

# **Analysis Report**

This report is structured as follow.

## **Contents**

Exploratory Factor Analysis .....	2
The construction of the Sustainability Index .....	2
References .....	10

SAMPLE REPORT - Rafael Data Analysis Portfolio

### **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). When the goal of the analysis is to look at the patterns of correlations among the variables, then the appropriate technique is Principal Component Analysis (PCA) (Tabachnick and Fidell, 2014)

Principal Component Analysis was performed on each scale 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).

The researcher should view the communalities to assess whether the variables meet acceptable levels of explanation. For example, a researcher may specify that at least one-half of the variance of each variable must be taken into account. Using this guideline, the researcher would identify all variables with communalities less than .50 as not having sufficient explanation (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).

### **The construction of the Sustainability Index**

30 variables measured across 56 schools were inserted in the PCA procedure. KMO's measure of sampling adequacy was 0.363 which indicates the items are not sufficiently correlated to generate a meaningful factor structure. An examination of the factor loadings showed that several items loaded poorly ( $< \pm 0.300$ ) on the single factor (table below).

*Component Matrix<sup>a</sup>*

	Component
	1
Energy Consumption KWh/m <sup>2</sup>	-.252
Solar Panel	.390
Insulation in walls and ceiling	.256
Shading in South elevation	.547
Shading in East elevation	.708
Shading in West elevation	.725
Water Consumption L/person.yr	-.198
Rain water Tank	.272
Sanitation	.249
Sanitation Efficiency	.340
Green Area %	.201
Shading Area %	-.418
Parking to Teacher %	.192
Closest to Public transportation m	.003
Accessibility	.354
Heat Island Effect	-.028
Waste Speration	-.192
Maintinance cost NIS/yr	.331
New Teachers #	.313
New Staff #	.389
Social activities	-.025
Low income students %	-.303
Disabeld students %	-.072
Teacher Participation (%)	-.019
Parents Participation (%)	-.389
Students Participation (%)	.455
Attendance Rates (%)	-.213
Sick Leave (%)	-.099
Success Rate (%)	-.027
Drop out rate (%)	.385

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Variables with loadings lower than  $\pm 0.300$  were subsequently dropped and new solutions generated until a solution with 11 variables was reached (table below).

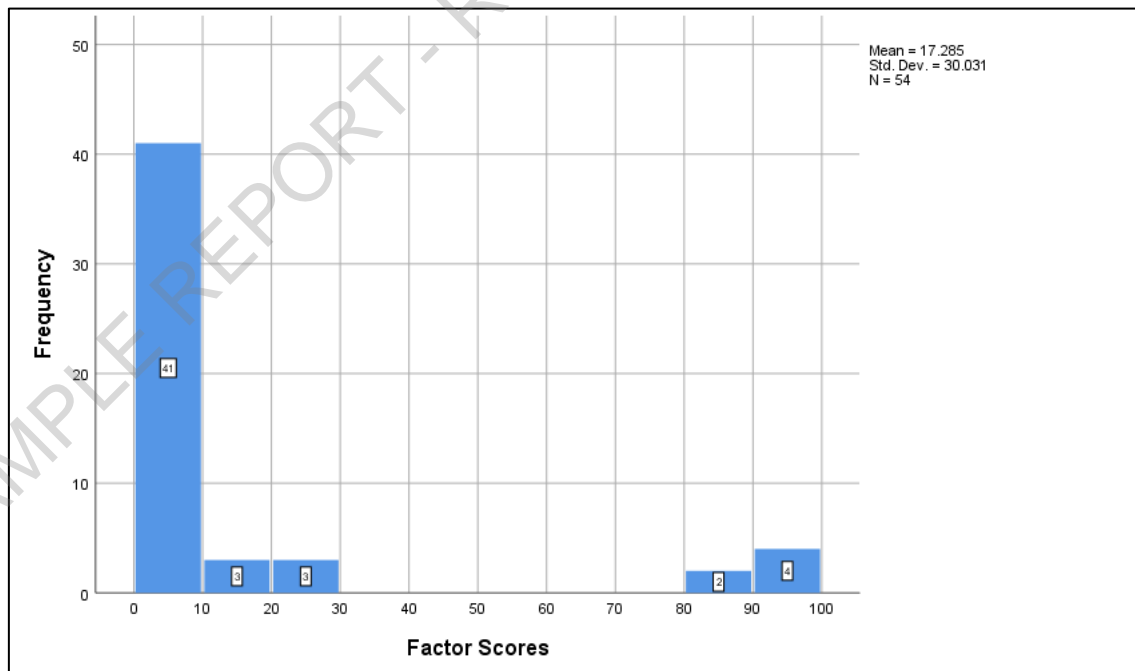
*Component Matrix<sup>a</sup>*

	Component
	1
Shading in South elevation	.662
Shading in East elevation	.764
Shading in West elevation	.829
Sanitation Efficiency	.366
Shading Area %	-.437
Accessibility	.339
Maintenance cost NIS/yr	.315
New Staff #	.348
Parents Participation (%)	-.411
Students Participation (%)	.420
Drop out rate (%)	.361

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Factor scores for each school were calculated using the Regression method in SPSS and their values were rescaled to fit in a 0-100 range. When examining the resulting factor scores, a substantial number of school had scores lower than 20 whereas a second group of school had very high scores (histogram below). The figure shows that 41 schools had scores lower than 10 while only 6 schools were scored above 80.



An examination of the variables shows that the final scores were very sensitive to the variables 'Shading in South elevation', 'Shading in East elevation' and 'Shading in West elevation' due to the binary nature of the variable (1 or 0). Therefore, these variables were dropped and a second factor analysis was conducted from the beginning. In the second solution, several variables showed poor loadings again (table below) and the sampling adequacy was low at 0.348 and Barlett's test of sphericity was non-significant ( $p = 0.064$ ).

*Component Matrix<sup>a</sup>*

	Component
	1
Energy Consumption KWh/m <sup>2</sup>	-.108
Solar Panel	.288
Insulation in walls and ceiling	-.095
Water Consumption L/person.yr	.419
Rain water Tank	.641
Sanitation	-.595
Sanitation Efficiency	-.156
Green Area %	.542
Shading Area %	-.295
Parking to Teacher %	.460
Closest to Public transportation m	.369
Accessibility	-.353
Heat Island Effect	.248
Waste Speration	-.136
Maintinance cost NIS/yr	.035
New Teachers #	.578
New Staff #	.185
Social activities	-.147
Low income students %	-.324
Disabeld students %	.378
Teacher Participation (%)	.139
Parents Participation (%)	.209
Students Participation (%)	.039
Attendance Rates (%)	.339
Sick Leave (%)	-.269
Success Rate (%)	.278
Drop out rate (%)	.043

Extraction Method: Principal Component Analysis.

a. 1 component extracted.

Items with loadings lower than  $\pm 0.300$  were subsequently dropped and new solutions generated until a solution with 10 variables was reached (table below shows factor loadings). The solution was acceptable since Barlett's test was not significant ( $p < 0.001$ ) and KMO's sampling adequacy was 0.581, barely below the acceptable threshold of 0.600. 25.6% of the variance on the global scale was retained by the factor solution.

*Component Matrix<sup>a</sup>*

	Component
	1
Water Consumption L/person.yr	.464
Rain water Tank	.605
Sanitation	-.688
Green Area %	.629
Parking to Teacher %	.373
Closest to Public transportation m	.451
Accessibility	-.474
New Teachers #	.537
Low income students %	-.330
Disabeld students %	.385

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

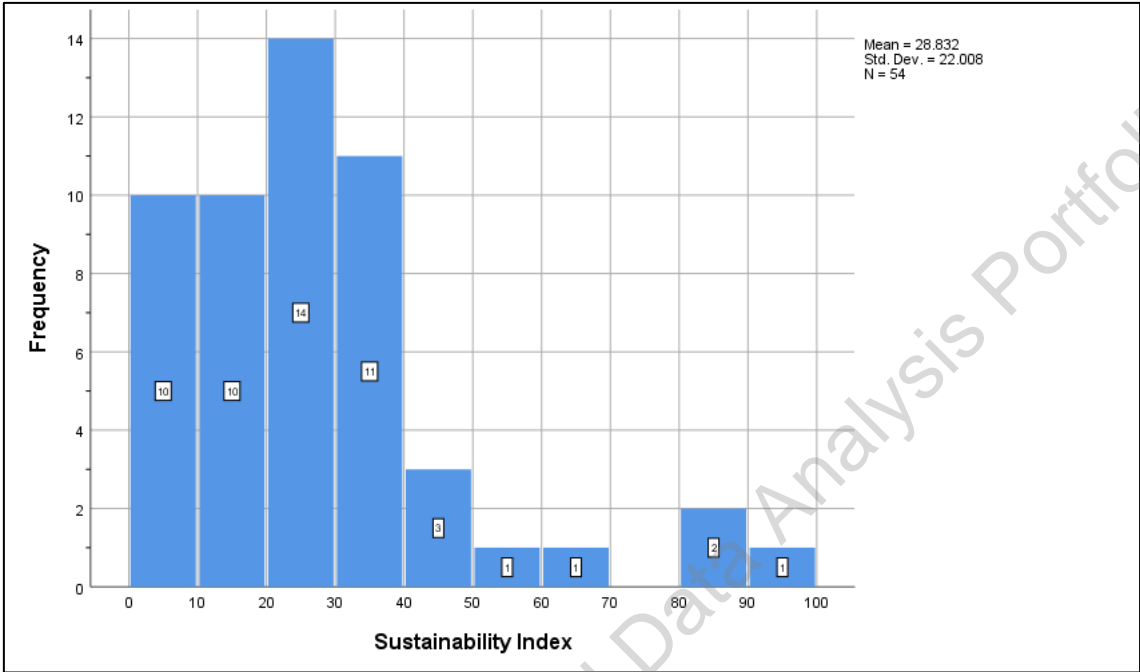
Factor scores were estimated using the regression method and rescaled to a 0-100 scale to represent the sustainability index. The index was calculated for each school and the result is showed in the table below.

School	Water Consumption L/person.yr	Rain water Tank	Sanitation	Green Area %	Parking to Teacher %	Closest to Public transportation m	Accessibility	New Teachers #	Low income students %	Disabled students %	Factor Scores	Factor Scores (rescaled)
1	1480	0	0	10	42	450	0	3	4	4	1.737	67.058
2	1406	0	1	14	18	1500	1	2	3	1	0.413	37.921
3	882	0	1	0	0	10	1	0	10	0	-1.078	5.117
4	1016	0	1	25	0	10	1	2	3	2	0.064	30.240
5	4657	0	0	6	0	2000	1	0	5	0	0.740	45.127
6	3085	0	1	0	0	100	1	0	4	3	-0.536	17.044
7	1130	0	1	1	0	1000	1	2	7	1	-0.408	19.848
8	1385	1	0	25	48	1200	1	5	7	0	2.365	80.882
9	877	0	1	0	0	1	1	0	10	1	-1.026	6.253
10	4924	0	0	10	0	20	1	1	1	0	0.449	38.717
11	2241	0	0	13	0	15	1	1	6	1	0.216	33.580
12	795	0	1	0	0	500	1	1	18	2	-0.866	9.765
13	1441	0	1	4	39	50	1	1	9	1	-0.364	20.829
14	741	0	1	5	0	150	1	2	9	4	-0.397	20.100
15	365	0	1	25	43	10	1	6	9	3	0.800	46.440
16	3976	1	1	1	13	0	1	0	4	0	0.119	31.454
17	1592	0	1	1	0	0	1	0	12	7	-0.582	16.028
18	3065	0	1	0	0	100	1	0	26	0	-1.202	2.371
19	1128	0	1	1	0	500	1	0	4	0	-0.763	12.030
20	795	0	1	10	0	600	1	2	4	2	-0.214	24.124
21	3924	0	1	15	0	2	1	1	6	1	-0.183	24.814
22	807	0	1	3	0	50	1	0	9	0	-1.015	6.500
23	565	0	1	3	8	0	1	0	6	0	-0.896	9.124
24	1738	0	1	13	0	1000	1	4	2	1	0.311	35.675
25	833	1	1	15	0	400	1	0	2	0	0.239	34.100
26	7500	0	1	1	0	1000	1	0	4	0	-0.013	28.555
28	455	0	1	15	0	200	1	1	16	1	-0.655	14.427
29	506	1	1	3	30	100	1	2	8	4	0.486	39.534
30	513	0	1	15	0	150	0	1	10	3	0.181	32.819

31	589	0	1	1	7	1	1	0	26	1	-1.310	0.000
32	3936	0	1	14	0	100	1	2	6	3	0.113	31.319
33	2529	0	1	13	0	150	1	1	5	0	-0.338	21.397
35	967	0	1	4	0	500	1	1	6	2	-0.513	17.541
36	431	1	1	3	0	10	1	0	8	0	-0.358	20.957
37	880	0	1	24	0	300	1	0	7	0	-0.286	22.545
38	4234	0	0	2	0	200	0	1	3	0	0.760	45.564
39	991	0	1	0	0	150	1	0	3	0	-0.908	8.849
40	387	0	0	12	0	1	1	0	8	1	-0.121	26.169
41	3726	0	0	2	0	150	1	0	8	1	-0.023	28.325
42	1188	0	1	15	0	200	1	0	2	1	-0.406	19.899
43	1236	0	1	0	0	50	1	2	14	0	-0.872	9.641
44	845	0	1	0	0	20	1	0	32	8	-1.005	6.705
45	602	0	0	2	0	500	1	4	2	2	0.406	37.761
46	1639	1	0	39	0	0	1	10	2	7	3.050	95.947
47	1070	0	1	6	0	5	1	1	6	2	-0.621	15.163
48	529	0	1	23	0	120	1	1	1	0	-0.180	24.878
49	533	0	0	9	0	0	1	1	10	0	-0.190	24.645
50	472	0	1	13	0	1	1	2	5	0	-0.476	18.361
51	544	0	1	0	0	2	1	2	10	2	-0.773	11.822
52	518	0	1	1	0	100	0	1	7	0	-0.304	22.136
53	413	0	1	16	0	0	1	2	1	0	-0.311	21.995
54	1996	0	0	1	58	30	1	4	2	4	1.152	54.184
55	9667	1	0	33	0	700	0	0	14	9	3.234	100.000
56	4405	1	0	10	0	3000	0	0	6	0	2.354	80.642



The histogram below shows the distribution of schools across the 0-100 scale of the sustainability index. It shows that most schools have an index of 50 or low. 14 of them show a score between 20 and 30.



## **References**

Blaikie, N. (2008). *Analyzing quantitative data* (First). SAGE Publications.

Hair, J. F., Black, W., Babin, B., & Anderson, R. (2014). *Multivariate data analysis* (Seventh). Pearson Education, Inc.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* / Barbara G. Tabachnick, Linda S. Fidell.

SAMPLE REPORT - Rafael Data Analysis Portfolio