control	.675	.072	-.001	.001	(.546, .832)

A comparison of the unstandardised parameter estimates shown in Tables 4.12 and 4.14 revealed minor variations in the values obtained. Estimates obtained using bootstrapping, except the Aff_Att coefficient (B = 0.068) were generally higher than those from the MLE approach. Standard errors obtained after performing bootstrapping were slightly larger than those obtained under the MLE approach. Inspection of the p-values showed similar significant results for all variables.

4.18 Original Multiple Regression Model: Composite variables and their descriptive statistics

Composite variables for use in the multiple regression model were obtained by taking the mean score of the variables making up the composite variable. Before running the multiple linear regression model, the descriptive statistics of *Aff_Att*, *Instr_Att*, *Norms*, *Perceived control* and *CdmUse_Intention* was conducted on the 794 participants. Table 4.15 shows the composite variables and their reliability and descriptive statistics values.

Table 4.15 Reliability and descriptive statistics for multi[le regression model variables

Composite Variable	Cronbach alpha	Mean	Std. Deviation	Skewness	Kurtosis
Aff_Att	0.661	1.93	0.75	0.47	-0.23
Instr_Att	0.876	4.39	0.84	-1.90	4.29
Norms	0.890	4.08	1.02	-1.26	1.05
Perceived_control	0.842	4.12	0.70	-0.91	1.33
CdmUse_Intention	0.847	4.41	0.81	-1.67	3.22

The reliability values for the composite variables were the same as those obtained for the variables in the final measurement model. Variable mean scores for *Instr_Att*, *Norms*, *Perceived control* and *CdmUse Intention* were all above 4 indicating more positive dispositions. *Aff_Att* had mean scores below the midpoint of 3. The statistics observed were consistent with those obtained for the measurement model. The standard deviations for all variable mean scores ranged from 0.70 to 1.02. The skewness and kurtosis indices showed acceptable ranges based on Kline's (2011) recommendations that the skewness and kurtosis indices should not exceed |3| and |10| respectively to ensure normality of the data. The data in this study was therefore regarded as univariate normal and excluding influential cases or outliers.

4.19 Linearity Assessment Results

The scatter plots in Figures 4.10, 4.11, 4.12 and 4.13 show the linear relationship between the DV and IVs.

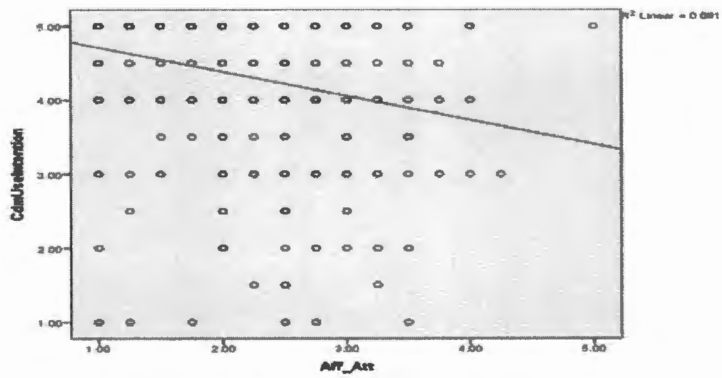


Figure 4.10 CdmUse_Intention against Aff_Att

Fig 4.10 shows a negative relationship between *CdmUse_Intention* and *Aff_Att.* Since the correlation between the two variables is significant, we conclude that there is a linear relationship between *CdmUse_ Intention* and *Aff_Att.*

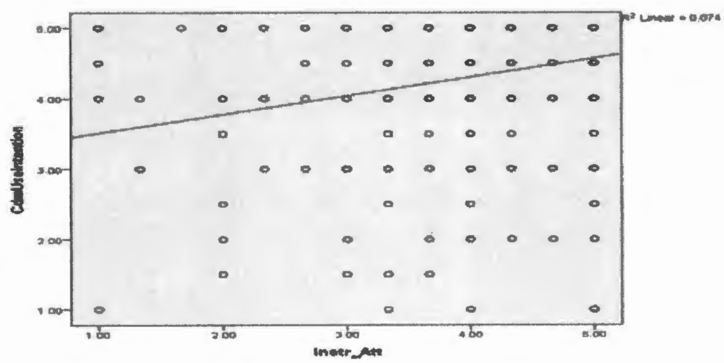


Figure 4.11 CdmUse_Intention against Instr_Att

Fig 4.11 shows a positive relationship between *CdmUse_Intention* and *Instr_Att.* Since the correlation between the two variables is significant, we conclude that there is a linear relationship between *CdmUse_Intention* and *Instr_Att.*

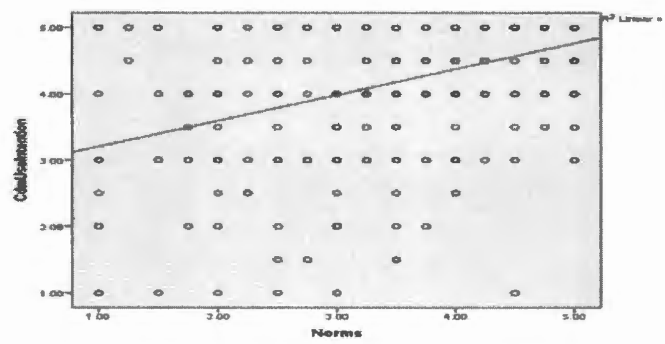


Figure 4.12   CdmUse_Intention against Norms

Fig 4.12 shows a positive relationship between *CdmUse_Intention* and *Norms.* Since the correlation between the two variables is significant, we conclude that there is a linear relationship between *CdmUse_Intention* and *Norms.*

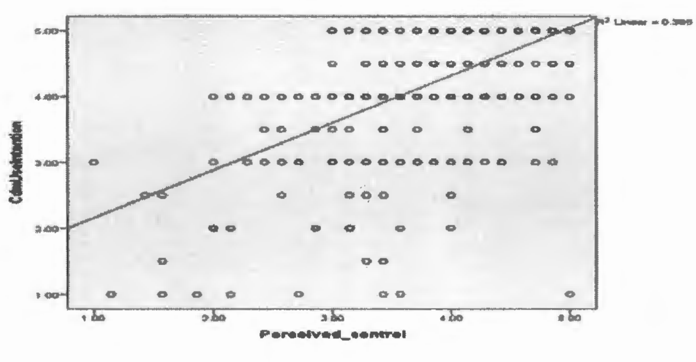


Figure 4.13   CdmUse_Intention against Perceived_control

Fig 4.13 shows a positive relationship between *CdmUse_Intention* and *Perceived_control.* Since the correlation between the two variables is significant, we conclude that there is a linear relationship between *CdmUse_Intention* and *Norms.* Based on the result of this assessment, there was no violation of the linearity assumption, as a result all IVs were retained in the model as they were significantly correlated with the dependent variable.

Table 4.16 shows the correlation matrix.

Table 4.16 Correlation Matrix

	1	2	3	4
1. Aff_Att	-			
2. Instr_Att	-0.239**	-		
3. Norms	-0.249**	0.202**	-	
4. Perceived control	-0.359**	0.314**	0.453**	-
5. CdmUse Intention	-0.302**	0.273**	0.490**	0.621**

The results indicated that there was significant correlation ($p < 0.01$) between all the variables. The strongest correlation relationship was between *Perceived control* and *CdmUse_Intention* (r = 0.621). *Aff_Att* and *Instr_Att* were moderately correlated with *CdmUse_Intention* whilst *Norms* and *Perceived_control* displayed strong correlation with the DV. Although the correlation coefficients were not exactly the same as in the SEM model, they did not differ much and had the same signs. *Aff_Att* was negatively correlated with all variables in both models.

4.20 Multiple Linear Regression Analysis Model

Since the initial checks did not indicate any problems with the data, fitting of the regression model was done. Table 4.17 shows the SPSS output of the model summary.

Table 4.17 Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.669[a]	.447	.445	.60329

a. Predictors: (Constant), Perceived_control, Instr_Att,Aff_Att, Norms

b. Dependent Variable: CdmUse_Intention

Results displayed in Table 4.17 show that 44.7% of the variation in *CdmUse_Intention* was explained by the variables *Perceived_control*, *Instr_Att*, *Aff_Att* and *Norms*.

Table 4.18 shows the MR model ANOVA results.

Table 4.18 Regression Model ANOVA Output

Source of variation	Sum of Squares	df	Mean Square	F	Sig.
Regression	232.592	4	58.148	159.763	.000
Residual	287.167	789	.364		
Total	519.760	793			

The ANOVA output shows that the regression model was significant($F(4, 789) = 159.763$, $p < 0.001$). Thus the regression model predicts *CdmUse_Intention* better than the mean of the response or the intercept-only model.

Table 4.19 shows the SPSS output of the multiple regression model coefficients.

Table 4.19 Multiple Regression Model Coefficients (n = 794)

Independent Variables	Unstandardised Coefficients		Standardised Coefficients	t	Sig.
	B	Std. Error			
Constant	1.212	0.195		6.205	.000
Aff_Att	-0.061	0.031	-0.056	-1.96	0.05
Instr_Att	0.06	0.027	0.062	2.192	0.029
Norms	0.201	0.024	0.253	8.452	.000
Perceived_control	0.542	0.037	0.467	14.738	.000

Intention to use condoms was statistically significantly predicted by instrumental attitudes (*Instr_Att*) ($\beta = 0.062$, $p < 0.05$), subjective norms (*Norms*) ($\beta = 0.253$, $p < 0.001$) and perceived behavioural control (*Perceived_control*) ($= 0.467$, $p < 0.001$). *Aff_Att* ($\beta =$ -0.056, $p = 0.05$) was not a statistically significant predictor of condom use intention. The final predictive regression model on substituting unstandardized estimates in equation (3.74) gives:

$$CdmUse\_Intention = 1.212 - 0.061Aff\_Att + 0.060Instr\_Att + 0.201Norms + 0.542Perceived\_control \tag{4.2}$$

4.21 Diagnostics test Results

Multicollinearity was further confirmed through the collinearity diagnostics while independence, normality of residuals and homoscedasticity assumptions were further checked through the plots in Fig 4.14, Fig 4.15 and Fig 4.16.

4.21.1 Multicollinearity Assessment

Multicollinearity was assessed using tolerance and VIF values, which were obtained as collinearity diagnostic results when using SPSS. The values are displayed in Table 4.20 below:

Table 4.20 Collinearity Statistics

	Collinearity Statistics	
	Tolerance	VIF
1. Aff_Att	0.846	1.118
2. Instr_Att	0.880	1.136
3. Norms	0.784	1.276
4. Perceived_control	0.698	1.432

As shown in Table 4.20 all the tolerance values > 0.2 while the VIF values are all below 10 thus indicating that there were no multicollinearity issues among the variables. Furthermore, the correlation matrix displayed in Table 4.16 confirmed the absence of multicollinearity issues among the variables. The zero-order correlations ranged from 0.202 to 0.621 and thus were all below 0.85.

4.21.2 Linearity, Normality, Homoscedasticity and Heteroscedasticity of Residuals

As indicated in section 3.8, examination of residuals scatterplots provides a test of assumptions of linearity, normality and homoscedasticity between predicted DV scores and errors of prediction.

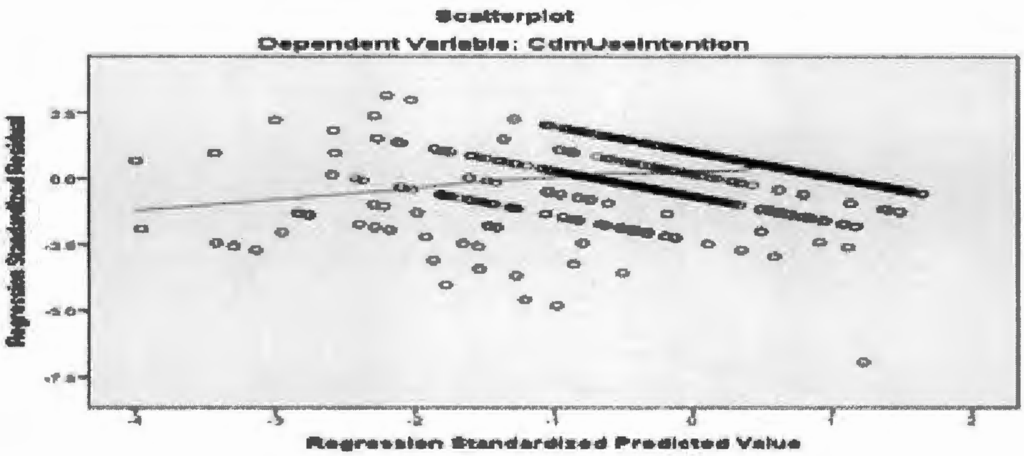


Figure 4.14 Scatterplot

Fig 4.14 shows the scatterplot of standardised residuals against standardised predicted values fitted with a loess line. The loess line fits Cohen et al. (2003)'s description of "a young child's freehand drawing of a straight line" thus confirming that the linearity assumption is met. Although the residuals are spread out more between -2 and 0 for the predicted value, the difference in spread is slight. Visual inspection of the scatterplot therefore suggests that heteroscedasticity is not a problem. The scatterplot, however, seems to suggest some departure from normality indicated by points lying around -5 and the single point lying at about -7.5 when considering vertical axis. The residuals histogram and P – P plot will probably give a better conclusion regarding the normality assumption.

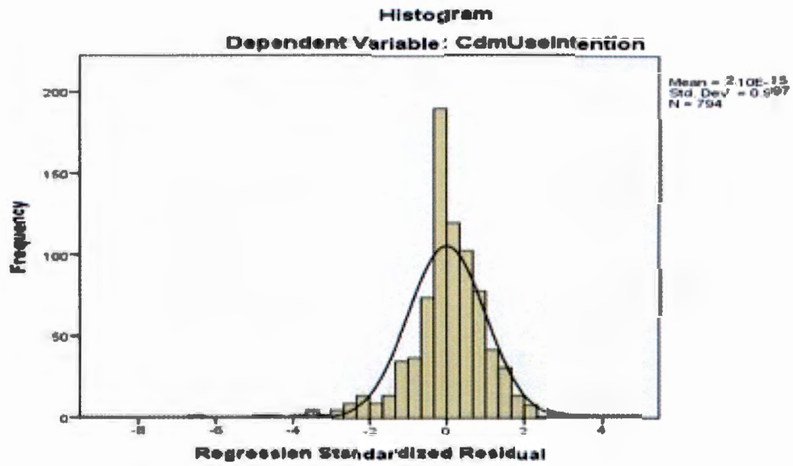


Figure 4.15 Scatterplot

Fig 4.15 shows the histogram of the regression standardised residuals with a normal curve overlay. Except for a few outliers to the left of the distribution, the histogram gives an indication of a symmetrical distribution.

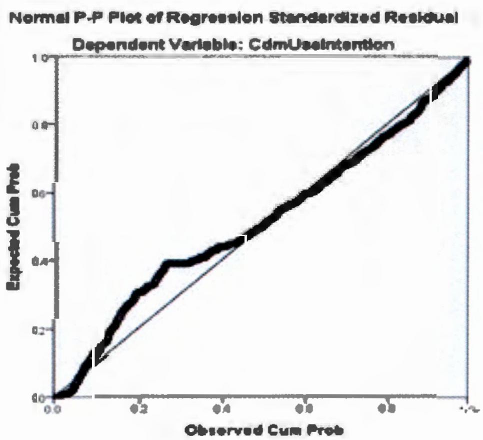


Figure 4.16 Normal Probability Plot

Fig 4.16 shows a normal P-P plot of regression standardised residuals. Although a substantial amount data in the P-P plot for the expected cumulative probability and observed cumulative probability seems to fall on the straight line, it is noticeable that taken as a totality, the residuals do not seem to be normally distributed. The P – P plot made it easier to confirm departure from normality of the residuals since it is easier to spot deviation from a straight unlike in the case of the normal curve (Cohen et al., 2003).

4.22 MR Bootstrapping Results

Bootstrapping was carried out using 1000 samples and the results are displayed in Table 4.21.

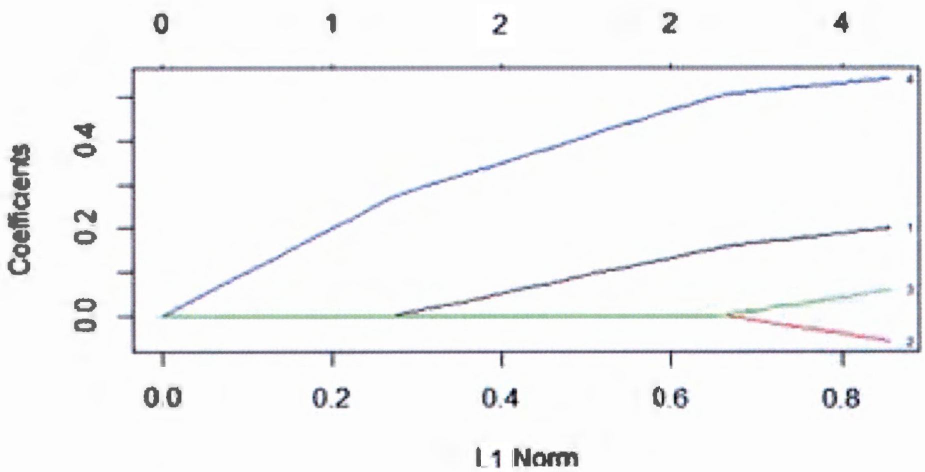
Table 4.21 MR Bootstrap Results (n = 794, B = 1000)

Parameter Estimate	Mean of estimates	S.E.	Bias	*P*	95% BCa Confidence Interval
Constant	1.212	.266	.004	.001	(.710, 1.723)
Aff_Att	-.061	.037	-.002	.108	(-.130, .006)
Instr_Att	.060	.031	.002	.052	(.004, .128)
Norms	.201	.032	.000	.001	(.141 .264)
Perceived_control	.542	.043	-.002	.001	(.456, .624)

Intention to use condoms was statistically significantly predicted by subjective norms (*Norms*) (B = 0.201, p = 0.001) and perceived behavioural control (*Perceived_control*) (B = 0.542, p = 0.001). Instrumental attitudes (*Instr_Att*) (B = 0.06, p = 0.052) and *Aff_Att* (B = -0.061, p = 0.108) were not statistically significant predictors of condom use intention.

4.23 LASSO Variable Selection Results

The Glmnet function in R was applied in conducting LASSO regression. Fig 4.17 and Fig 4.18 below show the results of the variable selection as well as determination respectively.



Key 1 – *Norms* 2 – *Aff_Att* 3 – *Instr_Att* 4 – *Perceived_Control*

Figure 4.17 Glmnet: all variables

Analysis of the plot in Fig 4.17 reveals that *Perceived_Control* most influences the model since it enters the model first and positively affects the response variable. The second important variable is *Norms* and it likewise affects the response variable positively. Furthermore, the plot shows that *Aff_Att* is the only variable that negatively affects the response variable.

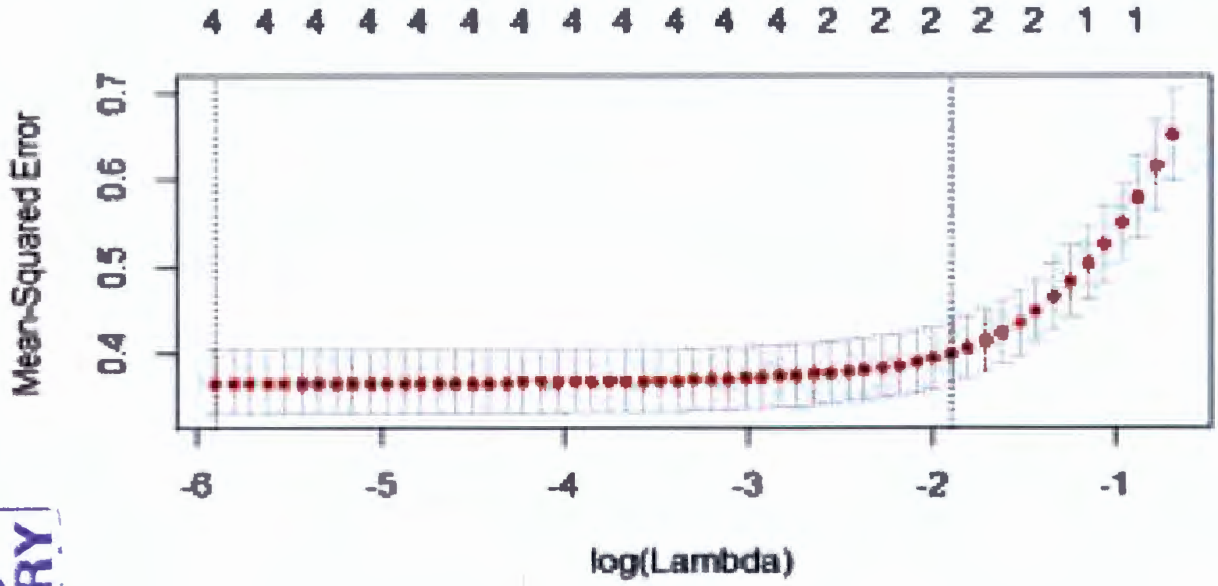


Figure 4.18 Cross Validation

Fig 4.18 shows the plot of the cross-validation fit results. It includes the cross-validation curve represented by the red dotted line as well as upper and lower standard deviation curves along the sequence (error bars). The vertical dotted line shows the selected λ. From the R output, $\lambda = 0.00274$ gives the minimum mean cross-validated error.

Table 4.22 shows the JMP Pro 13 output of the LASSO regression model coefficients

Table 4.22 LASSO Regression Model Coefficients (n = 794)

Term	Estimate	S.E.	Wald Chi-Square	*P*	95% CI
Intercept	1.212	0.195	38.507	.000	(.829, 1.595)
Aff_Att	-0.061	0.031	3.843	.050	(-.122, -.000)
Instr_Att	0.060	0.027	4.805	.0284	(.006, .113)
Norms	0.201	0.024	71.444	.000	(.154, .247)
Perceived_control	0.542	0.037	217.194	.000	(.470, .614)

Results for the LASSO regression model displayed in Table 4.22 show that intention to use condoms (CdmUse_Intention) was statistically significantly predicted by instrumental attitudes (Instr_Att) (B = 0.060, $p < 0.05$), subjective norms (Norms) (B = 0.201, $p < 0.001$) and perceived behavioural control (Perceived_control) (B = 0.542, $p < 0.001$). Aff_Att (B = -0.061, $p = 0.05$) was not a statistically significant predictor of condom use intention. The predictors explained 44.7% of the variation in CdmUse_Intention.

Regression analysis was also performed using GAM provided by PROC GAM procedure of the SAS software (release 9.4, SAS Institute incorporated, Cary, NC, USA) to assess the effects of the psycho-social factors on adolescent condom use intention. The model used condom use intention as the dependent variable and the four factors: affective attitude, instrumental attitude, normative beliefs and perceived control as the independent variables. The 'spline' function was utilised in the MODEL statement to request the additive model using a cubic smoothing spline with four degrees of freedom by default, in SAS software, for each psycho-social factor.

4.24 GAM Results

Table 4.23 GAM Regression Model Analysis Model Fit Statistics (n = 794)

Parameter	Estimate	S.E.	ft -value	**P-value**
Intercept	1.281	0.190	6.76	<.0001
Linear(Aff_Att)	-0.0398	0.0302	-1.32	0.1844
Linear(Instr_Att)	0.0611	0.0265	2.31	0.0211
Linear(Norms)	0.203	0.0231	8.79	<.0001
Linear(Perceived_control)	0.511	0.0357	14.33	<.0001

Table 4.24 Fit Summary for Smoothing Components (n = 794)

Component	Smoothing Parameter	DF	GCV	Num Unique Obs
Spline(Aff_Att)	0.980	3.00	0.366	15
Spline(Instr_Att)	0.953	3.00	0.116	13
Spline(Norms)	0.969	3.00	0.0900	17
Spline(Perceived_control)	0.997	3.00	0.463	27

Table 4.25 Approximate Analysis of Deviance

Source	DF	F value	Pr > F
Spline(Aff_Att)	3.00	2.90	0.0344
Spline(Instr_Att)	3.00	2.51	0.0575
Spline(Norms)	3.00	6.86	0.0001
Spline(Perceived_control)	3.00	11.25	<.0001

Table 4.23 provides analytical information about the fitted model, including parameter estimates for the linear portion of the model. Values displayed in Table 4.23a indicate that intention to use condoms (CdmUse_Intention) was statistically significantly predicted by instrumental attitudes (Instr_Att) (B = 0.061, $p < 0.05$), subjective norms (Norms) (B = 0.203, $p < 0.0001$) and perceived behavioural control (Perceived_control) (B = 0.511, $p < 0.001$). Aff_Att (B = -0.0398, $p = 0.18$) was not a statistically significant predictor of condom use intention. The predictors explained 47.9% of the variation in CdmUse_Intention. The final predictive GAM regression model equation is:

$$CdmUse\_Intention = 1.281 - 0.0398Aff\_Att + 0.061Instr\_Att + 0.203Norms + 0.511Perceived\_control \tag{4.3}$$

Smoothing parameters, degrees of freedom, the number of unique observations and the value of the generalised cross validation (GCV) are displayed in Table 4.24. One of the important parts of the PROC GAM results is the "analysis of deviance" displayed in Table 4.25. Table 4.25 presents F statistics for comparing the deviance between the full and reduced model (excluding non-parametric component). For this model, the analysis effects in three of the four continuous predictors were found to be significant at the 5% level of significance as shown by their corresponding p-values which were less than 0.05.

Fig 4.19 shows plots of the partial prediction for each of the continuous predictors considered in this study.

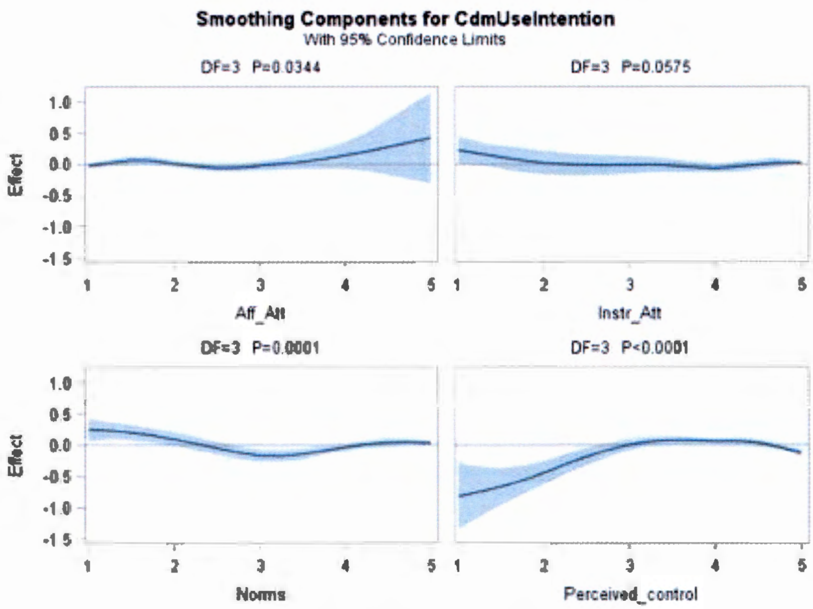


Figure 4.19 Partial prediction for each predictor

Fig 4.19 results from predictions from the non-linear part that are plotted against the respective predictor variables. Thus the figure above portrays the non-linear part of the relationship between CdmUse_Intention and each of the predictors, one at a time. The plots include a curve-wise Bayesian 95% confidence band for each smoothing component.

4.25 Summary

This chapter presented data analysis and results. Anderson & Gerbig's (1988) two-step approach was applied to the TPB based condom use intention structural equation model. The approach involves separate estimation of the measurement model before the estimation of the structural model. Prior to the estimation of parameters, data screening was carried out. Part of the data screening involved analysis of missing data. Listwise deletion was applied to the dataset since the proportion of missing data was very low($< 1\%$). Cases with unengaged responses were also deleted. Based on Stevens' (2009) suggestion of at least 15 cases per observed variable or indicator, the sample size in this study was found to be adequate. Descriptive statistics for the variables of interest were computed. All SEM assumptions except the multivariate normality assumption were met.

The measurement model was found to be over-identified ($df > 0$) and thus could be tested. Deletion of factors with standardised factor loadings < 0.5 led to the re-specification of the measurement model. Acceptable model fit was achieved without the use of modification indices. Discriminant and convergent validity for the constructs was confirmed. Parameters in the structural equation model were then estimated using an MLE procedure. Results from the SEM model revealed that norms and perceived control were significant predictors of intention to use condoms. The results obtained were confirmed using bootstrapping.

The SPSS linear regression procedure was used to estimate parameters in the multiple regression model. Prior to the regression models, composite variables were computed. Descriptive statistics for the composite variables were also computed. Data was examined for normality using both frequency histograms and measures of kurtosis and skewness. The two approaches were however not in agreement. All the variables showed that they were linearly related to the outcome variable (*CdmUse_Intention*). No multicollinearity problems were present among the variables. The Glmnet procedure for LASSO regression in R and the generalised regression option in JMP Pro 13 were used for variable selection as well as fitting the LASSO regression model. GAM regression was done using the GAM procedure in SAS as well as the *mgcv* and *gam* procedures in R. Results from all three models revealed that norms, perceived behavioural control and instrumental attitudes were significant predictors of intention to use condoms while affective attitudes were non-significant.

CHAPTER 5. SUMMARY,DISCUSSION AND CONCLUSIONS

5.1 Introduction

Chapter 4 presented the data and proceeded to analyse this data. This current chapter discusses the findings emanating from the data. It proceeds to establish implications for practice, pointing out the limitations in the process. The chapter terminates with conclusions and recommendations for further research. Most significantly, the chapter highlights conclusions of the entire study by reassessing the primary research questions in light of the data.

5.2 Discussion of the Findings

Research Question 1:

Which model, between the structural equation model, MR, LASSO regression model and GAM is more adequate for explaining Batswana in-school adolescents' intentions to use condoms?

Finding 1:

The structural equation model was more adequate for explaining Batswana in-school adolescents' intentions to use condoms than the GAM, MR and LASSO regression models. The structural equation model explained a higher proportion of the variance (57%) in condom use intention prediction than the GAM (47.9%), MR model and LASSO regression model which both explained 44.7% of the variance in condom use intention respectively. This finding is consistent with other studies (Boer & Mashamba, 2007; Eggers et al., 2016; Fazekas et al., 2001; Giles et al., 2005) conducted in African countries, which also revealed subjective norms and perceived behavioural control to be significant predictors of intention to use condoms. The total variance in intention to use condoms explained in these other studies ranged between 22% and 67%. The parameter estimates for the linear portion of the GAM were as good as the parameter estimates obtained in the MR model as well as the LASSO regression model. Table 5.1 shows a comparison of the MLE and bootstrap SEM results.

Table 5.1 SEM Maximum Likelihood Estimator and Bootstrap Result Comparison

	Maximum Likelihood Estimator			Bootstrap		
	Estimate	Std. Error	*P*	Estimate	Std. Error	*P*
CdmUse_Intention <— Aff_Att	-0.068	0.057	0.229	-0.068	0.068	0.256
CdmUse_Intention < — Instr_Att	0.049	0.038	0.201	0.050	0.040	0.191
CdmUse_Intention < — Norms	0.211	0.033	***	0.209	0.045	0.001
CdmUse_Intention <— Perceived_control	0.676	0.053	***	0.675	0.072	0.001

*** $p < .001$

A comparison of the unstandardised parameter estimates shown in Table 5.1 above revealed minor variations in the values obtained. Equal estimates for the *Aff_Att* coefficient (B = 0.068) were obtained for both the MLE and bootstrap approach. *Norms* and *Perceived_control* coefficient estimates obtained using the bootstrap were slightly higher than those from the MLE approach. A little lower *Instr_Att* coefficient estimate was obtained using the MLE method than when bootstrapping was applied. Standard errors obtained from performing bootstrapping were higher than those obtained under the MLE approach. Inspection of the p-values showed similar significant results for *Norms* and *Perceived_control* and non-significant results for *Aff_Att* and *Instr_Att*. Despite the slight variations, both approaches provided the same results. Thus, the bootstrap results confirmed the stability of the results obtained using the MLE approach.

Research Question 2:

Which TPB elements contribute significantly to explaining Batswana in-school adolescents' condom use intentions?

To assist in answering the foregoing question the following hypotheses as stated in section 1. were tested:

H1: *Instrumental attitude* has a positive and significant effect on *condom use intention*

H2: Affective attitude has a negative and significant effect on *condom use intention*

H3: Normative beliefs have a positive and significant effect on *condom use intention*

H4: Perceived behavioural control has a positive and significant effect on *condom use intention*

Finding 2:

Results pertaining to the hypothesis tests are summarised in Table 5.2 below.

Table 5.2 The hypothesis statement for every path and its conclusion

Hypothesis statement		**Result**	**Decision**
H1	*Instrumental attitude* has a positive and significant effect on *condom use intention*	Significant in all models except the structural equation model and MR bootstrap	Supported in all models except the structural equation model and MR bootstrap
H2	*Affective attitude* has a negative and significant effect on *condom use intention*	Not significant in all models	Not supported in all models
H3	*Normative beliefs* have a positive and significant effect on *condom use intention*	Significant in all models	Supported in all models
H4	*Perceived behavioural control* has a positive and significant effect on *condom use intention*	Significant in all models	Supported in all models

On the basis of the results from this study, the first hypothesis (*H1*) was accepted in the MR, LASSO and GAM models but rejected under SEM, including in the SEM bootstrap as well as MR bootstrap. Results indicated that instrumental attitude has a positive but insignificant effect on condom use intention when SEM and MR bootstrapping are carried out and a positive significant effect under the remaining models. This inconsistency thus shows the importance of choice of analysis method and how it affects the interpretation of results. In addition, the obtained results indicated that affective attitude has a negative yet insignificant effect on condom use intention for all models considered in the study. The second hypothesis (*H2*) was therefore not accepted for all models.

Normative beliefs (Norms) were a significant predictor of condom use intention among Batswana in-school adolescents in all models including the bootstrap models. In this study, normative beliefs were indicative of the adolescents' perceptions of the significant others' (father, mother, partner and friends) approval or disapproval of them using condoms. Results in this study indicated that normative beliefs had a positive and significant effect on condom use intention. This led to the acceptance of the third hypothesis (*H3*). The finding of this study is supported by previous studies (Guo et al., 2014; Sacolo et al., 2013) but contrary to Teye-Kwadjo et al. (2017b) and Jemmott et al. (2007) who found that normative beliefs were not statistically significantly associated with condom use intention.

Perceived control, on the other hand, was the strongest predictor of condom use intention in all models. Results displayed in Table 5.1 show that perceived control, which in this study referred to Batswana in-school adolescents' perceptions of how easy or difficult it is for them to use condoms, had a positive and significant effect on condom use intention. Thus the fourth and final hypothesis (*H4*) was accepted. This finding is consistent with results from studies carried out among South African Xhosa-speaking high school youths (Jemmott et al., 2007), in-school youth in Swaziland (Sacolo et al., 2013), 9^{th} – 12^{th} grade Ghanaian senior high school students aged 14 – 20 years (Teye-Kwadjo et al., 2017b) and Chinese college students (Guo et al., 2014). However, the finding was in contrast to Albarracín et al. (2001)'s meta-analysis results of 96 studies mostly conducted in Europe and United States, which established that attitude was the best predictor of condom use followed by perceived control. Additionally, the result obtained in this study was inconsistent with

Bennett & Bozionelos (2000), whose study discovered that perceived control of condom use had no effect on condom use intention.

Although Batswana culture, history, and language differ from those in western countries, the study results indicated that the TPB constructs, except for both affective and instrumental attitude in the SEM models and affective attitude in the GAM, MR and LASSO models, were predictive of intention to use condoms among Batswana in-school adolescents. While the signs on the coefficients for both models were similar, there was a difference in the magnitude of the coefficients as well as in the amount of variance explained by the predictors in each model.

Research Question 3:

What suggestions can be made towards intervention programs formulation?

The hypotheses displayed in Table 5.2 showed that both normative and perceived behavioural control beliefs have a positive and significant effect on condom use intention.

The findings of this study revealed that affective attitudes toward condom use were not predictive of intention to use condoms among Batswana in-school adolescents. Affective attitudes toward condom use in this study denoted Batswana adolescents' negative feelings or thoughts about condom use. The insignificant influence for affective attitudes on intentions to use condoms may be related to myths concerning condom use among Batswana in-school adolescents such as assumptions that condoms reduce sex pleasure, use of male condoms results in reduced fertility and condoms cause pain. According to Appiah et al. (2017), there are numerous documented reasons for lack of condom use. These include anxieties about the safety of condoms as well as cultural perceptions that relate condom use to promiscuity and unfaithfulness or cheating (Adebiyi & Asuzu, 2009; Chimbiri, 2007; Fehr et al., 2015). Kennedy et al. (2007) added that the feeling of embarrassment in purchasing condoms, societal pall that covers condom purchase and misperceptions about their effectiveness have all promoted low usage in many parts of SSA.

5.3 Contribution of the study

To the best of the researcher's knowledge, no studies have compared the adequacy of SEM, LASSO regression, GAM and MR models involving latent variables. This study thus does not only reveal the importance of selection of appropriate analysis methodology especially when dealing with data that has latent variables but also shows the existence of a variety of under-utilised, useful and feasible modern regression techniques. Nusair & Hua (2010:320) suggested that "SEM is most appropriate when the researcher has multiple constructs, each represented by several measured variables, and the constructs are distinguished based on whether they are exogenous or endogenous." In addition, Nusai & Hua (2010) noted that in instances where research questions to address interactions between latent variables in a study are raised, SEM is undoubtedly a good choice. On the other hand, when censored, truncated, time-series or panel data are involved or research questions are related to probability, MR is preferred. Moreover, GAMs would be a good choice in instances where linearity is not assumed or model is not specified a priori. LASSO regression is very handy in instances where the researcher needs to select a few variables from a myriad variables.

Furthermore, there is a dearth in literature and research on correlates or predictors of condom use among Batswana adolescents. This study consequently expanded upon the growing body of HIV/AIDS prevention literature with a new focus on the choice of a relevant statistical analysis methodology. To the best of the researcher's knowledge, this is the first TPB guided study that not only applied modern regression techniques but also investigated factors underlying intentions for condom use among Batswana adolescents. The TPB could have critical implications for community programmes that make use of social cognitive interventions for Batswana in-school adolescents. Interventions utilising the TPB are probably going to be the best in changing or affecting intention (Webb & Sheeran, 2006). The results of this study contribute towards knowledge regarding intentions of youth in engaging in protective sexual behaviour as well as literature on effective HIV/AIDS prevention among in-school adolescents that can be used by policy makers as a basis for the development of culturally sensitive and effective interventions for Botswana and other Sub-Saharan countries.

5.4 Implications for Practice

To investigate relationships among variables, researchers use statistical methods. However, each statistical method has its own strengths and limitations thus necessitating the need for researchers to consider such when dealing with statistical models of complex social phenomena in their research (Jeon, 2015). The study finding on SEM being more adequate for explaining and predicting Batswana in-school adolescents' intentions to use condoms points to the importance of selecting and applying this data analysis or modelling approach. This study shows that statistical methods should not only be primarily chosen based on their simplicity but rather in light of the appropriateness and adequacy of the method with regard to the type or nature of data. Since MR, LASSO and GAMs make use of a single DV, measurement error occurs. The problem of measurement error is however reduced when SEM is used. This study thus shows that when estimating abstract concepts such as attitudes or perceptions towards a certain behaviour, researchers should take SEM as a more appropriate approach than multiple regression and other modern regression techniques especially when dealing with data consisting of latent variables.

A second important implication of this study derives from the finding on the TPB elements that contribute significantly to explaining Batswana in-school adolescents' condom use intentions. Results from this study suggest that effective interventions for promoting condom use should aim at changing normative and perceived behavioural control beliefs over condom use. Given that the content of the beliefs may vary across cultures (Kok & Ruiter, 2014), designers of behavioural intervention programs and policy makers need to also take the context of the targeted populations into account thereby leading to development of culturally appropriate interventions.

5.5 Limitations

The study has the following limitations:

1. Secondary data was used in this study thus the researcher was limited to working with the available variables in building the models.

2. Findings may not be generalizable to other adolescent groups such as those living in rural areas, out of school adolescents and college or university adolescents since the sample consisted only of in-school adolescents from urban and semi-urban areas. A systematic review of literature conducted by Stroeken et al. (2012) covering the period 2000 – 2010 on young people who dropped out of school in Southern and Eastern Africa showed that out-of-school youths were likely to engage in risky sexual behaviour, had early sexual debut and reported more inconsistent condom use. Agyei & Abrefa-Gyan (2016) suggested that young people's location and living arrangements play a significant role in determining whether or not youth engage in risky sexual behaviours. Mutinta & Govender (2012) and Mutinta et al. (2013), concur and state that the college or university environment affords an opportunity for risky sexual behaviours since students are free from parental or guardian control. Although the findings are not generalizable, the modelling framework discussed in the thesis is applicable to any group of adolescents.

3. Self-reports were used to measure behavioural variables. Due the sensitivity of the sexual behaviour as well as social desirability concerns, there is a risk that participants may have either under-reported or over-reported their behaviour. It is well documented in literature that youth, particularly females, regularly under-report sexual behaviour whereas males occasionally over-report it (Beguy et al., 2009; Doyle et al., 2012; Marston & King, 2006; Plummer & Wight, 2011). While bias may be present in self-report measures, specifically when private sexual information is requested, an emergent body of research has revealed that the use of self-report data in sexual behaviour research presents no major problems (DiClemente et al., 2013; Goldberg et al., 2014; Schroder et al., 2003). Notwithstanding this, there is a possibility that the self-report data may have had some influence on the analysis.

5.6 Conclusions

The purpose of this study was to compare the adequacy of the structurall equation model, the MR model, LASSO and GAM in predicting condom use intention using Batswana in-school adolescents sample data while validating the TPB in studying Batswana in-school adolescents' attitudes and personal beliefs toward their intention to use condoms. The predictors of interest were: (a) attitudes (both affective and instrumental attitudes) toward condom use, (b) subjective norms about condom use and (c) perceived behavioural control of condom use, all derived from the TPB. Many studies have been conducted using the TPB in western countries, but few studies that apply the TPB to predict the intention to use condoms in Sub-Saharan Africa, specifically among Batswana adolescent populations, have been carried out. This study therefore extended the application of the TPB to Batswana in-school adolescents in order to identify the predictors that determined the intention to use condoms in this population. The study also introduced the application of modern regression techniques in analysing TPB related data. The study used secondary data from 794 adolescents which was collected as part of an adolescent research project on HIV/AIDS prevention conducted by the University of Botswana.

The study focussed on three research questions:

1. Which model, between the structural equation model, MR, LASSO regression model and GAM is more adequate for explaining Batswana in-school adolescents' intentions to use condoms?

2. Which TPB elements contribute significantly to explaining Batswana in-school adolescents' condom use intentions?

3. What suggestions can be made towards intervention programs formulation?

The first two questions were answered quantitatively through the use of the structural equation model, MR model, LASSO regression model and GAM. Findings from the first and second questions were used to inform the suggestions solicited through question three. A stimulating finding of this study was that instrumental attitudes were a significant predictor of condom use intention in the three models but were insignificant in the SEM model. Additionally, contradicting decisions were arrived at regarding the first hypothesis (*H1*) with the hypothesis being accepted under the GAM, MR and LASSO regression models while it was rejected under SEM. Instrumental attitudes, in this study, referred to preventive benefits of condom use. Since the same items were used to measure instrumental attitudes, the difference in significance as well as decision concerning the first hypothesis can therefore be attributed to the effect of different analysis approaches.

5.7 Recommendations for Further Research

While this study found SEM to be more adequate than MR, LASSO regression and GAM, additional studies that compare these modelling approaches specifically where latent variable data is concerned, need to be carried out. Results from further studies may possibly clarify whether the variance in adequacy can exclusively be attributed to the specific method. On the basis of the finding on the significant predictors of condom use intention made by this study, there is need for constructs and variables well-matched to the Botswana context to be further developed and validated for future studies. Furthermore, the TPB could be used as a framework to determine the predictors of intention to use condoms among Batswana in-school adolescents. It is recommended that any HIV education programs or interventions targeted at the adolescents should increase the intention to use condoms through promoting positive instrumental attitudes, subjective norms and perceived control of condom use. Additionally, an investigation of mediating and/or moderating variables in the case of Batswana adolescents could be yet another avenue of research to be pursued. Developing separate models based on gender could also be a worthwhile research pursuit.

REFERENCES

Adebiyi, A. O., & Asuzu, M. C. (2009). Condom use amongst out of school youths in a local government area in Nigeria. *African Health Sciences, 9*(2).

Agresti, A. (2007). *An Introduction to Categorical Data Analysis.* John Wiley & Sons Inc.

Agyei, W.K.A., & Abrefa-Gyan, T. (2016). Risky sexual behaviour and condom use among youth in Botswana. *Journal of Human Ecologyl, 55*(3), 141 – 151.

Anderson, D. R., Burnham, K. P., White, G. C. (1998). Comparison of Akaike information criterion and consistent Akaike information criterion for model selection and statistical inference from capture-recapture studies. Journal of Applied Statistics, 25(2), 263-282.

Airhihebuwa, C.O., & Obregon, R. (2000). A critical assessment of theories/models used in health communication for HIV/AIDS. *Journal of Health Communication,* 5, 5–15.

Ajzen, I. (2012). Martin Fishbein's legacy: The reasoned action approach. *The Annals of the American Academy of Political and Social Science, 640*(1), 11–27.

Ajzen, I. (2011). The theory of planned behaviour: Reactions and reflections. *Psychology and Health, 26*(9), 1113–1127. doi: 10.1080/08870446.2011.613995.

Ajzen, I. (2006). Theory of Planned Behaviour Diagram. http://people.umass.edu/aizen/tpb.diag.html

Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behaviour and Human Decision Processes, 50,* 179 – 211.

Ajzen, I. (1988). *Attitudes, Personality, and Behaviour.* Dorsey Press, Chicago, IL.

Ajzen, I. (1985). From intentions to actions: A theory of planned behaviour. In *Action control* (pp. 11–39). Springer Berlin Heidelberg.

Ajzen, I., & Fishbein, M. (1980). Understanding attitudes and predicting social behaviour. Englewood Cliffs, NJ: Prentice-Hall.

Albracín, D., Johnson, B. T., Fishbein, M., & Muellerleile, P. A. (2001). Theories of reasoned action and planned behaviour as models of condom use: a meta-analysis. *Psychological bulletin, 127*(1), 142.

Alvarez, C., Villarruel, A. M., Zhou, Y., & Gallegos, E. (2010). Predictors of condom use among Mexican adolescents. *Research and Theory for Nursing Practice, 24*(3), 187–196.

Anderson, J. C., & Gerbing, D.W. (1988). Structural equation modelling in practice: A review and recommended two-step approach, *Psychological bulletin, 103*(3), 411.

Appiah, A.B., Tenkorang, E.Y., & Maticka-Tyndale, E. (2017). Modelling Beliefs, Attitudes, and Intentions of Condom Use among Secondary School Students in Kenya. *Archives of Sexual Behaviour*, 1–12.

Arbuckle, J. L. (2010). *IBM SPSS Amos 19 user's guide.* Crawfordville, FL: Amos Development Corporation, 635.

Arifin, W.N. (2015). The graphical assessment of multivariate normality using SPSS. *Education in Medicine Journal, 7*(2), 71 – 75. doi: 10.5959/eimj.v7i2.361.

Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British journal of social psychology, 40*(4), 471–499.

Aron, A., & Aron, E.N. (1999). *Statistics for psychology* (2^{nd} ed.). Upper Saddle River, NJ: Prentice Hall.

Astrachan, C.A., Patel, V.K., & Wanzenried, G. (2014). A comparative study of CB-SEM and PLS-SEM for theory development in family firm research. *Journal of family business strategy, 5*, 116 – 128.

Bagozzi, R. P. (2007). On the meaning of formative measurement and how it differs from reflective measurement: Comment on Howell, Breivik, and Wilcox (2007).

Bagozzi, R. P., & Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the academy of marketing science, 40(1), 8-34.*

Bajpai, N. (2011*). Business Research Methods.* Pearson Education. India.

Ballinger, G.A. (2004). Using generalised estimating equations for longitudinal data analysis. *Organisational Research Methods, 7*(2), 127 – 150. doi: 10.1177/1094428104263672

Bandura, A. (1994). Social cognitive theory and exercise of control over HIV infection. In *Preventing AIDS* (pp. 25–59). Springer, Boston, MA.

Bandura, A. (1986). Social foundations of thought and action. *Englewood Cliffs, NJ, 1986.*

Bandura, A. (1977). *Self-efficacy. The exercise of control.* New York: Freeman.

Barrett, P. (2007). Structural equation modelling: adjudging model fit. *Personality and Individual Differences*, 42, 815 – 824.

Beguy, J.A., Kabiru, C.W., Ndera, E.N., & Ngeware, M.W. (2009). Inconsistencies in self-reporting of sexual activity among young people in Nairobi, Kenya. *Journal of Adolescent Health, 45*(6), 595 – 601.

Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand journal of public health, 25*(5), 464–469.

Bennett, P., & Bozionelos, G. (2000). The theory of planned behaviour as predictor of condom use: a narrative review. *Psychology, Health & Medicine, 5*(3), 307–326.

Bentler, P. M. (1988). Causal modelling via structural equation systems. In *Handbook of multivariate experimental psychology* (pp. 317–335). Springer US.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin, 107*(2), 238.

Bentler, P.M. (2005). *EQS 6 structural equations program manual.* Enci-no. CA: Multivariate Software (www.mvsoft.com).

Bentler, P.M. (2010). SEM with simplicity and accuracy. *Journal of consumer psychology: the official journal of the Society for Consumer Psychology, 20*(2), 215.

Bentler, P.M., & Bonnet, D.C. (1980). Significance Tests and Goodness of Fit Analysis of Covariance Structures. *Psychological Bulletin, 88*(3), 588–606.

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modelling. *Sociological Methods & Research, 16*(1), 78–117.

Berry, W. D., & Feldman, S. (1985). Multiple Regression in Practice. Quantitative Applications in the Social Sciences 50.

Blalock, H.M. (1964). *Causal inferences in non-experimental research.* Chapel Hill: University of North Carolina Press.

Boer, H., & Tshilidzi-Mashamba, M. (2007). Gender power imbalance and differential psychosocial correlates of intended condom use among male and female adolescents from Venda, South Africa. *British journal of health psychology, 12*(1), 51–63.

Bollen, K. A. (2007). Interpretational confounding is due to misspecification, not to type of indicator: Comment on Howell, Breivik, and Wilcox (2007).

Bollen, K. (1989a). *Structural equations with latent variables.* New York: Wiley.

Bollen, K. A. (1989b). A new incremental fit index for general structural equation models. *Sociological Methods & Research, 17*(3), 303–316.

Bollen, K.A., & Long, J.S. (Eds.). (1993). *Testing structural equation models.* Thousand Oaks, CA: Sage.

Boomsma, A., & Hoogland, J.J. (2001). The robustness of LISREL modelling revisited. In R. Cudeck, S. du Toit, & D. Sorbom (Eds.), *Structural equation models: Present and future. A Festschrift in honour of Karl Jöreskog,* (pp 139 – 168). Chicago: Scientific Software International.

Botswana. Statistics Bostwana. (2013). *Botswana AIDS Impact Survey 2013 Statistical Report.*

Bowen, N. K., & Guo, S. (2011). *Structural equation modelling.* Oxford University Press.

Bowman, K., & Shenton, L. (1975). Omnibus test contours for departures from normality based on $\sqrt{b1}$ and b2. *Biometrika, 62*(2), 243–250.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology,* 37, 1–21.

Browne, M.W., & Cudeck, R. (1993). *Alternative ways of assessing model fit.* In K.A. Bollen & J.S. Long (eds.), *Testing structural equation models* (pp 136 – 16). Newbury Park, CA: Sage.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioural Research, 24,* 445–455.

Bryan,A., Fisher, J.D., & Fisher, W.A. (2002). Tests of the Mediational Role of Preparatory Safer Sexual Behaviour in the Context of the Theory of Planned Behaviour. *Health Psychology, 21*(1), 71-80.

Bryan, A., Kagee, A., & Broaddus, M. R. (2006). Condom use among South African adolescents: Developing and testing theoretical models of intentions and behaviour. *AIDS and Behaviour, 10*(4), 387.

Bryman, A. (2012). *Social research methods* (5^{th} ed.). Oxford: Oxford University Press.

Bryman, A., & Cramer, D. (2005), *Quantitative Data Analysis with SPSS 12 and 13: A guide for social scientists.* East Sussex: Routledge.

Buhi, E.R., Goodson, P., & Neilands, T.B. (2007). Structural equation modelling: A primer for health behaviour researchers, *Am J Health Behaviour, 31*(1), 74 – 85.

Burdenski, T. (2000). Evaluating univariate, bivariate, and multivariate normality using graphical and statistical procedures. *Multivariate Linear Regression Viewpoints, 26*(2), 15 – 28.

Byrne, B.M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications and programming* (2^{nd} ed.). East Sussex: Routledge

Byrne, B.M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications and programming.* Mahwah, NJ: Lawrence Erlbaum.

Byrne, B. M. (1998). Structural equation modelling: basic concepts, application, and programming. Lawrence Earlbaum Associated. *Inc., Mahwah, NJ.*

Cai, Y., Ye, X., Shi, R., Xu, G., Shen, L., Ren, J., & Huang, H. (2013). Predictors of consistent condom use based on the Information-Motivation-Behaviour Skill (IMB) model among senior high school students in three coastal cities in China. *BMC Infectious Diseases, 13*(262), 1 – 8. doi: 10.1186/1471-2334-13-262.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin, 56*(2), 81.

Campbell, C., & Murray, M. (2004). Community health psychology: Promoting analysis and action for social change. *Journal of health psychology, 9*(2), 187–195.

Campbell, C., Nair, Y., & Maimane, S. (2007). Building contexts that support effective community responses to HIV/AIDS: a South African case study. *American Journal of Community Psychology,* 39, 347–363.

Carmack, C. C., & Lewis-Moss, R. K. (2009). Examining the theory of planned behaviour applied to condom use: the effect-indicator vs. causal-indicator models. *The journal of primary prevention, 30*(6), 659–676.

Carson, D., Gilmore, A., Perry, C., & Gronhaug, K. (2001). *Qualitative Marketing Research.* London: Sage.

Cha, E. S., Doswell, W. M., Kim, K. H., Charron-Prochownik, D., & Patrick, T. E. (2007). Evaluating the theory of planned behaviour to explain intention to engage in premarital sex amongst Korean college students: A questionnaire survey. *International journal of nursing studies, 44*(7), 1147–1157.

Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in medicine, 19*(9), 1141–1164.

Casson, R.J., & Farmer, L.D.M. (2014). Understanding and checking the assumptions of linear regression: a primer for medical researchers. Clinical and Experimental Ophthalmology, *42*, 590–596 doi: 10.1111/eco.12

Chambers, J.M., Cleveland, W.S., Kleiner, B., & Tukey, P. A. (1983). *Graphical Methods for Data Analysis.* Pacific Grove, CA: Wadsworth and Brooks/Cole.

Chatterjee, S., & Price, B. (1991). *Regression analysis by example* (2^{nd} ed.). New York: Wiley.

Chimbiri, A. M. (2007). The condom is an 'intruder' in marriage: evidence from rural Malawi. *Social science & medicine, 64*(5), 1102–1115.

Chin, W. (1998). The partial least squares approach for structural equation modelling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp 295 – 336). Mahwah, NJ: Lawrence Erlbaum Associates.

Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling.

Cohen, B. H. (2001). *Explaining psychological statistics* (2^{nd} ed.), New York: Wiley.

Cohen, J., Cohen, P., West, S.G., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioural sciences* (3^{rd} ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, L., Manion, L., & Morrison, K. (2007). Research methods in education (6^{th} ed.). London: Routledge.

Comrey, A.L., & Lee, H.B. (1992). *A first course in factor analysis* (2^{nd} ed.), Hillsdale, NJ: Lawrence Erlbaum Associates.

Conner, M., & Armitage, C. J. *(1998). Extending the theory of planned behaviour: A review* and avenues for further research. *Journal of applied social psychology, 28*(15), 1429–1464.

Cooperrider, D. L., & Srivastva, S. (1987). Appreciative Inquiry in Organizational Life. In W. A.Pasmore and R. W.Woodman (Eds.), *Research in Organizational Change and Development,* Vol. **1**. Greenwich, Conn.

Couture, M., Soto, J.C., Akom, E., Joseph, G., & Zunzunegui, M. (2010). Determinants of intentions to use condoms among clients of female sex workers in Haiti, *AIDS Care, 22*(2), 253–262.

Creswell, J.W. (2014). *Research Design Qualitative, Quantitative and Mixed Methods Approaches* (4^{th} ed.). Thousand Oaks, CA Sage.

Crisci, A. (2012). Estimation methods for the structural equation models: maximum likelihood, partial least squares and generalised maximum entropy. *Journal of applied quantitative methods,*

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Crotty, M. (1998). *The foundations of social research.* London: Sage.

Curran, P.J., West, S.G., & Finch, J.F. (1996). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. *Psychological Methods, 1,* 16 – 29.

Curtis, R.F., & Jackson, E.F. (1962). Multiple indicators in survey research. *American Journal of Sociology, 68*(2), 195–204.

Davis, D. (1996). Business research for decision making. Belmont, CA: Duxbury Press.

Denzin, N.K., & Lincoln, Y.S. (2011). Introduction: The discipline and practice of qualitative research. In N.K. Denzin & Y.S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (4^{th} ed., pp 1–19), London: Sage.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of business research, 61*(12), 1203–1218.

DiClemente, R. J., Swartzendruber, A. L., & Brown, J. L. (2013). Improving the validity of self-reported sexual behaviour: No easy answers. *Sexually Transmitted Diseases*, 40, 111–112.

Dilorio, C., Dudley, W.N., Soet, J., Watkins, J. & Maibach, E. (2000). A social cognitive model for condom use among college students, *Nursing Research*, 49, 208–214.

Dilorio, C., Dudley, W.N., Kelly, M., Soet, J.E., Mbwara, J. & Potter, J.S. (2001). Social cognitive correlates of sexual experience and condom use among 13- through 15-year-old adolescents, *Journal of Adolescent Health*, 29, 208–16.

Doornik, J. A., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics, 70*(s1), 927–939.

Doyle, A.M., Mavadzenge, S.N., Plummer, S.L., & Ross, D.A. (2012). The sexual behaviour of adolescents in sub-Saharan Africa: Patterns and trends from national surveys. *Tropical Medicine and International Health, 17*(7), 796–807.

Ebrahim, N.B., Davis, S., & Tomaka, J. (2017). Psychosocial determinants of intention to use condoms among Somali and Ethiopian immigrants in the U.S., *Psychology, Health & Medicine, 22*(5), 611–617. doi: 10.1080/13548506.2016.1204463

Edwards, J.R., & Bagozzi, R.P. (2000). Multidimensional constructs in organizational behavior research: an integrative analytical framework. *Organisational research methods, 4*(2), 144 – 192.

Efron, B. (1979). Bootstrap methods: Another look at jackknife. *The Annals of Statistics, 7*(1), 1 – 26. doi:10.1214/aos/1176344552.

Efron B. & Tibshirani, R. (1993). *An introduction to the bootstrap.* Chapman & Hall, New York.

Eggers, S. M., Aarø, L. E., Bos, A. E., Mathews, C., Kaaya, S. F., Onya, H., & de Vries, H. (2016). Socio-cognitive predictors of condom use and intentions among adolescents in three sub-saharan sites. *Archives of sexual behaviour, 45*(2), 353–365.

Enders, C.K. (2010). *Applied missing data analysis.* New York: The Guildford Press

Espada, J. P., Morales, A., Guillén-Riquelme, A., Ballester, R., & Orgilés, M. (2016). Predicting condom use in adolescents: a test of three socio-cognitive models using a structural equation modelling approach. *BMC public health, 16*(1), 35.

Fazekas, A., Senn, C. Y., & Ledgerwood, D. M. (2001). Predictors of intention to use condoms among university women: An application and extension of the theory of planned behaviour. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 33*(2), 103– 117.

Fehr, S. K., Vidourek, R. A., & King, K. A. (2015). Intra-and inter-personal barriers to condom use among college students: A review of the literature. *Sexuality & Culture, 19*(1), 103-121. doi 10.1007/s12119-014-9249-y.

Feilzer, M. Y. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research, 4*(1), 6–16.

Field, A. (2013). *Discovering statistics using IBM SPSS STATISTICS* (4^{th} ed.), Sage Publications Inc.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behaviour.* New York: Wiley.

Fishbein, M. (2000). The role of theory in HIV prevention. *AIDS Care*, 12, 273–8.

Fishbein, M., Bandura, A., Triandis, H. C., Kanfer, F. H., Becker, M. H., Middlestadt, S. E, & Eichler, A. (1992). *Factors influencing behaviour and behaviour change: Final report-theorist's workshop.* Rockville, MD: National Institute of Mental Health.

Fisher, J. D., & Fisher, A. W. (1992). Changing AIDS risk behaviour. *Psychological Bulletin, 111*(3), 455–474.

Fisher, W. A., Fisher, J. D., & Harman, J. (2003). The information-motivation-behavioural skills model: A general social psychological approach to understanding and promoting health behaviour. *Social psychological foundations of health and illness*, 82–106.

Fisher, J. D., Fisher, W. A., Bryan, A. D., & Misovich, S. J. (2002). Information-motivation-behavioural skills model-based HIV risk behaviour change intervention for inner-city high school youth. *Health Psychology, 21*(2), 177.

Fisher, J. D., Fisher, W. A., Misovich, S. J., Kimble, D. L., & Malloy, T. E. (1996). Changing AIDS risk behaviour: effects of an intervention emphasizing AIDS risk reduction information, motivation, and behavioural skills in a college student population. *Health Psychology, 15*(2), 114.

Flick, U. (2011). *Introducing research methodology: A beginner's guide to doing a research project.* London: Sage.

Fornell, C.G., & Bookstein, F.L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research, 19*(4), 440 – 452.

Fornell, C., Tellis, G. J., & Zinkhan, G. M. (1982). Validity assessment: A structural equations approach using partial least squares. In *Proceedings, American marketing association educators' conference* (pp. 1–5).

Gallagher, D., TING, L., & Palmer, A. (2008). A Journey into the unknown; taking the fear out of structural equation modelling with AMOS for the first-time user. *The Marketing Review, 8*(3), 255–275. doi: 10.1362/146934708X337672.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson and negative binomial models. Psychological Bulletin, 118, 392–404.

Garson, D.G. (2015). *Structural equation modelling (2015 ed.)* Statistical Associates Publishing.

Gaskin, J. (2012). Data screening. *Gaskination's StatWiki.*

Gebhardt, W. A., Kuyper, L., & Greunsven, G. (2003). Need for intimacy in relationships and motives for sex as determinants of adolescent condom use. *Journal of Adolescent Health, 33*(3), 154–164.

Gefen, D., Rigdon, E.E., & Straub, D. (2011). Editor's comments: an update and extension to SEM guidelines for administrative and social science research. *MIS Quarterly, 35*(2), III –XIV.

Gefen, D., Straub, D., Boudreau, M. C. (2000). Structural equation modeling and regression: Guidelines for research practice. Communications of the association for information systems, 4(1), 7.

Giles, M., Liddell, C., & Bydawell, M. (2005). Condom use in African adolescents: The role of individual and group factors. *AIDS care, 17*(6), 729–739.

Goddard, W., & Melville, S. (2004). *Research methodology: An introduction.* Juta and Company Ltd.

Godin, G., & Kok, G. (1996). The theory of planned behaviour: a review of its applications to health-related behaviours. *American journal of health promotion, 11*(2), 87–98.

Goldberg, S. K., Haydon, A. A., Herring, A. H., & Halpern, C. T. (2014). Longitudinal consistency in self-reported age of first vaginal intercourse among young adults. *The Journal of Sex Research*, 51, 97–106.

Goodboy, A. K., & Kline, R. B. (2017). Statistical and practical concerns with published communication research featuring structural equation modelling. *Communication Research Reports, 34*(1), 68–77.

Goodhue, D., Lewis, W., & Thompson, R. (2012). Comparing PLS to regression and LISREL: A response to Marcoulides, Chin and Saunders. *MIS Quarterly, 36*(3), 703–716.

Graham, J., Olchowski, A., & Gilreath, T. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8*(3), 206–213.

Grewal, R., Cote, J. & Baumgartner, H. (2004) Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing, *Marketing Science, 23*(4), 519–529.

Guadagnoli, E., & Velicer, W. (1988). Relation of sample size to the stability of component patterns. *PsychologicalBulletin, 103*, 265–275.

Guisan, A., Edwards, T.C., & Hastie, T. (2002). Generalised linear and generalised additive models in studies of species distributions: setting the scene. *Ecological Modelling, 157*, 89 – 100.

Guo, R., Mcaleese, W.J., Appleby, K.M., Guo, J., Zhang, W., Huang, Y., & Peterson, T. (2014). Predictors of intention to use condoms among Chinese college students. *J Community Health, 39*, 712–718. doi: 10.1007/s10900-013-9816-4.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2014a). Multivariate Data Analysis (7^{th} ed, Pearson new international edition). Harlow, Essex Pearson Education Limited

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate Data Analysis* (7^{th} ed.), Upper Saddle River, NJ: Prentice Education, Inc.

Hair, J.F., Hult, G.T.M., Ringle, C.M., & Sarstedt, M. (2014b). *A primer on partial least squares structural equation modelling (PLS-SEM).* Thousand Oaks: Sage.

Hancock, G.R. (2003). Fortune Cookies, Measurement Error and Experimental Design. *Journal of Modern Applied Statistical Methods, 2*(2) 293–305.

Harrison, D. A. (2002) Structure and timing in limited range dependent variables: Regression models for predicting if and when. In F. Drasgow & N. Schmitt (Eds.), Measuring and analyzing behaviour in organizations: Advances in measurement and data analysis (pp. 446–497). San Francisco: Jossey-Bass.

Hastie, T.J., & Tibshirani, R.J. (1990) *Generalized Additive Models.* Chapman and Hall, London, UK.

Hastie, T.J., & Tibshirani, R.J. (1986). Generalised additive models (with discussion). Statistical Science, 1, 297 – 318.

Hastie, T. and Tibshirani, R. (1984). Generalized additive models. Tech. Rept. 98, Dept. of Statistics, Stanford Univ.

Hayduk, L., Cummings, G.G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! Testing! One, two three – Testing the theory in structural equation models! *Personality and Individual Differences*, 42, 841 – 850.

Heeler, R. M., & Ray, M. L. (1972). Measure validation in marketing. *Journal of Marketing Research*, 361–370.

Hinkley, D.V. (1988). Bootstrap Methods. *Journal of the Royal Statistical Society B 50*(3), 321 – 337.

Holliday, A. (2002): *Doing and Writing Qualitative Research.* London: Sage Publications.

Hooper, D., Coughlan, J., & Mullen, M.R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods, 6*(1), 53–60. Available online at www.ejbrm.com

Horton, N. J., & Lipsitz, S. R. (1999). Review of software to fit generalized estimating equation regression models. *American Statistician, 53*, 160–169.

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007). Is formative measurement really measurement? Reply to Bollen (2007) and Bagozzi (2007).

Hox, J.J. & Bechger, T.M. (1998). An Introduction to Structural Equation Modelling. *Family Science Review, 11,* 354–373.

Hoyle, R.H. & Panter, A.T. (1995). Writing about structural equation models. In R.H. Hoyle (ed.), *Structural equation modelling: Concepts, issues and applications* (pp. 159 – 176). Thousand Oaks, CA: Sage.

Huberty, C. J. (1989). Problems with stepwise methods—better alternatives. *Advances in social science methodology, 1,* 43–70.

Hu, L. T., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modelling: a multidisciplinary journal, 6*(1), 1–55.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modelling: Sensitivity to under-parameterised model mis-specification. *Psychological Methods*, 3, 424 – 453.

Iacobucci, D. (2009). Everything you always wanted to know about SEM (structural equations modelling) but were afraid to ask. *Journal of Consumer Psychology, 19*(4), 673–680.

Idele, P., Gillespie, A., Porth, T., Suzuki, C., Mahy, M., Kasedde, S., & Luo, C. (2014). Epidemiology of HIV and AIDS among adolescents: current status, inequities, and data gaps. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 66*, S144-S153.

Izenman, A.J. (2013). *Modern Multivariate Statistical Techniques: Regression, Classification and Manifold Learning (2nd ed.),* Springer.

Jaccard, J., & Wan, C. K. (1996). *LISREL approaches to interaction effects in multiple regression* (No. 114). Sage.

Janner, M., Wolitski, R., Corby, N., & Fishbein, M. (1998). Using the theory of planned behaviour to predict intention to use condoms among female sex workers. *Psychological Health Review, 1,* 187–205.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of consumer research, 30*(2), 199–218.

Jemmott III, J. B. (2012). The reasoned action approach in HIV risk-reduction strategies for adolescents. *The annals of the American academy of political and social science, 640*(1), 150–172.

Jemmott III, J. B., Heeren, G. A., Ngwane, Z., Hewitt, N., Jemmott, L. S., Shell, R., & O'Leary, A. (2007). Theory of planned behaviour predictors of intention to use condoms among Xhosa adolescents in South Africa. *AIDS care, 19*(5), 677–684.

Jemmott III, J.B., Jemmott, L.S., & Fong, G.T. (1998). Abstinence and safer sex HIV risk-reduction interventions for African American adolescents: a randomized controlled trial. *Jama, 279*(19), 1529–1536.

Jeon, J. (2015). The Strengths and Limitations of the Statistical Modelling of Complex Social Phenomenon: Focusing on SEM, Path Analysis, or Multiple Regression Models. *Int J Soc Behav Educ Econ Bus Ind Eng, 9*, 1629–1637.

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2014). *The Gap Report.* Geneva, Switzerland: UNAIDS; 2014.

Joreskog, K.G. (2004). On chi-squares for the independence model and fit measures in LISREL. Retrieved December 15, 2017, from *www.ssicentral.com/lisrel/techdocs/ ftb.pdf*

Joreskog, K.G. (1973). A general method for estimating a linear structural equation system. In A.S. Goldberger & O.D. Duncan (Eds.), *Structural equation models in the social sciences.* New York: Seminar Press.

Joreskog, K.G. (1967). Some contributions to Maximum Likelihood factor analysis. *Psychometrika, 32*(4), 443–482.

Joreskog, K.G., & Sorbom, D. (1996). LISREL 8: User's Reference Guide. Chicago, IL: Scientific Software International Inc.

Joreskog, K. G., & Sorbom, D. (1984). LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood: User's guide. *University of Uppsala, Sweden.*

Kalolo, A., & Kibusi, S. M. (2015). The influence of perceived behaviour control, attitude and empowerment on reported condom use and intention to use condoms among adolescents in rural Tanzania. *Reproductive health, 12*(1), 105.

Kanekar, A., & Sharma, M. (2009). Using social cognitive theory to predict safer sex behaviours in African-American college students. Acta Didactica Napocensia, *2*(2), 51 – 56.

Kang, H. (2013). The prevention and handling of the missing data. *Korean J Anesthesiol.*, *64*(5), 402–406. doi.org/10.4097/kjae.2013.64.5.402

Kapogiannis, B. G., Legins, K. E., Chandan, U., & Lee, S. (2014). Evidence-based programming for adolescent HIV prevention and care: operational research to inform best practices. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 66*, S228-S235.

Kaplan, D. (2009). *Structural equation modelling: Foundations and extensions* (2^{nd} ed.). Thousand Oaks, CA: Sage Publications, Inc.

Kapoor, H.S., Jain, K & Sharma, S.K. (2016). Generalised Additive Model for Evaluation of Premium for Diabetic Patients. *Journal of Advances in Applied Mathematics, 1*(3), 149–159.

Katikiro, E., & Njau, B. (2012). Motivating factors and psychosocial barriers to condom use among out-of-school youths in Dar es Salaam, Tanzania: A cross sectional survey using the health belief model. *Isrn aids, 2012.*

Keith, T.Z. (2015). *Multiple Regression and Beyond: An Introduction to Multiple Regression and Structural Equation Modelling.* Routledge.

Kelloway, E. K. (1998). Using Lisrel for structural equation modelling: *A researcher's guide.* Thousand Oaks, CA: Sage.

Kennedy, C., O'reilly, K., Medley, A., & Sweat, M. (2007). The impact of HIV treatment on risk behaviour in developing countries: a systematic review. *AIDS care, 19*(6), 707–720.

Kenny, D.A. (2014). Measuring Model Fit. http://davidakenny.net/cm/fit.htm [accessed on 4 November 2015]

Khine, M.S., Ping, L.M., & Cunningham, D. (Eds.) (2013). *Application of structural equation modeling in educational research and practice, 3-21.* Sense Publishers, Rotterdam, NLD. Available from: ProQuest ebrary. [25 March 2015]

Klem, L. (2000). Structural equation modelling. In L.G. Grimm & P.R. Yarnold (Eds.), *Reading and understanding MORE multivariate statistics* (pp. 227 – 260). Washington, DC: American Psychological Association.

Kline, R.B. (2011). *Principles and practice of structural equation modeling.* New York: Guilford.

Kok, G., & Ruiter, R. A.C. (2014). Who has the authority to change a theory? Everyone! A commentary on Head and Noar. *Health psychology review, 8*(1), 61–64.

Krugu, J.K., Mevissen, F.E.F., Debpuur, C., & Ruiter, R.A.C. (2016). Psychosocial correlates of condom use intention among school students in the Bolgatanga Municipality of Ghana, *International journal of sexual health, 28*(1), 96 – 110. doi: 10.1080/19317611.2015.1124162.

Langdrigde D., Sheeran P. & Connolly K.J. (2007). Analysing Additional Variables in the Theory of Reasoned Action. *Journal of Applied Social Psychology, 37*(8), 1884–1913.

Lee, S.Y., & Song, X.Y. (2012). *Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences.* John Wiley & Sons.

Lei, P. W., Wu, Q. (2007). Introduction to structural equation modeling: Issues and practical considerations. Educational Measurement: issues and practice, 26(3), 33-43.

Liang, K.-Y., & Zeger, S.L. (1986). Longitudinal data analysis using generalized models. *Biometrika, 73*, 13–22.

Liew, J., & Forkman, J. (2015). A guide to generalized additive models in crop science using SAS and R. *Communications in Biometry and Crop Science, 10*, 41–57.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*, (2^{nd} ed.). Hoboken, NJ: Wiley.

Liu, Z., Wei, P., Huang, M., Liu, Y., Li, L., Gong, X., Chen, J., & Li, X. (2014). Determinants of consistent condom use among students in China: Application of the Information-Motivation-Behaviour Skills (IMB) Model. *PLoS ONE 9*(9): e108976. doi 10.1371/journal.pone.0108976

Loehlin, J.C. (2004). *Latent variable models: An introduction to factor, path and structural analysis* (4^{th} ed.). Hillsdale, NJ: Erlbaum.

Lowry, P. B., & Gaskin, J. (2014). Partial Least Squares (PLS) Structural Equation Modeling (SEM) for building and testing behavioral causal theory: When to choose it and how to use it. *Professional Communication, IEEE Transactions on, 57*(2), 123–146.

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annu Rev Public Health*, 23, 151 – 169.

Luszczynska, A., & Schwarzer, R. (2005). Social cognitive theory. *Predicting health behaviour, 2*, 127–169.

Macalino, G. E., Sachdev, D. D., Rich, J. D., Becker, C., Tan, L. J., Beletsky, L., & Burris, S. (2009). A national physician survey on prescribing syringes as an HIV prevention measure. *Substance abuse treatment, prevention, and policy, 4*(1), 13.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modelling in psychological research. *Annual review of psychology, 51*(1), 201–226.

MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modelling. *Psychological Methods*, 1, 130 – 149.

Maharaj, P., & Cleland, J. (2006). Condoms become the norm in the sexual culture of college students in Durban, South Africa. *Reproductive health matters, 14*(28), 104–112.

Manyaapelo, T., Nyembezi, A., Ruiter, R. A., Borne, B. V. D., Sifunda, S., & Reddy, P. (2017). Understanding the Psychosocial Correlates of the Intention to Use Condoms among Young Men in KwaZulu-Natal, South Africa. *International journal of environmental research and public health, 14*(4), 339.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*(2), 519–530.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. Psychological bulletin, 97(3), 562.

Marston, C., & King, E. (2006). Factors that shape young people's sexual behaviour: A systematic review. *Lancet,* 368, 1581–1586.

Mausbach, B. T., Semple, S. J., Strathdee, S. A., & Patterson, T. L. (2009). Predictors of safer sex intentions and protected sex among heterosexual HIV-negative methamphetamine users: an expanded model of the Theory of Planned Behaviour. *AIDS care, 21*(1), 17–24.

McDonald, R.P., & Ho, M.-H, R. (2002). Principles and Practice in Reporting Statistical Equation Analyses. *Psychological Methods, 7*(1), 64 – 82.

McEachan, R. R. C., Conner, M., Taylor, N. J., & Lawton, R. J. (2011). Prospective prediction of health-related behaviours with the theory of planned behaviour: A meta-analysis. *Health Psychology Review, 5*(2), 97–144.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2^{nd} ed.). London: Chapman and Hall.

Meyers, L.S., Gamst, G., & Guarino, A.J. (2013). *Applied multivariate research: Design and interpretation* (2^{nd} ed.). Thousand Oaks, CA: Sage.

Michielsen, K., Chersich, M., Temmerman, M., Dooms, T., & Rossem, R.V. (2012). Nothing as practical as good theory? The theoretical basis of HIV prevention interventions for young people in Sub-Saharan Africa: A systematic review. *AIDS Research and Treatment.* doi: 10.1155/2012/345327

Miles, J., & Shevlin, M. (2007). A Time and a place for incremental fit indices. *Personality and Individual Differences, 25*, 85 – 90.

Mimiaga, M.J., Reisner, S.L., Reilly, L., Soroudi, N., & Safren, S. (2009). Individual Interventions. *HIV Prevention,* 203–239.

Misovich, S. J., Martinez, T., Fisher, J. D., Bryan, A., & Catapano, N. (2003). Predicting breast self-examination: A test of the information-motivation-behavioural skills model. *Journal of Applied Social Psychology, 33*(4), 775–790.

Molla, M., Nordrehaug ÅstrØm, A., & Brehane, Y. (2007). Applicability of the theory of planned behaviour to intended and self-reported condom use in a rural Ethiopian population. *AIDs care, 19*(3), 425–431.

Muñoz-silva, A., Sánchez-García, M., Martins, A., & Nunes, C. (2009). Gender differences in HIV-related sexual behaviour among college students from Spain and Portugal. *The Spanish Journal of Psychology, 12*(2), 485–495.

Muthén, L.K., & Muthén, B.O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modelling, 4*, 599 – 620.

Mutinta, G., & Govender, K. (2012). The social environmental determinants of students' sexual risk behaviour and HIV prevention at the University of KwaZulu Natal. *The Journal of Human Ecology, 38*(1), 17 – 29.

Mutinta, G., Govender, K., Gow, J., & George, g. (2013). An investigation of students' risky sexual behaviour at KwaZulu-Natal University, Durban, South Africa. *American Journal of Sexuality Education,* 8(3), 121–139.

Myers, R.H. (1990). *Classical and modern regression with applications* (Vol. 2). Belmont, CA: Duxbury press

Nakagawa, S. (2015). Missing data: mechanisms, methods, and messages. *Ecological statistics: contemporary theory and application* (ed. by G.A. Fox, S. Negrete-Yankelevich and V.J. Sosa), pp. 81–105. Oxford University Press, Oxford.

Nelder, J. A., & Wedderburn, R.W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A, 135*, 370–384.

Ngulube, P., Mokwatlo, K., & Ndwandwe, S. (2015). Utilisation and prevalence of mixed methods research in library and information research in South Africa 2002 – 2008. *South African Journal of Libraries and Information Science, 75*(2), 105 – 116.

Noar, S. M. (2007). An interventionist's guide to AIDS behavioural theories. *AIDS care, 19*(3) 392–402. doi: 10.1080/09540120600708469.

Norton, W. E. (2009). Relative efficacy of a pregnancy, STI, or HV prevention-focused intervention on changing sexual risk behaviour among young adults. Unpublished doctoral dissertation. Storrs: University of Connecticut.

Nusair, K., & Hua, N. (2010). Comparative assessment of structural equation modelling and multiple regression research methodologies: E-commerce context. *Tourism Management, 31*(3), 314–324.

Ochieng, M. A., Kakai, R., & Abok, K. (2011). Knowledge, attitude and practice of condom use among secondary school students in Kisumu district, Nyanza province. *Asian Journal of Medical Sciences, 3*(1), 32–36.

Okonofua, F. (2013). Documenting new HIV/AIDS research and interventions in Sub-Saharan Africa. *African journal of reproductive health, (Special edition on HIV/AIDS, 17(4),* 9–12.

Olinsky, A., Chen, S., & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modelling. *European Journal of Operational Research, 151*(1), 53–79.

Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation, 8*(2)
Available online: http://PAREonline.net/getvn.asp?v=8&n=2.

Pearson, E. S. (1930). A further development of tests for normality. *Biometrika*, 239–249.

Pituch, K.A., & Stevens, J.P. (2016). *Applied multivariate statistics for the social sciences.* Routledge.

Plummer, M.L., & Wight, D. (2011). *Young People's Lives and Sexual Relationships in Rural Africa: Findings from a Large Qualitative Study in Tanzania.* Lanham, Maryland: Lexington Books.

Puth, M. T., Neuhäuser, M., & Ruxton, G. D. (2015). On the variety of methods for calculating confidence intervals by bootstrapping. *The Journal of animal ecology, 84*(4), 892–897.

Ramzan, S., Zahid, F.M., & Ramzan, S. (2013). Evaluating Multivariate Normality: A Graphical Approach. *Middle-East Journal of Scientific Research, 13*(2), 254 –263. DOI: 10.5829/idosi.mejsr.2013.13.2.1746

Rannie, K., & Craig, D. M. (1997). Adolescent females' attitudes, subjective norms, perceived behavioural control, and intentions to use latex condoms. *Public Health Nursing, 14*(1), 51–57.

Reinartz, W., Haenlein, M, & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of Market Research, 26*(4), 332–344.

Romero, L. M., Galbraith, J. S., Wilson-Williams, L., & Gloppen, K. M. (2011). HIV prevention among African American Youth: how well have evidence-based interventions addressed key theoretical constructs? *AIDS and Behaviour, 15*(5), 976–991.

Roth, P. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47,* 537–560.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63,* 581–592.

Rubin, R.B., Rubin, A. M., & Haridakis, P. (2010). *Communication Research: Strategies and Sources, 7th Edition.* Boston, MA: Cengage Learning.

Rutherford, A. (2001). *Introducing ANOVA and ANCOVA: A GLM Approach.* SAGE.

Sacolo, H.N., Chung M-H, Chu, H., Liao, Y-M., Chen, C-H., et al. (2013). High Risk Sexual Behaviours among the In-School Youth in Swaziland: A Structural Equation Modelling Approach. PLos 8(7): e67289. DOI: 10.1371/journal.pone0067289

Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research methods for Business Students.* (7^{th} ed.), Pearson Education Limited.

Saunders, M., Lewis, P., & Thornhill, A. (2012). *Research methods for Business Students.* (6^{th} ed.), Pearson Education Limited.

Savalei, V. (2008). Is the ML chi-square ever robust to non-normality? A cautionary note with missing data. *Structural Equation Modelling, 15*, 1–22.

Savalei, V., & Bentler, P.M. (2006). Structural equation modelling. In Grover R, Vriens M, (Eds.), *The handbook of marketing research: Uses, misuses, and future advances.* Sage; Thousand Oaks, CA. pp. 330–364

Schaalma, H., AarØ, L. E., Flisher, A. J., Mathews, C., Kaaya, S., Onya, H., Ragnarson, A., & Klepp, K. I. (2009). Correlates of intention to use condoms among Sub-Saharan African youth: the applicability of the theory of planned behaviour. *Scandinavian journal of public health, 37*(2 suppl), 87–91.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research, 8*(1), 3–15.

Schermelleh-Engel, K., Mossbrugger, H., & Muller, 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), 23–74.

Schroder, K. E. E., Carey, M. P., & Vanable, P. A. (2003). Methodological challenges in research on sexual risk behaviour: II. Accuracy of self-reports. *Annals of Behavioural Medicine, 26*(2), 104–123.

Schumacker, R., & Lomax, R.G. (2016). *A Beginner's Guide to Structural Equation Modelling* (4^{th} ed.). New York, NY: Routledge.

Schwager, S. J., & Margolin, B. H. (1982). Detection of multivariate normal outliers. *The annals of statistics, 10*(3), 943–954.

Scotland, J. (2012). Exploring the philosophical underpinnings of research: relating ontology and epistemology to the methodology and methods of the scientific, interpretive and critical research paradigms. *English Language Teaching, 5*(9), 9–16.

Shah, R., & Goldstein, S.M. (2006). Use of structural equation modeling in operations management research: Looking back and forward. *Journal of Operations Management, 24*(2), 148–169.

Shao, Y., & Zhou, M. (2014). A powerful test for multivariate normality. *Journal of Applied Statistics, 41*(2), 351–363.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 591–611.

Sheeran, P., Abraham, C., & Orbell, S. (1999). Psychosocial correlates of heterosexual condom use: a meta-analysis. *Psychol Bull., 125*(1), 90 – 132.

Sheeran, P., & Taylor, S. (1999). Predicting intentions to use condoms: a meta-analysis and comparison of the theories of reasoned action and planned behaviour. *Journal of Social Psychology, 29*, 1624–75.

Shook, C. L., Ketchen Jr, D. J., Hult, G. T. M., & Kacmar, K. M. (2004). An assessment of the use of structural equation modeling in strategic management research. *Strategic management journal, 25*(4), 397–404.

Silverman, D. (2013). Doing Qualitative Research: A practical handbook. London: Sage.

Singh, K., & Xie, M. (2008). Bootstrap: a statistical method. *Unpublished manuscript, Rutger University, USA. Retrieved from http://www.stat.rutgers.edu/home/mxie/RCPapers/bootstrap.pdf (Accessed: 20 October 2018).*

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate behavioural research, 25*(2), 173–180.

Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors. In *annual meeting of the Psychometric Society, Iowa City, IA* (Vol. 758, pp. 424–453).

Stevens, J.P. (2009). *Applied multivariate statistics for the social sciences* (5^{th} ed.), New York, NY: Routledge.

Stroeken, K., Remesb, P., De Kokerc, P., Michielsenc, K., Van Vossolec, A., & Temmermanc, M. (2012). HIV among out-of-school youth in Eastern and Southern Africa: A review. *AIDS Care 24*, 186–194.

Sutton, S., McVey, D., & Glanz, A. (1999). A comparative test of the theory of reasoned action and the theory of planned behaviour in the prediction of condom use intentions in a national sample of English young people. *Health Psychology, 18*, 72–81.

Tabachnick, B.G., & Fidell, L.S. (2014). *Using multivariate statistics.* (6^{th} ed.), Pearson Education, Inc.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. S. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 10–39). Newbury Park, CA: Sage.

Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology, 38*(2), 197–201.

Tao, M., Xie, P., Chen, J., Qin, B., Zhang, D., Niu, Y., Zhang, M., Wang, Q., & Wu, L. (2012). Use of a generalized additive model to investigate key abiotic factors affecting microcystin cellular quotas in heavy bloom areas of Lake Taihu. *PLoS One*, *7*(2), e32020.

Taylor, M., Jinabhai, C., Dlamini, S., Sathiparsad, R., Eggers, M. S., & De Vries, H. (2014). Effects of a teenage pregnancy prevention program in KwaZulu-Natal, South Africa. *Health Care for Women International, 35*(7–9), 845–858.

Tennhaus, M. (2008). Component-based Structural Equation Modelling. *Total quality management*, 19(7-8), 871–886.

Teye-Kwadjo, E., Kagee, A., & Swart, H. (2017a). Predicting the intention to use condoms and actual condom use behaviour: A three-wave longitudinal study in Ghana. *Applied Psychology: Health and Well-Being, 9*(1), 81 – 105.

Teye-Kwadjo, E., Kagee, A., & Swart, H. (2017b). Determinants of condom use among heterosexual young men and women in south-eastern Ghana: a mediation analysis, *Psychology & Sexuality, 8(4)*, 291–305.

Thompson, B. (2003). Guidelines for authors reporting score reliability estimates. *Score reliability: Contemporary thinking on reliability issues*, 91–102.

Thompson, B. (1989). Why won't stepwise methods die? *Measurement and Evaluation in Counselling and Development*, 21, 146–148.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Trafimow, D., Sheeran, P., Conner, M., & Finlay, K. A. (2002). Evidence that perceived behavioural control is a multidimensional construct: Perceived control and perceived difficulty. *British Journal of Social Psychology, 41*(1), 101–121.

Ullman, J.B. (2014). *Structural Equation Modeling.* In B.G. Tabachnick & L.S. Fidell, *Using multivariate statistics* (6^{th} ed., pp. 731-836). Pearson Education, Inc.

UNAIDS. (2012). Botswana 2012 Global AIDS response report. [online] Available at: http://www.bw.undp.org/content/botswana/en/home/library/hiv_aids/botswana-2012-global-aids-response-report.html

Van Rossem, R., & Meekers, D. (2011). Perceived social approval and condom use with casual partners among youth in urban Cameroon. *BMC Public Health, 11*(1), 632.

Villaruel, A.M., Jemmott III, J.B., Jemmott, L.S. & Ronis, D.L. (2004) Predictors of sexual intercourse and condom use intentions among Spanish-dominant Latino youth: a test of the planned behaviour theory. *Nursing Research, 53*(3), 172–181.

Wang, M. (2014). Generalised Estimating Equations in Longitudinal Data Analysis: A Review and Recent Developments. *Advances in Statistics, 2014.* doi: 10.1155/2014/303728.

Webb, T.L., & Sheeran, P. (2006). Does changing behavioural intentions engender behaviour change? A meta-analysis of the experimental evidence. *Psychological bulletin*, 132(2), 249.

Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modelling. *The Counselling Psychologist, 34*(5), 719–751.

Wheaton, B., Muthen, B., Alwin, D.F., & Summers, G. (1977). Assessing Reliability and Stability in Panel Models. *Sociological Methodology, 8*(1), 84–136.

Wilcox, J. B., Howell, R. D., & Breivik, E. (2008). Questions about formative measurement. *Journal of Business Research, 61*(12), 1219–1228.

Wise, D., Goggin, K. J., Gerkovich, M. M., Metcalf, K. A., & Kennedy, S. L. (2006). Predicting intentions to use condoms using gender, sexual experience, and the theory of planned behaviour. *American Journal of Health Education, 37*(4), 210–218.

Wood, S.N. (2006). *Generalised Additive Models: An Introduction with R.* Chapman & Hall/CRC, Boca Raton.

Wolitski, R. J., & Zhang, J. (2007). Understanding and motivating condom use among at-risk and HIV-seropositive persons: A review and demonstration of the applicability of the theories of reasoned action and planned behaviour. In I. Ajzen, D. Albarracın, & R. Hornik (Eds.), *Prediction and change of health behaviour: Applying the reasoned action approach* (pp. 173–196). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Wu, W., & Little, T.D.(2011). Quantitative Research Methods. *Encyclopaedia of Adolescence, 1*, 287–297.

Ybarra, M.L., Korchmaros, J., Kiwanuka, J., Bangsberg, D.R., & Bull, S. (2013). *AIDS Behav, 17*(3), 1116 – 1128. doi: 10.1007/s10461-012-0137-x.

Yee, T.W., & Mitchell, N.D. (1991). Generalised additive models in plant biology. *Journal of Vegetation Science 2*, 587 – 602.

Yi, S., Poudel, K.C., Yasuoka, J., Palmer, P.H., & Jimba, M. (2010). Role of risk and protective factors in risky sexual behaviour among high school students in Cambodia. *BMC Publ Health, 10*, 477.

Yuan, K. H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modelling. *British Journal of Mathematical and Statistical Psychology, 51*(2), 289–309.

Yung, Y-F., & Bentler, P. M. (1996). Bootstrapping techniques in analysis of mean and covariance structures. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 195–226). Mahwah, NJ: Erlbaum.

Zainudin, A. (2012). *A handbook on SEM: Structural equation modelling.*

Zhang, G. & Little, R. J. (2009). Extensions of the Penalized Spline of Propensity Prediction Method of Imputation. *Biometrics*, 65, 911–918.

APPENDIX A. Unmodified Measurement Model Fit Summary Results

Model	PRATIO	PNFI	PCFI
Default model	.889	.773	.800
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

NCP

Model	NCP	LO 90	HI 90
Default model	823.738	725.113	929.914
Saturated model	.000	.000	.000
Independence model	8220.310	7922.302	8524.679

FMIN

Model	FMIN	F0	LO 90	HI 90
Default model	1.403	1.039	.914	1.173
Saturated model	.000	.000	.000	.000
Independence model	10.776	10.366	9.990	10.750

AIC

Model	AIC	BCC	BIC	CAIC
Default model	1236.738	1241.109	1526.717	1588.717
Saturated model	702.000	726.744	2343.656	2694.656
Independence model	8597.310	8599.143	8718.914	8744.914

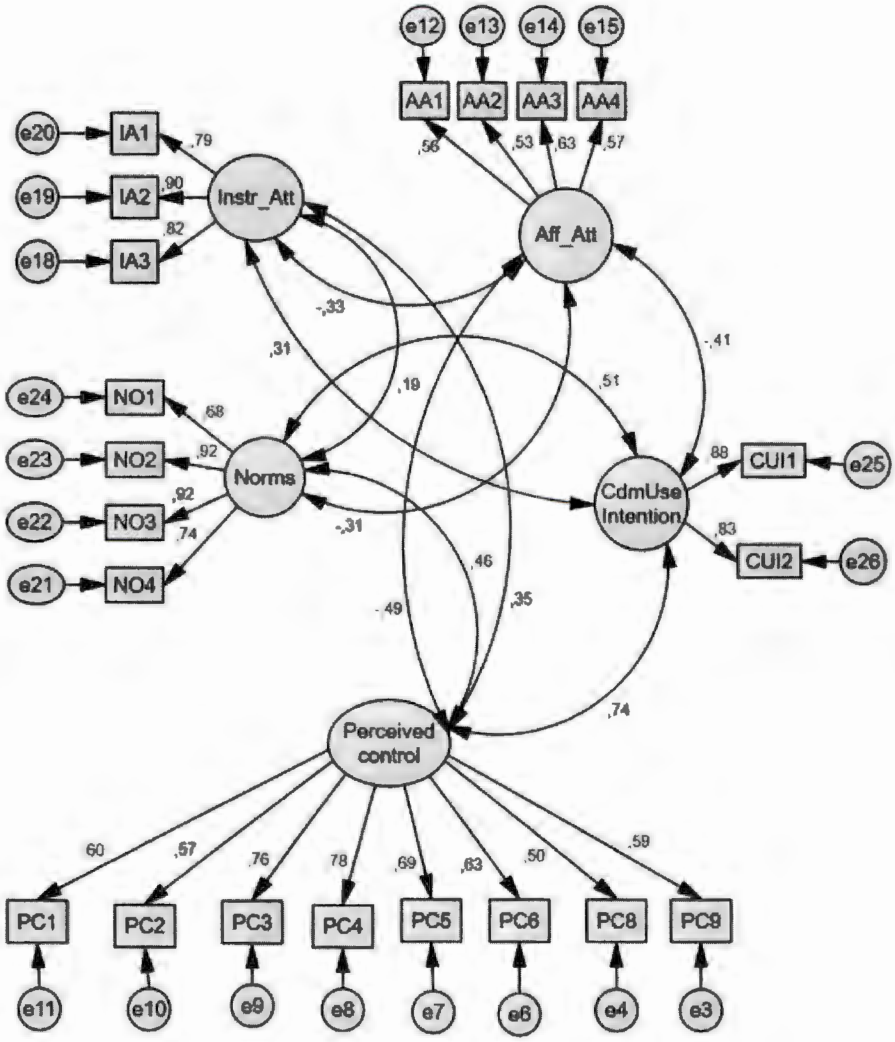
ECVI

Model	ECVI	LO 90	HI 90	MECVI
Default model	1.560	1.435	1.693	1.,565
Saturated model	.885	.885	.885	.916
Independence model	10.842	10.466	11.225	10.844

HOELTER

Model	HOELTER.05	HOELTER.01
Default model	235	248
Independence model	35	36

APPENDIX B. Re-specified Model 1



Re-specified Model 1 Fit Summary Results

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	52	671.467	179	.000	**3.751**
Saturated model	231	.000	0		
Independence model	21	7762.680	210	.000	36.965

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.051	**.921**	**.899**	.714
Saturated model	.000	1.000		
Independence model	.308	.337	.271	.307

Baseline Comparisons

Model	NFIDelta1	RFIrho1	IFIDelta2	TLIrho2	CFI
Default model	**.914**	.899	.935	.924	**.935**
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	**.059**	**.054**	**.064**	.001
Independence model	.213	.209	.217	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	.852	.779	.797
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

NCP

Model	NCP	LO 90	HI 90
Default model	492.467	416.854	575.650
Saturated model	.000	.000	.000
Independence model	7552.680	7267.982	7843.715

FMIN

Model	FMIN	F0	LO 90	HI 90
Default model	.847	.621	.526	.726
Saturated model	.000	.000	.000	.000
Independence model	9.789	9.524	9.165	9.891

AIC

Model	AIC	BCC	BIC	CAIC
Default model	775.467	778.434	1018.675	1070.675
Saturated model	462.000	475.183	1542.406	1773.406
Independence model	7804.680	7805.879	7902.899	7923.899

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Default model	.978	.883	1.083	.982
Saturated model	.583	.583	.583	.599
Independence model	9.842	9.483	10.209	9.843

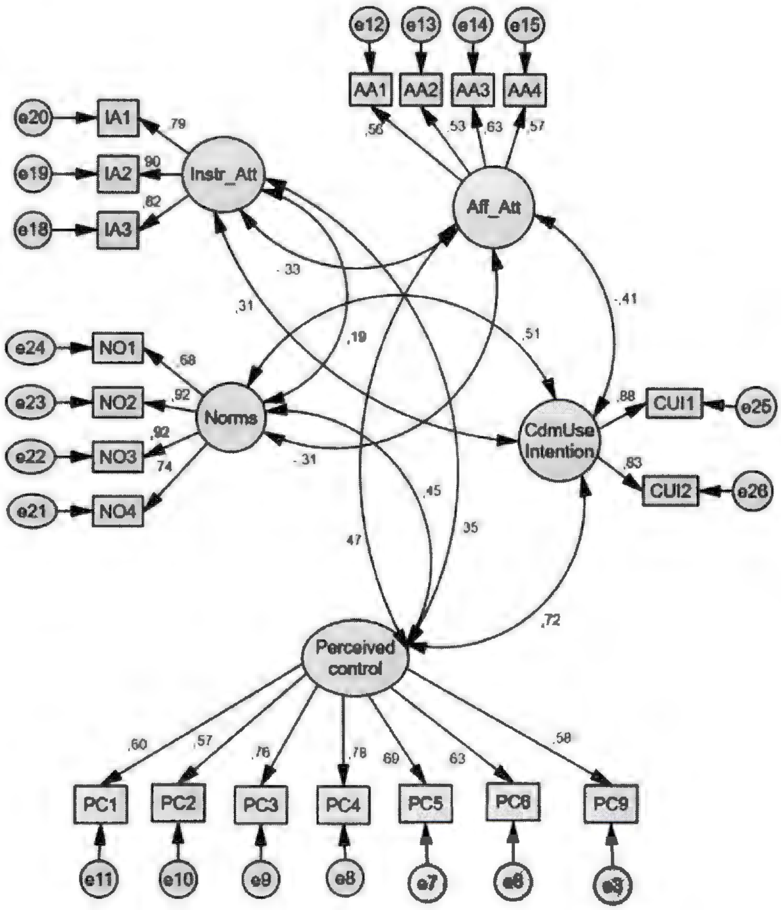
HOELTER

Model	HOELTER.05	HOELTER.01
Default model	250	267
Independence model	26	27

Standardized Regression Weights: (Group number 1 - Default model)

			Estimate
PC9	<---	Perceived_control	.592
PC8	<---	Perceived_control	.496
PC1	<---	Perceived_control	.602
AA1	<---	Aff_Att	.561
AA2	<---	Aff_Att	.530
AA3	<---	Aff_Att	.626
AA4	<---	Aff_Att	.568
CUI1	<---	CdmUse_Intention	.882
CUI2	<---	CdmUse_Intention	.833
IA3	<---	Instr_Att	.823
IA2	<---	Instr_Att	.903
IA1	<---	Instr_Att	.792
NO4	<---	Norms	.739
NO3	<---	Norms	.921
NO2	<---	Norms	.916
NO1	<---	Norms	.685
PC6	<---	Perceived_control	.630
PC5	<---	Perceived_control	.690
PC4	<---	Perceived_control	.778
PC3	<---	Perceived_control	.756
PC2	<---	Perceived_control	.567

APPENDIX C. Re-specified Model 2



Model Fit Summary for Final Measurement Model

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	50	**616.839**	160	.000	**3.855**
Saturated model	210	.000	0		
Independence model	20	7516.539	190	.000	39.561

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.052	**.925**	**.901**	.704
Saturated model	.000	1.000		
Independence model	.314	.343	.274	.310

Baseline Comparisons

Model	NFIDelta1	RFIrho1	IFIDelta2	TLIrho2	CFI
Default model	**.918**	.903	.938	.926	**.938**
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	.842	.773	.790
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	**.060**	**.055**	**.065**	.,001
Independence model	.221	.216	.225	.000

AIC

Model	AIC	BCC	BIC	CAIC
Default model	716.839	719.559	950.693	1000.693
Saturated model	420.000	431.425	1402.188	1612.188
Independence model	7556.539	7557.627	7650.081	7670.081

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Default model	.904	.813	1.005	.907
Saturated model	.530	.530	.530	.544
Independence model	9.529	9.176	9.890	9.530

Maximum Likelihood Estimates

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
PC9	<---	Perceived_control	.722	.045	16.054	***	par_1
PC1	<---	Perceived_control	.981	.058	16.798	***	par_2
AA1	<---	Aff_Att	1.000				
AA2	<---	Aff_Att	1.077	.108	9.942	***	par_3
AA3	<---	Aff_Att	.869	.080	10.801	***	par_4
AA4	<---	Aff_Att	.964	.093	10.330	***	par_5
CUI1	<---	CdmUse_Intention	1.000				
CUI2	<---	CdmUse_Intention	.905	.038	23.819	***	par_6
IA3	<---	Instr_Att	1.000				
IA2	<---	Instr_Att	1.204	.044	27.128	***	par_7
IA1	<---	Instr_Att	1.142	.046	24.662	***	par_8
NO4	<---	Norms	1.000				
NO3	<---	Norms	1.380	.052	26.331	***	par_9
NO2	<---	Norms	1.356	.052	26.219	***	par_10
NO1	<---	Norms	.862	.045	19.268	***	par_11
PC6	<---	Perceived_control	.900	.051	17.595	***	par_22
PC5	<---	Perceived_control	1.027	.052	19.626	***	par_23
PC4	<---	Perceived_control	1.000				
PC3	<---	Perceived_control	1.025	.047	21.870	***	par_24
PC2	<---	Perceived_control	.994	.063	15.765	***	par_25

Standardized Regression Weights: (Group number 1 - Default model)

Path	Estimate
PC9 <— Perceived_control	.580
PC1 <— Perceived_control	.604
AA1 <— Aff_Att	.561
AA2 <— Aff_Att	.528
AA3 <— Aff_Att	.628
AA4 <— Aff_Att	.566
CUI1 <— CdmUse_Intention	.884
CUI2 <— CdmUse_Intention	.831
IA3 <— Instr_Att	.823
IA2 <— Instr_Att	.903
IA1 <— Instr_Att	.792
NO4 <— Norms	.739
NO3 <— Norms	.921
NO2 <— Norms	.916
NO1 <— Norms	.685
PC6 <— Perceived_control	.630
PC5 <— Perceived_control	.694
PC4 <— Perceived_control	.784
PC3 <— Perceived_control	.764
PC2 <— Perceived_control	.570

Covariances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
Aff_Att <-> Instr_Att	-.144	.023	-6.343	***	par_12
Aff_Att <-> CdmUse_Intention	-.199	.027	-7.400	***	par_13
CdmUse_Intention <-> Norms	.337	.031	10.778	***	par_14
Instr_Att <-> Norms	.111	.024	4.686	***	par_15
CdmUse_Intention <-> Instr_Att	.169	.024	7.103	***	par_16
Aff_Att <-> Norms	-.160	.026	-6.061	***	par_17
Perceived_control <-> CdmUse_Intention	.382	.028	13.462	***	par_18
Perceived_control <-> Aff_Att	-.195	.024	-8.016	***	par_19
Perceived_control <-> Norms	.255	.026	9.631	***	par_20
Perceived_control <-> Instr_Att	.165	.021	7.846	***	par_21

Correlations: (Group number 1 - Default model)

	Estimate
Aff_Att <-> Instr_Att	-.333
Aff_Att <-> CdmUse_Intention	-.413
CdmUse_Intention <-> Norms	.513
Instr_Att <-> Norms	.189
CdmUse_Intention <-> Instr_Att	.306
Aff_Att <-> Norms	-.311
Perceived_control <-> CdmUse_Intention	.724
Perceived_control <-> Aff_Att	-.473
Perceived_control <-> Norms	.451
Perceived_control <-> Instr_Att	.347

Variances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
Perceived_control	.453	.036	12.557	***	par_26
Aff_Att	.377	.054	6.981	***	par_27
CdmUse_Intention	.612	.043	14.239	***	par_28
Instr_Att	.495	.037	13.441	***	par_29
Norms	.702	.059	11.882	***	par_30
e3	.467	.025	18.503	***	par_31
e6	.559	.031	18.077	***	par_32
e7	.514	.030	17.306	***	par_33
e8	.284	.018	15.411	***	par_34
e9	.339	.021	15.961	***	par_35
e10	.931	.050	18.571	***	par_36
e11	.760	.042	18.311	***	par_37
e12	.823	.052	15.829	***	par_38
e13	1.128	.068	16.476	***	par_39
e14	.438	.031	14.098	***	par_40
e15	.742	.047	15.705	***	par_41
e25	.171	.021	8.058	***	par_42
e26	.225	.019	11.541	***	par_43
e18	.235	.017	13.811	***	par_44
e19	.162	.019	8.348	***	par_45
e20	.384	.025	15.276	***	par_46
e21	.582	.032	18.043	***	par_47
e22	.238	.024	10.076	***	par_48
e23	.248	.023	10.626	***	par_49
e24	.590	.032	18.555	***	par_50

APPENDIX D. Model Fit Summary of Structural Model

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	50	616.839	160	.000	3.855
Saturated model	210	.000	0		
Independence model	20	7516.539	190	.000	39.561

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.052	.925	.901	.704
Saturated model	.000	1.000		
Independence model	.314	.343	.274	.310

Baseline Comparisons

Model	NFIDelta1	RFIrho1	IFIDelta2	TLIrho2	CFI
Default model	.918	.903	.938	.926	.938
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	.842	.773	.790
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

NCP

Model	NCP	LO 90	HI 90
Default model	456.839	384.363	536.882
Saturated model	.000	.000	.000
Independence model	7326.539	7046.312	7613.098

FMIN

Model	FMIN	F0	LO 90	HI 90
Default model	.778	.576	.485	.677
Saturated model	.000	.000	.000	.000
Independence model	9.479	9.239	8.886	9.600

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.060	.055	.065	.001
Independence model	.221	.216	.225	.000

AIC

Model	AIC	BCC	BIC	CAIC
Default model	716.839	719.559	950.693	1000.693
Saturated model	420.000	431.425	1402.188	1612.188
Independence model	7556.539	7557.627	7650.081	7670.081

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Default model	.904	.813	1.005	.907
Saturated model	.530	.530	.530	.544
Independence model	9.529	9.176	9.890	9.530

HOELTER

Model	HOELTER.05	HOELTER.01
Default model	245	263
Independence model	24	26

APPENDIX E. PROC GAM Output with four fixed degrees of freedom

The SAS System

The GAM Procedure

Dependent Variable: CdmUse_Intention

Smoothing Model Component(s): spline (Aff_Att) spline (Instr_Att) spline (Norms) spline (Perceived_control)

Summary of Input Data Set	
Number of Observations	794
Number of Missing Observations	0
Distribution	Gaussian
Link Function	Identity

Iteration Summary and Fit Statistics	
Final Number of Backfitting Iterations	7
Final Backfitting Criterion	3.3128439E-9
The Deviance of the Final Estimate	266.56426314

The backfitting algorithm converged.

Regression Model Analysis Parameter Estimates				
Parameter	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1.28141	0.18963	6.76	<.0001
Linear(Aff_Att)	-0.03978	0.03022	-1.32	0.1884
Linear(Instr_Att)	0.06113	0.02646	2.31	0.0211
Linear(Norms)	0.20284	0.02307	8.79	<.0001
Linear(Perceived_control)	0.51140	0.03569	14.33	<.0001

Smoothing Model AnalysisFit Summary for Smoothing Components				
Component	**Smoothing Parameter**	**DF**	**GCV**	**Num Unique Obs**
Spline(Aff_Att)	0.979694	3.000000	0.365796	15
Spline(Instr_Att)	0.953392	3.000000	0.115575	13
Spline(Norms)	0.968815	3.000000	0.090014	17
Spline(Perceived_control)	0.997398	3.000000	0.462875	27

Smoothing Model Analysis Approximate Analysis of Deviance			
Source	**DF**	**F Value**	**Pr > F**
Spline(Aff_Att)	3.00000	2.90	0.0344
Spline(Instr_Att)	3.00000	2.51	0.0575
Spline(Norms)	3.00000	6.86	0.0001
Spline(Perceived_control)	3.00000	11.25	<.0001