CHAPTER 3

RESEARCH METHODOLOGY

3.0 Introduction

This chapter presents the research methodology for the study. In this chapter, the researcher discussed the research paradigm and the methodology being used which incorporated quantitative as its research design. The discussion also involved the sampling, data collection, and analysis procedures.

3.1 Hypotheses Development

The primary objective of this study is to examine some selected factors that may influence the professionals' behaviour to comply with the tax law. To accomplish that, the study utilizes the MARS model of individual behaviour and performance which consists of motivation, ability, role perception, and situational factors as the basis to understand the individual taxpayer compliance behaviour. To further understand the behaviour to comply with the tax, the original model of MARS is extended with religiosity and ethical sensitivity variables. Based on the conceptual framework, hypotheses are then constructed. The hypotheses are developed concerning the relationship of MARS model elements (motivation, ability, role perception, and situational factors) with additional religiosity and ethical sensitivity variables in explaining the individual taxpayer compliance behaviour. To ensure consistency, the numbering of the subsequent hypotheses follows the precedinghypotheses.

3.1.1 The Influence of Motivation on Tax Compliance Behaviour

According to Campbell & Pritchard, (1976), motivation refers to a force within someone that pushes him or her direction, intensity, and persistence of voluntary behaviour. As McShane and Glinow, (2005) posited, to perform a job performance, individual employees must have the motivation to perform. As the theories of motivation are abundant in nature, the present study chooses to explain motivation to comply with the tax law with the concept of intrinsic and extrinsic motivation as suggested in the Self-Determination Theory. This concepthas been utilized widely by various researchers such as Herzberg (1959); Ryan & Deci (2000).

1. Does motivation have significant positive influence in tax compliance behaviour among professionals in Malaysia?

   H1: Motivation have a positive effect on tax compliance behaviour.

3.1.2 The Influence of Ability on Tax Compliance Behaviour

McShane & Glinow (2005) define ability as both natural aptitudes and learned capabilities required to successfully complete a task. Aptitudes here refer to natural talents that help an individual to learn specific tasks more quickly and perform them better while learned capabilities are the skills and knowledge that an individual currently possesses. Aptitudes and learned capabilities are always be associated with competencies. Competencies according to Spencer & Spencer (1993) are characteristics of a person that result in superior performance. Loo & Ho (2005) who study competencies among the taxpayers relate competencies with functional tax literacy. Functional tax literacy is defined as the ability of a taxpayer to file tax returns and calculate his or her tax liability independently and it encompasses the comprehension of some tax jargon and having fundamental tax knowledge on what constitutes taxable income, allowable deductions, relief, and rebates (Madi & Amrizah, 2003).

As past studies that adopted the MARS model, as well as competency study on tax, refers ability or competency as tax knowledge or tax literacy, therefore the present study also used the same path. This is relevant as other tax studies, for example, Siti Syaqilah Hambali & Amrizah Kamaluddin (2017) who studied taxpayers' competency levels on Goods and Service Tax (GST) also refers competency as literacy and awareness.

Based on the discussion in Chapter 2, no study analyzes the role of ability in tax compliance behaviour in Malaysia. Some studies in Malaysia examined the role oftax knowledge in understanding taxpayer behaviour in complying with the tax law.For instance, the study by Mohd Rizal Palil (2010) found a significant impact of tax knowledge on tax compliance. Another study is by Loo et al., (2009) who also confirmed a similar result. Therefore, the following hypothesis is suggested.

1. Does ability have significant positive influence in tax compliance behaviour among professionals in Malaysia?

   H2: Ability have a positive effect on tax compliance behaviour.

3.1.3. The Influence of Role Perception on Tax Compliance Behaviour

As motivation and abilities are two significant predictors of one behaviour, role perception as proposed by the MARS model is another influence. According to McShane & Glinow (2005), role perception refers to the extent to which people understand the job roles assigned to them or expected of them.

The discussion in Chapter 2 reveals that no study examines the influences of role perception towards tax compliance behaviour. This present study, therefore, molds role perception into the concept of civic duty of a citizen. The study of Orviska & Hudson, (2003) examined the role of civic duty in taxpayer compliance behaviour. The study found that civic duty tends to impact tax evasion primarily. Therefore, the following hypothesis is proposed.

1. Does role perception have significant positive influence in tax compliance behaviour among professionals in Malaysia?

   H_3: Role perception have a positive effect on tax compliance behaviour.

3.1.4. The Influence of Religion on Tax Compliance Behaviour

Religiosity is proved to be a significant determinant in the lifestyle of the Malaysian community. Some studies in the Malaysian context has demonstrated the significance of religiosity in various dimensions such as health problem (Yadollah Abolfathi Momtaz, Tengku Aizan Hamid, 2009), moral judgment (Muhamad & Ghani, 2006), social problems (Hazizan Md. Noon & Amin, 2003) and retail behaviour (Mokhlis, 2006). In the context of tax compliance, Mohd Rizal Palil et al., (2013) Nor Raihana Mohd Ali & Pope (2014), and Tekeli (2011) have confirmed the influence of religiosity in understanding tax compliance behaviour. The findings, however, are different in terms of its influence magnitude. Looking at the importance of religion in the Malaysian community, the present study, therefore, attempts to add in religiosity variable to the original MARS model. Hence, the following hypothesis is proposed.

1. Does religiosity have significant positive influence in tax compliance behaviour among professionals in Malaysia?

   H_4: Religiosity have a positive effect on tax compliance behaviour.

3.1.5. The Influence of Ethical Sensitivity towards Tax Compliance Behaviour

Another variable to be included in the framework of the present research is ethical sensitivity. Ethical sensitivity is found to be an important issue in Malaysia. Numbers of studies were carried out to investigate the impact of ethical sensitivity among auditors (Ahmed Mohamed Alteer, Sofri Yahaya, & Md Harashid Haron, 2013), accounting practitioners (Ismail & Ghazali, 2011), and tax practitioners (Suhaila Abdul Hamid, 2014). However, there is no study, to the researcher's knowledge, has analyzed the role of ethical sensitivity in predicting tax compliance behaviour among the professional groups in the Malaysian context. Hence, the current study aims to fill the gap by examining the aforementioned. The following hypothesis is proposed.

1. Does ethical sensitivity have significant positive influence in tax compliance behaviour among professionals in Malaysia.

   H_5: Ethical sensitivity have a positive effect on tax compliance behaviour.

3.1.6 The Moderating Effect: Situational Factor

The last element mentioned in the MARS model is situational factors. Situational factors refer to the factors that are beyond the immediate control of the people such as time, other people, budget, and physical work facilities. According to McShane & Glinow (2005), the situational factor is needed to strengthen the motivation and ability to influence performance. It is referred to as a situation in which the environment can influence the performance of an individual. Past tax compliance studies proved that, as presented in Chapter 2, situational factors might influence taxpayers to comply with the tax law. Few studies examined the role of situational factors towards complying with tax law such as Song & Yarbrough (2016) and Trivedi, et.al, (2003). These include factors like level of income, the unemployment rate, tax rates, the effectiveness of tax administration agencies, social norms, moral intensity, and expectations of audit. The present study, however, will only focus on the influence of financial strain and peer reporting behaviour as proxies to situational factors. These factors are chosen as there is a lack of investigation of the variables, especially in the Malaysian context. To date, there is no empirical evidence that associates situational factors to the behaviour to comply with tax in the Malaysiancontext.

The situational factor acts as a moderator in this study. The present study aims to test whether situational factor affects the strength of the relationship between individual behaviour construct and tax compliance. There are five hypotheses are proposed.

1. Does the situational factor moderate the relationship between motivation and tax compliance behaviour among professionals in Malaysia?

   H_6: The positive relationship between motivation and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.

2. Does the situational factor moderate the relationship between ability and tax compliance behaviour among professionals in Malaysia?

   H_7: The positive relationship between ability and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.

3. Does the situational factor moderate the relationship between role perception and tax compliance behaviour among professionals in Malaysia?

   H_8: The positive relationship between role perception and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.

4. Does the situational factor moderate the relationship between religiosity and tax compliance behaviour among professionals in Malaysia?

   H_9: The positive relationship between religiosity and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.

5. Does the situational factor moderate the relationship between ethical sensitivity and tax compliance behaviour among professionals in Malaysia?

   H_{10}: The positive relationship between ethical sensitivity and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.

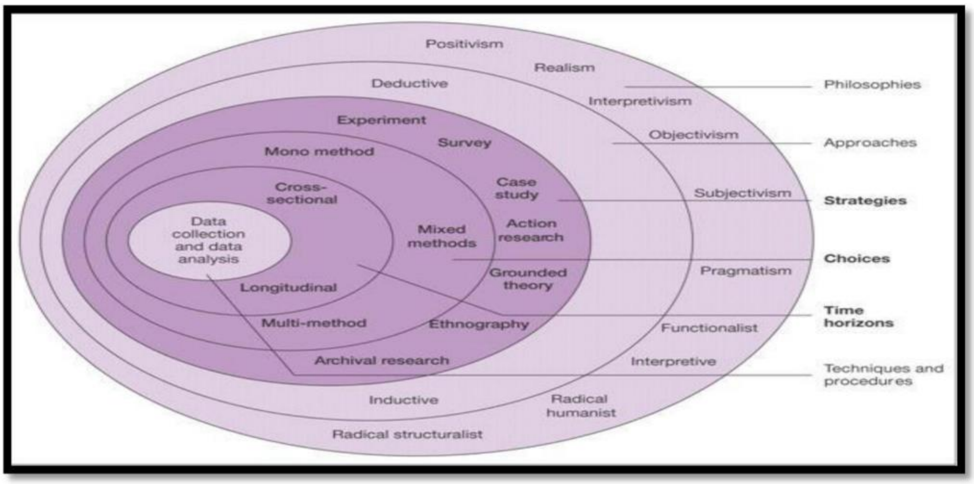
With reference to the conceptual framework, ten hypotheses are developed to be tested in the present study. Table 3.1 below presents a summary of the research hypotheses to be tested.

Table 3.1: Summary of research questions and research hypotheses

Research Questions	Research Hypotheses
Does motivation, ability, role perception, religiosity and ethical sensitivity have significant positive influence in tax compliance behaviour among professionals in Malaysia?	*H_1: Motivation have a positive effect on tax compliance behaviour.* *H2: Ability have a positive effect on tax compliance behaviour.* *H3: Role perception have a positive effect on tax compliance behaviour.* *H4: Religiosity have a positive effect on tax compliance behaviour.* *H5: Ethical sensitivity have a positive effect on tax compliance behaviour.*
Does the situational factor moderate the relationship between motivation and tax compliance behaviour among professionals in Malaysia?	*H6: The positive relationship between motivation and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.*
Does the situational factor moderate the relationship between ability and tax compliance behaviour among professionals in Malaysia?	*H7: The positive relationship between ability and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.*
Does the situational factor moderate the relationship between role perception and tax compliance behaviour among professionals in Malaysia?	*H8: The positive relationship between role perception and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.*
Does the situational factor moderate the relationship between religiosity and tax compliance behaviour among professionals in Malaysia?	*H9: The positive relationship between religiosity and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.*
Does the situational factor moderate the relationship between ethical sensitivity and tax compliance behaviour among professionals in Malaysia?	*H10: The positive relationship between ethical sensitivity and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.*

3.2. Research Paradigm

The idea of paradigm or worldview as an overarching framework is proposed by Kuhn (1962) in his work of The Structure of Scientific Revolutions. Kuhn (1962 pg.45), refers paradigm as “the set of common beliefs and agreements shared between scientists about how problems should be understood and addressed. The basic idea of the research paradigm is for the researcher to have a worldview of the problems that the study attempts to address. The identification of the research paradigm is crucial in steering the ways of doing research. As stated by Mackenzie & Knipe (2006), recognizing a paradigm as the first step is important as it is the step for the subsequent activities regarding methodology, literature, or research design. The identification of the theoretical paradigm or philosophical assumptions about the nature of reality is significant in understanding the overall perspective from which the study is designed and carried out. Saunders et al., (2009) develop a “research onion” (See Figure 3.1) which depicted the levels or layers of the research project and how every layer needs to be addressed while working towards the center. The choice of the research paradigm will influence the strategies and the methodology.



Source: Saunders 2006, pg 83

Figure 3.1: "The Research Onion"

Paradigm can be characterized through ontology (philosophy of reality), epistemology (philosophy of knowledge), and methodology (procedure to acquire knowledge). These characteristics create a holistic view of how we view knowledge. Ontology involves the philosophy of reality, epistemology on the other hand addresses the ways the researcher comes to know the reality, while methodology identifies practices to attain the knowledge (Krauss, 2005). Paradigm choice is a reflection of how the researchers view the world (ontology) and believes that knowledge is created (epistemology) (Mckerchar & Sharkey, 2008). According to Creswell (2013), there are four common paradigms used in social science research; positivism, constructivism, transformative, and pragmatism.

Positivism comes into existence through Auguste Comte's denunciation of metaphysics and his contention that barely only technical and scientific facts can disclose the reality concerning truth. It was later officially recognized as the dominant scientific method in the early 20th century by constituents of the Vienna Circle, with

Karl Menger, Otto Neurath, Rudolf Carnap, Gustav Bergmann, Philipp Frank, Herbert Feigl, and Moritz Schlick. The positivist paradigm believes that the world is external (Carson et al., 1988) and there is a single objective reality to any research phenomenon waiting to be discovered (Hudson & Ozanne, 1988). This paradigm focuses on describing, explaining, and uncovering facts, through stated theories and hypotheses (Jonker & Pennink, 2010).

The ontological position of positivism is one of realism. Realism is the view that objects have an existence independent of the knower (Cohen et al., 2007), while the epistemology for the positivist is one of objectivism (Scotland, 2012). The belief isthat the researcher should remain distant from the research to avoid bias in the research that can distort their objective view. In the positivist view, there is a clear distinction between science, personal experience, and value judgment. Positivist methodology is directed at explaining relationships. The purpose of positivism is to acquire an answer for the probability of causation that can be used to predict human activity (Mohamed, 2016). This can be achieved through empirical evidence and tested theories, which normally be conducted through observations or measurement of phenomena (Krauss, 2005). Positivist as a research philosophy normally be conducted through quantitative research methods using various statistical techniques. Positivist research can be used to generalize findings if there is a sufficiently large sample size (Mertens, 2010).

Interpretivist paradigm on the other hand began with the philosophy of Edmund Husserl's phenomenology and Wilhelm Dilthey's and other German philosophers' study of interpretive understanding called hermeneutics (Mertens, 2010).

Hermeneutics refers to the study of meaning and interpretation in the historical context. The purpose of research that the interpretive holds is to understand the experience of people in their everyday lives. The interpretivism is interested to uncover the meaning in a given situation. The interpretive researcher's role is to understand, explain, and interpret the social reality through the eyes of the participant. They believe that the world is complex, and it will constantly change. While positivist view in terms of cause and effect, interpretivism view such patterns as being created out of evolving meaning systems (Neuman & Dickinson, 2003). Positivists view that the researcher is external to the study which is in contrast with interpretivism that holds the view that the researcher and the people under study interact with each other (Hudson & Ozanne, 1988).

Interpretivist sometimes be referred to as anti-positivism is on the opinion that,the researcher could not be detached from the subjects being studied. This is because, to understand the reality behind every situation, the researcher should be part of the study. Only through interaction with the people under study, the researcher will be able to uncover the truth. Further, in seeking to answers the research, the personal values of the researcher influence the findings. The ontological position of interpretivism is relativism. Relativism is the view that reality is subjective and it differs from one individual to another (Scotland, 2012).

According to the interpretive, social reality is seen by multiple people andthese people interpret a situation differently. They deny the existence of one real world. The multiple realities can be explored, and to understand it, the researcher makes sense of thr social world by being part of the people and interacting with them.

Interpretive epistemology is subjectivism which is based on real phenomena. The world does not exist independently of our knowledge of it (Grix, 2004). The interpretive methodology is directed at the understanding phenomena from an individual's perspective.

As the primary purpose of interpretivism is to understand the meaning, the interpretative researchers appreciate the first-hand experience, truthful reporting, and direct quotations of the actual conversations from the insider's perspective (Tubey et.al., 2015). The researcher employed data gathering methods that are sensitive to a context which enable a rich description of social phenomena. The normal data collection method utilized by the interpretivism are interview, focus group discussion, and case study, which is considered qualitative method.

Unlike positivists which are able to generalize findings, the findings of the interpretative data cannot be generalized to its populations. This is because the primary aim of the interpretivism is to understand the situation and uncover the truth. While positivists follow a systematic research process to find the answer to the study, interpretive avoid rigid frameworks, they adopt a more flexible structure. They believe in anti-foundationalism which is no single route to discover meaning.

Applying the concepts of positivism and non-positivism to studies in taxation, McKerchar (2012) suggests that the use of deductive reasoning and quantitative approach reflect the positivism paradigm. In contrast, inductive reasoning and the qualitative approach imply the non-positivism research paradigm in understanding tax issues (McKerchar, 2012).

While positivism focus on quantitative methods and interpretive highlight on qualitative methods, the pragmatist paradigm lies in between the two paradigms. Incorporating the strength between the two approaches, pragmatists adopt a mixed-method approach. It was first introduced into philosophy by Charles Sanders Peirce in 1878. This method gained its position over the past 30 years. Mixed-method enables richness of understanding of phenomena through the confirmation of results and extension of knowledge (Wright & Losekoot, 2003).

Generally, most tax compliance studies revolve around positivism tradition, whereby the knowledge is created through an analysis of the causal relationship between taxpayer and tax compliance (See, for example, Braithwaite & Ahmed, 2005; Damayanti et al., 2015; Feld & Frey, 2009; Gilligan et al., 2008; Kirchler et al., 2003; Natrah Saad, 2010; Torgler, 2003c; Torgler & Schaltegger, 2006). However, the idea of mixed-method has gained its position in the last 5 years (See, for example, Nor Raihana Mohd Ali, 2013; Suhaila Abdul Hamid, 2014; Khadijah Mohd Isa, 2012; Lavermicocca & Mckerchar, 2013; Natrah Saad, 2011) intending to gain a comprehensive and deeper understanding on tax compliance behaviour.

The decision to apply whether positivism, interpretivism, or pragmatism is based on the philosophical belief of the study. As mentioned by Creswell (2013), elements such as influence, causes, factors, relationships, and effects lie within the positivist paradigm. As the core purpose of this study is to explore the influences of all the four elements of the MARS model of individual behaviour with additional variables of religiosity and ethical sensitivity, the positivism view is well suited as the research paradigm. On top of that, this research aims to generalize the findings of the research to the larger professional population of the taxpayer in Malaysia, hence the adoption of the positivist paradigm answers its purpose.

3.3. Research Method

As discussed in the previous section on the research paradigm, the present study adopted a quantitative approach as its methodology. This quantitative design reflects positivist philosophical assumptions. As defined by Cohen et al., (2007), quantitative research is defined as social research that utilized empirical methods and empirical statements. He further states that an empirical statement is referred to as a descriptive statement about what is the real world is rather than what the real world is ought to. Furthermore, Creswell (2013), defined quantitative research as a type of research that explains a phenomenon via analyzed numerical data using statistical methods. If compared to the qualitative method, which is less specific and less precise, quantitative studies are specific, well-structured, and have been proved their validity and reliability (Kumar, 2011). Quantitative research isolates variables and categorizes them. These variables are then linked together to form a hypothesis.

The present study adopted quantitative as its research methodology becausethe quantitative method is the best method that can be used to explain a phenomenon. It is a great method to further describe what constitutes a phenomenon, what are factors lie behind a phenomenon. As the current research interested to understand atax compliance phenomenon, the researcher aims to investigate whether motivation, ability, role perception, situational, religiosity factors, and ethical sensitivity do have influences on tax compliance behaviour, therefore, quantitative design is the best to suit its purpose.

In addition to that, the quantitative method is also useful to quantify opinions, attitudes, and behaviours to understand the population views towards certain phenomena. As this study aims to understand the taxpayer compliance behaviour, and to further investigate whether all related variables (as explained in the conceptual framework) do influence the taxpayer in complying with the tax law, hence, quantitative method is viewed as the best method. The quantitative method is also the most suitable method to test hypotheses. The primary objective of the current study is to test hypotheses of the various variables as discussed earlier.

Based on the justification discussed above, the researcher is on the view that the quantitative method is the best methodology that can be utilized in the current research.

Table 3.2: Review of Past Studies That Used Survey Method

Authors/ years	**Sample**	**Data Analysis**
Mohd Alif Anwar Abu Bakar et al (2021)	78 taxpayers	SmartPLS
Yusri Huzaimi et al (2021)	152 taxpayers	SmartPLS
Hardika, N. S., Wicaksana, K. A. B., & Subratha, I. N. (2021)	100 taxpayers	SmartPLS
Masunga, F. J., Kiria, J., & James, M. (2020).	109 tax students	SmartPLS
Al-Zaqeba, M. A. A., & Al-Rashdan, M. T. (2020).	480 taxpayers	SmartPLS
Fauziati, P., Minovia, A. F., Muslim, R. Y., & Nasrah, R. (2020)	300 taxpayers	SPSS
Agyeiwaa, H., Amankwaah, E., Abina, S., Agyei, K. N., & Antwi, K. (2019)	399 taxpayers	SPSS

Sritharan, N., & Salawati, S. (2019). .	419 taxpayers	SPSS
Malangkucecwara, S. T. I. E. (2018)	300 taxpayers	SmartPLS
Puspita, A. F., Subroto, B., & Baridwan, Z. (2016). 0.	136 taxpayers	SmartPLS
Fauziyyah, D. I., Septiani, T., & Musyaffi, A. M. (2016).	100 taxpayers	SmartPLS
Faizal, S. M., & Palil, M. R. (2015).	400 taxpayers	SPSS
Ocheni (2015)	300 taxpayers	SPSS
Zeng, T. (2014).	174taxpayers	SPSS
Modugu and Anyaduba (2014)	500 taxpayers	SPSS
Mohd Rizal Palil, Mohamad Abu Hamid, & Mohd Hizam Hanafiah (2014)	5500 taxpayers	SPSS
Suhaila Abdul Hamid (2013)	1500 taxpayers	SmartPLS
Modugu, P. K., Eragbhe, E., & Izedonmi, F. (2012).	600 taxpayers	SPSS
Alabede, J. O., Ariffin, Z. Z., & Idris, K. M. (2011).	382 taxpayers	SPSS
Mohd Rizal Palil & Mohd Fariq Mustapha (2011)	1075 taxpayers	SPSS
Murphy (2008)	658 taxpayers	SPSS
Chattopadhyay, S., & Das-Gupta, A. (2002).	172 taxpayers	SPSS

3.4. Research Design

Research design is the overall plan for connecting the conceptual research problems to the empirical research. It articulates the type of data, the methodology used to collect, and analyze the data and how the data can be used to answer the research question. As mentioned by Parahoo, (1997 pg. 142) a research design is “a plan that describes how, when and where data are to be collected and analyzed”. Therefore, the following discussion will revolve around the type of research followed by the type of data and methods used to collect and analyze the data.

3.4.1. Research Purpose

Research can be done in variously and most types of research can be classified according to how the researcher knows about the research problem before the investigation took place. According to Salkind et al., (2010) research can be grouped into exploratory, descriptive, explanatory, and evaluative. Exploratory studies areusually undertaken when little is known about a phenomenon. The goal is to determine whether a relationship exists among several variables. While descriptive research attempts to describe a group of people or a phenomenon. It is often be conducted to understand social problems and particular issues. On the other hand, explanatory research focuses on why variables are related to one another. Its primary purpose is to test a theory, and evaluative research purpose is to assess whether a particular intervention, process, or procedure able to change behavior.

As mentioned in the research objective earlier, the primary objective of the present research is to test the various elements described in the MARS model to understand the taxpayers' behaviour towards complying with the tax law. Therefore, the present study is explanatory. The researcher measures all the variables involved and then provides evidence to support refutes the contention that there is cause and effect relationship between the variables.

3.4.2. Research Process

This study followed a research process that comprises of defining the research issues, formulating hypotheses, designing the research tool, collecting and evaluating information, and discussing the summary of the results. Figure 3.2 outlines the research process of the study. First, the researcher defined the issues of the study by reviewing the literature. Based on the literature review conducted, a deeper knowledge of the variables is suggested.

Next, in the stage of the conceptual development stage, attention is given tothe model used in the current study. The framework of the study is adopted from the original MARS model of individual behaviour and performance (McShane & Glinow, 2005) where three independent variables, one moderator, and one dependent variable is adopted into the framework of the study. Based on the literature review, another two variables are incorporated into the research framework, which are religion and ethical sensitivity. In developing the research design, the researchers have considered different aspects such as instrument development, sampling design, and pretesting. The sampling design determines the process to identify sample size and sampling techniques. The measurement for the variables involved in the study is tailored from the past research and some of them are self-developed by the researcher. Based on the pretest, the validity of the research instrument was assessed. In the next stage, a preliminary assessment was performed to determine the reliability of the instrument. The final face-to-face survey was performed. The data obtained were screened using SPSS Version 23 software. The Partial Least Square- Structural Equation Modelling (PLS-SEM) was used to test all the hypotheses. Finally,the results were interpreted and presented, and the findings were discussed and documented.

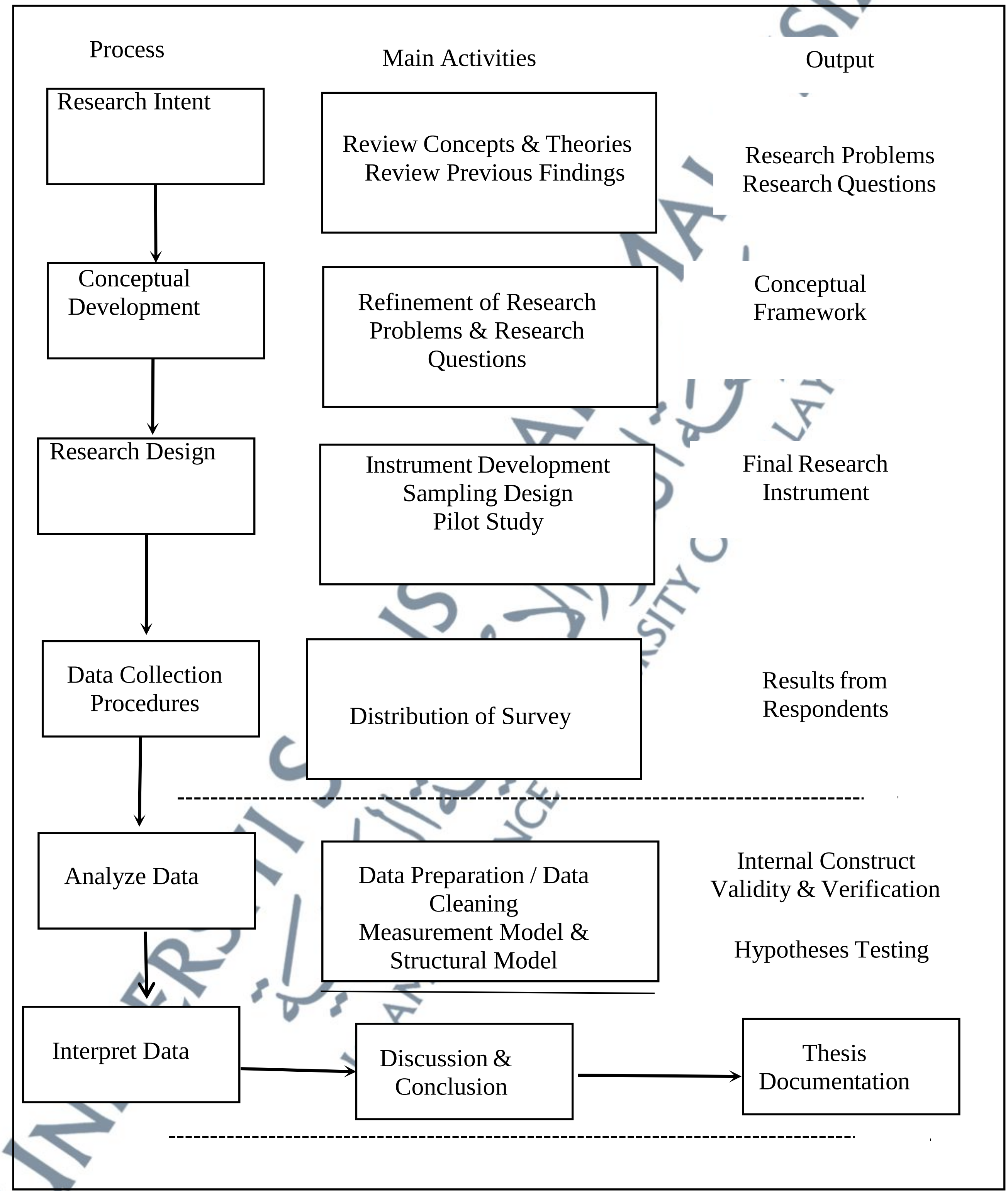


Figure 3.2: Research Process

3.5. Quantitative data collection

As illustrated in Figure 3.2, the Research Process diagram, after the research questions are developed, the researcher clarifies the data collection process which includes sampling, questionnaire design, and pilot study. This process is significant before the survey could be administered.

3.5.1. Sampling

The following discussion detailed the research's sampling frame, sample size, and sample selection.

3.5.1.1. Sampling Frame

According to Babbie, (2010) sampling frame refers to the list or quasi list of elements from which a probability sample is selected. This study involves understanding the compliance behaviour among the taxpayer a specific focus of the professional employees. The Employment Act of Malaysia (1955) provided no standard definition on professionals' employee, therefore the present study use the definition provides by the Inland Revenue Board (2019) where "professional" means a person who belongs to a profession and is a member of a professional body orregistered with a body which governs the profession. IRB also further defined "profession" as includes the profession of lawyers, accountants, architects, doctors, pharmacists, engineers, surveyors, or any other profession where the status of the profession is recognized by written law or statute in Malaysia.

In Malaysia, the most of the professionals are registered under their respective professional bodies. Among the registered professional body in Malaysia, are Lembaga Arkitek Malaysia, Lembaga Jurutera Malaysia, Malaysian Institute of Accountants (MIA), Malaysian Association of Certified Public Accountants (MICPA), The Institution of Engineers Malaysia, Pertubuhan Arkitek Malaysia, Majlis Perubatan Malaysia, Majlis Pergigian Malaysia, Lembaga Farmasi Malaysia, The Malaysian Bar and Institut Landskap Arkitek Malaysia and The Board of Quantity Surveyors Malaysia.

This study targeted professionals from six professions namely Accountants, Lawyers, Architects, Engineers, Doctors, and Surveyors. This is significant as IRB reported that RM1.111 million of additional assessment and penalties was levied on doctors, RM3.83 million engineers, RM1.81 million on architects, and RM7.39 million on lawyers. From this amount, 4,610 cases comprising 3,000 doctors, 182 lawyers, 141 architects, and 1,287 engineers (The Malaysian Reserve, 2017). Apart from that, these professions are selected because they are considered higher-earning professions in many countries such as in US and UK (Snowdon, 2011). As suggested by Kassipillai et. Al., (2003) study into a different group of taxpayers will help the tax authority in targeting groups which need extra education and attention.

3.5.1.2. Sample Size

Sample size can be defined as the subset of a population required to ensure thatthere is a sufficient amount of information to draw conclusions (Sekaran & Bougie, 2010). The sample size is significant to establish generalizability. It is a common process in research to determine appropriate sample size. The sample size must be adequate to ensure that the study is capable to produce useful results. It must be "big enough" so that it is statistical significance but not "too big" to avoid more resources than necessary.

As the current research involves the Malaysian Professional Group as its population, it is quite impossible for the researcher to conduct a nationwide pool. This is due to the geographical spread, extra effort, and cost which outweigh the benefits. Even there are some researchers that conduct nationwide surveys (See, for example, Christensen & Hite, 1997; Copeland & Harmelink, 1995; Hite & Roberts, 1991) the present study will only draw a sample from the population. As mentioned by Richardson & Sawyer (2001), it is common for the researchers to draw a sample from within a city or state, rather than considering a nationwide pool.

There are several approaches to determine sample size (See, for example, Cohen, 1970; Krejcie & Morgan, 1970; Roscoe, 1975). As the current study employed PLS-SEM as its data analysis technique, therefore determining sample size using Power analysis is crucial. According to Hairet al (2017) sample size should be based on power analysis, which is the minimum number of samples based on model complexity. Power analysis determines the minimum sample size by taking into account the part of a model with the largest number of predictors (Hair et al., 2014). It requires information related to power, effect size, and significance level to calculate the minimum required sample size (Hairet al., 2018). A value of 80 percent or more represents an adequate level of power in social science research (Cohen 1988, Hair et al., 2017). The G*Power application was used to determine the minimum sample size. According to Erdfelder, Faul, & Buchner(1996), G*Power is an all-in-one power analysis program which normally used in computer and social sciences research for statistical tests. For this study, the F test of regression was used via the G*Power application. To determine the accurate sample size, the Power analysis is set for multiple regression comprising eleven predictors, The test used alpha of 0.05, a power of 0.80, and a medium effect size of (f2=0.15).As 80 percent is considered as the minimum acceptable power in most social sciences studies (Gefen, Rigdon & Straub, 2011), the desired sample size was set to 118. Thisis also supported by 10 times rule (Barclay, Higgins & Thompson, 1995) which indicates the sample size should be equal to the larger of 10 times the largest number of formative indicators used to measure a single construct or 10 times the largest number of structural paths directed at a particular construct in the structural model. For this, a minimum sample size of 110 is needed.

As the table of Krejcie & Morgan (1970) offers more than 380 sample sizes, G*Power suggests 118 sample size and 10 times rule suggested 110 samples, the researcher decided to consider a larger sample size of 1000 considering the possibility of low response rates and minimum sample size requirements for PLS.

Table 3.3: Sample size recommendation in PLS-SEM for Statistical Power of 8

Maximum Number of Arrows Pointing at a Construct	Significance Level											
	1%				5%				10%			
	Minimum R^2				Minimum R^2				Minimum R^2			
	0.10	0.25	0.50	0.75	0.10	0.25	0.50	0.75	0.10	0.25	0.50	0.75
2	158	75	47	38	110	52	33	26	88	41	26	21
3	176	84	53	42	124	59	38	30	100	48	30	25
4	191	91	58	46	137	65	42	33	111	53	34	27
5	205	98	62	50	147	70	45	36	120	58	37	30
6	217	103	66	53	157	75	48	39	128	62	40	32
7	228	109	69	56	166	80	51	41	136	66	42	35
8	238	114	73	59	174	84	54	44	143	69	45	37
9	247	119	76	62	181	88	57	46	150	73	47	39
10	256	123	79	64	189	91	59	48	156	76	49	41

Source: Cohen (1992)

3.5.1.3. Sample Selection

There are two major groups of sampling techniques which is probability and non-probability sampling. The sample of the present study was selected using a cross between simple random sampling (probability sampling) and snowball sampling (non- probability sampling). Simple random sampling is a sampling design in which k distinct item is selected from the n items in the population in such a way that every possible combination of k item is equally likely to be the sample selected (Thompson, 2012). A database containing the names, addresses, and email addresses of the Professionals was developed by the researcher since there is a lack of information on the list of Professionals registered in Malaysia. The information was collected from the annual report and the websites of the respective registered professional bodies. From the list of the professionals, the researcher used Research Randomizer software in to generate a random number of respondents. The advantage of simple random sampling is it has the least bias and offers the most generalizability (Sekaran, 2003). However, the use of simple random sampling could become cumbersome and expensive (Sekaran, 2003). Therefore, in order to complement the process of data collection, the researcher has utilized snowball sampling. This is following the recommendation by Browne (2005), where when the subject of the research is sensitive then individual’s social networks can be applied particularly through snowballing. This is supported by Mahin Naderifar, Hamideh Goli and Freshteh Galjaie (2017) where snowball sampling is applied with the target characteristics are not easily accessible. This method is effective when studying hard-to-reach populations, such as professionals, who may not be easily accessible through traditional sampling method

(Freshteh Galjaie, 2017). In snowball sampling, one subject gives the researcher the name of another subject, who in turn provides the name of a third, and so on (Vogt & Johnson, 1999). This sampling method was advantageous because of the sensitive nature of the research topic. Respondents were understandably wary to answer questions on tax compliance. Therefore, it was much easier to access the population through referralsby their trusted professional colleagues.

However, since snowball sampling does not select units for inclusion in the sample based on random selection, it is impossible to determine the possible sampling error and make generalizations.

Since there is inadequate information on the list of professionals available in Malaysia, therefore the researcher develops a database containing the name list of the professionals. This is done by browsing through the list of professional bodies registered in Malaysia. A total of 303,512 professionals were acquired. These professionals are registered with their respective professional bodies. Therefore, to obtain a more robust sample, the researcher utilizes a snowball sampling approach by requesting those professionals who were selected from the database developed by the researcher, to distribute another two sets of the survey questionnaire packs to their peers. A similar approach has been used by Suhaila Abdul Hamid (2014) in selecting tax agents for samples in her study.

3.5.2. Questionnaire Design

A questionnaire is a pre-formulated written set of questions to which respondents record their answers (Sekaran, 2003). A survey is often used in studies that used individual people as the units of analysis. According to Babbie (2010), survey research is probably the best method for social science researcher who is interested to collect original data, where the population is too large to be observed directly. According to Jackson & Miliron (1986) and Richardson & Sawyer (2001), survey research was the most common approach in tax compliance studies.

The instrument of the study was developed by adopting, adapting items from existing validated measurements. The items were carefully modified to ensure they areunderstood by the respondents. There are also some of the items which are self-developed by the researcher.

Careful attention was paid to ensure that the wording of the items can be understood by the respondents. It is important to ensure that the questions are simple, easy to read, and straight to the point. The questionnaire is prepared in the English language with Malay translation.

The questionnaire is divided into three parts: Part 1 consists of hypothetical scenario questions; Part 2 consists of a close-ended statement, while Part 3 consists of questions on the respondent’s demographic information such as age, gender, ethnicity, education level, income level, and profession.

The items are categorized to 7-point items with the following response options 1 = Very Strongly Disagree", "2 = Strongly Disagree", "3 = Disagree", "4 = Neutral", "5= Agree", "6 = Strongly Agree", "7 = Very Strongly Agree". Several researchers have suggested that 7-point categories scale is the best that should be used to ensure validity and reliability (Finn, 1972; Foddy, 1994; Hudson & Miller, 2005). This is to support Artuna & Arslan (2016) who suggest that the reliability of scores can be optimized by the utilization of seven categories scale. It is argued that the human mind has a span ofapprehension capable of distinguishing about seven different items, which implies about seven on the number of categories that people are able to use in makingjudgments about the magnitudes of unidimensional stimuli.

An increasing renown method of data collection is the use of the online survey method. Online surveys offer a lower cost of data collection and higher speed; they are visual, interactive, and flexible. More importantly, they do not require interviewers to present and normally busy people will ignore taking part in a telephone survey and are more willing to answer questions posted on their computer screens (Kellner, 2004).The present study utilizes Survey Monkey as the medium of the online survey. Nevertheless, there are also some drawbacks of an online surveys. Online surveys are unable to collect a random sample of the population and if compared to the self- administered survey, the online survey does not enjoy a high response rate. Therefore, to ensure high response rates, an additional self-administered survey was used to complement the limitation of an online survey.

The main benefit of a self-administered questionnaire is that the researcher will be able to collect the completed responses within short periods of time. Any doubts that the respondents might have on any question could be clarified on the spot. The researcher is also afforded the opportunity to introduce the research topic and encourage the respondents to provide truthful answers. Compared to the interview method, personally administered questionnaires consume less time and cost (Sekaran, 2003). This method also assures a high response rate, accurate sampling, and minimum bias (Oppenheim, 1992).

Initially, the researcher will email the link of the survey to all respondents based on the email address in the database developed by the researcher. However, there are some respondents whose email address is not available. Therefore, the researcher approaches these respondents via face-to-face survey. Prior to that, phone calls were made for an appointment. In the case of respondents who are unable to complete the survey on the day, the researcher will collect the survey in two weeks' time and a follow-up telephone call reminder will be given.

The respondents will be provided with a questionnaire survey package. The questionnaire survey package contained a cover letter and one each for the questionnaire. The cover letter explained the purpose of the study, confirmation on the anonymity response for the survey, an invitation to participate in the survey, and ensured confidentiality of the survey data.

3.6. Pilot Study

Pilot testing or pre-testing refers to mini versions of a full-scale study as wellas the specific pre-testing of a particular research instrument. In the process, the questionnaire may have to be restructured and some items may have to be rewritten to fit the research objectives. At this stage, the focus is placed on the appropriate wording in questions, clarity of instructions, and the like rather than to report the result (Synodinos, 2003). Pre-testing aims to avoid poor data quality and to reduce or prevent the deletion of items during the measurement model evaluation (Ali Memon, Ting, Ramayah, Chuah, & Cheah, 2017).

In this study, the questionnaire was pre-tested in three stages. In stage one, the survey instrument was translated from the English language to the Malay language. This translation was performed by language experts to ensure wording and concepts are equivalent. The questionnaire was then improved based on the suggestion from the language experts.

To ensure the reliability and validity of the items, ten experts in the field of taxation in Malaysia were invited to provide feedback regarding the questionnaire items. The questionnaire was sent to ten tax experts in Malaysia. They are from Universiti Utara Malaysia, Universiti Islam Antarabangsa and Universiti Teknologi MARA, and three feedback were gained from this process. All experts were requested to comments on multiple elements, such as the appropriateness of the scale used, the accuracy of instructions provided, the clarity or uncertainty of construct definitions, and the representation of items. In summary, pre-testing tests the designed questionnaire for content validity. To enhance the reliability and validity of the questionnaire, the views of two tax practitioners were also acquired. The questionnaire is then improved after incorporating comments and suggestions.

In stage two, the questionnaires were sent out to Ph.D. students of Universiti Teknologi MARA. The involvement of postgraduate students is crucial with regard to their opinions on the terms or language used in the questionnaire. The study was only able to get seven students involved in the pilot study. Based on the feedback from the pilot testing, the wording of a few items were changed.

In the final stage, the survey instrument was validated by pre-testing to potential respondents. The questionnaire was distributed to 33 professional employees in Malaysia. Some of the items were rewritten based on their suggestions. Copies of the final survey questionnaires are included in Appendix 1.

Based on the data collected from the pilot survey, a reliability test is conductedto ensure the stability and consistency of the measurement. In this study, each construct is represented by different nature of items. Some are formative, while others are reflective. Hence, a multicollinearity test is conducted for formative constructs andCronbach alpha for reflective constructs.

3.7. Construct Development and Measurement

As mentioned earlier in the conceptual framework, to understand the taxpayer behaviour in complying with the tax law, the present study utilizes the MARS model of individual behaviour and performance with an extension of another two significant variables which are Religiosity and Ethical Sensitivity. Therefore, the proposed modelconsists of seven constructs, with four constructs derived from the MARS model of individual behaviour and performance (motivation, ability, role perception, and situational factor) with another two constructs namely religiosity and ethical sensitivity. These theoretical constructs are latent or unobservable. To measure a latent construct, the researcher captures indicators that represent the underlying construct. The indicators are directly observable, and the researcher believed thatthese indicators accurately represent the latent variable. Before discussing the specific measurement of these constructs, the following section discusses the differences between formative and reflective measures and explores the second-order factor whichbecomes important when constructs are defined at a more abstract level.

3.7.1. Reflective and Formative Measures

Generally, measurement models are using multiple indicators of latent constructs mentioned in the Structural Equation Modelling (SEM) literature namely the reflective model and the formative model. The main feature of these two models is depicted in Figure 3.3. In reflective measures, the latent variable or the principal factor is caused by the latent factor. As shown in Figure 3.3, in which each construct is represented by a circle with several arrows stemming from it to a set of indicators. The direction of causality is from the construct to the set of indicators (Jarvis, MacKenzie, & Podsakoff, 2003). On the contrary, in the formative model, if there are any changes be made in the measures it will cause changes in the underlying construct. Thedirection of causality moves from the indicators to the latent construct. The indicators act as a group to give meaning to the construct.

To determine the appropriate measurement model that could be used in the study, Hair et al., (2014) present a guideline to decide whether to measure a construct reflectively and formatively as depicted in Table 3.4.

Table 3.4: Guidelines for Choosing the Measurement Model Mode

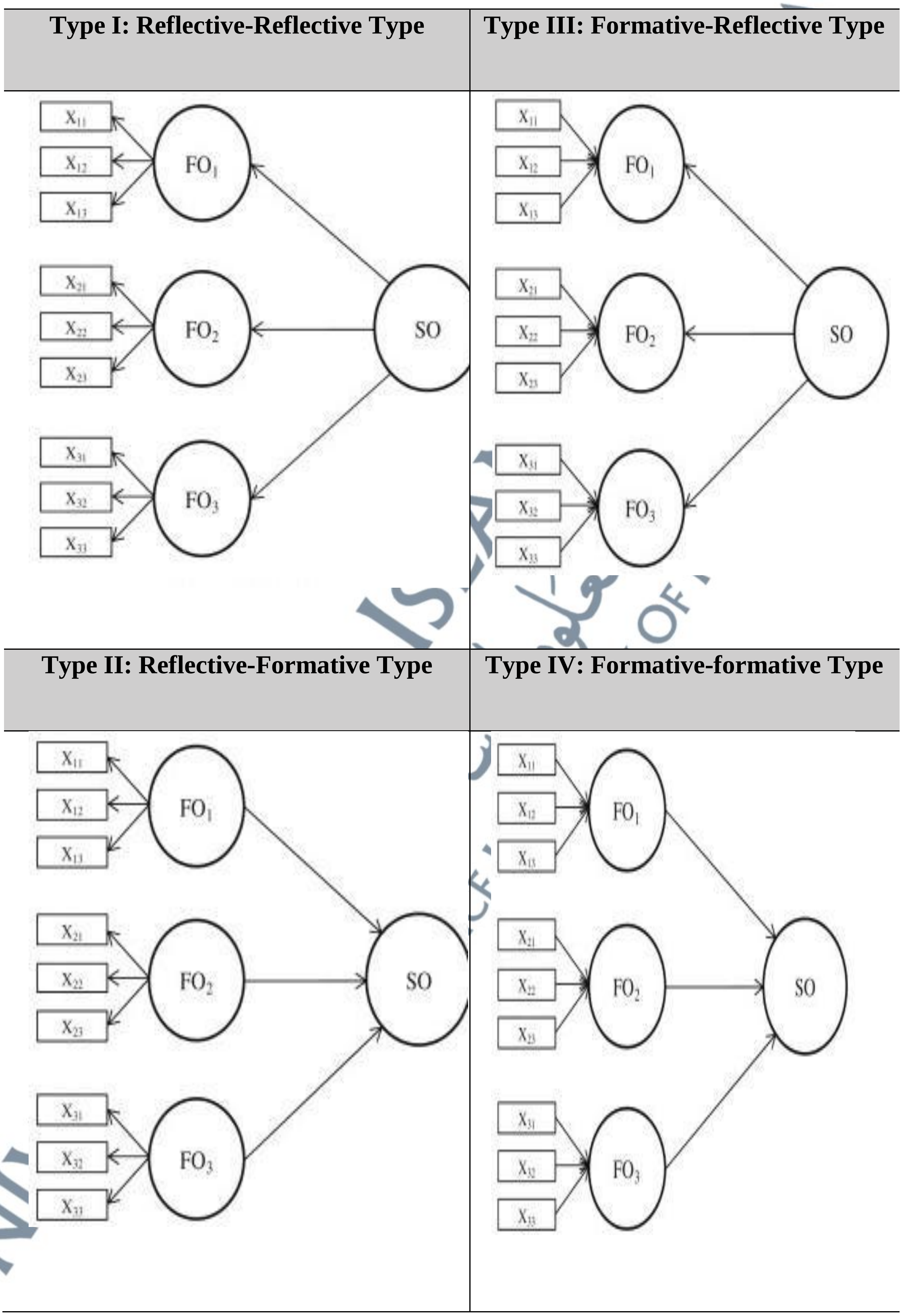
Criterion	Decision	Reference
Causal priority between the indicator and the construct?	• From the construct tothe indicators: reflective • From the indicators tothe construct: formative	Diamantopoulos & Winklhofer (2001)
Is the construct a trait explaining the indicators or rather a combination of the indicators?	• If trait: reflective • Ifcombination:formative	Fornell & Bookstein (1982)
Do the indicators represent consequences or causes of the construct?	• If consequences:reflective • If causes: formative	Rossiter (2002)
Is it necessarily true that if the assessment of the trait changes, all items will change in a similar manner (assuming they are equally coded)	• If yes: reflective • If no: formative	Chin (1998)
Are the items mutually interchangeable	• If yes: reflective • If no: formative	Jarvis, Mackenzie, & Podsakoff (2003)

3.7.2. Second-Order Factor

The previous discussion is only focused on the first-order latent construct. More often than not, the constructs are quite complex and can be operationalized at higher levels of abstraction. Figure 3.3 illustrates the four main types of the second-order factor model. These types of models have two elements: the first-order components, which capture the sub-dimensions of the abstract entity, and second- order components, which capture the more abstract entity.

In the first model, Type I, both first-order factor and second-order factor are conceptualized using reflective measures while in Type II, the first-order factor is constructed based on reflective measures and its second-order factor is developed using formative constructs. In Type III, the first-order factor model is described using formative measures and reflective constructs are used to explain the second-order factor. Finally, in Type IV, both the first-order and the second-order factor models are conceptualized using formative measures and constructs.

By referring to the literature review and conceptual framework of the study, the constructs and measures in this study applied Type III: Reflective first order, formative second order. The justification is, individual behaviour is formed by the constructs of motivation, ability, role perception, situational factors, religiosity, and ethical sensitivity. Therefore, these second-order constructs are formative, because each indicator form a formative construct captures a specific aspect of the construct's domain. Each item is an important construct, which implies that omitting an item potentially alters the nature of the construct. On the other hand, the first-order construct is reflective.



Source: Hair et. al., (2014)

Figure 3.3: Types of Hierarchical Components Models

3.7.3. Model Constructs and Measures

As explained in the Conceptual Framework, the proposed model of tax compliance behaviour has seven main constructs. These are behaviour to comply with tax law, motivation, abilities, role perception, situational factors, religiosity, and ethical sensitivity. These constructs are broad and complex, thus requiring a large number of items to adequately measure each of them.

Generally, the measurement items from past studies are used to measure the constructs except for some constructs where there are no available items from past research to be used. However, for religiosity constructs, there are available items from the past researcher, but the present study utilized different theoretical foundations hence different measurement items are developed. To guarantee the reliability and validity of the self-developed items, the researcher has acquired feedback from tax experts. The detail of reliability and validity discussion is presented in the 3.7 section. The following section discusses the model constructs and the indicators used to measure these constructs.

3.7.3.1. Behaviour to Comply with the Tax Law

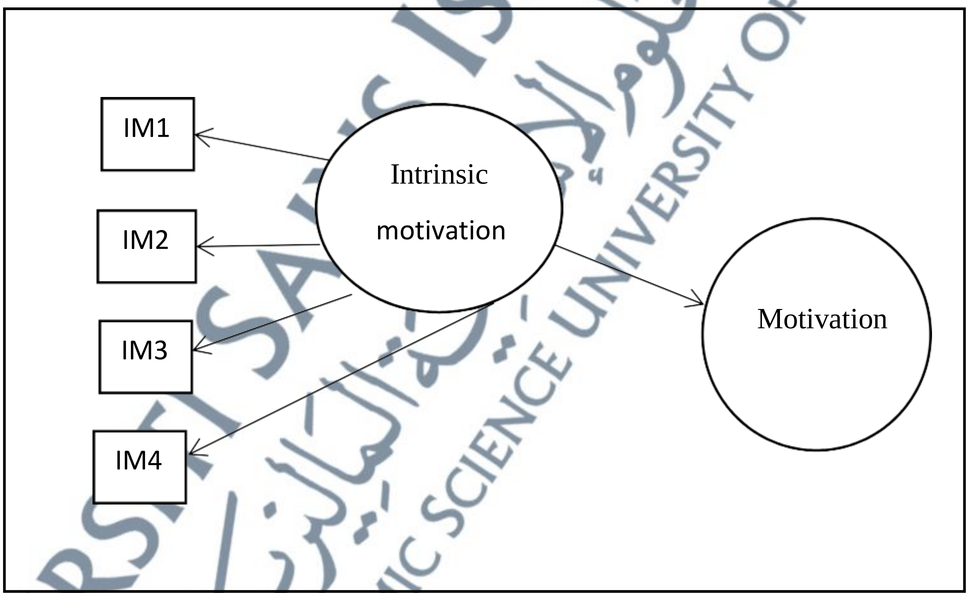
The behaviour to comply with tax law is operationalized as the dependent variable of the study. This study employed hypothetical scenarios to measure tax compliance behaviour. Scenarios are attempts to describe in some detail a hypothetical sequence of events that could lead plausibly to the situation envisaged (Kahn & Wiener, 1967). A hypothetical scenario has the ability to reduce social desirability bias (Ellfers et. Al., 2002; Kelly, 2015). For this study two hypothetical scenarios overstating income and understating income as well as fails to declare income were developed. This scenario is used to measure the MARS model items namely motivation, ability, role perception, and situational factors. As the natures of taxcompliance behaviour questions can be considered sensitive questions, the use of scenarios would be more likely to ease the respondents to provide more truthful responses. This scenario has been utilized by other studies like Noor Sharoja Sapiei et al (2014) and Alabede et al., (2011). Respondents are asked to read the scenarios and answer the questions provided. Notwithstanding the benefit that scenarios offer, it has also been criticized. One criticism is that the scenario may be divorced from reality as not only to be useless but also misleading (Kahn & Wiener, 1967). To increase the validity and reliability of the scenarios, the researcher had requested opinions from tax experts as well as tax practitioners. Following the scenario, a few statements tocapture the MARS model items were offered and respondents were required to state their opinions using a seven-point fully anchored scale.

3.7.3.2. Motivation

The motivation construct is operationalized as a second-order construct that is measured by two main components which are intrinsic motivation and extrinsic motivation. The following discussion discusses the measurement for intrinsic motivation and extrinsic motivation respectively.

3.7.3.2.1. Intrinsic Motivation

In the present study, the aspect of trust is used as a proxy to intrinsic motivation. Therefore, intrinsic motivation is operationalized as a first-order construct which is measured through a hypothetical scenario developed by the researcher. The respondents are requested to read the scenario and answer four sets of questions on seven-point scale to solicit the agreement from completely disagree and completely agree. The scenario is developed by the researcher. Two questions were adopted from Verboon & van Dijke, (2007), and another two questions are developed by the researcher.



*IM1, IM2, IM3 and IM4 represent intrinsic motivation items

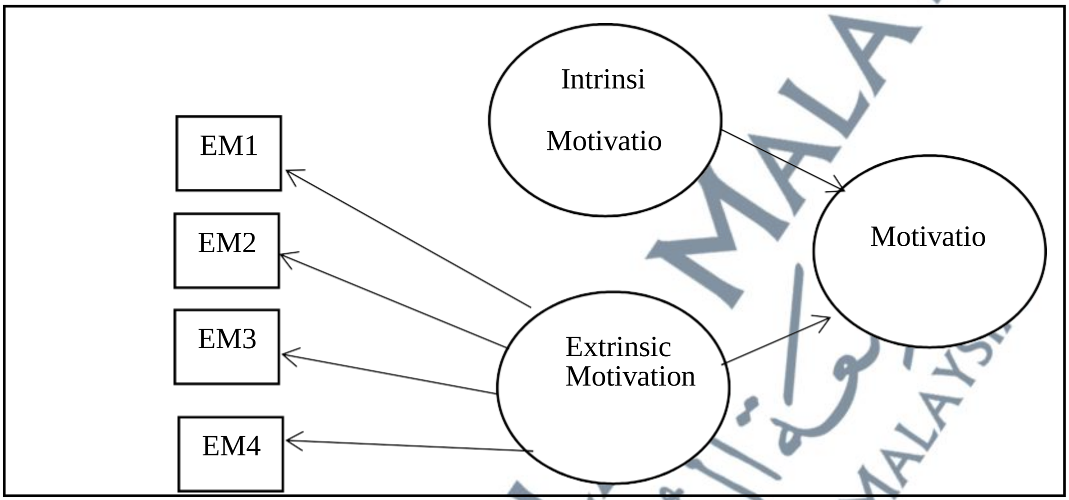
Figure 3.4: First and Second Order Measures of Intrinsic Motivation Construct

3.7.3.2.2. Extrinsic Motivation

In this study, the aspect of rewards is used to represent the construct of extrinsic motivation. Hence, extrinsic motivation is operationalized as the first-order construct which is measured by four items. Figure 3.5 provides an illustration for the second-order construct which is motivation and first-order construct is extrinsic motivation.

For the extrinsic motivation construct, to the best of the researcher's knowledge, there are no items developed by the previous researcher to measure extrinsic motivation in tax compliance behaviour. The past studies that analyze the role of extrinsic motivation examined it through an experimental method (See, Dwenger et. al., 2014; Feld, et.al., 2006; Farrar & Hausserman, 2016) and interview (See, for example, Rillstone, 2015). Therefore, self-developed items based on the past research that studies the role of rewards in other disciplines are offered for this study. The researcher modified the items from Bock & Kim (2002) and the semi-structured interview by Rillstone (2015). In Bock & Kim (2002), extrinsic reward is used to understand knowledge sharing behaviour while in Rillstone (2015), extrinsic reward is used to measure tax compliance behavior. Apart from the modifying questions from Bock & Kim (2001) and Rillstone (2015), the researcher also developed another one question.

Overall, the motivation construct is multi-dimensional and these dimensionsdo not tap into the same aspect, this construct is treated as a formative construct following the recommendation by Petter (2007). However, for the items in intrinsic and extrinsic motivation, the measures are likely to covary and thus measured as reflective. The following figure depicts examples of the conceptualization in this studyas first and second order levels.



*EM1, EM2, EM3, and EM4 represent extrinsic motivation items

Figure 3.5: First and Second-Order Measures of Extrinsic Motivation Construct

3.7.3.3. Ability

The ability factor is the second element mentioned in the MARS model. Ability is defined as both natural aptitudes and learned capabilities required to successfully complete a task. Aptitudes and learned capabilities are always be associated with competencies. As the primary objective of the current study is to examine the MARS model with tax compliance behaviour, abilities as one of the elements in the MARS model is represented by the tax knowledge variable as discussed in the literature review.

Here, the ability is operationalized as a first-order construct which is measured by seven items. All the items to measure tax knowledge are adopted from Natrah Saad (2011). The items are considered reflective indicators.

3.7.3.4. Role Perception

The third element of the MARS model is role perception. As discussed in the literature review, role perception refers to the extent to which people understand the job roles assigned to them or expected of them. MARS model is an organizational behaviour model used to understand the job performance behaviour of an employee. The present study, however, aims to test the model with tax compliance behaviour among the professional group. As in the organizational behaviour discipline, role perception is viewed as the extent to which the employees understands the job roles, hence in this tax study, role perception is viewed on the extent the taxpayer understand their job as a taxpayer, or the duty as a citizen or sometimes be referred to as a civic duty.

Here, role perception is operationalized as a first-order construct which is measured through a hypothetical scenario. There are five questions developed by the researcher to measure role perception towards tax compliance. The measures of civic duty are likely to covary and thus regarded as reflective.

3.7.3.5. Situational Factor

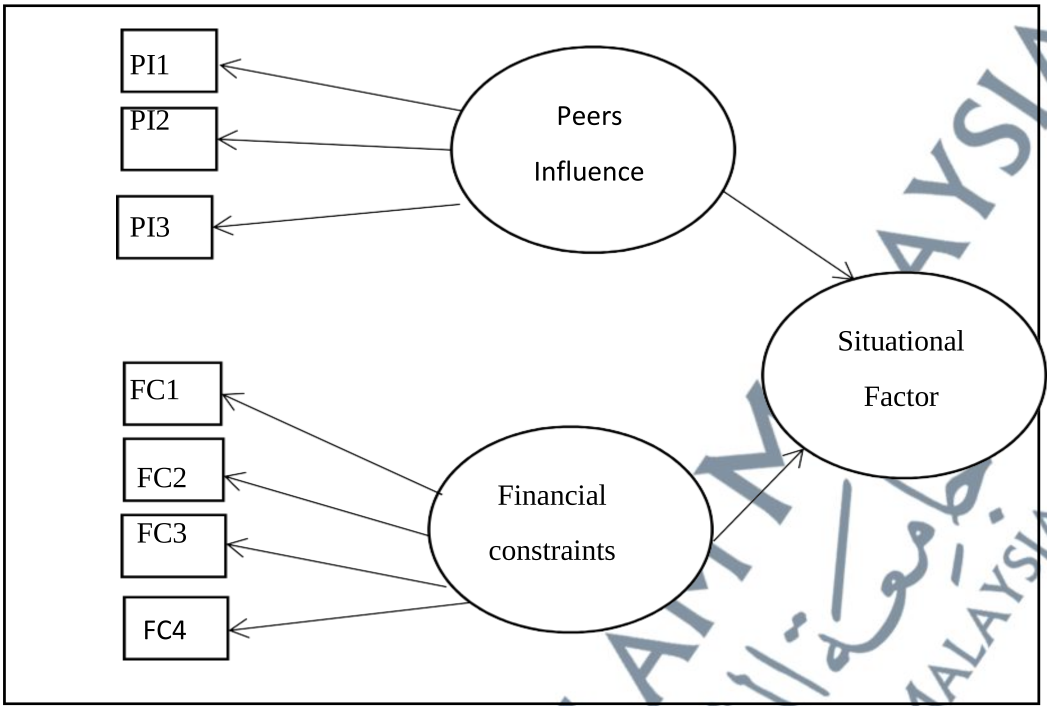
The final element of the MARS model is the situational factor. As discussed by McShane & Glinow (2005), employees' behaviour is also depends on the situation.Even an employee is a motivated employee, equipped with the right skills, and knows his or her role obligation but if some situational constraints hamper him or her, he or she will definitely perform poorly. A similar situation is expected for a taxpayer. Although a taxpayer has a strong motivation to pay tax, good knowledge on tax, and realize his civic duties towards the nation but if he or she faced financial constraints orbe influenced by the peers surrounding, he or she might try to evade to comply withthe tax law. This is the assumption that the present study attempt to examine.

Therefore, as mentioned above, the situational factor is viewed as a second-order construct while financial strain and peers influence is considered as a first-order construct. Financial strain and peers influence formed the situational factor construct which is the moderating variable of the framework.

The study developed a hypothetical scenario for both constructs. The scenario has the ability to provide a distance from obvious self-evaluation as depicted in the self-report scale (Prinstien et al, 2011). For financial strain, apart from adopting items from Mohd Rizal Palil, (2010), the researcher developed another two questions. On the other hand, items used to measure peers' influence construct were adopted from Mohd Rizal Palil (2010) as well as self-developed by the researcher.

The items to measure peers' influence are reflective as it is a trait used to explain the peers' influence construct. This is consistent with the suggestion of Hair et al., (2014) who said that, if the construct is a trait of the indicators then, the items are reflective. A similar measurement model was used to explain the other construct of situational factor which is financial constrain. All the items are reflective as well.

In this study, situational factor is considered a "moderating variable". Moderating variable is the variable that influences the strength and/or direction between two latent variables. In the context of the present study, the situational factor is assumed to influence between the exogenous variables (motivation, ability and role perception, religiosity, and ethical sensitivity) with the endogenous variable which is the behaviour to comply with the tax law. Based on Hair et al., (2014), a moderator effect requires the specification of the simple effect of the exogenous latent variable, the simple effect of the moderator variable, and the product term, which also be referred to as interaction term. In PLS-SEM, there are two approaches that are usually be employed to create the interaction term, which are the product indicator approach and the two-stage approach. The product-indicator approach involves multiplyingeach indicator of the exogenous latent variable with each indicator of the moderator variable. In the case of the exogenous latent variable or the moderator variable, having a formative measurement model, the product indicator approach cannot be applied (Hair et al., 2013). As there are formative constructs that existed in the present study, therefore two-stage approach will be used. The operationalization of the first and second-order levels for the situational constructs can be viewed in Figure 3.6.



* PI1, PI2, PI3 represent items for peer's influence, FC1, FC2, FC3, and FC4 represent items for financial constrain

Figure 3.6: First-Order Construct and Second-Order Construct for Situational Factor

3.7.3.6. Religiosity

The previous discussion on the constructs involves all variable mentioned in the MARS model. This study, however, aims to extend the MARS model to include another significant variable, which is religiosity. As discussed in the literature, religion is found to be a significant predictor of individual behaviour. Due to that, the present research aims to examine the role of religiosity towards the behaviour to comply with the tax law.

The past studies that examines the role of religiosity in understanding tax compliance behaviour measure religion in terms of its religious affiliation and the degree of religiosity and regressed it with the compliance behaviour (See, for example, Mohd Rizal Palil et al., 2013; Nor Raihana Mohd Ali & Pope, 2014; Torgler, 2005). The present study, however, is interested to develop items that directly measure religion towards tax compliance behaviour. To better comprehend religiosity as a construct towards understanding tax compliance behaviour, Hirschi's Social Control Theory is used as an underlying theory. Therefore, this religion variable mentioned in the theory is used as a reference to develop items for the religiosity variable. As the items are self-developed, the researcher has sought opinion and suggestions from the tax experts,to ensure the validity and reliability of the items. All the items to measure religiosity are reflective as the items are likely to covary with each other.

3.7.3.7. Ethical Sensitivity

Apart from extending the MARS model by including the religiosity variable, this study also aims to add another important variable which is ethical sensitivity. As mentioned in the literature, ethical sensitivity has been proven by the past researchers as a significant predictor in tax compliance behaviour (See, for example, Suhaila Abdul Hamid, 2014; Henderson & Kaplan, 2005; Ho & Wong, 2008; Reckers et al., 1994). Therefore, the present study is interested to analyze the role of ethical sensitivity in understanding compliance behaviour specifically within the framework of the MARS model.

As discussed in the literature that ethical sensitivity is viewed as a multi-dimension, therefore the construct for ethical sensitivity in this study is operationalized as second-order factor measured with twelve first-order items. The items were categorized into four groups which are moral equity, relativism, egoism, and utilitarianism. For each group, two items were developed to measure ethical sensitivity except for four items for the moral equity dimension. The dimension of contractualism was dismissed from the measurement model following the critiques from Skipper & Hyman (1993) that the contractualism items of MES are double- barrelled. For example, to answer violates/does not violate an unwritten contract assumes a Yes answer to Does a contract exist?

All of these items were adopted from Cruz et al., (2000). Respondents are requested to provide their responses through seven points Likert scale. These items were developed to answer a self-developed hypothetical scenario.

All the measures in the ethical sensitivity are reflective. At the second-order level, the multidimensional ethical sensitivity is measured using formative constructs since these constructs represent different underlying concepts.

3.7.3.8. Demographic Background Variables

The demographic details captured in this study include age; gender; ethnicity; education level; occupation; income level; profession and e-filing experience. The age of the respondents was measured in years. Respondents need to indicate their age by choosing the age range that is applicable to them, namely: (a) 25-29 (b) 30-34 (c) 35- 39 (d) 40-44 (e) 45-49 (f) 50-54 (g) 55-59 and (h) 60 and above.

For gender, the respondent needs to choose either (a) male or (b) female. To capture the ethnic group, the respondent needs to choose the appropriate group they belong to. The ethnic groupis either (a) Malay (b) Chinese (c) Indian (d) Other. In terms of educational level, respondents need to choose either (a) Secondary level and below (b) Certificate (c) Diploma (d) Degree (e) Masters (f) Ph.D. Another demographic background that the respondents are requested to answer is the respondent's profession. The respondents need to choose either (a) Lawyer (b) Doctor (c) Architect (d) Engineer (e) Accountant or (f) Surveyor. In terms of income level, it refers to the annual income earned by the respondents. The income range is as follows (a) RM40,001- RM 50,000 (b) RM 50,001-RM60,000 (c) RM60,001-RM 70,000 (d) RM 70,001-RM80,000 (e) RM80,000 and more. Finally, the respondents need to indicate their experience with e-filing. It is either (a) Never (b) One time (c) 2-5 times (d) More than 5 times.

The summary of the reference used to develop the survey questionnaire is presented in the following Table 3.5.

Table 3.5: Model Constructs and Measures

Constructs	Component measures	Represented by	Item measures	Source	Item codes
Tax compliance behaviour	Understating income		If faced with the same situation, I will not declare my sideincome.	Self-developed	UI1R
			The probability that I will not declare the side income I earnis high		UI2R
			If my friend is dealing with the same situation, my friendwill not even declare their side income.		UI3R
	Overstating expenses		If faced with a similar situation, I will claim my travelingexpenses more than 10% of the actual expense		OE1R
			The probability that I will make travel claims more than theactual traveling expenses is high		OE2R
			If faced with the same situation my friends will claim theirtraveling expenses more than the actual traveling expenses		OE3R
Motivation	Intrinsic motivation	Trust	I feel that the authority in the described scenario istrustworthy.	Adopted from Van Dijke & Verboon (2010)	IM1
			I would certainly declare tax money to the tax office		IM2
			Trust in tax authorities influences me to comply with the taxlaw		IM3
			Trust in tax authorities intrinsically motivates me to complywith the tax law		IM4
	Extrinsic motivation	Rewards	I expect to receive monetary rewards as an incentive forcomplying with tax law	Adapted from Bock & Kim (2001)	EM1

			I expect to receive an honour such as certificates that can beused as a credit for job promotion each time I successfully file a tax return	Adopted from Rillstone (2015**)**	EM2
			I believe that more focus on positive approaches encourageme to comply with tax law		EM3
			I believe a reward system is effective in increasing tax compliance (e.g monetary rewards, holiday package,acknowledgment in the workplace)		EM4
Ability	Tax knowledge	General knowledge	The income tax system is a way for the government tocollect revenue to manage an economy	Adopted from Natrah Saad, (2011)	GK1
			Individuals are subject to a single flat rate of income taxunder the current tax system		GK2R
		Legal knowledge	As far as I am aware, non-compliant taxpayers can beimprisoned, if found guilty of evading tax.		LK1
			Similar to other criminal offenses, individuals can also be prosecuted for not complying with the Malaysian income tax law		LK2
			I do not have to abide by the deadline for the submission of tax return form (s) (in case of having other income, such asrental and business income), as the deadline is only a guideline and does not result in penalties.		LK3R
		Technical knowledge	Everyone who earns income in this country is taxable,regardless of whether the person is a resident or not.		TK1
			I can deduct all personal expenses in calculating my personalliability.		TK2R

			I have little idea about the deductions that I can claim as ataxpayer in the computation of my tax liability.		TK3R
Role perception		Civic duties	If I face the same situation, I would not feel guilty for notdeclaring the side income.	Self-developed	RP1R
			If I face the same situation, I will not feel guilty of adding10% of the traveling expenses.		RP2R
			If I face the same situation, I will not feel guilty forunintentionally fail to declare the side income.		RP3R
			If I face the same situation, I would not feel guilty forunknowingly over claiming the traveling expenses		RP4R
			Based on the scenario above, I have a great moral responsibility for the country to file taxes honestly		RP5
Situational	Financial constraints		Based on the scenario above, financial position affects mydecision to complete the tax return	Self-developed	FC1R
			The probability that I would file taxes accurately is high		FC2
			I am satisfied with the financial situation of my household	Torgler (2003)	FC3
	Peers influence		Based on the scenario above, I think friends and relativesinfluence my decision not to declare my side income.	Self-developed	PI1R
			My friends and relatives strongly influence my life decision		PI2R
			The probability that I would reduce my chargeable incomebecause of peers and relative influence is high		PI3R
Religiosity			Religiosity appears to prevent deviant behaviour, includingtax evasion	Self-developed	R1

			I believe that, if an individual possesses a strong emotional connection with a religious institution (e.g mosque, church),the likelihood to involve in delinquent behaviour such as taxevasion is low		R2
			I believe that, if a person has a clear motivation and life goals, the likelihood to engage in delinquent behaviour like tax evasion is low		R3
			Individuals who are involved in religious activities are expected to protect individuals from delinquent behaviourlike tax evasion		R4
			Most religious belief encourages people to embrace the tenets of the religion with proscriptions against delinquent behaviour like tax evasion		R5
Ethical sensitivity			Based on the above scenario, the action of not declaring theside income and over claiming expenses are:	Adopted from Cruz et. al, 2000	
			Justice		ES1R
			Fair		ES2R
			Morally right		ES3R
			Acceptable to my family		ES4R
			Traditionally acceptable		ES5R
			Culturally acceptable		ES6R
			Not self-promoting for me		ES7
			Personally, satisfies me		ES8R
			Produce the greatest utility		ES9R
			Minimizing benefits and maximizing harms		ES10

3.8. Data Preparation and Pre-Analysis

This subsection discusses the pre-analysis performed for the quantitative part of the study. The discussions focused on data screening, response representativeness, nonresponse bias, common method bias, social desirability bias, descriptive analysis, and *t*-test analysis.

3.8.1. Data Screening

The data screening process constituted a careful stage for consideration in the social and management science research studies (Hair et al., 2013). A number of studies have highlighted the importance of the data screening process (Badara, Mu'azu Saidu & Saidin, Siti Zabedah (2014), Maiyaki & Mouktar, 2011 and Gorondutse & Hilman, 2014). Therefore, it is important to assess the possible errors in data entry andother possible mistakes to rectify and correct the identified error.

3.8.2. Blank Response

The first step in data cleaning is checking for blank responses among the collected questionnaires, and Microsoft Excel was used to conduct the procedure cases with more than 25% blank responses on the independent variables or any blank responses on the dependent variables are deleted from the sample. The study uses the formula of 155 COUNTBLANK to check on blank responses. The results are presented in Chapter 4.

3.8.3. Data Entry Error

In verifying the data error sampling checking method was used. For sampling checking, using SPSS 23, all fields were sorted in ascending or descending to see if there was an entry error or any entry that was outside of the range in the incorrect responses. In most cases, the incorrect value can be replaced with the correct value by checking back the original survey responses.

3.8.4. Missing Values

The next stage in data cleaning is to identify any missing part of the data set and missing data. According to Hair et al. (2017), missing data occurs when a respondent does not answer one or more questions, either intentionally or unintentionally (Hair et al., 2017). The effect of missing data is that it can diminish the statistical power of a study and may generate biased estimates that result in worthless conclusions (Kang, 2013). The gravity of missing data hinges on the configuration of missing data, the quantity of missing data, and the reason why it is missing. Missing data is considered a common problem in research. A check on the datarevealed that all 126 responses received have less than 10 percent missing data. Hair et al., (2010) suggest that any remedies of missing data could be applied if the missingdata is less than 10 percent. Following the recommendation of Hair et al (2010), the Expectation

Maximisation (EM) method was used to treat the missing data problem.

3.8.5. Outliers

Outliers are extreme responses to particular questions or extreme responses to all questions (Hair et al., 2017). An outlier may occur because of errors in participants'responses, errors in data entry, poorly worded survey items, or incorrect specification of the population or sample (West, Finch & Curran, 1995). The assessment of outliers is crucial in a PLS-SEM analysis because it can have a significant effect on data analysis (Hair et al., 2014). In the SEM technique, extreme data points will hurt the estimate of the parameters. The outliers were diagnosed based on the univariate and multivariate analysis.

Univariate outliers can be identified by an inspection of the frequency of the distribution or box plot for each variable. For continuous variables, several options exist for determining the threshold for outlier designation. For univariate outliers, the critical z-score value of 3.29 was used to identify potential extreme outliers.

For multivariate outliers, the Mahalanobis Distance (D2) was employed to analyzed outliers. The Mahalanobis measures allow for significant testing at 0.05 and 0.01 levels (Hair et., 2010). Both the univariate's Z-value (3.29) outliers and the multivariate Mahalanobis distance outliers detect only 3 cases and therefore the 3 cases were removed.

3.8.6. Normality

The first assumption in multivariate analysis is normality, which is embodied by the normal distribution assumption in each item and all linear combinations of items (Tabachnick & Fidell, 2012). The normality of the data was checked by examining the skewness and kurtosis of the data distribution. Hair et al. (2017) and Cain et al. (2016) suggested that the multivariate skewness and kurtosis are measured using the software available in statistical analysis websites available online. The results can be viewed in the following URL address: https://webpower.psychstat.org/models/kurtosis/results.php?url=ed734a6632b0b4b26a13f5881c2ce6a6.

The result shows that the majority of the data fell within the acceptable rangeof +/- 2 for skewness and +/- 3 for kurtosis. This is consistent with the suggestion of Hair et al., (1998) where values of skewness range between -2 to 2 and for kurtosis ranges between -3 to 3 is normal.

3.8.7. Response Rate

In any survey study, the survey response rate is important for the generalization of the findings to the whole population (Sekaran & Bougie, 2010). The survey was conducted between Jun 2018 to September 2018. A total of 1000 survey questionnaires were distributed to the respondents via online or self-administered. For the online survey, the researcher will email the respective respondents the link to the online survey. The purpose of the research and the anonymity of the respondent will also be stated. For self-administered survey, the researcher will send the questionnaire together with a cover letter that emphasized the research purpose and the anonymity of the respondent. A four-week return date will be requested. To increase the response rate, some measures are undertaken such as follow-up reminders, and a small gift to the respondent.

3.8.8. Non-Response Bias

Non-response is a challenging issue in using surveys as a data collection instrument since a high non-response rate could possibly result in bias in responses. According to Rogelberg & Luong (1998), non-response bias occurs when the individuals responding to a survey differ from non-respondents on variables relevant to the survey topic. Dillman (2000) on a similar note said that non-response bias occurs when a significant number of people in the survey sample fail to respond and have relevant characteristics.

Survey non-response is normally divided into four different categories which are inaccessibility, inability, carelessness, and noncompliance (Rogelberg et al., 2003). Inaccessibility means that the respondent never received the survey, while inability means that the respondents are not able to respond to the survey due to illness or the like. Carelessness on the other hand, respondents misplaced the survey, and active non-compliance means the respondent consciously decided not to participate in the study.

Non-response may reduce the sample size, and due to smaller sample size, the precision of estimators will be smaller, and the margins of error will be larger. A moreserious effect is that it can be selective. This occurs when specific groups are under or over-represented in the survey. As the present study, employed a self-administered method, the bias is expected to be lower.

To examine the existence of nonresponse bias in this study, the researcher followed the approach by Armstrong & Overton (1977), using late respondents, those who participated after the reminders being sent, as substitutes for nonrespondents. The responses were later divided into "early respondents" and "late respondents", and the means of the responses were compared to check whether the means between the two is significant at $p \leq 0.05$ using an independent *t-test* in SPSS. The independent *t*-test in SPSS provides two results, the Levene's Test for equality of variance and the *t*-test for equality of means.

Levene's test is used to determine whether or not there is equal variance between the two groups before deciding the appropriate t values in an independent *t*-test. This can be done by checking the significance of the F value at $p \leq 0.05$ in Levene's test. The t value is then determined depending on whether there is equal variance between the two groups or when the variance is not equal. Based on the t value, the significance of the means is determined at $p \leq 0.05$ (Gaur & Gaur, 2006; Field, 2009).

3.8.9. Common Method Variance

Researchers agree that common method bias is a potential problem in behavioural research (Podsakoff, et.al., 2003). A common method bias exists if the variation in responses is caused by the instrument rather than the actual predisposition of the respondents that the instrument attempts to uncover. In other words, a common method bias is a measurement error that is compounded by the tendency of the respondents to only provide an affirmative answer (Chang, van Witteloostuijn & Eden, 2010). Podsakoff et al., (2003) have discussed in their writing the source of common method bias. Generally, common method bias is caused as a result of using the same source, item characteristics, and item context. The effect caused by the item characteristics itself include complexity or ambiguity, scale format, and negatively worded items while effects that caused by item context includes priming effects, embeddedness, context-induced, and intermixing items of a different construct. Furthermore, effects caused by measurement constructs includes time and location, and common medium. Since the data of the present research is obtained from a survey, there is a possibility that common method bias may exist in the responses.

The procedural remedy to control common method bias has been taken into consideration when designing the questionnaire. Podsakoff et al., (2003) has proposed some ways which are obtaining measures from a different source, temporal, proximal, psychological, or methodological separation of measurement, protecting respondent anonymity, counterbalancing question order, and improving scale items. To ensure the robustness of the findings, common method bias is addressed in pre-distribution and post-distribution of the questionnaire. In the pre-distribution stage, these few methods are used, for example, obtaining measures from a different source, protecting respondent anonymity, and counterbalancing question order.

Apart from the previously mentioned remedies that can be utilized to minimize method biases, statistical remedies also can be employed. For statistical remedies, Harman's single factor test and full collinearity test were used to test the CMV. Harman's single-factor test is conducted to determine whether a single factor surfaces from the principal component analysis, or if a distinct factor explains the majority of covariance among the variables in the un-rotated factor analysis (Podsakoff et al., 2003). The result is presented in Chapter 4. Despite the acknowledgment of the Harman's test in most social science studies, few drawbacks have been discussed by some researchers. The test is said to be incomplete and insensitive (Podsakoff et al., 2003) and only provides information on the absence or presence of CMV (Tehseen et al., 2017).

Therefore, a full collinearity test was performed to determine whether any constructs reflect the Variance Inflation Factor (VIF) values of equal to or greater than 3.3 (Kock & Lynn, 2012). The result is presented in Chapter 4.

3.8.10. Social Desirability Bias

Social desirability refers to the need for social approval and acceptance and the belief that it can be attained by means of culturally acceptable and appropriate behaviours (Crowne and Marlowe, 1960). It is viewed on the part of the individuals as a tendency to portray themselves in a favorable light. This kind of bias affects the validity of the questionnaire (Huang, et.al., 1998) This is because the individual may respond in light of socially acceptable and not truthful responses. Normally, social desirability bias occurs in socially sensitive questions. According Tourangeau & Yan, (2007) income is also one of the sensitive questions. As one of the constructs in the survey involves financial constraints, social desirability bias may likely occur.

A common method used to detect Social Desirability Bias is Marlowe-Crowne Social Desirability Scale (MCSDS). It is a set of true and false questions. As the respondents of the study are the professional group, who are normally busy with their routine tasks, therefore the researcher adopted two-set questions by Cohen et al., (1993). For the items that involve sensitivity that is the question on the financial constraint, there will be two sets of questions on the same item. The first question asked the first person (Would you do it?) and the second question control for social desirability for the second person (Would your peers do it?). The responses for both questions were recorded using seven points scales ranging from High to Low, with the higher score indicating a low probability to perform described behaviour. The mean responses were then compared using Paired T-test in SPSS.

3.8.11. Descriptive Analysis

Descriptive analysis is important to present quantitative descriptions in a more manageable form. The present study will employ descriptive statistics to provide preliminary ideas on how the taxpayers perceive the income tax system. In this study, basic features of the survey data will be presented where frequency, mean, standard deviation, variance, minimum, and maximum value for each item and construct are calculated. All these measures are obtained from the descriptive statistics available from the SPSS software.

3.9 Introduction to Structural Equation Modelling and Partial Least SquaresMethods

3.9.1 Structural Equation Modelling (SEM)

For the past 20 years, researchers have shifted the technique of data analysis from the first-generation technique (multiple regression, logistic regression, and analysis of variance) to second-generation technique namely Structural equation Modelling (SEM). SEM is developed to overcome the first-generation technique weaknesses. SEM enables the researchers to incorporate unobservable variables measured indirectly by indicator variables (Hair et al., 2014). This powerful multivariate analysis technique includes applications like causal modeling or path analysis, confirmatory factor analysis, second-order factor analysis, regression models, covariance structure models, and correlation structure models. Apparently, SEM is preferred among researchers due to its flexibility to interact between the theory and data (Chin, et al., 2003).

SEM technique is a useful technique in developing and expanding theory, especially when second and even third-order factors which provide a better understanding of relationships that may not be apparent initially (Astrachan, et al., 2014). SEM consists of two components which are a measurement model linking a set of observed indicators to a smaller set of latent constructs and a structural model linking the hypothesized model of latent constructs (Hair et al., 2016). Latent construct is an unobserved construct and cannot be measured directly. Therefore, to measure the latent construct, the researchers develop indicators that capture the construct. The indicators are directly observable and the researcher believes that the indicators accurately measured the construct. Before the relationship between the latent construct is assessed through the path analysis, the validity and the reliability of the observed indicators will be addressed. Only after the validity and reliability of the measurement model are established, the relationship between the latent construct can be examined to test the hypothesis.

There are two complementary schools in the field of SEM which are co-variance based SEM and component-based SEM. Co-variance based SEM (demonstrated by its software such as LISREL, EQS, AMOS, SEPATH, and RAMONA) is considered as a generalization of path models, principal component analysis, and factor analysis to the case of several data tables connected by causal link (Tenenhaus, et al., 2005). Another school under the umbrella of SEM is PLS (Partial

Least Squares). It is a two-step method. First, latent variables scores are computed using the PLS algorithm, then OLS regressions are carried out on the latent variables scores for estimating the structural equations (Tenenhaus et al., 2005).

Despite the numerous benefits of CB-SEM, it is not without limitations. CB- SEM requires a larger sample size, preferably 200 and above. Inadequate sample size may result in poor parameter estimates. Secondly, CB-SEM also requires normal distribution of data. Thirdly, to obtain an adequate model for the CFA can result in the elimination of meaningful content for measuring the constructs. The recommended items per construct are only 3 items (Hair et al, 2011). The fourth limitation is the measured items of the latent variable should always be reflective. If there is any formative measure in any of the items, it would be problematic for the CB-SEM to account for all co-variances among its measures. A final significant issue with CB-SEM is, the analysis is expected to be performed using a strong theory with well-developed measures which have undergone a series of exploratory analyses. It is suggested that it is not appropriate when the theory is relatively tentative or when the measures are newly developed. The rules of thumb for selecting each method are listed in Table 3.6.

Considering the limitations of co-variance-based SEM, component-based SEMwas developed by Herman Wold in 1974. PLS-SEM is gaining its place due to its ability to handle problematic modeling issues that normally occur in research such as non-normal data and model that is highly complex. The following discussion is, therefore, will revolve around the background and overview of PLS.

Table 3.6: Rules of thumb for choosing SEM method

PLS-SEM	CB-SEM
The research objective is exploratory or confirmation of theory based on total variance	The research objective is confirmation of we developed structural and measurementtheory bas on common variance
The objective of the analysis is prediction	The measurement philosophy is estimation with the common factor model using only common variance (covariances)
The measurement philosophy is estimation with the composite factor model using total variance	The research requires a global goodness-of- fit criterion
The research objective is to explain the relationships between exogenous and endogenous constructs	The error terms require additional specification, such as covariation
The structural and/or measurement models are complex (many constructs ¼ 6+ and manyindicators¼ 50+)	The structural model specifies non-recursive relationships
Formatively measured constructs are specified in the research	The structural and/or measurement models are simple (5 or fewer constructs and 50 or fewer indicators)
Preferred method when the sample size is small (no 100). But PLS is also an excellent method for larger samples	
The error terms require additional specification, such as covariation	
The research objective is to use latent variablescores in subsequent analyses	
The data are not normally distributed	
The scaling of responses is ordinal or nominal	
The data is secondary/archival, particularly single-item measures	
The structural model will be estimated with a higher-order construct that has only two first- order constructs	
The analysis involves a continuous moderator. The investigation will examine the model for unobserved heterogeneity	

Adopted from Hair et al., (2017)

3.9.2 PLS-SEM

PLS-SEM is accepted by researchers as a viable methodology. It has gained itsmomentum through many business disciplines. From 1980 to 2011, there are morethan 300 studies that employed PLS-SEM as their methodology. Through the meta- analysis of Hair et al., (2014), reveals that the justification to use PLS is because of non-normal data, small sample size and formatively measured constructs.

PLS path model has three components, which are the structural model, or referred to as the inner model and the measurement models also referred to as outer models, and the weighting scheme. The inner models show the path between the latent constructs. PLS-SEM only permits recursive relationship which means the structural path between the latent constructs can only head in a single direction. In the structural model, there are two types of constructs which are exogenous and endogenous. Exogenous construct is the latent constructs that do not have any structural path relationship pointing to them. Endogenous construct, on the other hand, is the latent constructs in the structural model that are explained by other constructs.

The next component of the structural equation model is the measurement models or the outer models. This outer model is used to evaluate the relationships between the indicator variables and their corresponding construct. An extract of the path models of the study is presented in Figure 3.7. There are two types of measurement which are reflective construct and formative construct. The extensive discussion of reflective and formative measurement has been mentioned in sub-section 3.14.

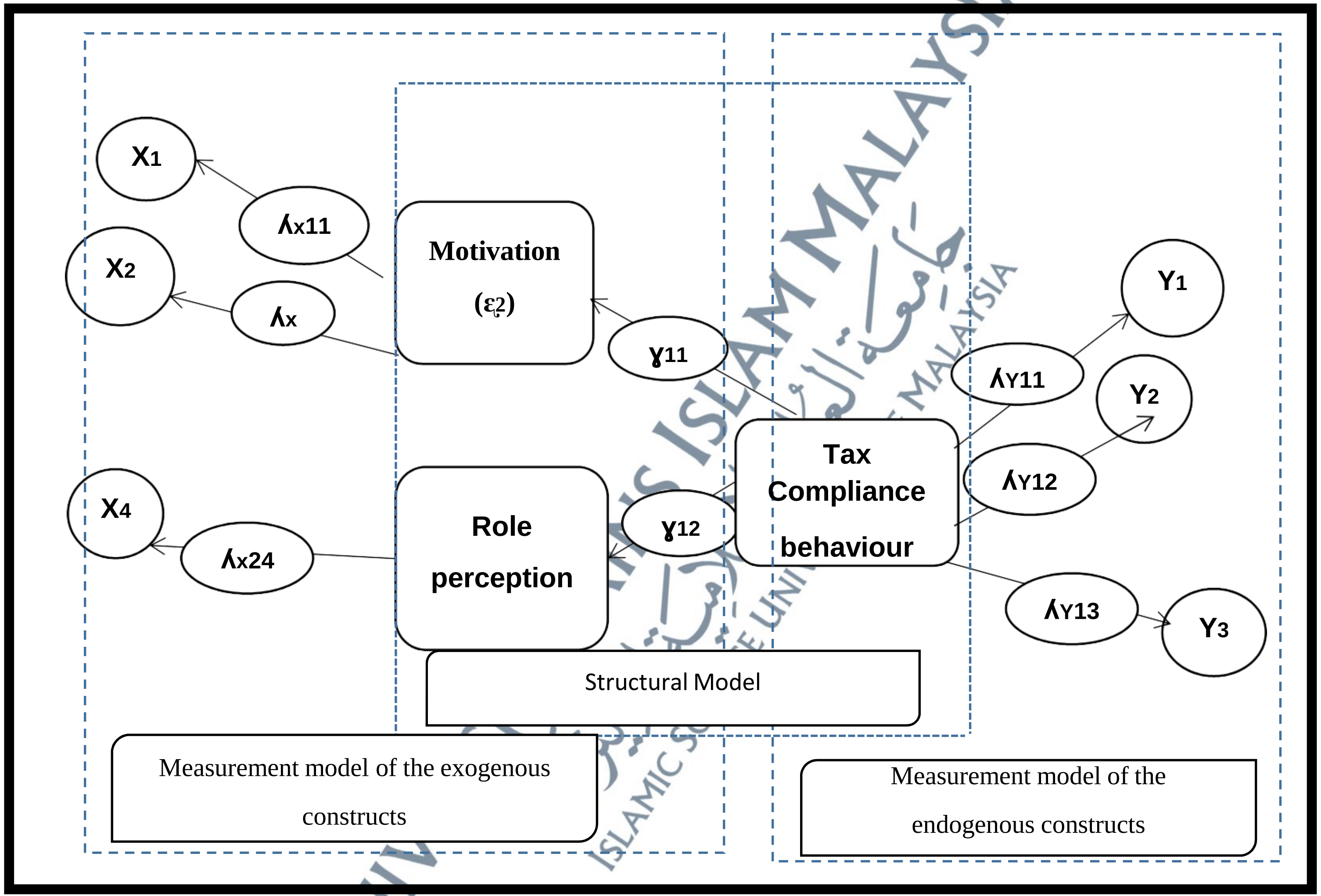


Figure 3.7: Path Models

3.9.3 Justification for Using PLS

The previous discussion explained on the PLS as a concept. Overall, PLS possesses many benefits compared to CB-SEM. Therefore, PLS is chosen as the data analysis technique due to the following reasons.

Firstly, the research objectives. As suggested by Hair et al (2017), the choiceof either to use PLS-SEM or CB-SEM is determined by the relevant researcher's research objective. In the case where the research goal is theory-testing or theory confirmation and the prior theory is strong and well developed, CB-SEM is appropriate, otherwise, PLS-SEM is the best alternative. With regards to the objective of the current research, the MARS model is a newly developed model with less than 10 empirical studies that have tested the model, therefore PLS-SEM is the best statistical technique to be applied.

Secondly, PLS-SEM is a less rigorous technique compared to PLS-SEM. In CB-SEM, a set of assumptions must be fulfilled, which are normality of distributions, minimum sample size, and maximum model complexity; otherwise, the results can be imprecise. If all these assumptions are violated, PLS-SEM is the best alternative technique. With all the violated assumptions, PLS is able to provide more robust estimations. As the present study have a small sample size, thus PLS-SEM suits this limitation. This has been mentioned in Hair et al (2017) that, where the most common reason for using PLS is due to small sample size.

Third, PLS provides a better prediction capability that suits the objective of thestudy which is to predict tax compliance behaviour.

Fourthly, PLS is suited for both reflective and formative constructs which is appropriate with the nature of the current study, which is a mix between formative and reflective constructs.

Fifthly, PLS is suitable to be used when a complex model is tested such as in this study. As the modelinvolves moderator and second-order factor variable, therefore PLS is well suited.This was mentioned in Hair et al (2017) where the two most common reasons for choosing PLS-SEM are model complexity and formative constructs. Finally, some of the items are newly developed items, so, therefore, PLS is the best-suited technique for this purpose.

3.9.4 Model Evaluation

A PLS path model consists of two elements which are the measurement model (also referred to as the outer models in PLS) and the structural model (also referred to as the inner model in the context of PLS-SEM). It is significant to perform a model evaluation in order to analyse the constructs. As the current study is utilizing PLS as its data analysis technique, model evaluation is one important step before further analysis is undertaken. There are 2 important processes involved which are (a) evaluation of the measurement model (b) evaluation of the structural model.

3.9.4.1 Evaluation of Measurement Model

The measurement model evaluates the association between measures and the latent construct. Analysis of PLS-SEM enables the researcher to evaluate the reliability and validity of the construct measures. When evaluating the measurement model, it is important to differentiate between reflective and formative constructs. These 2 constructs are based on different concepts and therefore require different consideration for each. Reflective measurement models should be assessed with regard to their reliability and validity. For the reflective measurement model, the composite reliability, convergent validity, and discriminant validity are assessed. While for formative measurement, content validity, convergent validity, andcollinearity are those assessments that must be taken into account.

3.9.4.1.1 Assessing the Result of Reflective Measurement Models

In the process of assessing the result of the reflective measurement models, some evaluations involve such as the reliability, convergent validity, and discriminant validity. The following discussion entails each criterion for the assessment of the result of the reflective models.

3.9.4.1.1.1 Reliability

The first criterion to be evaluated in assessing the result of reflective measurement models is reliability. Construct reliability for reflective measurement models normally focuses on composite reliability to estimate the internal consistency of a construct. It takes into account the different outer loading of the indicator variable. This type of reliability takes into account the different outer loadings of the indicator variables.

Composite reliability values of 0.60 to 0.70 in exploratory research and values from 0.70 to 0.90 in more advanced stages of research are regarded as satisfactory whereas a value below 0.60 indicates a lack of reliability. Likewise, the indicator’s loadings also need to be taken into account. Each indicator’s loadings should behigher than 0.70. Indicators with loadings between 0.40 and 0.70 should only be considered for removal from the scale if deleting this indicator leads to an increase in composite reliability (Hair et al., 2011).

3.9.4.1.1.2 Construct Validity

Construct Validity refers to the degree to which a measure relates to other variables as expected within a system of theoretical relationships (Babbie, 2010). According to Sarstedt & Mooi (2014), to evaluate construct validity, analysis of convergent validity discriminant validity, and criterion validity must be carried out.

3.9.4.1.1.2.1 Convergent Validity

Convergent validity is the extent to which a measure correlates positively to the alternative measure of the same construct. To establish convergent validity, researchers need to demonstrate that the constructs should be related, are related. Higher outer loadings indicate that the indicators have high similarities. A common rule of thumb for outer loading should be 0.708 or higher (Hair et al., 2014)

However, for social science research, an indicator with outer loadings of 0.70 is frequently observed especially when newly developed scales are used. Otherwise, outer loadings values ranging from 0.4 to 0.7 should be considered to be removed. A common technique to establish convergent validity is Average Variance Extracted (AVE). AVE refers to the grand mean value of the squared loadings of the indicator associated with the construct. AVE value of 0.5 or more indicates that on average, the construct explains more than half of the variance of its indicators.

3.9.4.1.1.2.2 Discriminant Validity

Discriminant Validity refers to the extent to which a construct is truly distinct from another construct by empirical standards (Hair et al., 2014). This is crucial to determine the uniqueness of each particular construct. There are 2 ways to analyze the discriminant validity which are cross-loading or the Fornell-Larcker criterion.

By examining the cross-loading of the indicators, the indicators of the outer loadings on the associated construct should be more than any other loadings for other constructs. If the cross-loading exceeds the indicator's outer loading it signifies a discriminant validity problem. On the contrary, The Fornell-Larcker criterion compares the square root of the AVE values with the latent variable correlations. The square root of each constructs' AVE should be greater than its highest correlation withany other construct.

Hair et al., (2014) summarized the criteria to assess the reliability and validity of the reflective construct.

Rules of Thumb for Evaluating Reflective Measurement Models
Internal consistency reliability: composite reliability should be higher than 0.708 (in exploratory research, 0.60 to 0.70 is considered acceptable). Consider Cronbach's alpha as a conservative measure of internal consistency reliability.
Convergent validity: the AVE should be higher than 0.50
Discriminant validity
An indicator's outer loadings on a construct should be higher than all its cross-loadings with other constructs.
The square root of the AVE of each construct should be higher than its highest correlation with any other construct.

Figure 3.8: Rules of Thumb for Evaluating Reflective Measurement Models

3.9.4.1.2 Assessing the Formative Measurement Model

In contrast to the reflective measurement models, formative measurement models are assumed to be error-free. Therefore, reliability measurement is not appropriate to be analyzed. In the formative measurement model, some assessments involve which are the assessment of content validity, convergent validity, collinearity, and lastly the outer weight.

3.9.4.1.2.1 Content Validity

In the formative measurement model, it is appropriate to establish content validity before all the measured constructs are evaluated. The rationale behind this is to ensure that the formative indicators captured the major facets of the constructs. The more items that represent the concept the greater the content validity (Sekaran, 2003). To establish this, experts' assessments can be used.

3.9.4.1.2.2 Convergent Validity

Convergent Validity is the extent to which a measure correlates positively withother measures of the same construct. It is important to analyze whether the formatively measured construct is highly correlated with the reflective measured construct. This type of analysis is known as redundancy analysis. The term redundancy analysis comes from the information of the model which is redundant in the sense that it is already included in the formative construct, and it is included then again in the reflective construct. A value from 0.80 to 0.90 is considered satisfied for the path between Ylformative and Ylrefelective. or R2 value of 0.64 to 0.81. If R2 value < 0.64, then it indicates that the formative indicator of the construct Ylformativedoes not contribute at a sufficient level to its intended content.

3.9.4.1.2.3 Indicator Collinearity

In the formative measurement models, high correlations between items are not expected. The high correlations between two formative indicators are referred to as collinearity problems. To assess the level of collinearity, the researchers have the option either to compute the tolerance or to utilize the Variance Inflation Factor (VIF). The tolerance represents the amount of variance of one formative indicator not explained by other indicators in the same block (Hair et al., 2014). The researchers should compute the tolerance to assess for the collinearity. In contrast, VIF is referred to as reciprocal of the tolerance. If the level of collinearity is high (tolerance value of < 0.20 or VIF value of >5), the researcher should consider removing one of the indicators.

3.9.4.1.2.4 Outer Weight

Another important criterion in assessing a formative indicator is its outer weight. The outer weight is the result of multiple regression. To determine each indicator's relative contribution and its relative importance, the values of the outer weights can be compared with each other. The estimated values of outer weights are normally smaller than the outer loadings of reflective indicators.

It is important to assess whether the formative indicators truly contribute to forming the construct. To establish this, it is crucial to analyze the outer weights which should be significantly different from zero. This can be achieved by means of bootstrapping.

In bootstrapping, subsamples are drawn from the original set of data. Each subsample is then used to estimate the model. The parameter estimates are used to derive standard errors for the estimates. t values are calculated to assess each indicator's weight significance. Figure 3.9 summarizes the rules of thumb for the assessment of formative measurement indicators.

Rules of Thumb for the Evaluation of Formative Measurement Indicators
• Assess the formative construct's convergent validity by examining its correlation with an alternative measure of the construct, using reflective measures or a global single item (redundancy analysis). The correlation between the constructs should be 0.80 or higher. • Collinearity of indicators: Each indicator's tolerance (VIF) value should be higher than 0.20 (lower than 5). Otherwise, consider eliminating indicators, merging indicators into a single index, or creating higher- order constructs to treat collinearity problems • Examine each indicator's outer weight (relative importance) and outer loading (absolute importance) and use bootstrapping to assess their significance • \When an indicator's weight is significant, there is empirical support to retain the indicator • When an indicator's weight is not significant, but the corresponding item loading is relatively high (ie>50), the indicator should be generally retained • If both outer weight and outer loading are nonsignificant, there is no empirical support to retain the indicator and it should be removed from the model

Figure 3.9: Rules of Thumb for the Assessment of Formative Measurement Indicators

3.9.4.2 Evaluation of Structural Model

Once the reliability and validity of the construct are confirmed, the next process is the evaluation of the structural model. In the structural model, exogenous constructs are distinguished. The term exogenous is used to describe latent constructs that do not have any structural path relationships pointing at them while the endogenous is described as the latent target of constructs that are explained by other constructs (Hair et al., 2011). Few procedures involve in the evaluation of thestructural model which are assessing the collinearity issues, the path coefficients, the level of R2, the effect sizes of f2, and the predictive relevance of Q2 and the q2 effect sizes.

3.9.4.2.1 R-Squares

The primary analysis to test the structural model is by observing the value of *R*-squares, which indicates the coefficient of determination. This coefficient is a measure of the model's predictive accuracy and is calculated as the squared correlation between a specific endogenous construct's actual and predicted values (Hair et al., 2014). The R^2 values range from 0 to 1 with higher levels indicating a higher level of predictive accuracy. According to Hair et al., (2014), it is quite difficultto provide rules of thumb for acceptable R^2 values as this depends on the model complexity and the research discipline. It is also commonly accepted that the value of 0.60 is considered as substantial while 0.30 is regarded as moderate and a value around 0.19 indicates a weak relationship (Henseler et al., 2009; Urbach & Ahlemann,2010).

3.9.4.2.2. Effect Size

Furthermore, the R^2 value can also be used to analyze the effect of an endogenous construct on exogenous construct. The formula to calculate effect size is as follows

$$f^2 = \frac{R^2_{included} - R^2_{excluded}}{1 - R^2_{included}}$$

The change in R^2 values is calculated twice. The values of 0.02, 0.15, and 0.35 respectively represent small, medium, and large effects (Cohen, 1988) of the exogenous latent variable.

3.9.4.2.3 Path Coefficient

After running the PLS-SEM algorithm, the structural model relationship estimates are obtained. This represents the hypothesized relationship among the constructs. The path coefficient value ranges from -1 to +1. If the value is close to +1, it is the indicator of a strong positive relationship among the constructs and vice versa for negative values. To test the significance of the coefficient, (Hair et al., 2014) suggest the use of bootstrapping.

As PLS-SEM does not assume normality of data that denied the utilization of parametric significance test, PLS-SEM, therefore, relies on the nonparametric bootstrap procedure. In bootstrapping, a large number of subsamples are drawn from the original sample with replacement.

3.9.4.2.4 Bootstrapping

Since PLS does not assume normality data distribution, it relies on nonparametric resampling technique in its evaluations which is bootstrapping procedure. In bootstrapping, a large number of subsamples are drawn from the original sample with replacement. This replacement means that each time an observation is drawn at random from the sampling population, it is returned to the sampling population before the next observation is drawn. In this study, a resampling with 5000 samples is used consistent with the suggestion by Hair et al., (2014).

3.9.4.2.5 Collinearity Assessment

In the assessment of collinearity, the same measures are taken as in the evaluation of formative measurement models. Tolerance and VIF are the two considerations that should be accounted for assessing the collinearity. As previously mentioned, tolerance levels of below 0.20 and VIF above 5.00 is the indicator of collinearity.

3.9.4.3 Moderating Effects

Moderating effects are evoked by variables that have influences the strength and/or the direction of a relationship between exogenous and endogenous variables (Baron & Kenny, 1986). The idea of a moderating effect is that the slope of the independent variable is no longer constant but depends linearly on the level of the moderator (Henseler & Fasott, 2010). The moderator effect requires the specification of the simple effect of the exogenous latent variable, the simple effect of the moderator variable, and the product term which is also be referred to as the interactionterm. This study has one moderator which is situational factor that moderates the relationship between motivation, ability, role perception, religion, and situational factor with tax compliance behaviour.

In PLS-SEM, two approaches are normally be used to create interaction term which are the product indicator approach and the two-stage approach. The product- indicator approach involves multiplying each indicator of the exogenous latent variable with each indicator of the moderator variable. In the case of the moderator variable having a formative measurement model, the two-stage is applied (Hair et al., 2014). The two stages are built up as follows:

Stage 1: In the first stage, the main effect PLS path model is run to obtain estimatesfor the latent variable scores. The latent variable scores are calculated and saved for further analysis.

Stage 2: In the second stage, the interaction term X M is built up as the element-wise product of the latent variable scores of X and M. This interaction term as well as the latent variable scores of X and M are used as independent variables in a multiple linear regression on the latent variable scores of Y. (Hair et al, 2014).

As the moderator for the present study only involves formative measurement, therefore the present study will employ the two-stage approach.

3.10 Conclusion

This chapter presents the proposed conceptual framework as a guide to test the variables under study. The proposed conceptual framework is later translated into hypotheses to be tested in the current study. Therefore, the discussion in this study is centralized on the operationalization of the central model of the research, MARS model of individual behaviour and performance, religiosity, and ethical sensitivity with tax compliance behaviour. The relationships of the variables proposed in the conceptual framework with tax compliance behaviour were then translated into 10 hypotheses.

The chapter also discussed the research paradigm used in the study, the methodology, the design of the research, and the research process. Subsequent to this, the discussion on sampling which includes sampling frame, sample size, and sample selection is presented.

The discussion also continued to explain on the data collection procedure and pilot testing. The discussion proceeded with construct development and measurement (reflective, formative, second-order factor, measures for variables tested in the study). The researcher also explains the procedure for data analysis which begins with the discussion of non-response bias, data screening process, common method bias, social desirability bias, and descriptive analysis. The discussions are then concentrated on structural equation modeling (SEM) applied in the study, explaining the types of SEMmodelling and model evaluation, both at the measurement of the first-order factor and second-order factor as well as the structural models. As the study involved moderatingvariables, the discussion on testing the moderating effects is also presented.