Supplementary material concerning
"Analyzing data from a fuzzy rating scale-based questionnaire. A case study"

This file presents the details for the statistical analysis of the data obtained from the conducted survey.

Descriptive analysis: frequency distributions of some aspects

In connection with the questions for which responses are only dichotomous, the frequency distributions are as follows:

QUESTIONNAIRE FORMAT	abs. freq.
Paper-and-pencil	24
Computerized	45

SEX	abs. freq.
Girl	31
Boy	38

Computer at home	abs. freq.
NO	1
YES	68

Own desk	abs. freq.
NO	8
YES	61

Own books	abs. freq.
NO	2
YES	67

Own room	abs. freq.
NO	25
YES	42
No answer	2

Internet connection	abs. freq.
NO	6
YES	63

By taking into account one of the tests to be detailed later, the factor 'mark in the scale $0-10$ taken in the last exam' has been assumed to act at the levels given by intervals

$$G1 = [0,6],\ G2 = (6,8],\ G3 = (8,9],\ G4 = (9,10].$$

Then, the 4-point Likert scale-based responses to the 9 questions concerning reading, maths and science would be distributed as follows:

		absolute frequencies				
		$G1$	$G2$	$G3$	$G4$	No answer
READING						
R.1	I like to read things that make me think	5	26	14	7	17
R.2	I learn a lot from reading	5	26	13	7	18
R.3	Reading is harder for me than any other subject	5	25	14	7	18
MATHS						
M.1	I like math	7	18	16	20	8
M.2	My teacher is easy to understand	7	18	16	19	9
M.3	Math is harder for me than any other subject	7	19	16	20	7
SCIENCE						
S.1	My teacher taught me to discover science in daily life	9	23	10	16	11
S.2	I read about science in my spare time	8	22	10	17	12
S.3	Science is harder for me than any other subject	10	23	10	17	9

Table 1: Absolute frequency distribution of the number of respondents to each of the 9 questions in accordance with the group associated with their marks in the area

Although the rows in the last table would often coincide for the three questions in each area, this has not been the case, since some students have not provided either their marks or their response to the question.

With respect to the questions for which responses are given in accordance with a 4-point Likert scale, the frequency distributions are as follows:

Use frequency of computer at home	abs. freq.
Never or almost never	9
Once/twice a month	16
Once/twice a week	20
Every/almost every day	22
No answer	2

Use frequency of computer at school	abs. freq.
Never or almost never	0
Once/twice a month	24
Once/twice a week	31
Every/almost every day	11
No answer	3

R.1	abs. freq.
Disagree a lot (**A1**)	3
Disagree a little (**A2**)	13
Agree a little (**A3**)	39
Agree a lot (**A4**)	13
No answer	1

R.2	abs. freq.
Disagree a lot (**A1**)	0
Disagree a little (**A2**)	6
Agree a little (**A3**)	24
Agree a lot (**A4**)	37
No answer	2

R.3	abs. freq.
Disagree a lot (**A1**)	40
Disagree a little (**A2**)	16
Agree a little (**A3**)	5
Agree a lot (**A4**)	6
No answer	2

M.1	abs. freq.
Disagree a lot (**A1**)	3
Disagree a little (**A2**)	20
Agree a little (**A3**)	20
Agree a lot (**A4**)	24
No answer	2

M.2	abs. freq.
Disagree a lot (**A1**)	2
Disagree a little (**A2**)	4
Agree a little (**A3**)	19
Agree a lot (**A4**)	41
No answer	3

M.3	abs. freq.
Disagree a lot (**A1**)	15
Disagree a little (**A2**)	15
Agree a little (**A3**)	10
Agree a lot (**A4**)	29
No answer	0

S.1	abs. freq.
Disagree a lot (**A1**)	7
Disagree a little (**A2**)	12
Agree a little (**A3**)	28
Agree a lot (**A4**)	19
No answer	3

S.2	abs. freq.
Disagree a lot (**A1**)	27
Disagree a little (**A2**)	27
Agree a little (**A3**)	9
Agree a lot (**A4**)	2
No answer	4

S.3	abs. freq.
Disagree a lot (**A1**)	22
Disagree a little (**A2**)	19
Agree a little (**A3**)	16
Agree a lot (**A4**)	9
No answer	3

Finally, the frequency distributions of the fuzzy rating scale-based responses would make sense, but they would not be much more informative than the corresponding datasets. This is due to the fact that, as a rather natural consequence from the continuity of the rating scale and the freedom in responding, almost all the answers differ for each posed question.

Estimating some means, medians and variances

In connection with the estimates of the mean, median and variance of the responses to the 9 considered questions in this sample of 69 students, one can summarize them as follows (with ELikert meaning the re-scaled to interval $0-10$ usual integer encoding $1-4$ Likert responses – i.e., $\mathbf{A1} \equiv 0$, $\mathbf{A2} \equiv 10/3$, $\mathbf{A3} \equiv 20/3$, and $\mathbf{A4} \equiv 10$ – and FRS meaning the fuzzy rating scale-based ones):

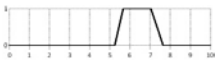
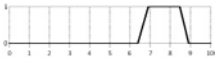
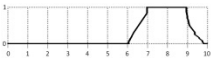
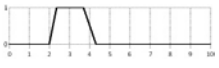
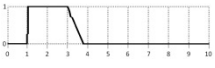
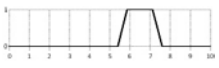
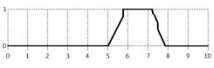
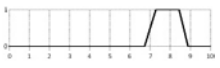
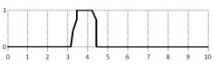
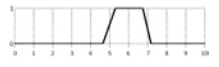
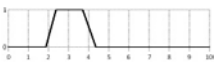
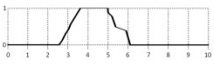
	ELikert mean	FRS Aumman-type mean	Likert median	FRS 1-norm median	ELikert variance	FRS *D*-variance	*# Valid responses*
R.1	6.3738		Agree a little		6.2191	4.7650	*68*
R.2	8.2099		Agree a lot		4.8236	3.1600	*67*
R.3	2.1885		Disagree a lot		10.2874	8.2858	*67*
M.1	6.5672		Agree a little		9.4243	7.0894	*67*
M.2	8.3341		Agree a lot		6.2381	5.2719	*66*
M.3	5.8935		Agree a little		16.3918	12.2326	*69*
S.1	6.2572		Agree a little		9.9034	6.6063	*65*
S.2	2.6553		Disagree a little		7.1199	5.1205	*64*
S.3	3.9392		Disagree a little		12.2823	8.2269	*66*

Table 2: Sample mean, median and variance of the fuzzy rating scale- and the, re-scaled to 0−10, 4-point Likert scale-based response along different questions

Conclusions from the estimation analysis in Table 2 (which could be also viewed as a descriptive summary since no specific comments have been made on some of the valuable properties held by the estimators, like consistency and others) are more informative and clarifying when using the fuzzy rating scale. This is due to the ability of the latter scale to distinguish the outputs in a more visible way. For instance, we can notice that

- for the two columns regarding means (the ELikert and the Likert-based responses), one can see that two very close mean responses with respect to the first one (e.g. those for M.3 and S.1) are not so 'close' with the second; or the answers to the same question are not very close when we vary the scale (e.g. those for R.3);
- for the two columns concerning the median, the conclusions are even more evident; one can look, for example, what happens with the median responses to M.3 and S.1: they coincide for the Likert scale-based responses, but they are quite different for the fuzzy rating scale ones;
- the last two columns concerning variance corroborate what has been empirically asserted by De la Rosa de Sáa *et al.* (2015). Although the fuzzy rating scale incorporates a much larger diversity of values, the mean squared deviation is substantially reduced in passing from the Likert to the fuzzy rating, as one can expect because of using a continuum.

Testing the equality of some means

To examine the influence of the factor 'mark in the scale $0-10$ taken in the last exam' (the factor acting at the 4 possible levels given by intervals $G1 = [0,6]$, $G2 = (6,8]$, $G3 = (8,9]$, $G4 = (9,10]$) on the response to each of the 3 questions posed for each of the 3 considered curricular areas, conclusions will be drawn on the basis of the available sample of 69 students. Results for both possible types of responses are gathered in Table 3, and they will be detailed a bit more in Table 4.

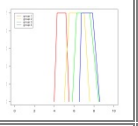
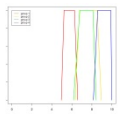
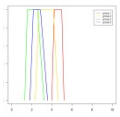
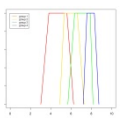
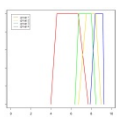
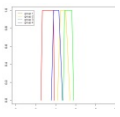
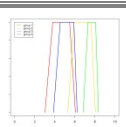
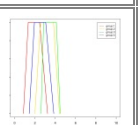
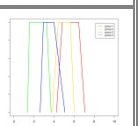
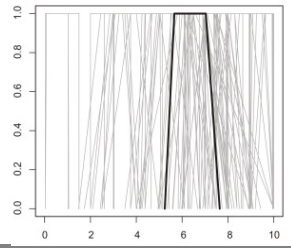
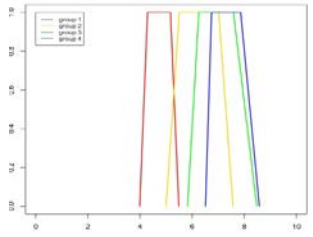
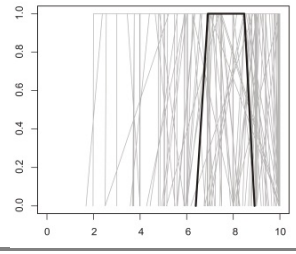
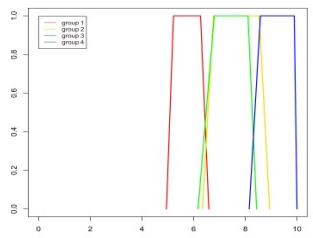
QUESTION	KW Likert *p*-value	FRS means	FRS *p*-value
R.1	.026*		.084
R.2	.012*		.000***
R.3	.045*		.100
M.1	.005**		.000***
M.2	.167		.000***
M.3	.462		.067
S.1	.008**		.001**
S.2	.457		.030*
S.3	.006**		.001**

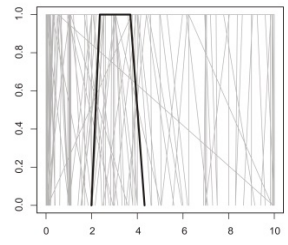
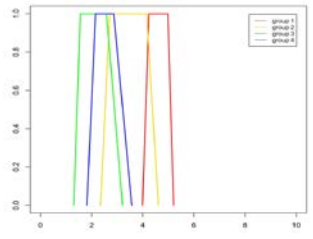
Table 3: *p*-values in testing the equality of mean responses for 4 different levels of the mark taken in the last exam

	FRS data	FRS group means	FR *p*-value	Likert data	KW *p*-value	ELikert means	ELikert *p*-value
R.1			.084		.026*	ELikert-mean $G1 = 4.67$ ELikert-mean $G2 = 6.41$ ELikert-mean $G3 = 6.67$ ELikert-mean $G4 = 8.10$	.128

	A1	A2	A3	A4	NA
G1	0	3	2	0	0
G2	0	5	18	3	1
G3	0	2	10	2	0
G4	1	0	1	5	0
NA	2	3	8	3	0

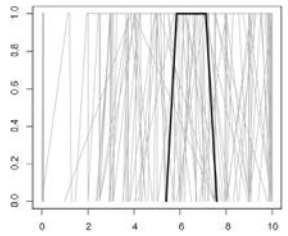
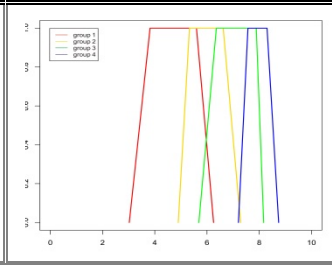
	FRS data	FRS group means	FR *p*-value	Likert data	KW *p*-value	ELikert means	ELikert *p*-value
R.2			.000***		.012*	ELikert-mean $G1 = 6.00$ ELikert-mean $G2 = 8.46$ ELikert-mean $G3 = 7.18$ ELikert-mean $G4 = 10.00$	.000***

	A1	A2	A3	A4	NA
G1	0	2	2	1	0
G2	0	1	10	15	1
G3	0	3	5	5	1
G4	0	0	0	7	0
NA	0	0	7	9	0

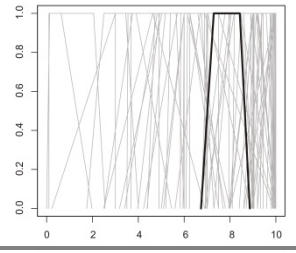
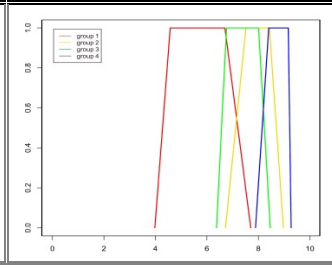
	FRS data	FRS group means	FR *p*-value	Likert data	KW *p*-value	ELikert means	ELikert *p*-value
R.3			.100		.045*	ELikert-mean $G1 = 4.67$ ELikert-mean $G2 = 2.80$ ELikert-mean $G3 = 1.67$ ELikert-mean $G4 = 0$	.000***

	A1	A2	A3	A4	NA
G1	2	1	0	2	0
G2	11	8	5	1	2
G3	9	4	0	1	0
G4	7	0	0	0	0
NA	11	3	0	2	0

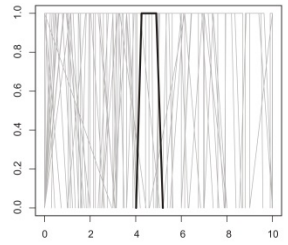
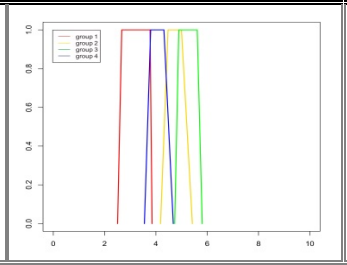
Table 4: Fuzzy rating scale-based datasets (in black the global sample mean), means and *p*-values in testing the equality of mean responses for different levels of the mark taken in the last exam, Likert scale-based datasets, 0−10 re-scaled means and *p*-values in testing the equality of mean responses for different levels of the mark taken in the last exam

	FRS data	FRS group means	FR *p*-value	Likert data	KW *p*-value	ELikert means	ELikert *p*-value
M.1			.000***		.005**	ELikert-mean $G1$ = 4.28 ELikert-mean $G2$ = 5.74 ELikert-mean $G3$ = 7.50 ELikert-mean $G4$ = 8.17	.000***

	A1	A2	A3	A4	NA
G1	1	4	1	1	0
G2	0	6	11	1	1
G3	0	5	2	9	0
G4	0	3	5	12	0
NA	2	2	1	1	1

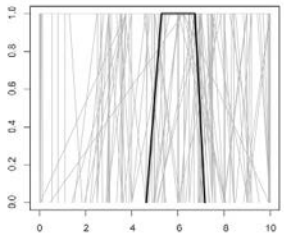
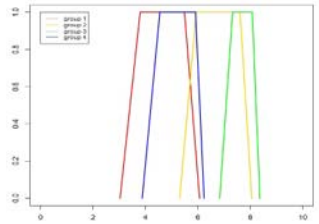
	FRS data	FRS group means	FR *p*-value	Likert data	KW *p*-value	ELikert means	ELikert *p*-value
M.2			.000***		.167	ELikert-mean $G1$ = 6.67 ELikert-mean $G2$ = 8.52 ELikert-mean $G3$ = 7.71 ELikert-mean $G4$ = 8.95	.006**

	A1	A2	A3	A4	NA
G1	0	2	3	2	0
G2	0	1	6	11	1
G3	2	0	5	9	0
G4	0	1	4	14	1
NA	0	0	1	5	1

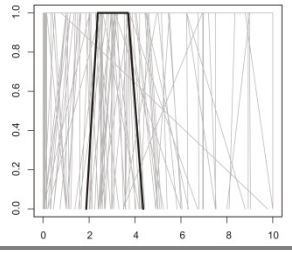
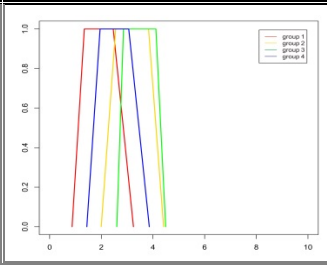
	FRS data	FRS group means	FR *p*-value	Likert data	KW *p*-value	ELikert means	ELikert *p*-value
M.3			.067		.462	ELikert-mean $G1$ = 4.28 ELikert-mean $G2$ = 5.09 ELikert-mean $G3$ = 6.46 ELikert-mean $G4$ = 6.50	.068

	A1	A2	A3	A4	NA
G1	2	3	0	2	0
G2	4	6	4	5	0
G3	3	3	2	8	0
G4	5	2	2	11	0
NA	1	1	2	3	0

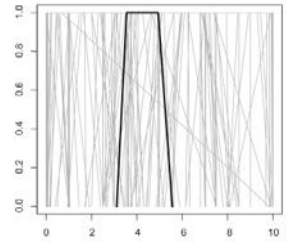
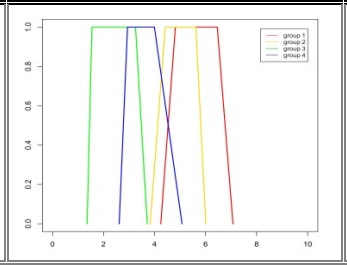
Table 4 (continuation)

	FRS data	FRS group means	FR *p*-value	Likert data	KW *p*-value	ELikert means	ELikert *p*-value
S.1			.001**		.008**	ELikert-mean $G1 = 4.07$ ELikert-mean $G2 = 7.10$ ELikert-mean $G3 = 8.00$ ELikert-mean $G4 = 5.21$	.000***

	A1	A2	A3	A4	NA
G1	2	3	4	1	0
G2	1	2	13	7	1
G3	0	2	2	6	0
G4	3	3	8	2	1
NA	1	2	1	3	1

	FRS data	FRS group means	FR *p*-value	Likert data	KW *p*-value	ELikert means	ELikert *p*-value
S.2			.030*		.457	ELikert-mean $G1 = 1.25$ ELikert-mean $G2 = 2.88$ ELikert-mean $G3 = 3.00$ ELikert-mean $G4 = 2.35$	.056

	A1	A2	A3	A4	NA
G1	6	3	0	0	1
G2	8	10	3	1	2
G3	3	6	0	1	0
G4	8	6	3	0	0
NA	2	2	3	0	1

	FRS data	FRS group means	FR *p*-value	Likert data	KW *p*-value	ELikert means	ELikert *p*-value
S.3			.001**		.006**	ELikert-mean $G1 = 6.00$ ELikert-mean $G2 = 4.64$ ELikert-mean $G3 = 1.33$ ELikert-mean $G4 = 2.94$	.000***

	A1	A2	A3	A4	NA
G1	0	5	2	3	0
G2	5	7	8	3	1
G3	7	2	1	0	0
G4	8	4	4	1	0
NA	2	1	1	2	2

Table 4 (continuation)

Table 3 includes, in addition to the p-values obtained with the well-known Kruskal-Wallis non-parametric test applied on the Likert scale-based responses, the Aumann-type sample means of the fuzzy rating scale-based (FRS) responses for each of the 4 groups ($\boldsymbol{G1}$, $\boldsymbol{G2}$, $\boldsymbol{G3}$ and $\boldsymbol{G4}$) and the corresponding p-values approximated by the bootstrap method in González-Rodríguez *et al.* (2012).

The preceding tests indicate that marks affect more associated questions in Science than in Math, and in Math than in Reading. From Table 3 one can easily conclude that statistical results often differ, and one cannot state a general assertion about how this is made. As distances between variable values are usually bigger between the integer encoding of the Likert responses than between the fuzzy rating ones, the variation achieved in the first case is almost always bigger too. Hence, since most of the considered statistics involve standardization, no general conclusion can be established, but of course the FRS column in Table 3 exploits more expressive and rich information than the KW Likert one.

Table 4 displays not only these computations, but also some additional illustrative information. For each question, the following outputs have been specified in Table 4:

- the graphical display of the fuzzy rating scale-based data for each of the 9 questions (in grey) and the global sample mean (in black),
- the graphical display of the (sample) Aumann-type mean of the fuzzy rating scale-based response for each of the four groups $\boldsymbol{Gi}$ ($i = 1,2,3,4$) and each of the 9 questions,
- the p-values obtained from the ANOVA especially developed for fuzzy number-valued data (FR p-values),
- the sample frequency distribution of the ELikert responses for each of the four groups $\boldsymbol{Gi}$ ($i = 1,2,3,4$) and each of the 9 questions,
- the p-values obtained from the Kruskal-Wallis test (KW p-values),
- the (sample) means of the re-scaled to 0-10 integer encoding of the Likert responses for each of the four groups $\boldsymbol{Gi}$ ($i = 1,2,3,4$) and each of the 9 questions.

A different grouping has been considered to verify the importance of the grouping. Indeed, this affects the final conclusions, so that if instead of the preceding 4 levels, we consider 2 levels, $G'1 = [0,8]$, $G'2 = (8,10]$, we get the p-values in Table 5, the Likert case being examined through the Man-Whitney-Wilcoxon test (MWW).

QUESTION	MWW Likert p-value	FRS p-value
R.1	.065	.099
R.2	.738	.416
R.3	.016*	.081
M.1	.001**	.001**
M.2	.326	.253
M.3	.118	.762
S.1	.880	.825
S.2	.827	.733
S.3	.002**	.012*
* p < .05, ** p < .01, *** p < .001		

Table 5: p-values in testing the equality of mean responses for 2 different levels of the mark taken in the last exam

An alternative study has been considered to verify the influence of the strength in Math&Science background on the given responses. The analysis consider other two levels, for the average mark taken in the last exams for Math and Science, the groups corresponding to the averages ranging over $G^*1 = [0,8.5]$, $G^*2 = (8.5,10]$. The p-values are collected in Table 6, the Likert case being examined through the Man-Whitney-Wilcoxon test (MWW).

QUESTION	MWW Likert p-value	FRS p-value
R.1	.843	.747
R.2	.834	.883
R.3	.513	.644
M.1	.025*	.027*
M.2	.121	.037*
M.3	.648	.760
S.1	.608	.768
S.2	.820	.392
S.3	.018*	.016*
* $p < .05$, ** $p < .01$, *** $p < .001$		

Table 6: p-values in testing the equality of mean responses for 2 different levels of the average mark taken in the last exams Math and Science

Multiple two-sample independent comparisons have been performed in association with the 4 level grouping, the conclusions for the Likert scale-based responses being based on the Mann-Whitney-Wilcoxon pairwise comparison test. The outputs have been collected in Table 7

	p-value					
	$G1 \leftrightarrow G2$	$G1 \leftrightarrow G3$	$G1 \leftrightarrow G4$	$G2 \leftrightarrow G3$	$G2 \leftrightarrow G4$	$G3 \leftrightarrow G4$
R.1	.115	.107	.073	.747	.043*	.094
R.2	.071	.443	.018*	.187	.090	.024*
R.3	.516	.298	.106	.239	.023*	.197
M.1	.198	.055	.011*	.075	.004**	.626
M.2	.141	.308	.063	.695	.538	.350
M.3	.651	.308	.341	.317	.270	.937
S.1	.013*	.013*	.388	.343	.079	.041*
S.2	.185	.203	.406	.952	.604	.639
S.3	.363	.002**	.031*	.010*	.126	.264
*$p < .05$, **$p < .01$, ***$p < .001$						

Table 7: p-values in the pairwise comparisons of Likert responses for levels of the mark taken in the last exam

whereas, by using the two-sample procedure in Montenegro *et al.* (2001), we have obtained the pairwise comparison of the fuzzy rating scale-based responses gathered in Table 8.

	p-value					
	$G1 \leftrightarrow G2$	$G1 \leftrightarrow G3$	$G1 \leftrightarrow G4$	$G2 \leftrightarrow G3$	$G2 \leftrightarrow G4$	$G3 \leftrightarrow G4$
R.1	.091	.059	.144	.300	.512	.766
R.2	.106	.242	.047*	.427	.000***	.011*
R.3	.625	.217	.414	.078	.506	.821
M.1	.169	.046*	.002**	.203	.010*	.222
M.2	.010*	.063	.001**	.578	.225	.138
M.3	.323	.208	.558	.667	.512	.351
S.1	.007**	.008**	.557	.283	.073	.024*
S.2	.090	.145	.612	.915	.204	.308
S.3	.372	.003**	.059	.012*	.194	.236
* $p < .05$, ** $p < .01$, *** $p < .001$						

Table 8: *p*-values in the pairwise comparisons of the fuzzy means of levels of the mark taken in the last exam

From Tables 7 and 8, along with the results from Tables 3 and 4, one can easily conclude that statistical results differ due to the fact that the use of the fuzzy rating scale and related methodology exploit more expressive and rich information than the ordinal one. Anyway, although one cannot give a general comparative assertion, one can say that when there exist very significant differences with the two scales, these differences are mostly more clearly detected for the fuzzy rating scale-based data than for Likert's ones.

Also from Table 7, one can deduce in connection with the six questions showing lowest *p*-values that:

- ✓ for question R.2 ("I learn a lot from reading"), the main significant differences are those associated with the comparison between the group of students with the highest marks ($G4$) and any of the other three groups of students;
- ✓ for M.1 and M.2 ("I like Maths" and "My Maths teacher is easy to understand"), the main differences correspond to those between the group of students with the highest marks and the two groups of students with lowest marks ($G1$ and $G2$);
- ✓ for S.1, S.2 and S.3, that is, the questions related to Science, the clearest differences are those shown between the students with rather high marks ($G3$) and the groups with either the lowest or the highest marks ($G1$ and $G4$).

Alternative analyses have been developed by pairwise (since there are only two levels per 'factor') comparing responses to the 9 questions above with respect to the filled questionnaire format, sex and room independence.

To test the influence of each of the three previous factors on the response to the three questions posed for each curricular area, on the basis of the considered sample of 69 students, two-sample tests have been carried out, and conclusions have been gathered in Tables 9, 10 and 11.

QUESTIONNAIRE FORMAT	FRS *p*-value	Likert Mann-Whitney-Wilcoxon *p*-value
R.1	.135	.886
R.2	.051	.452
R.3	.601	.105
M.1	.311	.950
M.2	.014*	.572
M.3	.069	.001**
S.1	.032*	.445
S.2	.615	.707
S.3	.083	.256

Table 9: *p*-values in testing the equality of the mean response between the two formats for the filled questionnaire

SEX	FRS *p*-value	Likert Mann-Whitney-Wilcoxon *p*-value
R.1	.643	.596
R.2	.653	.582
R.3	.388	.591
M.1	.052	.080
M.2	.620	.748
M.3	.440	.347
S.1	.262	.303
S.2	.566	.760
S.3	.794	.774

Table 10: *p*-values in testing the equality of the mean response between girls and boys

ROOM INDEPENDENCE	FRS *p*-value	Likert Mann-Whitney-Wilcoxon *p*-value
R.1	.446	.032*
R.2	.038*	.016*
R.3	.751	.647
M.1	.177	.245
M.2	.678	.909
M.3	.473	.127
S.1	.803	.763
S.2	.710	.910
S.3	.455	.105

Table 11: *p*-values in testing the equality of the mean response between having or not own room

Whereas we saw before that the mark taken in the last exam has a clear influence on the responses to questions related to reading, math and science, most of differences due to the questionnaire format, sex and room independence are not significant for the usual significance levels.

The study has served not only to show the potential of the fuzzy rating scale and the considered methodology to analyze data based on it, as well as the differences in statistical conclusions depending on the considered scale.

Furthermore it confirms that, although not being as immediate to fill as Likert-based questionnaires – what makes it not suitable for being conducted in certain frameworks (like on the street, by phone, etc.) –, the required training and background to answer fuzzy rating scale-based questionnaires are not really deep. Therefore, it is an especially advisable scale when one aims to have more accurate and informative conclusions.