

Analysis Report

This report is structured as follow.

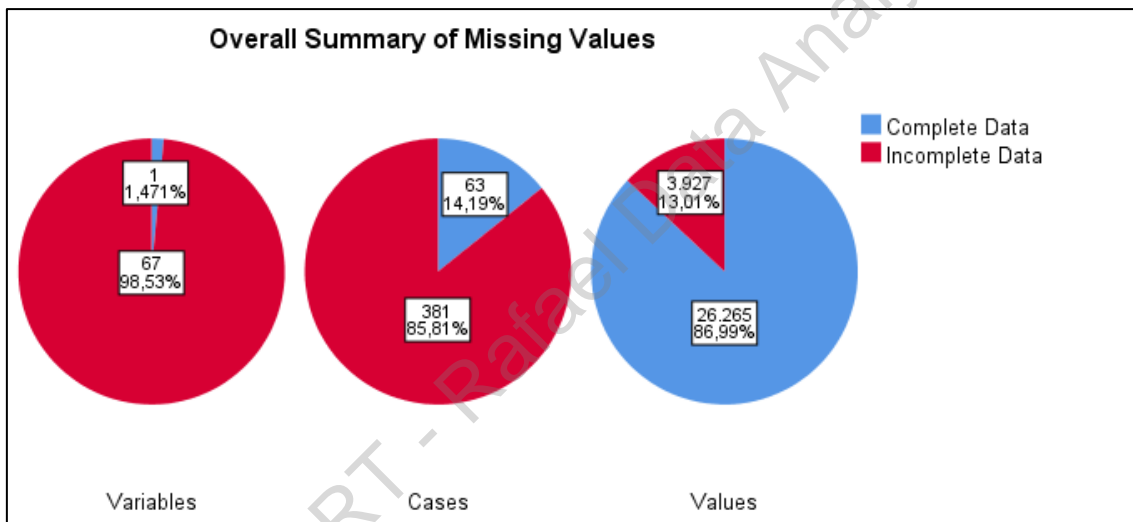
Contents

Missing Value analysis	2
Exploratory Factor Analysis.....	4
Dimension 1	5
Dimension 2.....	7
Confirmatory Factor Analysis	8
Dimension 1	10
Convergent Validity and Reliability	10
Discriminant Validity	11
Dimension 2	13
Convergent Validity and Reliability	13
Discriminant Validity	13
References	14

The report begins with the analysis of missing data, which is followed by the results of an Exploratory Factor Analysis and a subsequent Confirmatory Factor Analysis.

Missing Value analysis

This section aims at analysing the amount of missing data present in the sample. The figure below shows a summary of missing values. 381 participants (85.81%) showed blank responses to at least one variable. 67 out of the 68 variables had incomplete answers in at least one participant. In terms of all the cells present in the data, 86.99% of them were filled (figure below).



The table below shows the amount of missing data per variable. The variable with the greatest number of missing cases was 'ID.1369_REVERSE' (45.7% of missing cases), which is quite substantial (Fowler, 2009).

Univariate Statistics

	N	Mean	Std. Deviation	Missing	Percent
Id.1294_REVERSE	437	3.93	0.867	7	1.6
Id.1295	421	4.08	0.723	23	5.2
Id.1296	418	3.63	1.148	26	5.9
Id.1297	405	4.14	0.770	39	8.8
Id.1298	335	3.82	0.847	109	24.5
Id.1299	392	2.82	1.098	52	11.7
Id.1300	325	3.42	0.915	119	26.8
Id.1301	304	3.39	0.853	140	31.5
Id.1302_REVERSE	335	3.24	0.910	109	24.5
Id.1303_REVERSE	307	3.32	0.856	137	30.9
Id.1305	429	3.75	0.956	15	3.4
Id.1306	400	3.89	0.760	44	9.9
Id.1307	436	4.28	0.771	8	1.8
Id.1308	401	3.67	0.925	43	9.7
Id.1309	399	3.47	0.915	45	10.1
Id.1310_REVERSE	352	2.66	0.944	92	20.7
Id.1312	421	3.89	0.958	23	5.2
Id.1313	333	3.85	0.757	111	25.0
Id.1314	407	3.78	0.813	37	8.3
Id.1315	418	4.34	0.622	26	5.9
Id.1316	422	4.43	0.634	22	5.0
Id.1317	420	4.19	0.752	24	5.4
Id.1318	382	3.78	0.825	62	14.0
Id.1319	402	4.06	0.861	42	9.5
Id.1320	388	4.09	0.835	56	12.6
Id.1321	393	4.26	0.753	51	11.5
Id.1322	312	3.81	0.851	132	29.7
Id.1323	316	3.97	0.786	128	28.8
Id.1324	275	3.64	0.747	169	38.1
Id.1325	286	3.55	0.708	158	35.6
Id.1326	258	3.43	0.715	186	41.9
Id.1327	403	3.78	0.773	41	9.2
Id.1328	365	3.63	0.820	79	17.8
Id.1330	418	3.97	0.866	26	5.9
Id.1331	408	4.24	0.775	36	8.1
Id.1332	390	4.03	0.664	54	12.2
Id.1335	438	4.11	0.746	6	1.4
Id.1336	417	3.68	0.914	27	6.1
Id.1337	363	3.58	0.877	81	18.2
Id.1338_REVERSE	395	3.33	0.994	49	11.0
Id.1339	376	3.34	0.956	68	15.3

Id.1340	384	3.18	0.961	60	13.5
Id.1341	366	3.34	0.957	78	17.6
Id.1342	409	3.05	1.054	35	7.9
Id.1353	433	4.03	0.719	11	2.5
Id.1354	444	4.27	0.703	0	0.0
Id.1355_REVERSE	422	4.08	0.782	22	5.0
Id.1356_REVERSE	409	4.04	0.783	35	7.9
Id.1358	386	4.29	0.680	58	13.1
Id.1359	381	4.29	0.645	63	14.2
Id.1360	386	4.26	0.784	58	13.1
Id.1361	402	3.27	1.492	42	9.5
Id.1363	435	4.20	0.838	9	2.0
Id.1364	429	4.42	0.778	15	3.4
Id.1365	429	3.91	0.998	15	3.4
Id.1366	302	4.32	0.807	142	32.0
Id.1367	430	4.15	0.724	14	3.2
Id.1368	323	4.28	0.742	121	27.3
Id.1369_REVERSE	241	4.37	0.857	203	45.7
Id.1370_REVERSE	335	4.36	0.775	109	24.5
Id.1372	428	4.13	0.707	16	3.6
Id.1373	427	4.04	0.751	17	3.8
Id.1374	412	3.73	0.925	32	7.2
Id.1375	428	4.45	0.604	16	3.6
Id.1376	432	4.50	0.558	12	2.7
Id.1377	433	4.37	0.644	11	2.5
Id.1378_REVERSE	418	4.61	0.525	26	5.9
Id.1379	439	4.77	0.507	5	1.1

In order to perform EFA and CFA with a complete dataset, which is a requirement for the Maximum-Likelihood method, the Expectation-Maximization algorithm was performed to impute values on the missing cases (SPSS v26). After this technique was employed, the resulting sample size for all variables was 444 (N = 444).

Exploratory Factor Analysis

Factor analysis (FA) is a technique used to identify underlying factors present in the pattern of correlations among a set of measures. Where there is a large set of measures, factor analysis can determine whether there are subsets of items forming separate scales (Blaikie, 2008). This procedure can yield very useful results, making a further analysis more profound and easier to interpret. What should be noted, however, is that the

technique makes no reference to the conceptual meaning of a factor. This should be assessed by the researcher when looking at the empirical associations given by FA (Babbie, 1990).

The Maximum-Likelihood method was performed on each dimension under study to examine the factor structure underlying the data. Two assumptions were tested before proceeding to the analysis: the sampling adequacy and the test of Sphericity. Pallant (2010) states that the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy should be higher than 0.600, while Barlett's test of sphericity should indicate a significant value ($p < .05$).

A possible number of factors on the data structure was examined using the 'Eigenvalue higher than 1' criteria suggested by Hair et al. (2014).

Although factor loadings of $\pm.30$ to $\pm.40$ are minimally acceptable, values greater than $\pm.50$ are generally considered necessary for practical significance (Hair et al., 2014).

Next, coefficients were examined to identify items with cross-loadings (loading in more than 1 factor). If one item presented loadings higher than 0.400 on more than one factor, this item was excluded. Additionally, if the item showed a substantially higher loading on any secondary factor, meaning that the item is not measuring the same concept as the other items, it was also excluded. After excluding items (when appropriate), the procedure was executed again.

Dimension 1

The examination of the eigenvalues of a first non-rotated solution indicated that any solution from 6 to 10 factors would be feasible. Since a 10-factor solution would make more theoretical sense, this solution was generated using Varimax rotation and an iterative process of dropping items and re-running the analysis took place following two criteria: (1) if an item showed no factor loading above 0.400; and (2) if an item showed cross-loadings of more than 0.400 on 2 or more factors. Four solutions were iteratively generated, until arriving at a final solution which is shown in the table below. This solution has passed the test of sphericity ($p < .001$) and sampling adequacy ($KMO = .882$).

Items	Factor									
	1	2	3	4	5	6	7	8	9	10
Id.1297	0.490	0.038	0.211	0.288	0.163	0.049	0.177	0.130	0.060	0.026
Id.1298	0.509	0.055	0.233	0.305	0.192	0.086	0.089	0.081	0.093	0.057
Id.1300	0.125	0.160	0.142	0.832	0.136	0.017	0.146	0.078	0.096	0.126
Id.1301	0.213	0.146	0.183	0.806	0.132	0.021	0.147	0.122	0.083	0.083
Id.1302_REVERSE	0.114	0.086	0.084	0.052	0.024	0.943	0.044	-0.015	0.083	-0.005
Id.1303_REVERSE	0.110	0.085	0.082	-0.006	0.023	0.911	0.016	-0.049	0.065	0.003
Id.1307	0.401	0.071	0.132	0.074	0.358	-0.010	0.200	0.200	0.126	-0.066
Id.1308	0.136	0.805	0.126	0.170	0.114	-0.026	0.123	0.034	0.049	0.160
Id.1309	0.135	0.871	0.160	0.102	0.079	0.093	0.092	0.003	0.096	0.162
Id.1310_REVERSE	0.157	0.560	0.103	0.050	0.069	0.162	0.118	-0.054	0.175	0.121
Id.1312	0.704	0.114	0.127	-0.015	0.064	0.064	0.116	0.123	0.023	0.081
Id.1313	0.548	0.262	0.242	0.219	0.173	0.068	0.099	0.008	0.116	0.191
Id.1314	0.606	0.287	0.140	0.061	0.170	0.107	0.139	0.054	0.212	0.134
Id.1315	0.570	0.053	0.103	0.096	0.364	0.084	0.140	0.055	0.169	0.081
Id.1316	0.315	0.088	0.058	0.118	0.689	0.002	0.181	0.086	0.098	0.109
Id.1317	0.262	0.161	0.144	0.181	0.651	0.000	0.157	0.277	0.103	0.046
Id.1318	0.335	0.221	0.109	0.113	0.416	0.080	0.181	0.073	0.193	0.314
Id.1319	0.419	0.080	0.200	0.098	0.360	0.146	0.103	0.034	0.161	0.195
Id.1320	0.124	0.021	0.138	0.102	0.099	-0.062	0.186	0.949	0.013	0.062
Id.1321	0.168	-0.081	0.186	0.105	0.234	-0.026	0.205	0.608	0.053	0.133
Id.1322	0.369	0.170	0.219	0.308	0.108	0.040	0.171	0.292	0.065	0.418
Id.1325	0.261	0.326	0.295	0.059	0.082	0.001	0.242	0.077	0.169	0.678
Id.1326	0.089	0.391	0.243	0.184	0.147	-0.037	0.124	0.131	0.092	0.757
Id.1327	0.198	0.148	0.184	0.074	0.152	0.121	0.108	0.040	0.652	0.111
Id.1328	0.162	0.150	0.175	0.108	0.126	0.055	0.081	0.035	0.931	0.074
Id.1330	0.249	0.135	0.279	0.095	0.163	0.073	0.508	0.181	0.145	0.117
Id.1331	0.212	0.137	0.085	0.165	0.134	0.040	0.597	0.224	0.035	0.065
Id.1332	0.186	0.177	0.103	0.157	0.222	0.003	0.739	0.132	0.117	0.198
Id.1339	0.272	0.071	0.597	0.224	0.021	0.027	0.194	0.181	0.042	0.078
Id.1340	0.204	0.089	0.679	0.037	0.153	0.007	0.002	0.093	0.097	0.122
Id.1341	0.120	0.160	0.573	0.072	0.115	0.142	0.086	0.128	0.105	0.154
Id.1342	0.134	0.172	0.502	0.190	-0.005	0.078	0.137	0.002	0.230	0.091

Extraction Method: Maximum Likelihood.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

A subjective assessment of the solution detected that item 'ID. 1307' should be excluded due to not making theoretical sense to be part of Factor 1 whatsoever. The low factor loading for this item ($\lambda = .401$) corroborated this idea. The rest of the factor structure made theoretical sense. Thus, the factor structure presented above will be transferred to the CFA procedure (without item 1307).

Dimension 2

The process was repeated to dimension 2. Eigenvalues of an initial non-rotated solution indicated that any solution from 5 to 8 factors would be adequate (eigenvalues around 1). After five iterations following the exclusion criteria presented earlier, 17 items were kept (from the original 24) to form the final factor structure (table below). The solution passed the tests for sample adequacy (KMO = .847) and sphericity ($p < .001$).

	Factor				
	1	2	3	4	5
Id.1353	0.107	0.265	0.634	0.234	0.023
Id.1354	0.099	0.065	0.784	0.139	0.219
Id.1355_REVERSE	0.080	0.076	0.542	0.064	0.217
Id.1359	0.193	0.190	0.521	0.354	0.111
Id.1363	0.159	0.148	0.165	0.867	0.079
Id.1364	0.272	0.075	0.178	0.704	0.036
Id.1365	0.110	0.296	0.213	0.488	0.092
Id.1369_REVERSE	0.209	0.040	0.381	0.015	0.654
Id.1370_REVERSE	0.201	0.132	0.227	0.156	0.880
Id.1372	0.289	0.641	0.117	0.255	0.071
Id.1373	0.298	0.838	0.131	0.179	-0.008
Id.1374	0.218	0.738	0.222	0.069	0.155
Id.1375	0.651	0.178	0.195	0.041	0.175
Id.1376	0.701	0.253	0.138	0.081	0.088
Id.1377	0.626	0.183	0.071	0.207	-0.031
Id.1378_REVERSE	0.613	0.124	0.062	0.142	0.282
Id.1379	0.523	0.101	0.047	0.162	0.070

Extraction Method: Maximum Likelihood.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

The next section presents the CFA's, which was conducted to validate the solutions suggested by the EFA's.

Confirmatory Factor Analysis

The main objective of testing a measurement model is to test construct validity. Construct Validity is defined as “the extent to which a set of measured variables actually represent the theoretical latent construct they are designed to measure” (Hair et al., 2014, p. 543). Confirmatory Factor Analysis (CFA) was used in the analysis, as it is an adequate method to be used as evidence of construct validity of theory-based instruments (Li, 2016). This type of analysis is used when a researcher wishes to confirm a specific pattern of variables that are predicted based on theory or previous analytical studies (DeVellis, 2012). That is, based on knowledge of the theory, he or she assumes the a priori factorial structure and then tests this hypothetical arrangement statistically (Byrne, 2016).

CFA was conducted using SPSS AMOS software, which uses Maximum Likelihood (ML) algorithm to estimate the results. ML is the most common method used to estimate parameters in CFA, because of its attractive statistical properties (i.e., asymptotic unbiasedness, normality, consistency, and maximal efficiency) (Li, 2016). After defining the model in the software and executing the analysis, four main phases were conducted to examine construct validity (1) assessment of model fit; (2) assessment of convergent validity; (3) assessment of discriminant validity and (4) respecification of the model (if necessary). The statistics that were used to assess model fit and their rules of thumb are presented in the table below.

Fit index	Rules of thumb
Normed chi-square (χ^2/df)	The division between the chi-square value and the model's degrees of freedom should be less than 4.
Root mean square error of approximation (RMSEA)	RMSEA < 0.08
Comparative fit index (CFI)	CFI > 0.90
Normed fit index (NFI)	NFI > 0.85

After the assessment of model fit, convergent and discriminant validity were examined. Convergent validity refers to the “extent to which indicators of a specific construct converge or share a high proportion of variance in common” (Hair et al. 2014, p. 601), while discriminant validity is defined as the “extent to which a construct is truly distinct from other constructs both in terms of how much it correlates with other constructs and how distinctly measured variables represent only this single construct (Hair et al. 2014,

p. 601). A second important concept to test is construct's reliability. This was done using the Composite Reliability index (which is based on factor loadings) and Cronbach's Alpha (which is based on correlations). A summary of the indicators used to measure constructs' validity and reliability are detailed below (table below).

Indicator of convergent validity	Definition	Rules of thumb
Factor loadings (λ)	Correlation between the original variables and the factors, and the key to understanding the nature of a particular factor. Squared factor loadings indicate what percentage of the variance in an original variable is explained by a factor.	In the case of high convergent validity, high one-factor loadings would indicate that they converge on a common point, the latent construct. At a minimum, all factor loadings must be statistically significant. Because a significant load can still have quite weak strength, a good rule of thumb is that standardized loading estimates should be 0.5 or higher and ideally 0.7 or higher.
AVE	A summary measure of convergence among a set of items representing a latent construct. It is the average percentage of variation explained (variance extracted) among the items of a construct.	An AVE of 0.5 or higher is a good rule of thumb suggesting adequate convergence. An AVE of less than 0.5 indicates that, on average, more error remains in the items than variance explained by the latent factor structure imposed on the measure.
Indicator of discriminant validity	Definition	Rules of thumb
AVE and correlations (p)	The squared variance extracted estimates for a construct should be greater than the correlation estimates between this and other constructs.	Squared AVE > p (Fornell & Larcker, 1981).
Indicator of internal consistency	Definition	Rules of thumb
Construct Reliability (CR)	Measure of reliability and internal consistency of the measured variables representing a latent construct. Must be established before construct validity can be assessed. It is computed from the squared sum of factor loadings for each construct and the sum of the error variance terms for a construct.	0.7 or higher suggests good reliability. Reliability between 0.6 and 0.7 may be acceptable, provided that other indicators of a model's construct validity are good.
Cronbach's Alpha	Cronbach's Alpha is a coefficient that represents the proportion of total variance among items that are due to the construct that they intend to measure	0.7 is the minimum acceptable level (Pallant, 2010).

Dimension 1

Convergent Validity and Reliability

The first solution, derived from the EFA, showed AVE lower than 0.500 for Factor 1 (AVE = .466) and 7 (AVE = .462). The item with the lowest factor loading for each factor was deleted and the analysis was executed again. The deleted items were 'ID.1297' for factor 1 ($\lambda = .635$) and 'ID.1342' for factor 7 ($\lambda = .635$). The second solution, without those items, also showed AVE's lower than 0.500 for factor 1 (AVE = .478) and factor 7 (AVE = .488). Following the same logic of the first iteration, items 'ID.1298' ($\lambda = .615$) and 'ID.1341' ($\lambda = .687$) were dropped. The third solution showed acceptable convergent validity according to AVE and good reliability for most factors according to the CR and Alpha coefficients (higher than 0.700). The final solution, with indicators of convergent validity and reliability is shown below, along with the labels created for each factor. The factor 'Supporting environment: addressing inconveniences' (factor 7) still showed coefficients slightly below acceptable, but since there were no more items to be dropped (only two items left), this factor was kept in the final structure as it is.

Items	Construct	Loadings	AVE	CR	Cronbach's Alpha
Id.1312	<---	0.634			
Id.1313	<---	0.757			
Id.1314	<---	0.782	0.501	0.833	0.822
Id.1315	<---	0.700			
Id.1319	<---	0.654			
Id.1308	<---	0.858			
Id.1309	<---	0.938	0.676	0.860	0.845
Id.1310_REVERSE	<---	0.643			
Id.1339	<---	0.759	0.497	0.662	0.658
Id.1340	<---	0.646			
Id.1300	<---	0.877	0.819	0.900	0.898
Id.1301	<---	0.932			
Id.1316	<---	0.746			
Id.1317	<---	0.782	0.570	0.799	0.789
Id.1318	<---	0.736			
Id.1302_REVERSE	<---	0.969	0.895	0.944	0.943
Id.1303_REVERSE	<---	0.922			
Id.1320	<---	0.820	0.706	0.828	0.825
Id.1321	<---	0.860			
Id.1330	<---	0.720	0.573	0.800	0.784
Id.1331	<---	0.710			
Id.1332	<---	0.835			
Id.1327	<---	0.864	0.746	0.854	0.854
Id.1328	<---	0.863			

Id.1322	<---		0.703			
Id.1325	<---	Commitment towards and from	0.914	0.701	0.874	0.857
Id.1326	<---	external parties	0.879			

The final model presented in the table below was acceptable fit (χ^2 (279, N = 444) = 979.801; $p < .001$; $\chi^2/df = 3.512$; RMSEA = 0.075; CFI = 0.902; NFI = 0.870).

Discriminant Validity

After determining convergent validity, the discriminant validity was assessed. The table below shows the squared AVE value (diagonal), along with correlations among constructs, obtained through CFA (non-diagonal values).

The only factor that lacked discriminant validity (squared AVE lower than one or more correlations) was 'Management Commitment', which correlates quite strongly with 'Dealing with accidents and supervisor/management commitment'. All other constructs showed good discriminant validity.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dealing with accidents and supervisor/management commitment (1)	0.708									
Management Commitment (2)	0.805	0.755								
Commitment towards and from external parties (3)	0.628	0.594	0.837							
Employee commitment (4)	0.521	0.455	0.648	0.822						
Supporting environment: safety education (5)	0.635	0.687	0.631	0.452	0.757					
Dealing with near-misses (6)	0.483	0.521	0.474	0.395	0.512	0.905				
Supporting environment: addressing inconveniences (7)	0.647	0.534	0.620	0.391	0.560	0.524	0.705			
Victim blaming (8)	0.294	0.153	0.104	0.209	0.154	0.123	0.171	0.946		
Safety department commitment (9)	0.378	0.530	0.415	0.114	0.561	0.367	0.517	-0.016	0.840	
Supporting environment: time and people (10)	0.559	0.500	0.461	0.388	0.431	0.351	0.439	0.244	0.246	0.864

Dimension 2

This section presents the results of the CFA conducted for dimension 2. It followed the same logic as the analysis for dimension 1.

Convergent Validity and Reliability

AVE indices of the first solution were not acceptable for Factor 1 and Factor 3. Thus, items 'ID.1379' ($\lambda = .551$) and 'ID.1355' ($\lambda = .568$) were dropped. A second solution still showed unacceptable convergent validity for Factor 1 and item 'ID.1377' ($\lambda = .648$) was dropped. The third solution reached sufficient convergent validity and reliability (Table below).

Item	Construct	Loadings	AVE	CR	Cronbach's Alpha
Id.1378_REVERSE	<---	0.675			
Id.1376	<---	0.750	0.515	0.761	0.758
Id.1375	<---	0.726			
Id.1374	<---	0.789			
Id.1373	<---	0.909	0.674	0.860	0.847
Id.1372	<---	0.757			
Id.1359	<---	0.727			
Id.1354	<---	0.727	0.527	0.770	0.769
Id.1353	<---	0.724			
Id.1365	<---	0.612			
Id.1364	<---	0.772	0.578	0.801	0.772
Id.1363	<---	0.873			
Id.1370_REVERSE	<---	0.873			
Id.1369_REVERSE	<---	0.815	0.713	0.832	0.831

Discriminant Validity

The solution also showed good discriminant validity, as shown in the table below. All squared values of AVE (diagonal values) are higher than inter-construct correlations.

	(1)	(2)	(3)	(4)	(5)
Personal priorities & safety responsibilities (1)	0.718				
Intention for proactive safety behaviour (2)	0.609	0.821			
Safety knowledge and competence (3)	0.481	0.506	0.726		
Trust in the organization (4)	0.422	0.471	0.604	0.760	
Intention to behave safely/unsafely (5)	0.547	0.310	0.554	0.333	0.844

The goodness-of-fit indices for the final model (χ^2 (67, N = 444) = 297.099; $p < .001$; $\chi^2/df = 4.434$; RMSEA = 0.088; CFI = 0.917; NFI = 0.897). indicate a RMSEA slightly higher than the acceptable threshold of 0.080. A standardized Root Mean Square Residual (SRMR), which is an alternative to RMSEA as an indicator of models' parsimony, showed evidence of good fit (SRMR = .058). Values lower than 0.08 are acceptable (Hair et al., 2014).

References

- Babbie, E., 1990. Survey Research Methods, 2nd ed. Wadsworth Publishing Company, Belmont, California.
- Blaikie, N., 2008. Analyzing quantitative data, First. ed. SAGE Publications, London, UK.
- Byrne, B.M., 2016. Structural equation modeling with AMOS: Basic concepts, applications, and programming. Routledge.
- DeVellis, R.F., 2012. Scale development: Theory and Applications, Third. ed. SAGE Publications, Thousand Oaks, California.
- Fornell, C., Larcker, D.F., 1981. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. J. Mark. Res. 18, 39. <https://doi.org/10.2307/3151312>
- Fowler, F.J., 2009. Survey Research Methods, 4th ed. ed. SAGE Publications.
- Hair, J.F., Black, W., Babin, B., Anderson, R., 2014. Multivariate data analysis, Seventh. ed. Pearson Education, Inc., Edinburgh.

Li, C.H., 2016. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* 48, 936–949. <https://doi.org/10.3758/s13428-015-0619-7>

Pallant, J., 2010. *SPSS Survival Manual*, 4th ed. McGraw-Hill, Berkshire, England.

SAMPLE REPORT - Rafael Data Analysis Portfolio