Assessing Data Sharing's Model Fitness Towards Open Data by using Pooled CFA

Siti Nur'asyiqin Ismael[1], Othman Mohd[2], Yahaya Abd Rahim[3]

Faculty of Information and Communication Technology

UTeM, Melaka, Malaysia

Abstract—This study demonstrates the step-by-step procedure to perform Pooled Confirmatory Factor Analysis (CFA) in the measurement part of Structural Equation Modelling (SEM). CFA is crucial for the SEM measurement model to obtain the acceptable model fit before modeling the structural model. There are two techniques in CFA; individual CFA and Pooled-CFA. Usually, Pooled-CFA is done due to the high number of constructs and items. If the model is too complicated and has so many constructs and items, then it is recommended to perform Pooled-CFA to simplify the model's looks yet easy to understand. The perception of Malaysia Technical University Network (MTUN) academics on data sharing towards open data was analysed by using pooled-CFA. There are three main constructs: data sharing with its 4 sub-constructs; (technological factor, organizational factor, environmental factor, and individual factor), mediator construct (open data licenses), and open data construct was analyzed in this research. Furthermore, second-order constructs' factor loadings towards their corresponding sub-constructs were investigated. This research collected the primary data of 442 respondents using a stratified random sampling technique. This paper will explain the theoretical framework before revealing the results of Pooled-CFA on data sharing towards open data.

Keywords—Pooled CFA; data sharing; open data; measurement model; validity

I. INTRODUCTION

Open data initiatives have become ubiquitous in every country. According to [1], Malaysia has embarked on the open government data framework by The Malaysian Administrative Modernisation and Management Planning Unit (MAMPU) in the year 2015. The initiative is then leveraged to be implemented at the ministries' and agencies' levels. It seems crucial to have an open data framework within the higher education environment as [2] has mentioned that the higher institutions play a significant role and are among the most significant contributors that support the citizen's needs in the education world. In the meantime, [3] has stated that the data producer is reluctant to share data might because it possesses challenges at many levels such as cultural, ethical, financial, and technical. Adding to these challenges, [4] has highlighted that the reluctance of data sharing perhaps due to disinterest from the universities. Thus, this study employs quantitative techniques; survey to Malaysia Technical University Network (MTUN) academics. There were 442 feedbacks received and there was a need to perform Confirmatory Factor Analysis (CFA) to confirm the factor that influence MTUN academics on data sharing.

This research aims to identify the factor influence MTUN academics on data sharing and analysing the open data license, which will act as a mediator between data sharing and open data. This paper will explain in details the theoretical framework developed for this research, the component in structural equation modeling (SEM), the determination of sample size, the fitness indexes of technological, organizational, environmental, and individual construct that determine data sharing and how the procedure of pooled-CFA is done. The reliability and validity indexes also will be measured to indicate the acceptance state.

II. LITERATURE REVIEW

Open data might change the relationship between the government and the public in terms of transparency [5]. This intention can be perceived by accessing the government's data through an open format datasets form. Furthermore, [6] has emphasized that open data enactment will address the existing legal challenges. The challenges include the scope of accessing the data and data ownership.

Since open data has been announced, it has created a sensation worldwide. In [7], the statement is supported by highlighting the potential of open data to improve organization services to the public. Besides that, citizen participation is encouraged in open data towards having a transparent government. As in [8], the approach to embark on open data will be different as every country has a different governance structure, and the organization has its policies regarding open data.

Ministry of Education Malaysia (MOE) has specified that Malaysia's Higher Education Institutions (HEIs) are categorized as Public Universities, Private Higher Educational Institutions, Polytechnics, and Community Colleges [9]. As for this research's scope, MTUN universities include 4 public universities (UMP, UTEM, UTHM, and UNIMAP) where it focuses on the Technical and Vocational Education and Training (TVET) approaches.

In Malaysia, according to [9], 12 National Key Economic Areas (NKEA) have been identified under the government's Economic Transformation Programme (ETP). In [10], it is highlighted that the programme has demanded an additional 1.3 million TVET workers by 2020. This demand has strengthened the need to have MTUN open data framework as this initiative will help the potential worker make an informed decision from the data shared.

As in [11], the open data demand keeps growing in public universities. The demand ensues due to the open data capabilities of removing the barriers to reuse and redistributing the data. On the other hand, it will help the public to make informed decisions. As the results in [2], the economy's deception can be reduced, and universities' accountability can be expected while embarking on open data. To conclude, the data shared are valuable in creating innovations towards having a better university in the future.

This research endeavors to develop an open data framework for MTUN academics. Before that, the factors that influence data sharing, the roles of open data license as a mediator, and the indexes of open data components are measured. This paper will explain in detail the results of factors that influence data sharing towards open data by running pooled-CFA.

III. THEORETICAL FRAMEWORK

This study integrates the technological, organizational, and environmental (TOE) framework with the Theory of Planned Behavior (TPB) to determine factors that influence data sharing. According to [12], The TOE framework is an organization-level theory that explains organization structures from 3 perspectives. These 3 perspectives are technological, organizational, and environmental. These contexts were adopted and integrate with TPB theory that examines individuals' perspectives and were analysed as the factors that contribute to the data sharing.

The TPB theory has been useful and considered one of the most influential models in predicting social behaviors [13], [14]. The TOE framework and TPB theory integration were used to develop the MTUN open data theoretical framework as Fig. 1.

Fig. 1 shows how the technological, organizational, and environmental sub-constructs are derived from the TOE framework. Whereby the individual sub construct is derived from TPB theory. From the framework, it can be seen that all of these 4 sub-constructs (technological, organizational, environmental, and individual) are the factors that contribute to the data sharing construct. The open data licenses (ODL) will act as the mediator between data sharing and open data (MTUN_OD). This paper will explain how the pooled-CFA is conducted and how the constructs and items were analyzed by using IBM SPSS AMOS software (version 2.4).
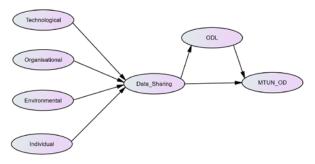

Fig. 1. MTUN Open Data Theoretical Framework.

IV. STRUCTURAL EQUATION MODELLING

Structural equation modeling (SEM) is a powerful, multivariate technique that has been used widely in scientific investigations to assess and evaluate multivariate causal relationships. According to [15], [16], sometimes, it is also called a statistical methodology where the confirmatory approach is used to analyse the structural theory.

There are 2 main components in SEM: measurement and structural models. These 2 components are used to examine variables in different ways. The measurement model section will relate the measured variables to the latent variables. On the other hand, the structural model section relates the latent variables to one another. SEM combines 2 statistical methods; confirmatory factor analysis (CFA) and path analysis. There are 2 CFA techniques: Individual CFA and Pooled-CFA. This paper focused on the SEM measurement model, assessed through the CFA.

The first step to run pooled-CFA was having to perform the individual CFA for each construct. In [17], [18] has mentioned that the second-order construct was validated using the CFA procedure separately before it's been simplified into first-order constructs to reduce the model's complexity. As [19] suggested, the pooled-CFA for all constructs was important to perform to assess the discriminant validity among the model's constructs. Thus, this study analyzed the feedback from the MTUN academics survey of data sharing towards open data.

There was 1 main construct (second-order construct) involved in this study: data sharing with its sub-constructs: technological factor, organizational factor, environmental factor, and individual factor. According to [20], that second-order CFA is employed in this study as it involved the assessment of a second-order variable's factor loadings towards its corresponding sub-constructs. By running a second-order CFA, the relationships of data sharing towards its sub-constructs were examined as well. The ODL construct and MTUN_OD construct were identified as first-order constructs as they do not have the sub-construct and was analyzed directly without the need to simplifying it anymore.

V. MATERIAL AND METHOD

A. Preface

This study's population target covers MTUN academics from various educational backgrounds and working experiences. Based on [21], the total population of MTUN academics in 2018 was 3818. According to [22], as Table I, the sample needed for this study was 351 for the population of 4000.

As per shown in Table I, the population sample obtained was 442, which was higher than the number of samples required. A total of 442 respondents were chosen randomly. The comprehensive questionnaire for the field study was constructed which was derived from an exploratory factor analysis (EFA). The EFA was executed by using IBM SPSS software. The data collection for the field study is done by distributing the questionnaire to MTUN academics using a stratified random sampling technique.

TABLE I.   DETERMINING SAMPLE SIZE FOR A FINITE POPULATION

N	S	N	S	N	S
10	10	220	140	1200	291
15	14	230	144	1300	297
20	19	240	148	1400	302
25	24	250	152	1500	306
30	28	260	155	1600	310
35	32	270	159	1700	313
40	36	280	162	1800	317
45	40	290	165	1900	320
50	44	300	169	2000	322
55	48	320	175	2200	327
60	52	340	181	2400	331
65	56	360	186	2600	335
70	59	380	191	2800	338
75	63	400	196	3000	341
80	66	420	201	3500	346
85	70	440	205	4000	351
90	73	460	210	4500	354
95	76	480	214	5000	357

N = number of population

S = Sample Size

IBM statistical package for social science (SPSS) and IBM SPSS analysis of moment structures (AMOS) version 24.0 were used to build and analyze the model in this study.

B. Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) is a method of factor analysis, most commonly used in social research. It is usually used to examine the consistency of a construct with a researcher's understanding of that construct's factor. CFA's objective is to examine whether the data fit a hypothesized measurement model. This hypothesized model is based on theory or previous analytic research. In CFA, several things need to be tested: reliability, validity, and unidimensionality of the measurement model. The results must meet the stated requirement before modeling the structural model. According to [18], [23], the theorized model must pass 3 types of validities: Construct Validity, Convergent Validity, and Discriminant Validity. The details of validity and reliability indexes are shown in Table II.

Table II shows the validity and reliability test that need to be passed. The construct validity is assessed through the fitness indexes of the measurement model. The convergent validity is assessed by computing the Average Variance Extracted (AVE). The Discriminant Validity is evaluated by developing the Discriminant Validity Index Summary.

Adding to this, several fitness indexes need to be examined as well to evaluate the model fitness. Absolute fit, incremental fit, and parsimonious fit are three types of model fit categories. Below are the fitness of indexes as shown in Table III.

As shown in Table III, [24] has mentioned that the names of indexes that are frequently reported in many research are

Root Mean Square Error Approx (RMSEA), Comparative Fit Index (CFI), and Chi-square/degrees of freedom (Chisq/df).

C. Discriminant Validity

The discriminant validity needs to be assessed to ensure no construct redundancy occurs in the model. Construct redundancy might occur when any pair of constructs in the model are highly correlated. This redundancy also can happen when one or more constructs assess the same variable. In other words, discriminant validity tests whether the concepts of measurements that are not supposed to be related are unrelated. According to [24], if the redundancy occurs, that particular redundant items in a model need to be deleted. The deletion should start from the lowest value of factor loading until the model is fit.

Besides that, correlation coefficients are used to measure the strength of the relationship between 2 variables. As mentioned in [25], it is also acted as evidence of discriminant validity. A correlation between variables indicates that if one variable changes in value, the other variable tends to change in a specific direction. The variables should not be highly correlated to each other, or else the multi-collinearity problem will exist. Besides, [24] has highlighted that the correlation value among the exogenous variables should not exceed 0.85 to achieve the variables' discriminant validity.

D. Summary

There are 2 techniques of CFA in SEM's measurement model: Individual CFA and Pooled-CFA. Individual CFA runs each unobserved construct in the research individually; whereas Pooled-CFA runs all construct simultaneously [26]. Before performing Pooled-CFA, the individual CFA for all constructs need to be done separately. The results must achieve the indexes' fitness as Table II and Table III to make them reliable and validated. The AVE's results were recalculated to get the mean score and were used in Pooled-CFA.

TABLE II.   VALIDITY AND RELIABILITY INDEXES

Name of Category	Name of Index	Level of Acceptance	Literature
Convergent Validity	Average Variance Extracted	$AVE \geq 0.5$	Zainudin (2015)
Internal Reliability	Cronbach Alpha	$A \geq 0.5$	Zainudin (2015)
Construct Reliability	Composite Reliability	$CR \geq 0.6$	Zainudin (2015)

TABLE III.   FITNESS OF INDEXES

Name of category	Name of index	Level of acceptance
Absolute Fit Index	RMSEA	$RMSEA < 0.1$
	GFI	$GFI > 0.90$
Incremental Fit Index	AGFI	$AGFI > 0.90$
	CFI	$CFI > 0.90$
	TLI	$TLI > 0.90$
	NFI	$NFI > 0.90$
Parsimonious Fit Index	Chi-sq/df	Chi-Square/ df < 5.0

VI. RESULTS AND DISCUSSIONS

A. Individual CFA

The analysis started with performing Individual CFA. It ran the latent construct one after another to achieve the required model fitness. The CFA can only be performed if the constructs have more than 3 items with no model identification problem. Fig. 2 shows that all these 4 constructs (technological, organizational, environmental, and individual) have met the initial requirement to run CFA. All of the constructs must achieve the fitness indexes required.

Fig. 2. The CFA Results for Technological Factor Construct.

Fig. 2 shows that the technological factor construct has 3 components; technical infrastructure (4 items), usability (3 items), and standard (10 items). The model fitness of the technological factor construct was overall met the fitness indexes. The value for RMSEA shown was .067, the CFI was .959, and Chisq/df was 2.989.

For the Convergent Validity (CV) assessment, the study needs to calculate the AVE. According to [19], [26], the construct achieved the CV if its AVE exceeds the threshold value of 0.5. Besides in [24], there was a need to compute the CR, and the value should exceed the threshold value of 0.6 for this reliability to achieve. The AVE and CR for the primary constructs and their respective components were computed and presented in Table IV.

Table IV shows that each item's factor loading was high, which above 0.6. The CR value for the technological factor was 0.948, and AVE was 0.859. Meanwhile, the CR value for technical infrastructure was 0.921 and AVE was 0.745. In addition to that, the CR value for usability was 0.754 and AVE was 0.506. Meanwhile, the CR value for a standard was 0.940, and AVE was 0.610.

From these results, we can conclude that technological factors construct together with its components and items have met the CR's requirement, which must above 0.6, and AVE, which must above 5.0. Fig. 3 shows the CFA results for the organizational factor construct.

The organizational factor construct has 4 components; norms (10 items), data sharing policy (3 items), governance (3 items), and resources (5 items). The model fitness of organizational factor constructs was overall met the fitness indexes. The value for RMSEA shown was .067, the CFI was .945, and Chisq/df was 2.988. Table V shows the AVE and CR for the organizational factor construct.

TABLE IV. THE AVE AND CR FOR TECHNOLOGICAL CONSTRUCT

Construct	Items	Factor Loading	CR (>0.6)	AVE (>0.5)
Technological Factor	Technical Infrastructure	0.95	0.948	0.859
	Usability	0.94		
	Standard	0.89		
Technical Infrastructure	TFTI1	0.90	0.921	0.745
	TFTI2	0.88		
	TFTI3	0.86		
	TFTI4	0.81		
Usability	TFU21	0.69	0.754	0.506
	TFU22	0.67		
	TFU23	0.77		
Standard	TFS31	0.67	0.940	0.610
	TFS32	0.70		
	TFS33	0.82		
	TFS34	0.83		
	TFS35	0.77		
	TFS36	0.80		
	TFS37	0.78		
	TFS38	0.76		
	TFS39	0.79		
	TFS310	0.87		

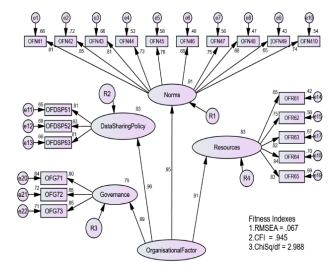Fig. 3. The CFA Results for Organisational Factor Construct.

TABLE V. THE AVE AND CR FOR ORGANISATIONAL CONSTRUCT

Construct	Items	Factor Loading	CR (>0.6)	AVE (>0.5)
Organisational Factor	Norms	0.95	0.961	0.861
	Data sharing policy	0.96		
	Governance	0.89		
	Resources	0.91		
Norms	OFN41	0.81	0.927	0.562
	OFN42	0.85		
	OFN43	0.81		
	OFN44	0.73		
	OFN45	0.76		
	OFN46	0.69		
	OFN47	0.75		
	OFN48	0.68		
	OFN49	0.65		
	OFN410	0.74		
Data sharing policy	OFDSP51	0.81	0.857	0.667
	OFDSP52	0.83		
	OFDSP53	0.81		
Resources	OFR61	0.65	0.886	0.610
	OFR62	0.75		
	OFR63	0.82		
	OFR64	0.84		
	OFR65	0.83		
Governance	OFG71	0.80	0.872	0.695
	OFG72	0.85		
	OFG73	0.85		

Based on Table V, it can be concluded that the CR value for the organizational factor was 0.961, and AVE was 0.861. The CR value for norms was 0.927, and AVE was 0.562. The CR value for the data sharing policy was 0.857, and AVE was 0.667. Meanwhile, The CR value for resources was 0.886, and AVE was 0.610. Finally, the CR value for governance was 0.872 and AVE was 0.695.

From these results, it can be concluded that the organizational factor constructs and their components and items have met CR's requirement, which must above 0.6, and AVE, which must above 5.0. Fig. 4 shows the CFA results for the environmental factor construct.

The environmental factor construct has 2 components; data sharing culture (3 items) and research practice (3 items). The model fitness of the environmental factor construct was overall meet the fitness indexes. The value for RMSEA shown was .066, the CFI was .990, and Chisq/df was 2.925. Table VI shows the AVE and CR for the environmental factor construct.
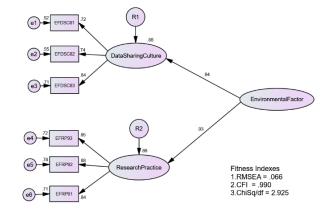

Fig. 4. The CFA Results for Environmental Factor Construct.

TABLE VI. THE AVE AND CR FOR ENVIRONMENTAL CONSTRUCT

Construct	Items	Factor Loading	CR (>0.6)	AVE (>0.5)
Environmental Factor	Data sharing culture	0.94	0.933	0.874
	Research Practice	0.93		
Data sharing culture	EFDSC81	0.72	0.812	0.591
	EFDSC82	0.74		
	EFDSC83	0.84		
Research practice	EFRP91	0.84	0.892	0.734
	EFRP92	0.88		
	EFRP93	0.85		

Table VI shows that each item's factor loading was high, which above 0.6. The CR value for the environmental factor was 0.933, and AVE was 0.874. The CR value for data sharing culture was 0.812, and AVE was 0.591. The CR value for research practice was 0.892, and AVE was 0.734. It can be concluded from these results that the environmental factor construct and its components and items have met CR requirements, which must above 0.6 and AVE must above 5.0. Fig. 5 shows the CFA results for the individual factor construct.
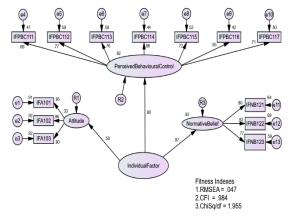

Fig. 5. The CFA Results for Individual Factor Construct.

The individual factor construct has 3 components, which are attitude (3 items), perceived behavioral control (7 items), and normative belief (3 items). The model fitness of the individual factor construct was overall met the fitness indexes. The value for RMSEA shown was .047, the CFI was .984, and Chisq/df was 1.955. Table VII shows the AVE and CR for the individual factor construct.

Table VII shows each item's factor loading was high above 0.6. The CR value for the individual factor was 0.868, and AVE was 0.696. Meanwhile, the CR value for attitude was 0.956, and AVE was 0.878. In addition to that, the CR value for perceived behavioral control was 0.875 and AVE was 0.502. The CR value for normative belief was 0.842, and AVE was 0.641. From these results, we can conclude that individual factors construct together with its components and items have met CR's requirement, which must above 0.6, and AVE, which must above 5.0. Fig. 6 shows the CFA results for the open data license factor construct.

In Fig. 6, the ODL construct has 5 items. Thus, the model fitness of the ODL construct was overall met the fitness indexes. The value for RMSEA shown was .053, the CFI was .995, and Chisq/df was 2.257. Table VIII shows the AVE and CR for ODL construct.

Table VIII shows that each item's factor loading was high, which above 0.6. The CR value for ODL was 0.898, and AVE was 0.639. Thus, from these results, we can conclude that ODL construct and items have met CR requirements that must above 0.6 and AVE, which must above 5.0. Fig. 7 shows the CFA results for the open data (MTUN_OD) construct.

TABLE VII.   THE AVE AND CR FOR INDIVIDUAL FACTOR CONSTRUCT

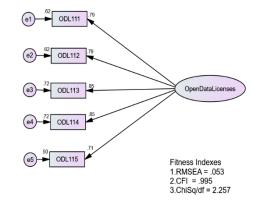| Construct	Items	Factor Loading	CR (>0.6)	AVE (>0.5)
Individual Factor	Attitude	0.58	0.868	0.696
	Perceived Behavioral Control	0.90		
	Normative Belief	0.97		
Attitude	IFA101	0.95	0.956	0.878
	IFA102	0.96		
	IFA103	0.90		
Perceived Behavioral Control	IFPBC111	0.68	0.875	0.502
	IFPBC112	0.77		
	IFPBC113	0.76		
	IFPBC114	0.66		
	IFPBC115	0.72		
	IFPBC116	0.65		
	IFPBC117	0.71		
Normative Belief	IFNB121	0.80	0.842	0.641
	IFNB122	0.83		
	IFNB123	0.77		

Fig. 6.   The CFA Results for Open Data Licenses Factor Construct.

TABLE VIII.   THE AVE AND CR FOR OPEN DATA LICENSES CONSTRUCT

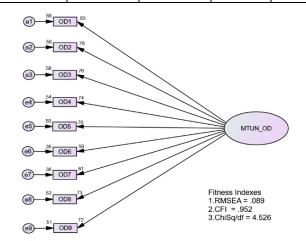| Construct	Items	Factor Loading	CR (>0.6)	AVE (>0.5)
Open Data Licenses	ODL111	0.79	0.898	0.639
	ODL112	0.79		
	ODL113	0.85		
	ODL114	0.85		
	ODL115	0.71		

Fig. 7.   The CFA Results for Open Data Construct.

In Fig. 7, the MTUN_OD construct has 9 items. The model fitness of the MTUN_OD construct was overall met the fitness indexes. The value for RMSEA shown was .089, the CFI was .952, and Chisq/df was 4.526. Table IX shows the AVE and CR for the open data construct.

Table IX shows each item's factor loading was high, above 0.6. The CR value for the open data construct was 0.909, and AVE was 0.528. Thus, from this result, we can conclude that MTUN_OD constructs and their items have met CR's requirement, which above 0.6, and AVE must above 5.0.

An overall, the technological construct, organizational construct, environmental construct, individual construct, ODL construct, and MTUN_OD construct has met the fitness indexes and passed the measurement of AVE and CR.

TABLE IX. THE AVE AND CR FOR OPEN DATA CONSTRUCT

Construct	Items	Factor Loading	CR (>0.6)	AVE (>0.5)
MTUN_OD	OD1	0.83	0.909	0.528
	OD2	0.76		
	OD3	0.77		
	OD4	0.75		
	OD5	0.71		
	OD6	0.60		
	OD7	0.63		
	OD8	0.74		
	OD9	0.72		

The study needed to simplify the overall measurement model from the first-order construct and pool them together to undergo the CFA procedure at once. This procedure is called Pooled-CFA.

B. Pooled-CFA for all Measurement Model of Constructs

In this pooled-CFA model, as suggested in [24], the measurement model for the second-order constructs was validated using the CFA procedure separately and simplified into first-order constructs to reduce complexity. The reason to perform the pooled-CFA was to assess the discriminant validity among constructs in the model [17], [19], [23], [26]. Fig. 8 shows the pooled-CFA that consists of data sharing (DS) construct, ODL, and MTUN_OD in 1 model.

The result of the analysis shown 3 types of values; fitness indexes for all constructs in the model, the factor loading for every component to the main construct, and the correlation between constructs.

In determining the fitness indexes, the values should meet the threshold as shown in Table III. As in [17], [19], [23], [26] have highlighted that the factor loading for every item should not less than 0.6 and the correlation coefficient of any two constructs should not exceed 0.85. The multicollinearity problem will occur if the correlation between any two constructs exceeds 0.85. In this study, none of the values found to be greater than 0.85. Thus, the multicollinearity problem does not arise.
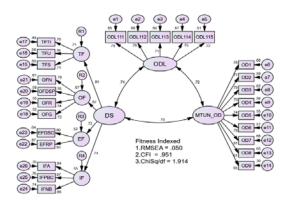

Fig. 8. The 3 Constructs are Pooled Together for the Pooled-CFA Procedure.

The Pooled CFA has merged 3 constructs. From Fig. 8, the model looks much more straightforward and easy to understand. The pooled CFA was also performed to avoid violating regression assumptions. The correlation between DS to ODL was 0.74. Then, the correlation between ODL to MTUN_OD was 0.72, and the correlation between DS to MTUN_OD was 0.70. Thus, no multicollinearity occurs as the correlation between each construct was below 0.85. Besides, pooled CFA's model fitness was overall met the fitness indexes. The value for RMSEA shown was .050, the CFI was .951, and Chisq/df was 1.914. Table X shows the AVE and CR for Pooled-CFA.

TABLE X. THE AVE AND CR FOR POOLED-CFA

Construct	Items	Factor Loading	CR (>0.6)	AVE (>0.5)
Data Sharing	Technological Factor	0.84	0.862	0.611
	Organizational Factor	0.82		
	Environmental Factor	0.72		
	Individual Factor	0.74		
Technological Factor	Technical Infrastructure	0.70	0.767	0.524
	Usability	0.76		
	Standard	0.71		
Organisational Factor	Norms	0.78	0.837	0.563
	Data Sharing Policy	0.76		
	Resources	0.74		
	Governance	0.72		
Environmental Factor	Data Sharing Culture	0.80	0.858	0.752
	Research Practice	0.93		
Individual Factor	Attitude	0.84	0.850	0.655
	Perceived Behavioural Control	0.87		
	Normative Beliefs	0.86		
ODL	ODL111	0.74	0.855	0.542
	ODL112	0.76		
	ODL113	0.71		
	ODL114	0.75		
	ODL115	0.72		
MTUN_OD	OD1	0.78	0.922	0.570
	OD2	0.77		
	OD3	0.79		
	OD4	0.75		
	OD5	0.74		
	OD6	0.76		
	OD7	0.81		
	OD8	0.81		
	OD9	0.55		

Based on the AVE and CR values in Table X, the study found that all AVE and CR exceed their threshold values of 0.5 and 0.6 respectively. Thus, the study can conclude that the AVE and CR for all latent constructs in the model have been achieved.

Proceed to the next validity test, the study needs to measure discriminant validity. The table discriminant validity index summary is developed as shown in Table XI. The diagonal values in bold were the square root of the AVE of the respective constructs while the other values indicate the correlation coefficient between the pair of the individual constructs.

TABLE XI. THE DISCRIMINANT VALIDITY INDEX SUMMARY FOR ALL CONSTRUCTS

Construct	Data Sharing	ODL	MTUN_OD
Data Sharing	**0.782**		
ODL	0.74	**0.740**	
MTUN_OD	0.70	0.720	**0.755**

TABLE XII. THE ASSESSMENT OF NORMALITY FOR ALL

Variables	Skew	CR	Kurtosis	CR
IFA	-.759	-5.886	1.261	4.892
IFPBC	-.616	-4.774	1.022	3.964
IFNB	-.809	-6.276	1.309	5.076
EFDSC	-.326	-2.531	-.213	-.827
EFRP	-.528	-4.096	.541	2.099
OFN	-.381	-2.958	.106	.412
OFDSP	-.429	-3.328	-.071	-.275
OFR	-.061	-.470	-.254	-.986
OFG	-.188	-1.460	-.339	-1.316
TFTI	-.247	-1.914	-.009	-.036
TFU	-.066	-.513	-.420	-1.628
TFS	.024	.188	-.144	-.560
OD9	-.483	-3.744	.690	2.676
OD8	-.353	-2.737	.201	.779
OD7	-.390	-3.022	-.052	-.202
OD6	-.154	-1.196	-.189	-.731
OD5	-.160	-1.242	-.054	-.209
OD4	-.069	-.532	-.422	-1.637
OD3	-.143	-1.113	-.288	-1.115
OD2	-.080	-.622	-.252	-.977
OD1	-.375	-2.907	.280	1.085
ODL115	-.115	-.889	-.158	-.611
ODL114	-.184	-1.424	.009	.035
ODL113	-.009	-.071	-.239	-.928
ODL112	-.265	-2.057	-.099	-.385
ODL111	-.035	-.272	-.285	-1.106
Multivariate			57.862	14.406

Table XI shows the discriminant validity index summary for all constructs. The discriminant validity has been achieved when the diagonal values (in bold) are higher than any other values in its row and column. Since SEM employs the parametric statistical approach of modeling, the study needs to assess all items' normality distribution measuring their respective constructs. According to [17], [19], [23], [24], [26], the value of skewness should fall within the range of -1.5 to 1.5 to make it normally distributed.

Table XII shows the values of skewness for all components in the model fell within the range between -1.5 and 1.5. It means that the distribution does not depart from normality and there were no outliers' data. Thus, the data distribution meets the normality distribution requirement for employing parametric statistical analysis in SEM.

VII. CONCLUSION

Data sharing in this study that was defined through the combination of technological, organizational, environmental, and individual components. The components were derived from the literature review. However, in this study, the exact components that form data sharing were investigated through the process of survey distribution to MTUN academics that were then be confirmed through CFA. The investigations were then extended to the ODL construct and MTUN_OD construct. As for this research purposes, this study examined the factor influence data sharing and the impact of data sharing on ODL construct and MTUN-OD construct.

In a conclusion, Pooled-CFA is recommended to perform on a complicated model in making it simpler to analyse and easy to understand. The model is considered complicated when it involves many second-order constructs and items. There are 3 important types of validities in this study; CR, Cronbach Alpha, and AVE. The CR is important in this study as its measure of internal consistency in scale items. On the other hand, the AVE is important to employ in this study to confirm that the construct should correlate with related variables but it should not correlate with dissimilar, unrelated ones. In determining the value of CR and AVE for each construct, the analysis results of Pooled-CFA are recalled and it can be concluded that 4 components influence data sharing; technological, organizational, environmental, individual construct.

The technical factor has a CR value of 0.948; which above the minimum accepted value of CR; 0.6 and 0.859; which above the minimum accepted value for AVE; 0.5. The fitness indexed for this construct was achieved with the value of RMSEA was 0.067, which less than 0.1 to make it accepted. CFI was 0.959, which above 0.9 to make it accepted and Chi-sq was 2.989, which less than 5.0 to make it accepted. Meanwhile, the result of organizational factors shown the CR value of 0.961 and 0.861 for AVE. The fitness indexed for this construct was achieved with the value of RMSEA was 0.067, which less than 0.1 to make it accepted. CFI was 0.945, which above 0.9 to make it accepted and Chi-sq was 2.988, which less than 5.0 to make it accepted.

On the other hand, the environmental factor has a CR value of 0.933; and 0.874 for AVE. The fitness indexed for this

construct was achieved with the value of RMSEA was 0.066, which less than 0.1 to make it accepted. CFI was 0.990, which above 0.9 to make it accepted and Chi-sq was 2.925, which less than 5.0 to make it accepted. In the meantime, the result of the individual factor shown the CR value of 0.868; and 0.696 for AVE. The fitness indexed for this construct was achieved with the value of RMSEA was 0.047, which less than 0.1 to make it accepted. CFI was 0.984, which above 0.9 to make it accepted and Chi-sq was 1.955, which less than 5.0 to make it accepted.

Besides that, the ODL construct has a CR value of 0.898 and 0.639 for AVE. The fitness indexed for this construct was achieved with the value of RMSEA was 0.053, which less than 0.1 to make it accepted. CFI was 0.995, which above 0.9 to make it accepted and Chi-sq was 2.257, which less than 5.0 to make it accepted.

Furthermore, the result of the MTUN_OD construct shown the CR value of 0.909 and 0.528 for AVE. The fitness indexed for this construct was achieved with the value of RMSEA was 0.067, which less than 0.1 to make it accepted. CFI was 0. 952, which above 0.9 to make it accepted and Chi-sq was 4.526, which less than 5.0 to make it accepted.

An overall, the technological construct, organizational construct, environmental construct, individual construct, ODL construct, and MTUN_OD construct for CFA distinctively has met the fitness indexes and passed the measurement of AVE and CR.

To ensure the overall fitness indexes of the model, this study employed Pooled-CFA. From the result of pooled CFA, the fitness indexed for this overall model was achieved with the value of RMSEA was 0.050, which less than 0.1 to make it accepted. CFI was 0.951, which above 0.9 to make it accepted and Chi-sq was 1.914, which less than 5.0 to make it accepted.

Based on Table XI, it can be shown that the model achieved discriminant validity when the diagonal values (in bold) are higher than any other values in its row and column. As stated, the data sharing value for discriminant validity was 0.782, ODL was 0.740, and MTUN-OD was 0.755. It indicates that each of the constructs was measure distinctively and not related to each other.

Finally, for the normality test, the distribution does not depart from normality and there were no outliers' data. Thus, the data distribution meets the normality distribution requirement for employing parametric statistical analysis in SEM which will be discussed in the next paper.

In a conclusion, the step-by-step to do pooled-CFA must start with performing an individual CFA for every constructs to make it simpler and easy to understand for the complicated model. All the results must follow the table of indexes (Table II and Table III) to indicate that the results are reliable and validated.

These results of CFA will be used to be modeled in SEM. However, for future work, it is advisable to add 1 more component to be measured; data quality which should be determined under technological construct.

REFERENCES

[1] S. N. Ismael, O. Mohd, and Y. A. Rahim, "Implementation of open data in higher education: A review," J. Eng. Sci. Technol., vol. 13, no. 11, pp. 3489–3499, 2018.

[2] M. Krumova, "Higher Education 2 . 0 and Open Data : a Framework for University Openness and Co-creation Performance," pp. 562–563, 2017.

[3] A. S. Figueiredo, "Data Sharing: Convert Challenges into Opportunities," Front. Public Heal., vol. 5, no. December, pp. 1–6, 2017.

[4] K. Shamash, J. P. Alperin, and A. Bordini, "Teaching Data Analysis in the Social Sciences: A case study with article level metrics," Open Data as Open Educ. Resour. Case Stud. Emerg. Pract., pp. 50–56, 2015.

[5] Reale, Giuseppe, "Opportunities and Differences of Open Government Data Policies in Europe," Athens J. Soc. Sci., vol. Volume 1, no. Number 3, pp. 195–206, 2014.

[6] E. Tran and G. Scholtes, "Open Data Literature Review," 19th Annu. BCLT/BTLJ Symp. Open Data Addressing Privacy, Secur. Civ. Rights Challenges 2, pp. 1–33, 2015.

[7] J. Manyika, M. Chui, P. Groves, D. Farrell, S. Van Kuiken, and E. A. Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information," McKinsey, no. October, p. 24, 2013.

[8] M. W. AlRushaid and A. K. J. Saudagar, "Measuring the Data Openness for the Open Data in Saudi Arabia e-Government - A Case Study," Int. J. Adv. Comput. Sci. Appl., vol. 7, no. 12, pp. 113–122, 2016.

[9] Mi. of E. Malaysia, Malaysia Education Blueprint 2015-2025 (Higher Education). 2015.

[10] StudyMalaysia.com, "TVET in Malaysia." [Online]. Available: Technical and vocational education and training (TVET) in Malaysia - StudyMalaysia.com.

[11] X. Shacklock, From Bricks to clicks. The Potential of Data and Analytics in Higher Education. 2016.

[12] C.-Y. Chiu, S. Chen, and C.-L. Chen, "An integrated perspective of TOE framework and innovation diffusion in broadband mobile applications adoption by enterprises," Int. J. Manag. Econ. Soc. Sci., vol. 6, no. 1, pp. 14–39, 2017.

[13] L. Chen and X. Yang, "Using EPPM to Evaluate the Effectiveness of Fear Appeal Messages Across Different Media Outlets to Increase the Intention of Breast Self-Examination Among Chinese Women," Health Commun., vol. 34, no. 11, pp. 1369–1376, 2019.

[14] I. Ajzen, "The Theory of Planned Behavior," pp. 179–211, 1991.

[15] S. Bauldry, "Structural Equation Modeling," Int. Encycl. Soc. Behav. Sci. Second Ed., pp. 615–620, 2015.

[16] S. Abrahim, B. A. Mir, H. Suhara, F. A. Mohamed, and M. Sato, "Structural equation modeling and confirmatory factor analysis of social media use and education," Int. J. Educ. Technol. High. Educ., vol. 16, no. 1, 2019.

[17] Z. Awang, U. Sultan, and Z. Abidin, "Modeling and Analyzing Second Order Model in Structural Equation Modeling The Second Order Confirmatory Factor Analysis (CFA)," no. August 2015, 2016.

[18] Z. Awang, A. Afthanorhan, M. Mamat, U. Sultan, and Z. Abidin, "The Likert scale analysis using parametric based Structural Equation Modeling (SEM)," Comput. Methods Soc. Sci., vol. 4, no. 1, pp. 13–21, 2016.

[19] Jawaria Nasir, Rashidah Mohamad Ibrahim, M. A. Sarwar, and Raja Irfan Sabir, "Impact Of High Involvement Work Practices On Employee

Performances In Health Sector, Pakistan," J. Manag. Theory Pract., vol. 1, no. 2, pp. 22–32, 2020.

[20] I. Gede Mahatma Yuda Bakti and S. Sumaedi, "An analysis of library customer loyalty:The role of service quality and customer satisfaction, a case study in Indonesia," Libr. Manag., vol. 34, no. 6–7, pp. 397–414, 2013.

[21] MOE, "Statistik Pendidikan Tinggi 2018 : Public Universities," pp. 8–45, 2018.

[22] R. Walker, "Another round of globalization in San Francisco," Urban Geogr., vol. 17, no. 1, pp. 60–94, 1996.

[23] Z. Awang, A. Afthanorhan, M. Mohamad, and M. A. M. Asri, "An evaluation of measurement model for medical tourism research: The confirmatory factor analysis approach," Int. J. Tour. Policy, vol. 6, no. 1, pp. 29–45, 2015.

[24] Awang, "Overview of Structural Equation Modeling (SEM)," A Handb. SEM, pp. 1–17, 2012.

[25] N. Ahmad and A. Sabri, "Assessing the unidimensionality, reliability, validity and fitness of influential factors of 8th grades student's mathematics achievement in Malaysia," Int. J. Adv. Res., vol. 1, no. 2, pp. 1–7, 2013.

[26] W. Mohamad, A. Bin, and W. Afthanorhan, "International Journal of Asian Social Science Pooled Confirmatory Factor Using Structural Equation Modeling On Volunteerism Program : A Step By Step Approach Sabri Ahmad Ibrahim Mamat Contribution / Originality," vol. 4, no. 5, pp. 642–653, 2014.