

EduLai Skills Checker
Statistical Analysis
May 23, 2020

edulai

Notes and Contacts

All calculations are carried out using the statistical software R¹ in the version 3.5.1 (2018-07-02) installed on the system x86_64-apple-darwin15.6.0.

The specific libraries used for psychometric analyses are Psy, Multilevel and Alpha coefficient.

For information not available in this document, or for further information please write to lucilla.crosta@smarthink.eu.

¹ R Core Team (2018). R: A language and environment for statistical computing.

R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>

Table of content

1	Introduction	4
2	Data	5
2.1	Item and Skill	5
2.2	The sample	6
2.3	Descriptive statistics on the Items	6
3	Skills analysis	11
3.1	Descriptive statistics	11
3.2	Correlation analysis	11
3.3	Cronbach's Alpha Coefficient	13
4	Conclusions	15

Chapter 1

Introduction

Edulai Skills Checker is a test aimed at assessing individuals' soft skills through a questionnaire.

This analysis is intended as subsequent to the construction of the questionnaire and to its administration on a sample of individuals.

The statistical analysis of the responses of the selected sample are analyzed here.

In Chapter 2, the *Edulai Skills Checker* questionnaire is presented, and a distinction between *Items* and *Skill* is made. In the same chapter the selected sample for the test is described and some descriptive statistics related to the *Items* are discussed.

In Chapter 3, *Skills* are analyzed first with descriptive statistics, then with Correlation Analysis and Cronbach's Alpha Coefficient.

Chapter 4 reports some final considerations.

Chapter 2

Data

In this chapter, it is first presented the questionnaire and then the sample considered. There are also some calculated descriptive statistics on the responses of the subjects belonging to the sample.

2.1 Item and Skill

The *EduLai Skills Checker* questionnaire consists of 10 questions, from now on called *Items*. Latent variables, namely what it wants to be measured by the test, are seven

- Collaboration;
- Problem Solving;
- Leadership;
- Communication;
- Interculturalism;
- Learning;
- Critical Thinking.

These seven latent variables, from now on, will be called *Skills*.

The set of *Skills* is then used to describe the construct, that is the *Soft Skills* of the subject submitting the questionnaire. Each *Item* is designed to include within it references two or three latent variables.

Each *Item* is made up of a question with 5 possible answers. Any answer has associated a discrete value on a scale from 1 to 5. These values are therefore combined to obtain the evaluation of the *Skills*.

The associations between *Items* and *Skills* are:
Item 01 Communication and Leadership;
Item 02 Critical thinking and Problem solving;
Item 03 Critical thinking and Communication;
Item 04 Leadership and Critical thinking;
Item 05 Collaboration and Interculturalism;
Item 06 Collaboration and Learning;
Item 07 Interculturalism and Problem solving;
Item 08 Learning and Leadership;
Item 09 Collaboration and Communication;
Item 10 Learning and Interculturalism.

2.2 The sample

The sample considered consisted of 64 subjects, 28 women and 36 men.

All subjects answered all the questions in the questionnaire. The sample has a sufficient number of subjects to carry out some statistical evaluations. However, with 64 subjects the results can be affected by a considerable component of variability. The results are therefore to be understood as correct "on average", and may vary with the growth of the sample.

2.3 Descriptive statistics on the Items

Table 2.1 shows the cumulative responses observed for each *Item* on the whole sample. The number is therefore calculated for each *Item* of valid observations and some descriptive statistics, that is, the standard deviation, median, minimum, maximum, the range (the difference between the value maximum and minimum value), asymmetry and kurtosis. Those values are reported on the total observations in Table 2.2 while are distinguished according to two genders in Table 2.3.

From a first analysis, Table 2.1 seems to suggest that in some *Items* the sample responses selected are not equally distributed among the available answers.

Item	Score 1	Score 2	Score 3	Score 4	Score 5
Item 01					
Item 02	1	5	4	11	43
Item 03	3	10	14	18	19
Item 04	2	10	13	3	36
Item 05	3	2	42	5	12
Item 06	0	1	20	1	42
Item 07	6	15	7	13	23
Item 08	3	4	4	43	10
Item 09	1	3	15	27	18
Item 10	4	10	10	11	29

Table 2.1: The number of answers calculated for the whole sample is shown for each *Item*.

	n	media	ds	med	min	max	range	skew	kurtosis
Item 01	64.00	3.75	1.25	4.00	1.00	5.00	4.00	-0.88	-0.34
Item 02	64.00	4.41	1.02	5.00	1.00	5.00	4.00	-1.65	1.69
Item 03	64.00	3.62	1.20	4.00	1.00	5.00	4.00	-0.45	-0.88
Item 04	64.00	3.95	1.30	5.00	1.00	5.00	4.00	-0.68	-1.09
Item 05	64.00	3.33	0.98	3.00	1.00	5.00	4.00	0.23	0.25
Item 06	64.00	4.31	0.97	5.00	2.00	5.00	3.00	-0.74	-1.31
Item 07	64.00	3.50	1.43	4.00	1.00	5.00	4.00	-0.36	-1.38
Item 08	64.00	3.83	0.94	4.00	1.00	5.00	4.00	-1.49	2.26
Item 09	64.00	3.91	0.92	4.00	1.00	5.00	4.00	-0.66	0.19
Item 10	64.00	3.80	1.34	4.00	1.00	5.00	4.00	-0.66	-0.97

Table 2.2: For each *Item* it is reported the number of valid observations (**n**), the mean, the standard deviation (**ds**), the median (**med**), the minimum value (**min**), the maximum value (**max**), the difference between the maximum value and the minimum value (**range**), asymmetry (**skew**) and kurtosis. The descriptive statistics are calculated on the whole sample.

(a) Group: Women

	n	media	ds	med	min	max	range	skew	kurtosis
Item 01	28.00	3.86	1.30	4.00	1.00	5.00	4.00	-1.02	-0.06
Item 02	28.00	4.54	0.92	5.00	2.00	5.00	3.00	-1.74	1.69
Item 03	28.00	3.75	1.04	4.00	2.00	5.00	3.00	-0.27	-1.20
Item 04	28.00	4.00	1.22	5.00	1.00	5.00	4.00	-0.71	-0.79
Item 05	28.00	3.50	1.00	3.00	1.00	5.00	4.00	0.21	-0.29
Item 06	28.00	4.11	1.07	5.00	2.00	5.00	3.00	-0.38	-1.72
Item 07	28.00	3.86	1.18	4.00	1.00	5.00	4.00	-0.79	-0.53
Item 08	28.00	3.82	1.02	4.00	1.00	5.00	4.00	-1.47	1.89
Item 09	28.00	3.82	1.02	4.00	1.00	5.00	4.00	-0.87	0.33
Item 10	28.00	3.75	1.48	5.00	1.00	5.00	4.00	-0.51	-1.45

(a) Group: Men

	n	media	ds	med	min	max	range	skew	kurtosis
Item 01	36.00	3.67	1.22	4.00	1.00	5.00	4.00	-0.75	-0.68
Item 02	36.00	4.31	1.09	5.00	1.00	5.00	4.00	-1.50	1.22
Item 03	36.00	3.53	1.32	4.00	1.00	5.00	4.00	-0.42	-1.10
Item 04	36.00	3.92	1.38	5.00	1.00	5.00	4.00	-0.61	-1.37
Item 05	36.00	3.19	0.95	3.00	1.00	5.00	4.00	0.20	0.57
Item 06	36.00	4.47	0.88	5.00	3.00	5.00	2.00	-1.03	-0.93
Item 07	36.00	3.22	1.55	3.00	1.00	5.00	4.00	-0.00	-1.67
Item 08	36.00	3.83	0.88	4.00	1.00	5.00	4.00	-1.41	2.07
Item 09	36.00	3.97	0.84	4.00	2.00	5.00	3.00	-0.23	-1.02
Item 10	36.00	3.83	1.23	4.00	1.00	5.00	4.00	-0.77	-0.54

Table 2.3: For each *Item* it is reported the number of valid observations (**n**), the mean, the standard deviation (**ds**), the median (**med**), the minimum value (**min**), the maximum value (**max**), the difference between the maximum value and the minimum value (**range**), asymmetry (**skew**) and kurtosis. The descriptive statistics are calculated separately on the responses of the two sexes.

Item	Gini Coefficient
Item 01	0.445
Item 02	0.711
Item 03	0.312
Item 04	0.609
Item 05	0.695
Item 06	0.805
Item 07	0.328
Item 08	0.672
Item 09	0.523
Item 10	0.398

Table 2.4: Gini concentration index on the *Items*.

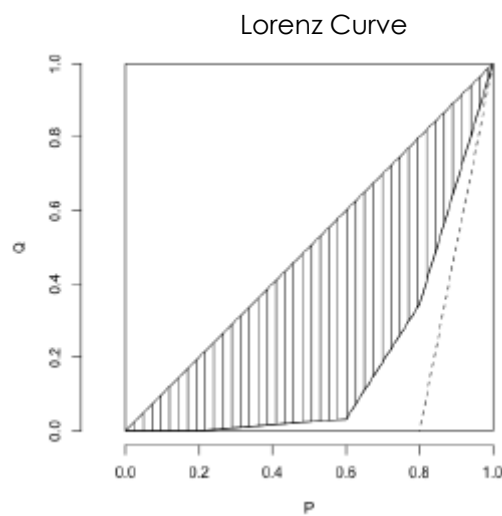


Figure 2.1: Lorenz Curve referred to *Item 06*

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Q1	1.00									
Q2	0.13	1.00								
Q3	0.13	0.02	1.00							
Q4	-0.00	0.29	-0.01	1.00						
Q5	0.00	0.04	-0.19	0.12	1.00					
Q6	0.05	0.07	0.06	-0.08	-0.30	1.00				
Q7	0.01	-0.10	0.10	-0.04	0.02	-0.06	1.00			
Q8	-0.08	0.01	0.10	-0.08	0.01	-0.14	0.06	1.00		
Q9	0.21	0.14	-0.03	0.06	0.06	0.25	-0.07	0.02	1.00	
Q10	0.01	0.03	0.16	0.01	-0.31	0.35	0.05	0.16	0.01	1.00

Table 2.5: Correlations calculated by applying *Kendall's Tau* on the *Items* (indicated as Q for display purposes). The correlations in bold indicate pairs of *Items* that contribute to determine the same *Skill*.

The dispersion indices presented in Table 2.2 e Table 2.3 concern measures of standard deviation and range. The standard deviation ranges between 0.92 and 1.43. However, it must be considered that the standard deviation is not necessarily indicative of dispersion, as the variables analyzed here numerically are collected as categorical variables. The correlation coefficients of Gini are shown in Table 2.4 and are equal to 1 in the case of maximum concentration and equal to 0 in the case of equal distribution. Although a certain level of concentration is inevitable in the answers to the various questions, we notice a concentration index higher than 80% in *Item 06*. The Lorenz Curve is also reported for this *Item* in Figure 2.1 (the Gini concentration index is intended as a measure of the colored area in this in this graph). In fact, for this question, answers are mainly focused on values 3 and 5.

The range is a measure corresponding to the maximum value recorded minus the minimum value, in this case represented by 5 and 1. The range is equal to 4 for all questions except for *Item 06*. No one has selected the answer corresponding to score 1. This implies that for this question the level 1 was never observed in the sample responses.

Descriptive statistics show a higher average of responses than the arithmetic mean, which in theory would be equal to 3.5, hence we notice a right asymmetry (negative skew values). This is valid for all *Items* except for number 5.

Table 2.5 shows the correlations calculated through the application of the Kendall's Tau on the *Items*, indicating in bold the pairs of *Items* that contribute to determining the same *Skill*. There are no correlations high enough to believe that the information contained in a pair of *Item* is redundant. However, some negative correlations are observed even in couples who can compete for the same *Skill*.

Chapter 3

Skills analysis

In this Chapter the analyzes are carried out with respect to the *Skills*. First the *Skills* values are summarized with descriptive statistics, then the correlations are analyzed and the internal consistency of the questionnaire is assessed.

Analyzes on the internal consistency of a questionnaire are typically carried out on the *Items*, coinciding with the questions, which define the construct to be measured. In this case, however, the questions (*Item*) help to define the *Skills* which in turn describe the construct, in particular, the individual's *Soft Skills* profile. For this reason, the analyses on the coherence of the questionnaire are carried out for this level, that is on *Skills*.

3.1 Descriptive statistics

Table 3.1 shows the number of answers calculated for each *Skill* on the whole sample. For each *Skill* the number of observations and some descriptive statistics are calculated, that is, the standard deviation, median, minimum, maximum, range (the difference between the maximum value and the minimum value) and asymmetry. A negative asymmetry of can be observed for all the *Skills*.

3.2 Correlation analysis

The correlation analysis allows to verify how correlated the *Skills* are between them and with the general theme investigated, in this case the set of *Soft Skill* of the subject. The Pearson correlation coefficient allows to check the degree of association between each *Skill* and each other, taking them two by two, or with respect to the scale as a whole.

The analyzes that are selected are therefore two: a first analysis of the correlations between *Skill*, hence each *Skill* compared to the scale as a whole.

The Pearson correlation coefficients are shown in Table 3.2 on each pair of *Skills*. These correlations are all positive.

This effect is attributable to two different elements. The first one is by construction of the *Skills'* values: every *Item* observed contributes to the calculation of two or three *Skills*. The second one is that the different *Skills* are all related to *Soft Skill* features, reasonably thought to be related to them to a more or less pronounced extent. However, there are no *Skills* with a correlation to be selected as redundant. There are three cases with correlations greater than 0.5, namely

- *Interculturalism* and *Problem solving* with 0.64;
- *Leadership* and *Problem solving* with 0.58;
- *Leadership* and *Communication* with 0.54.

Correlation values with respect to the scale as a whole are reported in Table 3.3. Values above 0.3 indicates *Skills* with consistent measurements with the scale. Overall, the *Skills* are consistent with the global measurement scale of *Soft Skills* except for *Collaboration*, which is little below the threshold value.

	n	media	ds	med	min	max	range	skew
Collaborazione	64.00	3.85	0.58	4.00	2.00	5.00	3.00	-0.62
Problem Solving	64.00	3.95	0.86	4.00	1.00	5.00	4.00	-0.75
Leadership	64.00	3.84	0.66	4.00	2.33	5.00	2.67	-0.62
Comunicazione	64.00	3.76	0.75	3.83	1.33	5.00	3.67	-0.87
Interculturalità	64.00	3.54	0.67	3.50	2.33	5.00	2.67	-0.04
Apprendimento	64.00	3.98	0.72	4.00	2.00	5.00	3.00	-0.54
Pensiero Critico	64.00	3.99	0.75	4.17	1.67	5.00	3.33	-0.88

Table 3.1: For each Skill are reported the number of valid observations (**n**), the mean, the standard deviation (**ds**), the median (**med**), the minimum value (**min**), the maximum value (**max**), the difference between the maximum value and the minimum value (**range**) and asymmetry (**skew**). The descriptive statistics are calculated on the whole sample.

	Col.	P.S.	Lea.	Com.	Int.	App.	P.C.
Col.	1.00						
P.S.	0.05	1.00					
Lea.	0.16	0.15	1.00				
Com.	0.39	0.14	0.54	1.00			
Int.	0.21	0.64	0.14	0.12	1.00		
App.	0.25	0.03	0.24	0.23	0.40	1.00	
P.C.	0.04	0.40	0.58	0.44	0.07	0.09	1.00

Table 3.2: Correlations calculated by applying the coefficient of Pearson correlation on the *Skills*. For display purposes the names of the *Skills* are abbreviated as follows: Col. for Collaboration, P.S. for Problem Solving, Lea. for Leadership, Com. for Communication, Int. for Interculturalism, App. For Learning, P.C. for Critical Thinking.

<i>Skill</i>	Correlation
Collaboration	0.2850
Problem Solving	0.3799
Leadership	0.5088
Communication	0.4998
Interculturalism	0.4528
Learning	0.3087
Critical Thinking	0.4632

Table 3.3: Pearson correlation coefficient for each *Skill* with respect to the total scale.

3.3 Cronbach's Alpha Coefficient

Cronbach's Alpha Coefficient is a measure of internal consistency widely used for the analysis of questionnaires. This indicator is built to evaluate the consistency of the concept that the questionnaire in question tries to measure.

Cronbach's Alpha is calculated by comparing the variability of *Skills* with the total variability. It is calculated as

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum_{i=1}^k \sigma_{y_i}^2}{\sigma_x^2} \right)$$

where k is the Skill number in question, $\sigma_{y_i}^2$ is the variance associated with the i -th *Skill* and σ_x^2 is the observed total variance.

The α value ranges between 0 and 1 and is usually considered satisfactory for values greater than 0.7. In this case, Cronbach's Alpha index is equal to 0.7019.

This value is presented together with the calculation of the same index with removal of each *Skill*, in order to evaluate the presence of dimensions that are detrimental to the global consistency. The results are shown in the table 3.4. You can see that there are no values higher than the Alpha during the removal of each *Skill*.

<i>Skill</i>	Alpha
Collaboration	0.70
Problem Solving	0.68
Leadership	0.65
Communication	0.64
Interculturalism	0.66
Learning	0.69
Critical Thinking	0.65

Table 3.4: Cronbach's Alpha coefficient obtained by removing the individual Skills.

Chapter 4

Conclusions

The questionnaire is made up of ten *Items* which are coherent with each other in terms of average, mostly unbalanced towards the high values of the scale. The answers are typically spread across all available values, except of *Item 06* for which there is a strong polarization on two of the 5 possible answers.

Still regarding the *Items*, some negative correlations are observed also in pairs that can compete for the same skill. This element must be compared with a qualitative analysis on the construction of the questions, in order to understand if this phenomenon is compatible with the ideas followed during the construction of the questionnaire. In particular, we refer to questions 5 and 6 with respect to the *Collaboration Skill* and to questions 5 and 10 with respect to the *Interculturalism Skill*.

Please also note that the *Problem solving Skill* is built on values of the answers to questions 2 and 7, while the *Critical thinking Skill* on values of questions 2, 3 and 7: these two skills are very similar in construction. Their correlation is equal to 0.4, not in values that suggest redundancy of information.

As for *Skills*, they are consistent with each other according to the correlation analysis, avoiding overlaps and redundancies. The internal reliability calculated by Cronbach's Alpha coefficient is within the allowed values, although the validity of this result may be impaired from the absence of the *Tau equivalence* hypothesis. Anyway, the analysis of the Alpha coefficients on *Skill* removal does not suggest removal of any *Skill* to improve internal consistency.

It is recommended to repeat the analyzes with a larger sample to get more stable results.