

CHAPTER 5

MODEL EVALUATION

5.0 Introduction

This chapter presents the results of the model evaluation conducted as part of this study. Specifically, the discussion of the chapter focuses on the measurement model's first-order factor by examining the reliability and validity of the indicators and constructs. The discussion also covers the second-order factor model and the structural model. Finally, the results of the hypothesis testing are presented. The chapter ends with a summary.

5.1 Measurement Model

The research model for this study was tested using Smart PLS 3.0 (Ringle, Wende & Becker, 2015). This study examined both the measurement model, which assessed the validity and reliability of the measures, and the structural model, which tested the hypothesized relationships. To predict the significance of the path coefficients and loadings, a bootstrapping method was employed using 5000 samples. It should be noted that all constructs in the research model are multi-item constructs and conceptualized as reflective and formative.

5.1.1 Reliability of Reflective Constructs

The reliability of reflective constructs, as discussed in Chapter 3 (Research Methodology), can be determined at two stages: the individual level and the construct level. At the individual level, the measure is tested on its factor loadings, while at the construct level, composite reliability is used.

5.1.1.1 Internal Consistency Reliability

The first assessment conducted in this study evaluated the internal consistency and reliability of the measures. Two tests were performed to measure reliability: Cronbach's Alpha and Composite Reliability Index.

According to Table 5.1, the constructs of intrinsic motivation, role perception, religion, moral equity, and relativism all met the threshold of 0.7 as suggested by Hair et al. (2010). For the constructs of extrinsic motivation and egoism, the suggestion of Wim et al. (2018) and DeVellis (2003) was applied, where a Cronbach Alpha of 0.6 is still considered acceptable. However, due to low Cronbach Alpha readings for the ability and utilitarianism constructs, items with low loadings were deleted with caution. For the ability construct, item GK2R was deleted, and the Cronbach Alpha score later increased to 0.605, which was deemed acceptable. For the utilitarianism construct, item ES10 was retained due to the conceptualization of the theory.

There have been debates on the use of Cronbach's Alpha as a tool to measure reliability due to its unrealistic assumptions, as pointed out by Hair et al. (2017). As an alternative, McNeish (2017) recommended using omega reliability or composite reliability. While both measures assess internal consistency, composite reliability takes into account the loadings of the indicators, making it a more accurate measure of reliability. A composite reliability scores higher than 0.7 indicates adequate internal consistency, according to Hair et al. (2011).

After some items were deleted, all constructs showed a minimum cutoff value of composite reliability of 0.7 except for Extrinsic Motivation, which had a composite reliability of 0.662. However, as suggested by Bagozzi and Yi (1988), a composite reliability score of 0.6 is still considered acceptable. These results suggest that the measurement model had acceptable reliability.

5.1.1.2 Indicator reliability

Once the internal consistency reliability has been achieved, the indicator reliability is then measured. As shown in Table 5.1, some items had to be deleted due to low Average Variance Extracted (AVE) values or low outer loading values. As shown in Table 5.2, only items that achieved the threshold value set by Byrne (2016) with AVE scores higher than 0.5 were retained.

5.1.1.3 Convergent Validity

Convergent Validity refers to the extent to which individual indicators reflect the constructs in comparison to indicators measuring other constructs (Urbach & Ahlemann, 2010). To analyse Convergent Validity, the Average Variance Extracted (AVE) is measured. The value of AVE should be higher than 0.5, which explains 50 percent of the assigned indicator's variance (Chin, 2010; Hair, Hult, Ringle & Sardstedt, 2017). Using the PLS algorithm in SmartPLS 3.0, the AVE value is calculated. Table 5.2 shows the AVE values of all the constructs. All constructs recorded AVE values higher than 0.5 for each group data. The lowest AVE value reported is for Religion (0.501), Role Perception (0.546), followed by Extrinsic Motivation (0.559), Ability (0.596), Egoism (0.754), Intrinsic Motivation (0.782), Moral Equity (0.822), Relativism (0.932) and Utilitarianism (0.541).

Table 5.1: Results Summary for Reflective Models (Before deletion)

Constructs	Constructs	Items	Indicator Reliability	Convergent Validity	Internal Reliability	Consistency	
			Outer Loadings	AVE	Composite Reliability	Cronbach's Alpha	
			>0.60	>0.50	>0.70	>0.70	
Motivation	Intrinsic motivation	IM1	0.345	0.606	0.849	0.753	
		IM2	0.856				
		IM3	0.899				
		IM4	0.875				
	Extrinsic motivation	EM1	0.469				0.356
		EM2	0.299				
		EM3	0.835				
		EM4	0.647				
Ability		GK1	0.537	0.283	0.698	0.549	
		GK2R	-0.068				
		LK1	0.662				
		LK2	0.752				
		LK3R	0.740				
		TK1	0.229				
		TK2R	0.566				
		TK3R	0.221				
Role Perception		RP1R	0.794	0.481	0.817	0.723	
		RP2R	0.621				
		RP3R	0.869				
		RP4R	0.591				
		RP5	0.533				
Religion		R1	0.701	0.498	0.826	0.802	
		R2	0.942				
		R3	0.529				
		R4	0.693				
		R5	0.591				
Ethical Sensitivity	Moral Equity	ES1R	0.914	0.822	0.949	0.927	
		ES2R	0.925				
		ES3R	0.931				
		ES4R	0.854				
	Relativism	ES5R	0.971	0.939	0.969	0.935	
		ES6R	0.968				
	Egoism	ES7R	0.831	0.754	0.860	0.679	
		ES8R	0.904				
	Utilitarianism	ES9R	0.986	0.541	0.317	-0.409	
		ES10	-0.334				

Table 5.2: Results Summary for Reflective Models (After deletion)

Constructs	Constructs	Items	Indicator Reliability	Convergent Validity	Internal Reliability	Consistency
			Outer Loadings	AVE	Composite Reliability	Cronbach's Alpha
			>0.60	>0.50	>0.70	>0.70
Motivation	Intrinsic Motivation	IM2	0.860	0.782	0.849	0.861
		IM3	0.912			
		IM4	0.881			
	Extrinsic Motivation	EM3	0.865	0.559	0.662	0.658
		EM4	0.608			
Ability		LK1	0.749	0.596	0.744	0.676
		LK2	0.819			
		LK3R	0.747			
Role Perception		RP1R	0.794	0.546	0.817	0.711
		RP2R	0.661			
		RP3R	0.876			
		RP4R	0.591			
Religion		R1	0.702	0.501	0.826	0.802
		R2	0.940			
		R3	0.538			
		R4	0.695			
		R5	0.597			
Ethical Sensitivity	Moral Equity	ES1R	0.914	0.822	0.949	0.927
		ES2R	0.925			
		ES3R	0.931			
		ES4R	0.854			
	Relativism	ES5R	0.971	0.939	0.969	0.935
		ES6R	0.968			
	Egoism	ES7R	0.814	0.754	0.860	0.679
		ES8R	0.904			
	Utilitarianism	ES9R	0.990	0.537	0.645	0.290
		ES10R	0.312			

5.1.1.4 Discriminant Validity

Discriminant validity refers to the degree to which indicators differentiate across constructs or measure distinct concepts by examining the correlations between measures of potentially overlapping constructs. In other words, it refers to the extent to which the constructs under investigation are truly distinct from one another. In SmartPLS 3.0, there are three criteria to assess discriminant validity: cross-loading criterion, Fornell & Larcker's (1981) criterion, and Heterotrait-Monotrait ratio of correlations (HTMT). Following the suggestion of Ramayah et al. (2019) that any one method should be adequate for establishing discriminant validity, this study uses Fornell & Larcker's (1981) criterion. According to this criterion, a latent variable should explain the variance of its own indicators better than the variance of other latent variables. The AVE of a latent variable should be higher than the squared correlation between the latent variable and all other variables or the square root of AVE on the diagonal should be higher than the correlation on the off-diagonal. Based on Table 5.3, the AVE of all constructs is higher than the correlations between the constructs and other constructs in the model.

Table 5.3: Discriminant Validity

Constructs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1.Ability	0.773														
2.Egoism	0.291	0.868													
3.Ethical Sensitivity	0.319	0.878	0.826												
4.Extrinsic Motivation	0.029	-0.169	-0.119	0.679											
5.Financial Constraints	0.353	0.106	0.145	-0.034	0.948										
6.Intrinsic motivation	0.245	0.073	0.123	0.222	0.157	0.885									
7.Moral Equity	0.296	0.754	0.949	-0.077	0.104	0.122	0.907								
8.Motivation	0.234	0.025	0.084	0.464	0.134	0.967	0.094	0.602							
9.Peers Influence	0.328	0.19	0.213	0.016	0.192	0.249	0.151	0.233	0.897						
10.Relativism	0.258	0.692	0.863	-0.137	0.223	0.088	0.735	0.048	0.212	0.969					
11.Religion	-0.035	0.049	0.11	0.196	-0.038	0.016	0.132	0.065	-0.186	0.093	0.686				
12.Role Perception	0.519	0.337	0.343	0.064	0.186	0.401	0.291	0.385	0.375	0.305	-0.024	0.801			
13.Situational factor	0.411	0.203	0.237	0.001	0.526	0.271	0.169	0.249	0.936	0.264	-0.175	0.392	0.723		
14.Tax Compliance Behaviour	0.5	0.379	0.406	-0.048	0.257	0.263	0.326	0.232	0.483	0.404	-0.075	0.785	0.511	0.687	
15.Utilitarianism	0.291	0.756	0.797	-0.055	0.073	0.171	0.689	0.145	0.292	0.584	0.068	0.322	0.28	0.393	0.787

5.2. Second-Order Factor Model

At this stage, the reliability and validity of measures in the Measurement first order model have been adequately satisfied. Since two constructs namely motivation and ethical sensitivity are developed as a second order factor model, there is a need to test the second order factor model. In this study, two stage approach is used to assess second order factor test. This is due to the different number of indicators across the lower order components and the involvement of formative measures in the model (Henseler & Chin, 2010). In the first stage, the measurement model of the first-order constructs is assessed to ensure the reliability and validity of the constructs. In the second stage, the measurement model of the second-order construct is assessed to confirm the validity of the overall model. The validity of the second order construct is tested using the Partial Least Squares (PLS) algorithm in SmartPLS 3.0. The PLS algorithm tests the significance of the outer weights of the second-order construct and the path coefficients of the second-order construct to other first-order constructs.

5.2.1 Motivation

The first stage of the analysis involves running a main effect PLS path model to obtain estimates for the latent variable scores. The measurement model is then assessed for convergent validity using factor loadings, Cronbach Alpha, Composite Reliability (CR), and Average Variance Extracted (AVE), as recommended by Hair et al. (2017), Hair et al. (2014), and Hair et al. (2006). Internal consistency of the

constructs is measured using Cronbach Alpha and Composite Reliability. The results, as shown in Table 5.4, indicate that both constructs pass the Internal Consistency Reliability test, with Cronbach Alpha values of 0.658 and 0.861, meeting the acceptable and very good reliability thresholds suggested by Ursachi et al. (2015). Additionally, the Composite Reliability Test shows that both constructs have adequate internal consistency, with readings above 0.7, as recommended by Hair, Ringle, and Sardstedt (2011). Convergent validity of the constructs is assessed by analysing the factor loadings and AVE, with factor loadings between 0.6 and 0.7 being considered acceptable in social science studies according to Hair et al. (2017). Similarly, an AVE value above 0.5 suggests an adequate convergent validity (Hair et al., 2017; Bagozzi & Yi, 1988). The results of these tests are also shown in Table 5.4.

Subsequently, the discriminant validity of the constructs is assessed. It is measured using Fornell Lacker's Criterion where the square root of the AVE of each of the latent variables should be greater than its correlation with other latent variables. As shown in Table 5.5, the square root of the AVE of each of the latent variables was greater than its correlation with another latent variable.

At the second stage, outer weights, outer loadings, t-values, and VIF are assessed. Outer weights are the results of a multiple regression of a construct on its set of indicators. Weights are the primary criterion used to assess each indicator's relative importance in formative measurement models. The bootstrapping procedure was carried out using 5000 resamples to assess the significance of weights. Lohmöller (1989) recommended a weight of >0.1 for an indicator. The results reveal that the weights of the intrinsic motivation indicators are more than 0.1, but the weight of the

extrinsic motivation indicators is less than 0.1. Looking at the significance levels, it was found that the extrinsic motivation indicators are non-significant. Based on Table 5.6, the t-values of the intrinsic motivation indicators are more than 2.57, which indicates the significance of the outer loading. However, the t-values for the extrinsic motivation indicators show that they are non-significant. Despite the weights of extrinsic motivation being found not significant, prior research and theories on motivation provide support and relevance for these indicators in capturing the motivation dimension (Ryan & Deci, 2000; Reiss, 2005). Thus, all indicators are retained even though one of the outer weights is not significant. In terms of collinearity between formative items, Variance Inflation Factor (VIF) was examined. According to Table 5.6, the VIF values for both constructs are 1.091, which fall below the threshold value of 5. It can be concluded that collinearity does not reach critical levels in any of the formative constructs, and it is not an issue for the estimation of the PLS path model.

Subsequently, the discriminant validity of the constructs at the second stage is also assessed. It is measured using Fornell and Larcker Criterion where the square root of the AVE of each of the latent variables should be greater than its correlation with another latent variable. As shown in Table 5.7, the square root of the AVE of each of the latent variables was greater than its correlation with another latent variable.

Table 5.4: Measurement Model for Motivation (Stage One)

Construct	Item	Loading	Cronbach Alpha	Composite Reliability	Average Variance Extracted (AVE)
Intrinsic Motivation	IM2	0.860	0.861	0.711	0.559
	IM3	0.912			
	IM4	0.881			
Extrinsic Motivation	EM3	0.864	0.658	0.915	0.782
	EM4	0.609			

Table 5.5: Fornell Larcker Criterion for Motivation (Stage One)

Constructs	1	2
1. Extrinsic motivation	0.748	
2. Intrinsic motivation	0.288	0.885

Table 5.6: Measurement Model for Motivation (Stage 2)

Construct	Item	Weights	Loadings	T-Values	VIF	p-values
Motivation	Intrinsic motivation	0.991	1.000	3.736**	1.091	0.000
	Extrinsic Motivation	0.029	0.314	0.704	1.091	0.475

Note: >2.57*

Table 5.7: Fornell Larcker Criterion for Motivation (Stage 2)

	1	2
1. Motivation	0.708	
2. Tax Compliance Behaviour	0.304	0.669

5.2.2 Ethical Sensitivity

At the first stage of the analysis, a main effect PLS path model was run to obtain estimates for the latent variable scores. The measurement model was assessed for convergent validity, which was examined through factor loadings, Cronbach Alpha, Composite Reliability (CR), and Average Variance Extracted (AVE) (Hair et al., 2017; Hair et al., 2014; Hair et al., 2006). Internal consistency of the constructs was measured using Cronbach Alpha and Composite Reliability. Table 5.8 shows that all constructs passed the Internal Consistency Reliability test, with Cronbach Alpha values ranging between 0.679 and 1.000, meeting the suggestion of Ursachi et al. (2015) that 0.6-0.7 indicates an acceptable level of reliability, and 0.8-0.94 indicates a very good reliability. To ensure a more rigorous estimate, a Composite Reliability test was carried out, with all constructs achieving a reading of more than 0.7, indicating adequate internal consistency, according to Hair, Ringle & Sardstedt (2011). The convergent validity of the constructs was assessed by analyzing the factor loadings and AVE, with the factor loadings in this study being acceptable between 0.6 and 0.7, as suggested by Hair et al. (2017). Likewise, the AVE value of the study, which is above 0.5, suggests an adequate convergent validity (Hair et al., 2017; Bagozzi & Yi, 1988). The results are presented in Table 5.8.

Subsequently, the discriminant validity of the constructs is assessed. It is measured using Fornell Larcker Criterion where the square root of the AVE of each of the latent variables should be greater than its correlation with other latent variables. As shown in Table 5.9, the square root of the AVE of each of the latent variables was greater than its correlation with other latent variables.

At the second stage, outer weights, outer loadings, t-values and VIF are assessed. The bootstrapping procedure was carried out using 5000 resamples to assess the significance of weights. Lohmöller (1989) recommended >0.1 weight for an indicator. The results reveal that all the indicator's weights are more than 0.1. Looking at the significance levels, it was found that all indicators are significant. Based on Table 5.10 also, the t-values of all indicators are more than 2.57 which indicate the significance of the outer loading. In terms of collinearity between formative items, Variance Inflation Factor (VIF) was examined. According to Table 5.9, the VIF values for all constructs are below the threshold value of 5 except for utilitarianism.

Subsequently, the discriminant validity of the constructs at the second stage is also assessed. It is measured using Fornell Larcker Criterion where the square root of AVE of each of the latent variables should be greater than its correlation with other latent variables. As shown in Table 5.11, the square root of AVE of each of the latent variables was greater than its correlation with other latent variables.

Table 5.8: Measurement Model for Ethical Sensitivity (Stage One)

Construct	Item	Loading	Cronbach Alpha	Composite Reliability	Average Variance Extracted (AVE)
Moral Equity	ES1RC	0.914	0.927	0.949	0.822
	ES2RC	0.925			
	ES3RC	0.931			
	ES4RC	0.854			
Relativism	ES5RC	0.971	0.935	0.969	0.939
	ES6RC	0.967			
Egoism	ES7RC	0.830	0.679	0.860	0.754
	ES8RC	0.905			
Utilitarianism	ES9RC	0.986	0.290	0.654	0.541
	ES10RC	0.332			

Table 5.9: Fornell Larcker Criterion for Ethical Sensitivity (Stage One)

	1	2	3	4
1. Egoism	0.868			
2.Moral Equity	0.754	0.907		
3.Relativism	0.692	0.735	0.969	
4.Utilitarianism	0.756	0.689	0.584	1.000

Table 5.10: Measurement Model for Ethical Sensitivity (Stage 2)

Construct	Item	Weights	Loadings	T-Values	VIF	p-values
Ethical Sensitivity	Egoism	0.288	0.909	11.047**	3.318	0.000
	Moral Sensitivity	0.250	0.894	7.776***	3.099	0.000
	Relativism	0.309	0.860	9.071***	2.406	0.000
	Utilitarianism	0.289	0.858	8.868***	5.523	0.000

Table 5.11: Fornell Larcker Criterion for Ethical Sensitivity (Stage Two)

	1	2
1. Ethical Sensitivity	0.881	
2.Tax Compliance Behaviour	0.437	0.686

5.3 Structural Model

After the measurement model was established, the analysis continued with structural model evaluation. Assessment of the structural model is used to determine the model's capabilities to predict one or more target constructs.

5.3.1 Assessment of the Structural Model

The first step in the structural model is to assess collinearity issues. It is crucial to safeguard against collinearity issues between the constructs before performing a latent variable analysis in the structural model. The VIF value is used to measure the collinearity between the constructs. The threshold value for assessment is 5, following the suggestion of Hair, Ringle, & Sarstedt (2011) or 3.3, following Diamantopoulos & Siguaw (2006). In this study, as shown in Table 5.12, all the inner VIF values for the constructs are within 1.029 and 3.388, which are less than 5 (Hair, Ringle, & Sarstedt, 2011), indicating collinearity is not a concern in this study.

5.3.2 Assessment of the Structural Model Relationships

In order to test the hypotheses of the study, the bootstrapping procedure is utilized to produce t-value results for each path relationship in the model as shown in Table 5.12. Bootstrapping in PLS is a nonparametric test which comprises repeated random sampling with replacement from the original sample to produce a bootstrap sample and to attain standard errors for hypothesis testing (Hair et al, 2011).

In this study, five hypotheses were developed for the constructs, excluding the moderator. To test the significance level, t-statistics for all paths were generated using SmartPLS 3.0 bootstrapping function. The bootstrapping was set at a 0.05 significance level, one-tailed test, and 1000 subsamples, following the suggestion of Chin (2010). The critical values or a significance level of 1 per cent ($\alpha = 0.01$), 5 per

cent ($\alpha = 0.05$), and 10 per cent ($\alpha = 0.1$) are 2.33, 1.645, and 1.28, respectively, for the one-tailed test (Ramayah et al., 2018).

Based on the assessment of the path coefficient, as shown in Table 5.12, only two relationships were found to have a t-value > 1.645 , thus significant at the 0.05 level of significance. Specifically, the predictors of role perception ($\beta=0.674$, $p<0.01$) and ethical sensitivity ($\beta=0.129$, $p<0.01$) are positively related to tax compliance behavior, which explains 67.6% of variances in tax compliance behavior. Thus, H3 and H5 are supported. The R^2 value of 0.676 is above the 0.26 value suggested by Cohen (1988), which indicates a substantial model.

5.3.3 The Coefficient of Determination (R^2)

The next stage is to assess the model's predictive accuracy through the coefficient of determination (R^2). The R^2 computes the model's predictive power and the value ranges between 0 to 1, with a higher value indicating a higher level of predictive accuracy (Hair, Hult, Ringle & Sarstedt, 2017). Using the SmartPLS algorithm, the R^2 is calculated. As there are various sets of rules on the acceptable R^2 , this study follows the guideline by Chin, 1998. By referring to Table 5.12, motivation, ability, role perception, religiosity and ethical sensitivity explain 67.6% of variance in tax compliance behaviour, which indicate a substantial level of predictive accuracy.

5.3.4 Assessment of the Effect Size (f^2)

In this stage, the effect sizes (f^2) are analyzed. As stated by Sullivan and Fein (2012), while p-values can inform the reader whether an effect exists, they do not reveal the size of the effect. The f^2 measure computes the relative impact of an exogenous construct on an endogenous construct. Specifically, it assesses how strongly an exogenous construct contributes to explaining a certain endogenous construct in terms of R^2 . To measure the effect size, the guideline by Cohen (1988) is followed. According to Cohen (1988), values of 0.02, 0.15, and 0.35 represent small, medium, and large effects, respectively. From Table 5.12, ethical sensitivity has a small effect on producing R^2 for tax compliance behavior, while role perception has a substantial effect in producing R^2 for tax compliance behavior. On the other hand, motivation, ability, and religiosity do not predict tax compliance behavior.

5.3.5 Assessment of Predictive Relevance (Q^2)

Finally, the predictive relevance of the model is assessed through the blindfolding procedure as suggested by Hair et al (2017). The Q^2 is larger than 0, indicating that the model has sufficient predictive relevance. The Q^2 or predictive relevance analysis was conducted by using a distance value of 7. Based on blindfolding assessment, the predictive relevance Q^2 values for motivation, ethical sensitivity and tax compliance behaviour are 0.479, 0.671 and 0.305 respectively.

Table 5.12: Structural Model Assessment

Relationship	Path Coefficient β	Std Error	BCI LL	BCI UL	t-value	p-value	Decision	R ²	f ²	Effect Size	VIF
M→TCB	-0.082	0.063	-0.187	0.016	1.306	0.096	Not supported		0.018	None	1.029
AB→TCB	0.037	0.066	-0.063	0.145	0.564	0.286	Not supported		0.003	None	1.515
RP→TCB	0.674	0.073	0.538	0.779	9.173**	0.000	Supported	0.676	0.887	Substantial	1.661
RLG→TCB	-0.019	0.074	-0.123	0.121	0.261	0.397	Not supported		0.001	None	1.067
ES→TCB	0.129	0.057	0.050	0.230	2.245**	0.013	Supported		0.045	Small	1.202

5.4. Moderation Analysis

After testing the direct effect, the moderation hypothesis is tested. A moderator can be visualized as a third variable that changes the relationship between the independent variable and dependent variable (Yong et al., 2019). As situational factor variable which is the moderator of the study is developed as second order factor model, there is a need to analyze the second order factor before the interaction effects can be analyzed. In this study, a two-stage approach is used to analyze the second order factor. This is due to the different number of indicators across lower order components and the involvement of formative measures in the model (Hanseler & Chin, 2010).

At the first stage, the main effect PLS path model is run to obtain estimates for the latent variable scores. First, the measurement model was assessed for convergent validity. This was examined through factor loadings, Cronbach Alpha, Composite Reliability (CR), and Average Variance Extracted (AVE) (Hair et al., 2017; Hair et al., 2014; Hair et al., 2006). Internal consistency of the constructs was measured using Cronbach Alpha and Composite Reliability. Based on Table 5.13, both constructs pass Internal Consistency Reliability. The values of Cronbach Alpha are 0.888 and 0.878 which meet the threshold of 0.7 as suggested by Hair et al. (2010). In order to ensure a more rigorous estimate, the Composite Reliability Test is carried out. According to Hair, Ringle & Sardstedt (2011), a value of more than 0.7 indicates adequate internal consistency; therefore, this study fulfills this. Furthermore, the convergent validity of the constructs was assessed by analyzing the factor loadings and the AVE. According to Hair et al. (2017), factor loadings are acceptable between 0.6 and 0.7 in social science studies; therefore, the factor loadings in this study are

acceptable. Likewise, the AVE value of the study above 0.5 suggests adequate convergent validity (Hair et al., 2017; Bagozzi & Yi, 1988).

Subsequently, the discriminant validity of the constructs is assessed. It is measured using Fornell Larcker Criterion where the square root of the AVE of each of the latent variables should be greater than its correlation with another latent variable. As shown in Table 5.14, the square root of the AVE of each of the latent variables was greater than its correlation with other latent variables.

At the second stage, the outer weights, outer loadings, t-values, and VIF are assessed. Outer weights are the results of a multiple regression of a construct on its set of indicators. Weights are the primary criterion to assess each indicator's relative importance in formative measurement models. The bootstrapping procedure was carried out using 5000 resamples to assess the significance of weights. Lohmöller (1989) recommended a weight of >0.1 for an indicator. The results reveal that the weights of the intrinsic motivation indicators are more than 0.1, but the weight of the extrinsic motivation indicators is less than 0.1. Looking at the significance levels, it was found that the extrinsic motivation indicators are non-significant. Based on Table 5.15, the t-values for both constructs are more than 2.57, indicating the significance of the outer loading. In terms of collinearity between formative items, the Variance Inflation Factor (VIF) was examined. According to Table 5.15, the VIF values for both constructs are 1.038, which is below the threshold value of 5. It can be concluded that collinearity does not reach critical levels in any of the formative constructs, and it is not an issue for the estimation of the PLS path model.

The discriminant validity of the constructs at the second stage is also

assessed. It is measured using Fornell Larcker Criterion where the square root of the AVE of each of the latent variables should be greater than its correlation with other latent variables. As shown in Table 5.16, the square root of the AVE of each of the latent variables was greater than its correlation with other latent variables.

Table 5.13: Measurement Model for Situational Factor (Stage One)

Construct	Item	Loading	Cronbach Alpha	Composite Reliability	Average Variance Extracted (AVE)
Financial Constraints	FC2	0.944	0.888	0.947	0.899
	FC3	0.953			
Peers Influence	PI1RC	0.877	0.878	0.925	0.804
	PI2RC	0.894			
	PI3RC	0.920			

Table 5.14: Fornell Larcker Criterion for Situational Factor (Stage One)

Constructs	1	2
1.Financial Constraints	0.948	
2.Peers Influence	0.192	0.897

Table 5.15: Measurement Model for Motivation (Stage 2)

Construct	Item	Weights	Loadings	T-Values	VIF	p-values
Situational factor	Financial Constraints	0.398	0.560	4.238**	1.038	0.000
	Peers Influence	0.844	0.921	18.784**	1.038	0.000

Note: >2.57*

Table 5.16: Fornell Larcker Criterion for Motivation (Stage 2)

	<i>1</i>	<i>2</i>
1. Situational factor	0.762	
2. Tax compliance behavior	0.554	0.676

Once the assessment of second order factor is satisfied, then the interaction effects of the moderator variable can be analyzed. In order to assess the moderation effects, a Two- Stage Approach is used. The idea of a two-stage approach was initially proposed by Chin et al. (2003) and elaborated further by Fassot, Henseler & Coelho (2016) as well as Henseler & Fassott (2010). As formative indicators are not assumed to reflect the same underlying construct, the product indicator approach is not suitable to be used. Instead, a two-stage approach is more suitable to estimate the moderating effects.

This study aims to test the effects of situational factor which acts as moderator towards the relationship between individual behaviour construct and tax compliance. There are five hypotheses proposed for the moderator.

5.4.1. Motivation

This study tests the influence of situational factor (moderator) towards the relationship between motivation (independent variable) and tax compliance behaviour (dependent variable).

This study hypothesized that:

H6: The positive relationship between motivation and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.

The moderation assessment follows a Two-Stage approach (Chin et al,2003). This approach takes advantage of PLS path modelling's ability to explicitly estimate latent variable scores. The first step is to obtain estimates for the latent variable scores, which is done by using an algorithm. Before proceeding with the algorithm, the assessment of collinearity issues in the formative measurement model is carried out. Since the indicators are not interchangeable, high correlations are not expected between them in formative measurement models. High correlations between the formative indicators indicate a collinearity issue. To assess collinearity issues, Variance Inflation Factor (VIF) is examined. According to Table 5.17, the VIF values for both constructs are 1.038, which is below the threshold of 5 suggested by Hair, Ringle & Sardstedt (2011). Therefore, collinearity issues do not reach critical levels in any of the formative constructs. Then, the significance and relevance of the indicators are assessed using outer weights. The bootstrapping procedure is carried out using 5000 resamples to assess the significance of weights. Lohmöller (1989) recommended a weight of >0.1 for an indicator. It was found that the indicator's weights for both constructs are more

than 0.1, indicating that both indicators are significant and relevant.

Once the measurement model scores are satisfied, the next stage is to calculate the interaction term. The R^2 for the main model without the interaction is 0.336 and with the interaction effect model, the R^2 is 0.337.

Based on Kenny's (2016) guidelines, effect sizes of 0.005, 0.001, and 0.025 indicate small, medium, and large effects, respectively. Therefore, as the f^2 effect size in this study is 0.0015, it is considered none. Next, to determine the significance of the relationship, bootstrapping procedures were conducted with cutoff values of 1.645 ($\alpha= 0.05$) and 2.33 ($\alpha= 0.01$). As shown in Table 5.18, MTVN*SF is not significant (t-value= 0.342), leading to the rejection of hypothesis H6.

Table 5.17: Measurement model for Situational Factor

Construct	Item	Weights	Loadings	t-values	VIF	p-values
Situational factor	Financial Constraints	0.267	0.443	1.615	1.038	0.053
	Peers Influence	0.914	0.965	11.441	1.038	0.000

Table 5.18: Moderation Model Assessment for Motivation

Hypothesis	Relationship	Std. Beta	Std. Error	t-value
H6	Motivation (MTVN) Situational Factor (SF) *	0.023	0.105	0.342

5.4.2. Ability

This study tests the influence of situational factor (moderator) towards the relationship between ability (independent variable) and tax compliance behaviour (dependent variable).

This study hypothesized that:

H7: The positive relationship between ability and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.

An interaction effect was created between the ability construct and tax compliance behavior construct. The R² for the main model without the interaction is 0.375, and with the interaction effect model, the R² is 0.759. The R² change of 0.384 indicates that adding one interaction term changes the R² by about 38.4%. The effect size was calculated and found to be 1.5934, which is considered large according to Kenny's (2016) suggestion that an f² greater than 0.025 is a large effect size. To determine the significance of the relationship, bootstrapping procedures were conducted, with a cutoff value of 1.645 ($\alpha = 0.05$) and 2.33 ($\alpha = 0.01$). As shown in Table 5.19, AB*SF is not significant (t-value= 0.570). Therefore, hypothesis H7 is rejected.

Table 5.19: Moderation Model Assessment for Ability

Hypothesis	Relationship	Std. Beta	Std. Error	t-value
H7	Ability (AB) * Situational Factor (SF)	0.048	0.076	0.570

5.4.3. Role Perception

This study examines the influence of situational factor as the moderator towards the relationship between ability as the independent variable and tax compliance behaviour as the dependent variable.

This study hypothesized that:

H8: The positive relationship between role perception and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.

The interaction effect between the role perception construct and tax compliance behavior construct was created. The R² for the main model without the interaction is 0.666, and with the interaction effect model, the R² is 0.675. The R² change of 0.009 indicates that adding one interaction term changes the R² by about 9%. The effect size was calculated and found to be small with a value of 0.0277, following the suggestion by Kenny (2016) that an f² greater than 0.005 is considered a small effect size. To determine the significance of the relationship, bootstrapping procedures were conducted, with cutoff values of 1.645 ($\alpha= 0.05$) and 2.33 ($\alpha= 0.01$).

As shown in Table 5.20, RP*SF is significant as the t-value is 11.121. Therefore, hypothesis H8 is accepted.

Table 5.20: Moderation Model Assessment for Role Perception

Hypothesis	Relationship	Std. Beta	Std. Error	t-value
H8	Role Perception (RP) * Situational Factor (SF)	0.093	0.063	11.121**

Next, as suggested by Dawson (2014), to further elaborate on the moderating interaction effect of the situational factor, the pattern of interaction effect is plotted to see how the moderator changes the relationship between role perception and tax compliance behaviour. As seen in Figure 5.1, the line labelled for high situational factor has a steeper gradient when compared to a low situational factor. This indicates that the positive relationship is indeed stronger when the situational factor is high. Therefore, based on the hypothesis, it can be concluded that higher situational factors will strengthen the positive relationship between role perception and tax compliance behaviour.

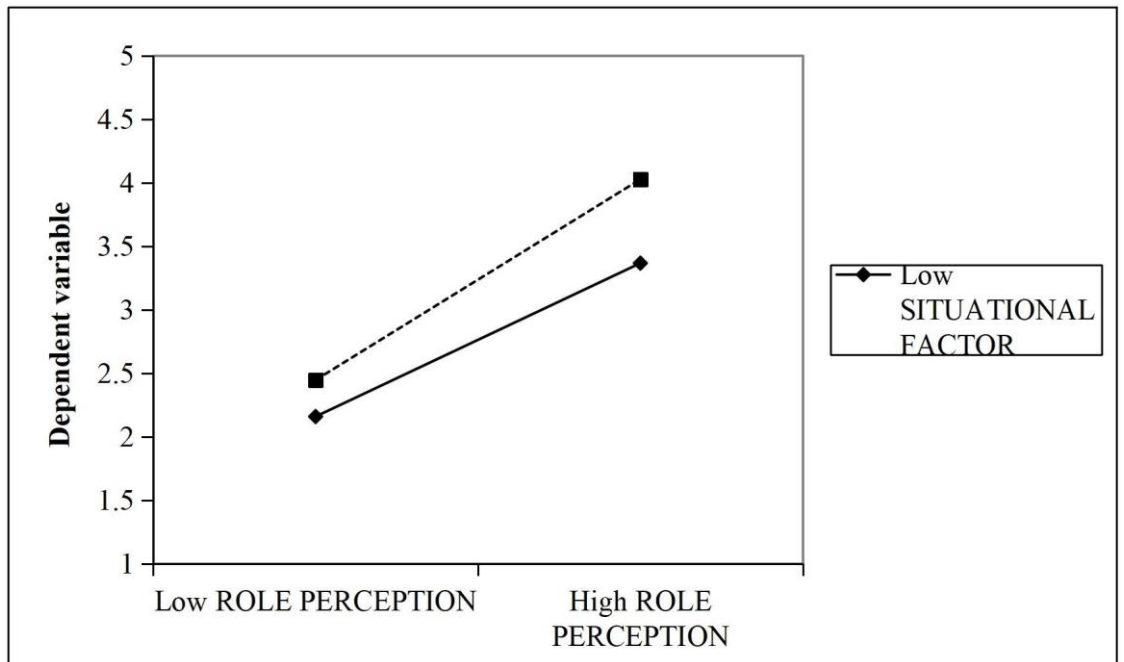


Figure 5.1: Interaction Plot

5.4.4. Religion

This study tests the influence of situational factor as the moderator towards the relationship between religion as the independent variable and tax compliance behaviour as the dependent variable.

This study hypothesized that:

H9: The positive relationship between religiosity and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.

The interaction effect between religiosity construct and tax compliance behaviour construct was created. The R^2 for the main model without the interaction is

0.308 and with the interaction effect model, the R^2 is 0.310. The R^2 changes of 0.002 indicates that with the addition of one interaction term, the R^2 changes about 2%. Next, the effect size is calculated, and it is found out that the effect size is of 0.029 which is considered as none. This is following the suggestion by Kenny (2016) where an f^2 of more than 0.005 is considered as small effect size. Next, in order to determine the significant relationship, the bootstrapping procedures are conducted. The cut off value for this test is 1.645 ($\alpha= 0.05$) and 2.33 ($\alpha= 0.01$). As shown in Table 5.21 RLGN*SF is not significant as the t-value=0.316. Due to this, hypothesis H₉ is rejected.

Table 5.21: Moderation Model Assessment for Religiosity

Hypothesis	Relationship	Std. Beta	Std. Error	t-value
H9	Religiosity (RLGN) * Situational Factor (SF)	0.007	0.085	11.121**

5.4.5. Ethical Sensitivity

This study examines the influence of situational factor as the moderator towards the relationship between ethical sensitivity as the independent variable and tax compliance behaviour as the dependent variable.

This study hypothesized that:

H10: The positive relationship between ethical sensitivity and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.

The interaction effect between the ethical sensitivity construct and tax compliance behaviour construct was created. The R² for the main model without the interaction is 0.367, and with the interaction effect model, the R² is 0.373. The R² change of 0.006 indicates that with the addition of one interaction term, the R² changes by about 6%. Next, the effect size is calculated, and it is found to be none (0.0096), following the suggestion by Kenny (2016) that an f² of more than 0.005 is considered to have no effect size.

Next, to determine the significant relationship, bootstrapping procedures are conducted. The cutoff value for this particular test is 1.645 ($\alpha = 0.05$) and 2.33 ($\alpha = 0.01$). As shown in Table 5.22, ES*SF is not significant, as the t-value is 0.929. Therefore, hypothesis H10 is rejected.

Table 5.22: Moderation Model Assessment for Ethical Sensitivity

Hypothesis	Relationship	Std. Beta	Std. Error	t-value
H10	Ethical Sensitivity (RLGN) * Situational Factor (SF)	-0.064	0.082	0.929

5.4 Summary of Hypotheses Testing

Based on the previous evaluation of the structural model, the assessment of the path coefficient and the t-value are used to assess the hypotheses of the study. Table 5.23 summarizes all the hypotheses tested in this study.

Table 5.23: Summary of Hypotheses Testing

No.	Hypothesis Statement	Decision
H ₁	<i>Motivation has a positive effect on tax compliance behaviour</i>	Not supported
H ₂	<i>Ability has a positive effect on tax compliance Behavior</i>	Not supported
H ₃	<i>Role perception has a positive effect on tax compliance behaviour</i>	Supported
H ₄	<i>Religiosity has a positive effect on tax compliance Behavior</i>	Not supported
H ₅	<i>Ethical sensitivity has positive effect on tax compliance behavior</i>	Supported
H ₆	<i>The positive relationship between motivation and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.</i>	Not supported
H ₇	<i>The positive relationship between ability and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.</i>	Not supported
H ₈	<i>The positive relationship between role perception and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.</i>	Supported
H ₉	<i>The positive relationship between religiosity and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.</i>	Not supported
H ₁₀	<i>The positive relationship between ethical sensitivity and tax compliance behavior among professionals in Malaysia will be stronger when situational factor is high.</i>	Not supported

5.6. Conclusion

This chapter explains in detail all the analyses that have been conducted both on the measurement and structural models. Firstly, the structural model demonstrates the reliability and validity of the measures. Constructs that demonstrate low cut off value are treated with precaution. Secondly, the validation of the structural model is tested using R_2 values. Based on the findings, three hypotheses are supported. The next chapter provides the discussion of the findings and the overall contribution of this study.

