Optimization of Data Processing and Presentation in Social Surveys: From Likert-Means to "Yes Percentage"

Gediminas Merkys[1(✉)] and Daiva Bubeliene[2]

[1] Vytautas Magnus University, 44248 Kaunas, Lithuania
kadagysgm@gmail.com
[2] Kaunas University of Applied Sciences,
Pramones pr. 20, 50468 Kaunas, Lithuania
daivabubeliene@gmail.com

Abstract. The article presents a replication study of a significant empirical study carried out earlier. Borg, Gabler (2002) showed that there is an extremely high correlation between the Means of Likert Scale Items and the percentage of agreement (Yes%). On this basis, they suggested to use not traditional Likert mean, but Yes % in survey reports. The latter are easier to interpret and do not have the same problem as Equidistance of Likert-type scales. It was decided, based on the big date (N ≈ 9000), to carry out the replication study in another historical time, in another culture and measuring another construct. If the statistical regularity detected by Borg, Gabler is repeated, it is universal, then it is really appropriate to move to a wider use of Yes% when preparing the survey report. The replication study showed that there is an extremely high correlation ($R^2 = 0.948$) between the primary Likert items means and Yes% of items, approximating to the linear function. Used the classic 5-grade Likert scale. The verification is carried out only at the level of single items without passing to the level of analysis of additive indexes. It also turned out that the Likert items-mean correlation with the No% is lower ($R^2 = 0.865$), which negates the postulate of symmetry of the scale. In addition, the Likert means correlation with "neutral category%" is even lower - $R^2 = 0.340$. When preparing Survey studies reports, give priority to "yes %" instead of Likert items means.

Keywords: Likert scale · Equidistance · Agreement percentage · Correlation · Reports of social survey

1 Introduction: Theoretical and Practical Context of the Study

Social sciences have been using mathematical methods for a long time. The use of measurement, experiment and mathematical modelling is one of the reasons why the social sciences have diverged from the humanities and became independent fields of study. Social surveys are one such area where mathematical-statistical methods are applied. However, the methodology employed in the creation and the use of surveys is still characterised by the lack of development of any original mathematical models of

N. Agarwal et al. (Eds.): MSBC 2019, CCIS 1079, pp. 12–25, 2019.
https://doi.org/10.1007/978-3-030-29862-3_2

the social phenomenon. Data are simply analyzed by well-known and accepted statistical tools such as normal distribution or chi-squared and nonparametrical functions, linear regression, factor and cluster analysis, latent class analysis, Rasch-model etc. Recently more mathematically-inclined social scientists have started to employ the elegant mathematical apparatus in their questionnaires and tests which allows them to comprehensively model the behaviour of the survey response. Such attempts fall under the term Item Response Theory (IRT), [8, 11, 25].

Evidently, mathematical-statistical methods must be applied correctly such that the conditions of the model used are satisfied. Contrary to soft sciences, where many observations can be explained away by pluralism and a multitude of paradigms, the use of mathematics and statistics remains strictly in the domain of hard sciences. If a mistake is made in the model, then the output will increase it tenfold simultaneously masking it under the authority of mathematics. Thus, the moral is simple: either mathematical-statistical methods are not to be used at all or they ought to be used correctly. The point is driven further by the fact that fully legitimate qualitative methods are spreading massively in the practice of social research [2, 19].

This paper wishes to attract attention to one specific problem of applied statistics: the use of some classical statistical parameters of interest, widely used and accepted in social surveys, is of questionable correctness. We hope that such doubt will encourage a discussion and possibly even the revision of established practices and views. In particular, we address the massively-used Likert scale, named after the American psychologist Likert (1903–1981) [20]. The aforementioned scale is universal and easy to use. These are probably the reasons for the non-decaying popularity and application of the scale to this day.

The respondents are given a set of sensible statements (psychometric stimuli) and are asked to register their answers in a uniform format next to each statement. The uniform format encompasses several gradations of the answers: from complete agreement to complete disagreement. The most common categories are: “I strongly disagree”, “I disagree”, “I have no opinion”, “I agree”, “I strongly agree”. The classical version of the scale has five levels of gradation where two correspond to agreement, two to disagreement and the middle category corresponds to a neutral answer “no opinion”. The answers are then quantified by agreement into scores from 1 to 5 or from 0 to 4.

Over time many more modifications of the scale were introduced. Such as scales that have the middle neutral category removed and scores encoded from 1 to 4 or from 1 to 6 points, or the version with 7 categories including the neutral one [7, 14, 15, 21]. Nevertheless, despite the number of categories, the distinguishing attributes of the Likert scale are its bipolarity and symmetry. The respondents are given an equal number of positive and negative (of agreement and disagreement) gradations.

Practice shows that researchers sometimes work with estimates derived from single statements. However, usually an additive index is computed from the answers of multiple statements that are unified by a common theme. Either the sum or the mean of the scores is found. In this case we speak of psychometric approach. It is postulated that the given diagnostic statements are of equal weight chosen randomly from the set of all statements. The estimates of additive indices comprise a manifest variable which is then used to indirectly infer the latent variable - the consciousness-construct of the individual or group in question [13, 26].

In an ideal case the calculated additive index must correctly differentiate the subjects by the latent construct. The creation of the additive index is controlled by psychometric statistics tools, reliability control, factor validation and, currently, Item Response Theory. Yet even in respectable research a statement or an estimate derived from it is taken for granted. Notwithstanding, the validity of a psychometric test begins with the validity of a single item.

Statements in Likert scale can be formulated about a subject's inner mental state or behaviour patterns, in which case we speak of self-evaluation. That is the case in all clinical or self-concept questionnaires, research on happiness, personnel questionnaires (on work motivation, job satisfaction) and so on. Further, statements can be on: (a) concrete objective outside objects - a brand logo, holiday destination, political entity etc. (b) on views and attitudes on universal values etc. Nowadays we could hardly find a thematic direction in social research and social surveys where the Likert scale is not used. It is used in clinical, work and organizational psychology, pedagogical psychology, socialization research, educational research, surveys on values, healthy lifestyle, politics and consumer behaviour etc. Unsurprisingly, many papers are published on the usage of Likert scale, a methodological discourse is ongoing [5, 12, 16, 18].

A correct use of a Likert scale requires that a few formal conditions be satisfied which, unfortunately, are not always guaranteed in practice: the statistical independence of statements (items). If we permute the statements, that is, change the order in which the statements are given in a survey, it should, ideally, not have any impact on the respondent's measured opinion or view in the result. This problem can easily be solved for online surveys. Every time a survey window is opened, a random number generator produces a random order of the statements for each subject. Then the data fall into a matrix by a predefined order. Lamentably, many social survey platforms do not have this function, even though they should [1].

However, there is one problem of the Likert scale that has not been solved to this day. It may happen that it never will be completely. It is obvious that the 5 or 7 grade scale is a typical ordinal scale. In this case equal distances between the grades of the scale, which is something that social scientists have been dreaming about for a long time and what an interval scale formally requires, cannot be guaranteed in any way. There is no such equal grading in a Likert scale, it is simply postulated. Thus, the estimates of a Likert scale do not differ in any way from, say, a quasi-scale of school grades which was not taken seriously even at the turn of 19th - 20th centuries. Hence there is a real problem that is usually ignored. It is not entirely correct to compute the mean and the variance - the parameters of a normal distribution - from an ordinal scale. Unfortunately, by cloaking behind the equal grading postulate and the established practice ("it has always been done this way"), the social survey reports, papers and dissertations give the means and standard deviations of the data in a Likert scale. This is done with single-item estimates as well as the estimates of additive indices.

The aforementioned problem of using ordinal data as if it were of an interval is not the only one. Quite honestly, this problem can even be relaxed. We can assume that the measured and existing latent constructs (characteristics of the human psyche), their expression, can in principle be properly measured quantitatively on an interval or even a ratio scale. It is just that currently the social and behavioural sciences do not yet have the necessary technological capabilities. The development of neurosciences shines

expectantly in this context. Even now there are data, bold hypotheses, that the content of human social attitudes and their expression has a connection with deep brain processes [6, 17].

The latter, as is known, can already be observed by objective study methods of brain activity or neurochemical reactions. Maybe one day a reliable bridge between traditional surveys and the objective data of neuroscience will be established?

In principle, it might be possible to postulate that the Likert scale generally reflects the measured characteristics of a human mind, it just does it approximately, in a slightly distorted way. Some mathematical formalities are not held. In the spirit of the history of natural sciences we could say that Pascal's mercury thermometer is also an imprecise measurement device since some of the heat energy is consumed by the thermometer itself. If science could not measure something exactly, then it persistently and consistently tried to do it approximately. And there is no constructive alternative to this attitude of scientists. Analogously it is possible to tolerate the limitations of the Likert scale and it is no secret that the majority of researchers do. There is another argument in favour of the Likert scale. It is known from the central limit theorem in probability theory and statistics that if one has a large enough sample, then the sample mean converges in distribution to a normal distribution [9]. It is constantly encountered that in the statistical analysis of the survey data. If an additive index of a Likert scale is formed from at least 15–20 primary items and the number of studied subjects is $N > 500$, then the resulting distribution is close to a normal distribution.

Thus, in this context the authors of this paper have raised the problems of Likert scale's equal distances and normality, they have likewise vindicated its use. It is possible to discuss the impact that these problems have on the quality of data analysis of social surveys. The existence of the two aforementioned problems cannot be denied. Whatever the case, besides the two discussed Likert scale problems, there exists yet another problem of no less importance. It, apart from a few exceptions, is not even discussed. We shall look at this problem in more detail.

No measurement in the practice of life, nor in science, is an end in itself. The measurement procedure is always a part of some concrete meaningful theoretical context. It means that the result of a measurement is always interpreted. The most universal definition of a measurement is this: the assignment of numerical values to an object, phenomenon or a relation between them. The point is that the assigned numerical value must be interpreted. Ideally, the zero, the minimum and the maximum values of the scale ought to have an objective meaning. In the absence of such objectivity the zero of a scale can be chosen by convention.

Psychometric measurements are ontologically and epistemologically probably most closely related to biometric measurements. The latter have a big advantage since the norm of a given value is more or less known to everyone. Suppose we have the values of a person's weight, height, blood pressure, blood sugar level and various other parameters. Their essence is easily understood and can immediately be sensibly interpreted. The same can be said of measurements that are based on an accepted currency (in Euros, US Dollars). If we needed to interpret the measurement results based on Venezuelan Bolivar (Fuerte Bolivar), whose nominal value drastically changes every day, then we would immediately face difficulties interpreting. The result of a given measurement would not be so clear outright. The problem is that the

estimates of a Likert scale, from the point of view of measurement interpretation, are more complicated than the Venezuelan Bolivar.

If a simple additive index is constructed, then its score depends on the number of statements involved and on the number of possible answers next to each statement. The more primary items and the more categories there are, the larger the maximum possible value of the additive index will be. The authors of this paper has once had to form an additive Likert scale index which was composed of 160 (!) primary items and five categories. Cases where an additive Likert index is comprised of 15–20 primary items are encountered tens of thousands of times in the practice of social surveys. In the prior case we speak of a widely used psychometric inventory "safe school - safe child" which addresses older students [4, 22].

From the logical and factor analysis of 160 primary items 38 subscales are formed. Secondary factor analysis shows that we can freely form a total score. This way the raw total score can range from 160 to 800. If computed an average should fit between 1 and 5 or 0 and 4 points, depending on the encoding. As is known, the raw score and an average are related by a linear transformation on a scale while the correlation coefficient between these two values is always one.

The bounds of the scale's range are not entirely clear. In the case of a Likert scale neither the researcher nor the client who commissioned the survey, nor the consumer is sure what the computed value of an estimate from a single individual or the mean of a group really means. Of course, it is possible to see whether the estimate of an individual or a group has crossed (or not crossed) the formal zero reference point of the scale. The problem will be reduced if the raw scores are transformed into the z-score of a standard normal distribution. We encourage social scientists to do this since then by inductive statistical explanation we could at least find a temporary zero of the scale that corresponds to the mean of the standardized sample. With each increment of the standardized sample this zero becomes more and more accurate. But this is a topic for another paper (on the necessity to apply statistical standardization more widely when presenting the results of surveys).

Whatever the case, a question arises: why do social scientists so often report the mean and the standard deviation from raw Likert scores? Such estimates, first of all, are not entirely correct from a purely formal mathematical perspective and, secondly, are hard to interpret in practice. Is this due to tradition? It is clear that this tradition is not entirely adequate if not to say malicious.

Some researchers have taken notice that the results of a Likert scale are best presented in the form of "agreement percentage" or "yes percentage" [3]. For clarity the so called "yes %" on a five-grade Likert scale is found by summing over the response frequency of the last two categories (corresponding to agreement). Borg, Gabler (2002) have done statistical reanalysis of four large surveys, each having a sample of tens of thousands. The surveys were on personnel and organizational psychology. The authors gave empirical proof that there exists a tight linear relationship ($r^2 \approx 0.98 \approx 0.99$) between Likert means and the yes percentages approximating a functional dependence. Such strong repeated systemic relationship practically means that Likert mean and "yes %" are essentially the same thing. Well, formally speaking, the values are not the same, however, it is obvious that what they measure or reflect is common to both. This conclusion opens an innovative overturn in the process of social survey result analysis

and, especially, in their presentation to the professional or public audience. What is this optimistic interpretation based on? Contrary to the mean of a Likert scale, an estimator such as agreement percentage is impeccable from a formal mathematical perspective. The data arranged in a non-decreasing order is a manifest variable that has an objective origin. It shows which percentage of subjects reacted to a written stimulus (statement) with agreement. Yes % forms a dichotomic scale "yes/no". As is known, a dichotomic scale satisfies the conditions of a "higher-level" scale - an interval scale. Furthermore, if a sample is large enough, the relative data are approximately normally distributed, See: Moivre-Laplace Theorem and Central Limit Theorem, (Normal Approximation to the Binomial Distribution) [10].

Moreover, if 50% of respondents in a population X and 25% of respondents in a population Y agree with a statement-stimulus "A", then we are justified in performing the operation of division and saying that one population has twice the frequency of another population. We can let go of the notion of frequency and claim that the quantitative expression of a measured attribute in a population X is expressed twice as strongly than in population Y. In this instance a meaningfully applied operation of division would even let us talk of a ratio (absolute) scale. On the contrary, we could not say anything like that for a ranked scale. For example, if on a 10-point scale of high-school grades where 10 is the highest mark one student receives a score of 10 and another a score of 5, we cannot conclude that one knows twice as much as the other.

It is especially important that the agreement percentage is very simple, clear and easy to interpret for the researcher, the client who commissioned the survey and its results for the public consumer. It is characteristic of most social surveys that their results do not lie dormant in the archives of the researchers or the university, but falls into public discourse through mass media and forms the societies' self-concept and public opinion. It is an important attribute of a democratic society and a public good. Note that there are not social surveys in China, North Korea, Turkmenistan and other autocratic countries. We can say that the conversion of mathematically expressed survey results into layman's language is a public good.

Let us return to the inventory example "Safe school - safe child". It has a subscale on "Conflict&Aggression in the school yard" which is composed of 9 primary items. It is pointless to give the raw score of this subscale since it is uninterpretable. The averages of single Likert items are similarly senseless. There is a possibility to use "yes percentages" and to express the result of the measurement in a much more informative fashion. For instance, to 9 stimuli, indicating the lack of safety in the school yard, the respondents reacted with an average agreement reaching 13%. Individual scores of the 9 indicators of unsafe school yard environment varied between 7,9 and 19,1%.

Whatever the case, the doubtful practice of submitting Likert scale means in the reports of social research is not waning. On the other hand, attempts to use "yes %" in the report do not catch on.

The author's of the paper has accumulated a vast archive of social surveys. Some of which may be considered as big data. It was decided to do something similar to a repeated study. We wanted to see how universal the statistical law of a very high correlation between Likert means and yes % that Borg, Gabler noticed in 2002 was. How does this law hold in a different historical context, in a different culture, measuring a different (not workplace psychology) construct? If the repeated study could

find identical or similar results, we would justifiably speak of a universal statistical law whose knowledge and application could help in the development of social survey statistical analyses and methodological culture.

2 Methods

The total sample contains 9 thousand subjects. The latter have been questioned in Lithuania (2016–2019) by standard survey techniques. A unified 5-grade Likert scale with graphical elements was used (Table 1). The survey asked citizens if they were satisfied/dissatisfied with public services.

The analysis was done on single items as well as on the level of psychometric scales. The large sample size is a result of the survey being given in 9 municipalities. The number of respondents in the separate samples (separate municipalities) reached 1 thousand subjects. The survey instrument was developed since 2002. It and its use can be found in [24].

Table 1. An example of a Likert scale Items used

EVALUATE					**Very bad – very good**
Street management under critical weather conditions (freezing, strong winds)					
Management of remote roads and drive-ways under critical weather conditions					
Street, sidewalk, road repairs (road pit fixes, asphalting)					
Road and street traffic sign condition					
Street lighting during dark daytime					
Evaluation categories	**Disfavorable** (Bad/Somewhat bad)		**Neutral evaluation**/ Don't know	**Favorable**(Somewhat good/Good)	
Evaluation encoding scores	1	2	3	4	5
Survey response scale					

As is customary when a survey is done by public funds and for public needs, the reports of the applied study were (and still are) distributed freely on the internet [23]. They questionnaire contains around 190 primary variables which correspond to various indicators of public service: utilities, public transportation, environmental protection and recreation, education, culture, healthcare and social security, public safety and so on. Around 40 psychometric indices have been formed by logical and factor validation methods from services falling under a common theme, reflecting different areas of public service. The number of primary variables in every municipality questioned is almost identical. It may differ only by a few elements[1].

[1] For example, a particular urban centre does not contain a railway station and a hospital whose owner is the local government. Whereas the standard questionnaire includes questions on the railway station's environmental management and the services of a local hospital.

While looking for statistical dependencies it is important that the variables have a large variance. This condition is met. On a 190 variable list a natural clustering emerged. Some indicators tended to have positive evaluations and some negative, still others neutral. Out of a sample of 9 thousand a secondary database was formed consisting of 1695 items. It was composed of several variables: Likert scale means and agreement percentages of primary indicators, as well as disagreement and neutral evaluation percentages. The means of a Likert scale were defined by an independent variable - a predictor, while the "yes %" were defined by a dependent variable. Disagreement and neutral percentages were also defined as dependent variables.

Several research questions have been raised: what type of and how strong the statistical relation exists between the predictor and the dependent variables, that is between the means of primary Likert items and (a) Likert yes %; Likert no % and percentages falling under the "I have no opinion" category. The question of whether saturation of "I have no opinion" category has any impact on the strength between Likert means and yes % was also raised.

Accordingly, several main hypotheses were raised and tested:

1. There is a very strong correlation between Likert means and yes percentages approaching a linear dependence.
2. A large, saturated middle neutral category "I have no opinion" has a weak negative effect on the correlation between Likert means and "yes %".

3 Results

The means of both variables are distributed normally, both in the total sample and in the clusters of separate municipalities (Table 1, Figs. 1 and 2). It is to be expected by the central limit theorem as the sample size is large. However, the empirical distributions had to be tested. In the case of Kolmogorov-Smirnov test a natural paradox of large samples emerged: judging by the large K-S (two-tailed) significance test p-values, the normality of the sample is more strongly expressed in the separate municipality clusters rather than in the total sample. The p-value of the total sample is much smaller: for the Likert means $p = 0.052$ and in the case of yes % $p = 0.055$. Since the sample size was large, it was reasonable to ask for a large significance $p \leq 0.001$. In any case, skewness and kurtosis parameters have also been investigated. If the absolute values of these parameters do not cross 1.00, it is said by convention that the empirical distribution does not differ from the theoretical distribution significantly. Both in the total sample and in clusters small values of both parameters testify in favour of normality. Furthermore, as it should be, the total sample skewness and kurtosis parameters do not seem worse than the parameters of separate clusters.

There is a very strong linear relationship between the predictor and the dependent variable - yes % (Table 2, Figs. 3 and 4). $R^2 = 0.948$ for the total sample while for the 9 clusters it varies between $R^2 = 0.936$ and $R^2 = 0.968$. $p = .000$ in all cases.

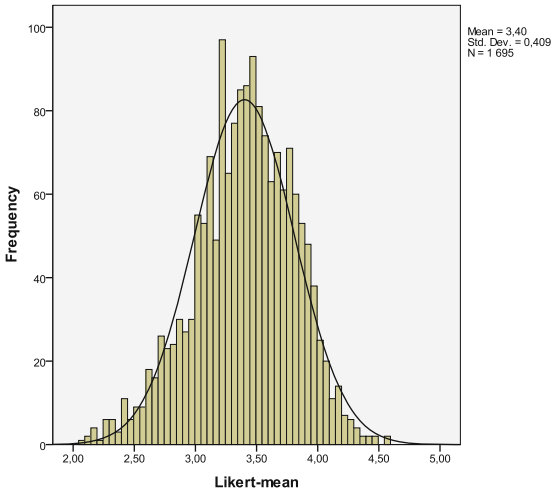

Fig. 1. Empirical distribution of Likert mean, N = 1695

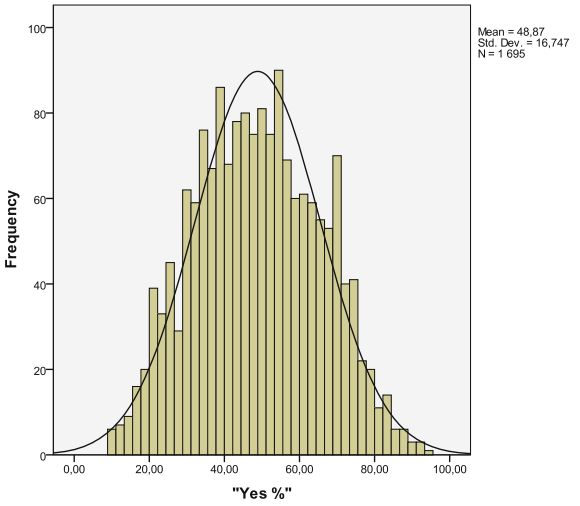

Fig. 2. Empirical distribution of agreement percentages (Yes %), N = 1695

It is symptomatic that the analogous relationship between Likert means and no% is clearly weaker. $R^2 = 0.865$ for the total sample and ranges from $R^2 = 0.772$ to $R^2 = 0.931$ for the 9 clusters, p = .000 in all cases. An even weaker relationship exists between Likert means and the neutral category "no opinion". $R^2 = 0{,}340$ for the total sample and in the 9 clusters ranges form $R^2 = 0{,}093$ to $R^2 = 0{,}622$, p = .000.

It became evident that increased saturation of the middle (neutral) category slightly weakens the relationship between Likert means and yes-%. The data corresponding to neutral evaluation frequencies were divided into groups by quartiles. The determination

Table 2. Normality tests of empirical distributions in different sample clusters and in the total sample; coefficients of determination R^2 (independent Variable- Means of Likert items, dependent variable - Yes %)

Sample	N_{sampl}	Value	K-S sig. (2 tailed)	Skewness	Kurtosis	R^2	p < 0.000
Sample 1 Ukmergė district	187	Likert mean	0,539	−0,383	0,254	0,940	****
		Yes %	0,819	0,140	−0,563		****
Sample 2 Radviliškis district	193	Likert mean	0,781	−0,188	−0,302	0,954	****
		Yes %	0,623	0,223	−0,58		****
Sample 3 Lazdijai district	190	Likert mean	0,077	−0,678	−0,302	0,951	****
		Yes %	0,738	−0,138	−0,693		****
Sample 4 Klaipėda reg. district	189	Likert mean	0,581	−0,313	−0,432	0,949	****
		Yes %	0,731	0,35	−0,887		****
Sample 5 Kedainiai district	189	Likert mean	0,794	−0,411	−0,061	0,950	****
		Yes %	0,600	0,069	−0,792		****
Sample 6 Druskininkai district	179	Likert mean	0,128	−0,408	0,223	0,958	****
		Yes %	0,058	−0,359	−0,554		****
Sample 7 Alytus district	181	Likert mean	0,176	−0,658	0,283	0,950	****
		Yes %	0,954	−0,142	−0,483		****
Sample 8 Jonava. district	195	Likert mean	0,983	−0,081	−0,243	0,968	****
		Yes %	0,870	0,123	−0,602		****
Sample 8 Jonava._2 district	192	Likert mean	0,945	0,046	−0,204	0,936	****
		Yes %	0,197	0,295	−0,603		****
Sample total	**1695**	Likert mean	**0,052**	**−0,327**	**0,054**	**0,948**	****
		Yes %	**0,055**	**0,057**	**−0,635**		****

coefficients of the first three quartiles are, respectively, $R^2 = 0.980$; $R^2 = 0.991$; $R^2 = 0.985$; they are large and close to each other. On the other hand the determination coefficient of the fourth quartile drops lower: $R^2 = 0.929$.

The fourth quartile contains the category “no opinion” which includes more than 39.1% of all frequencies. We are speaking of a rather large category of neutral answers. For reference, the various estimated parameters of neutral response frequencies: mean = 32.77; median = 32.8; mode = 31.5; min = 4.90; max = 60.40.

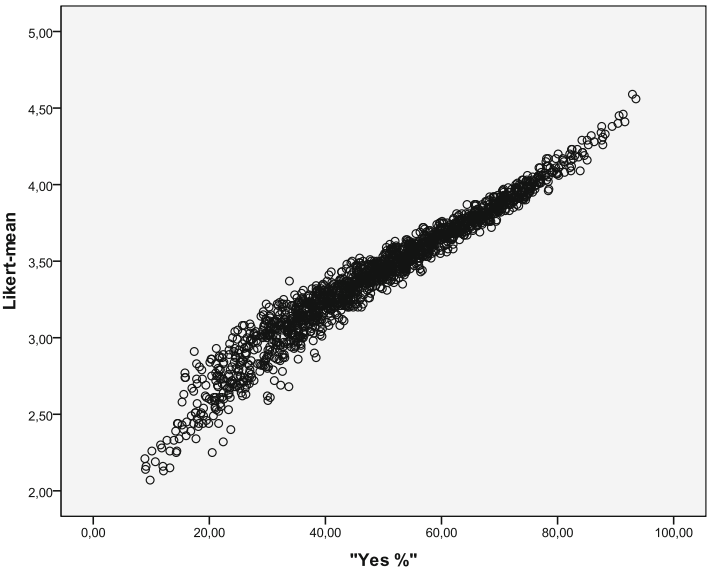

Fig. 3. Scatterplot (bivariat): independent Variable - Means of Likert items, dependent variable - Yes %), N = 1695

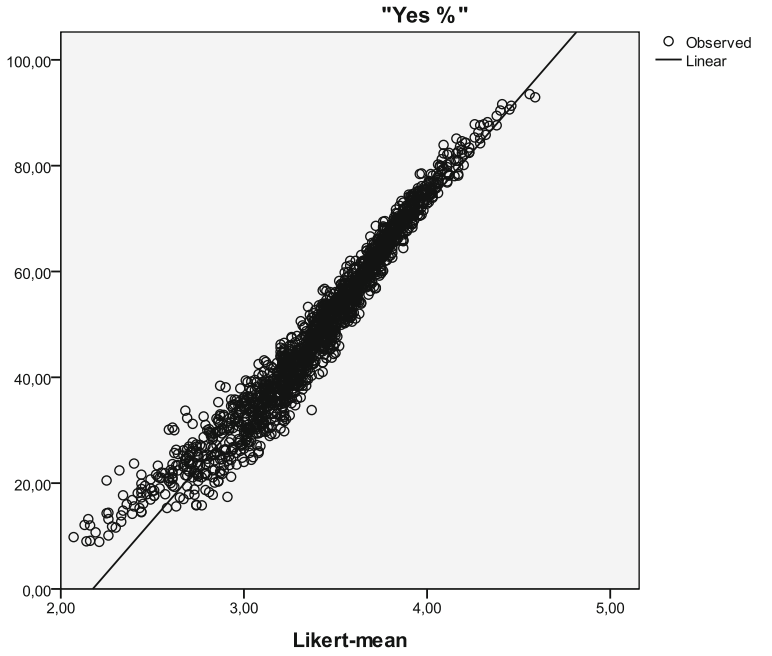

Fig. 4. Curve Estimation: Dependent Variable: "Yes %"; R^2 = 0,948; F = 31133,217, df1 = 1; df2 = 1693; p = 0,000; Parameter Estimates: Constant = −86,675; b1 = 39,857.

4 Conclusion and Discussion

There exists a very strong correlation between Likert means and Yes % which converges to a linear functional dependence. It means that these values correspond to the same underlying object. Thus, Yes % should take precedence since it has several advantages over Likert means. First of all, Yes % are more easily understood and interpreted. Second, there is no doubt over their mathematical "purity" whereas Likert means have problems: equal distances between scale orders are not guaranteed but postulated and the calculation of normal distribution parameters - mean and variance - from ordered data is not justified.

1. Correlation between Likert means and No % is palpably weaker. It shows that the Likert scale is not entirely symmetric, contrary to theoretical postulate. There exist some latent dependencies between stimulus-text comprehension and the preferences of answers. In the methodology of social attitude, measurement exists a textbook norm: if there is a long scale, it is necessary to formulate some statements positively and some negatively. Then it is likely that a randomization of semantic stimuli occurs which prevents any possible bias effects where, say, the respondents systematically start to pick "yes".
2. A very long category "no opinion" has a weak negative effect on the correlation between Likert mean and Yes %, even though the correlation still remains strong in general. This means that the modification of the Likert scale in which the researcher consciously decides to omit the neutral category "no opinion" does not necessarily imply a methodological opinion or error. It might sometimes be advantageous.
3. A question arose, why the correlation coefficient between the variables measured was slightly lower in this study than in the analogous Borg, Gabler study. Of course, we could rehabilitate the somewhat lower determination coefficient by hiding behind the confidence interval of the correlation coefficient. However, if the sample size is large enough, then the confidence interval is extremely small. Thus, we must take the fact at face value.

   It appears that the somewhat lower correlation between Likert mean and Yes % could have manifested as a side effect of this particular questionnaire and this particular sample. A hypothetical reason could be the high occurrence of "no opinion" answers. In the long and diverse list of public services some services are well known by the residents, their good or poor condition is seen by everyone (waste and environmental management, supply of drinking water, centralized heating) whereas some services are very specific and not encountered by everyone (services for the disabled, unemployed and so on). The "no opinion" answer for many specific services naturally rises in frequency. As has been shown, this slightly weakens the correlation. It is likely that if we had an instrument or a sample that did not have this systemic problem, then the correlation between Likert means and Yes % would be equally high as in Borg, Gabler study. It is simply necessary to have more data from repeated studies that would mitigate the possible side effects of a questionnaire or a sample.
4. This time the analysis was done only on the level of single items. The correlation between psychometric indices and average Yes % values on the scale was not

studied. It is worthwhile to do so in other studies, although it is not hard to guess that, given extremely high correlation on the primary items level, it will likewise repeat itself among additive psychometric indices.

5. It would be much more interesting to make an analogous repeated (or similar) study with 4–6 category Likert scale which does not include "no opinion" option. A no less meaningful analysis could be done with a 7 rank Likert scale that has a neutral category. A hypothesis would be that we should not expect any significant difference from Borg, Gabler or this study.
6. It is clear that social scientist who make survey reports to their colleague scientists, commissioners, consumers and media should use the Yes % more often and more widely rather than Likert means.

References

1. Balch, V.Ch.: Internet Survey Methodology. Cambridge Scholars Publishing, Tyne (2010)
2. Baros, W., Rost, J.: Natur-und kulturwissenschaftliche Perspektiven in der Psychologie. Methodologie, Methoden, Anwendungsbeispiele. Regener, Berlin (2012)
3. Borg, I., Gabler, S.: Zustimmungsanteile und Mittelwerte von Likertskalierten Items. ZUMA Nachrichten **26**(50), 7–25 (2002)
4. Bubelienė, D., Merkys, G.: School's cultural diversity: what is the difference between "school of happiness" and "school prison"? In: Lubkina, V., Danilane L., Usca, S. (eds.) International Scientific conference on Society, Integration, Education - SIE2019, vol. II, pp. 53–69. Rezekne Academy of Technologies, Rezekne (2019)
5. Carinio, J., Perla, R.J.: Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. J. Soc. Sci. **3**, 106–116 (2007)
6. Cunningham, W.A., Zelazo, P.D.: Attitudes and evaluations: a social cognitive neuroscience perspective. Trends Cogn. Sci. **11**(3), 97–104 (2007). https://doi.org/10.1016/j.tics.2006.12.005
7. Dawes, J.: Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. Int. J. Mark. Res. **50**(1), 61–77 (2008)
8. De Ayala, R.J.: The Theory and Practice of Item Response Theory. The Guilford Press, New York (2009)
9. Fischer, H.: A History of the Central Limit Theorem: From Classical to Modern Probability Theory. Springer, New York (2011)
10. Georgii, H.O.: Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik, vol. 4. de Gruyter, Auflage (2009)
11. Hambleton, R.K., Swaminathan, H., Rogers, H.J.: Fundamentals of Item Response Theory. Sage Press, Newbury Park (1991)
12. Heiberger, R.M., Robbins, N.B.: Design of diverging stacked bar charts for likert scales and other applications. J. Stat. Softw **57**(5), 1–32 (2014)
13. Hofmans, J.: Editorial: applications of functional measurement in psychology. Psichologica **31**(3), 431–439 (2010)
14. Hofmans, J., Theuns, P., Mairesse, O.: Impact of the number of response categories on linearity and sensivity of self-anchoring scales. Methdology **3**(4), 160–169 (2007)
15. Jacoby, J., Matell, M.S.: Three-point likert scales are good enough. J. Mark. Res. **8**(4), 495–500 (1971)

16. Jamieson, S.: Likert scales: how to (ab)use them. Med. Educ. **38**(12), 1217–1218 (2004)
17. Kanai, R., Feilden, T., Firth, C., Rees, G.: Political orientations are correlated with brain structure in young adults. Curr. Biol. **21**(8), 677–680 (2011)
18. Kemper, Ch., Brähler, E. Zenger, M.: Psychologische und sozialwissenschaftliche Kurzskalen. Standardisierte Erhebungsinstrumente für Wissenschaft und Praxis. Medizinisch Wissechaftliche Verlagsgesellschaft, Berlin (2014)
19. Lamnek, S., Krell, C.: Qualitative Sozialforschung 6., überarbeitete Auflage. Beltz Verlag, Weinheim (2016)
20. Likert, R.: A technique for the measurement of attitudes. Arch. Psychol. **22**(140), 55 (1932)
21. Lozano, L.M., Carcia-Cueto, E., Muniz, J.: Effect of the number of response categories on the reliability and validity of rating scales. Methdology **4**(2), 73–79 (2008)
22. Merkys, G., Bagdonas, A., Bubelienė, D.: Vaiko ir mokyklos saugumo vertinimo klausimynas: indikatoriai ir faktorinė validacija. Socialinis ugdymas **3**(35), 52–61 (2013)
23. Merkys, G., Bubelienė, D.: Radviliškio rajono savivaldybė "Radviliškio rajono gyventojų pasitenkinimo viešosiomis paslaugomis apskaičiavimo indeksas" (2017). (Unprinted report). http://www.radviliskis.lt/lit/Ar-radviliskio-rajono-gyventojai-patenkinti-teikiamomis-viesosiomis-paslaugomis-kvieciame-susipazinti-su-tyrimo-ataskaita
24. Merkys, G., Brazienė, R.: Evaluation of public services provided by municipalities in Lithuania: an experience of applying a standardized survey inventory. Soc. Sci. **4**(66), 50–61 (2009)
25. Rost, J.: Lehrbuch Testtheorie - Testkonstruktion (2., vollst. überarb. u. erw. Aufl.). Bern, Huber (2004)
26. Steyer, R., Schmitt, M., Eid, M.: Latent state-trait theory and research in personality and individual differences. Eur. J. Pers. **3**, 389–408 (1999)