Dental Journal
Majalah Kedokteran Gigi
Vol. 42. No. 1 January–March 2009

Literature Review

# The transformation of ordinal scale for parametric statistic analysis on dental health questionnaire

**Adi Hapsoro**
Department of Dental Public Health
Faculty of Dentistry Airlangga University
Surabaya - Indonesia

***ABSTRACT***

***Background:*** *Questioner Measurement is very customize, it means that researchers make very individual based on the aim of the research. Frequently, the results of questioner measurements are numeric rank and index or both of them. Numeric rank and index are categorical data. Beside of validity and reliability problems, they have analytical problems as well. So, they need transformation to scale type (ratio, interval) data in order to minimize their problems.* ***Purpose:*** *This article reviews the effectiveness of two type data transformation in dental health research measurement.* ***Reviews:*** *There are two type data transformations, i.e. interval equivalency and sum of rating transformation method.* ***Conclusion:*** *Interval equivalency transformation method is more effective for the data come from index, and sum of rating transformation method is more effective for the data come from numeric rank.*

***Key words****: dental health, measurement, transformation method*

*Correspondence:* Adi Hapsoro, c/o: Departemen Ilmu Kesehatan Gigi Masyarakat, Fakultas Kedokteran Gigi Universitas Airlangga. Jl. Mayjend. Prof Dr. Moestopo No. 47 Surabaya 60132, Indonesia.

## INTRODUCTION

There are many studies on dental health using questionnaire as measurement tools, especially in measuring either patients, clinic visitors' respon, or society's respons towards an object about dental health services. Those responses can be used in measuring either opinion, reaction, interest, or level of knowledge. If the answers of the respondents are put into qualitative data/semi quantitative data (ordinal data) and analyzed with qualitative approach (non parametric), there would be no problem. However, it is often necessary to take qualitative data from a constructive phenomenon into quantitative one.

In questionnaire, the measurement tools either on index or on scale can also be conducted. If the items of the questions and the options of the answers are not conducted into ranks, this measure is usually used in descriptive study.[1–3]

If the items of the questions and the options of the answers are not conducted into ranks, the result of this measurement on questionnaire is generally an ordinal data. This ordinal data only as codes instead of as values or scores. Thus, the codes can not be put into quantitative data and can not describe parameter continuities. If the codes are analyzed with parametric statistics, the result will be bias and the conclusion will also be bias.[4,5]

Based on those problems, two different methods will be discussed in this study as alternatives for researchers in conducting questionnaire in order to obtain quantitative data from an abstract phenomenon. It means that those two following methods can obtain values or scores not only on codes but on ordinal scales as well.

### Conducting questionnaire with ordinal and interval scales

The measurement on ordinal scale is shown with a qualitative grade or rank. If the options of the answers are scored with: 1, 2, and 3, the interval between 1 and 2 will not be the same as the interval between 2 and 3. Nevertheless, it may be concluded that 2 is bigger than 1, or 3 is bigger than 2 and 1.

The example of using ordinal scale, moreover, is in measuring the responses of the service quality with the options of the answers categorized into: Good, Average, and Poor. Good category is scored with 3, Average category is

scored with 2, and Poor category is scored with 1. In other way, Good category may also be scored with 9, Average category may also be scored with 7, and Poor category may also be scored with 3. It is indicated that the scores can be changed, but the ranks of the options must be the same.

The result of measurement on interval scales, on the other side, is the same as the result of the measurement on ordinal scales, but it has the same intervals among the ranks. Therefore, based on the example above, it may be concluded that the interval of 1–2 is the same as the interval of 2–3.

Interval scales, moreover, has no score, 0 as an absolute score. Thus, if the results of the measurement are 1, 2, 3, and 4, score 4 can not be said as twice as score 2. For instance, if in an exam A gets 80 and B gets 40, it cannot be concluded that intelligence is as twice as the B.

Nevertheless, on interval scales the scores have already had their own values so that their average can statistically be for statistic parametric analysis purpose.

In measuring an object with interval scales, the interval of the options of the answers in a questionnaire should also be measured. It means that the interval between the options Good and Average should be measured, and so the interval between the options Average and Poor.

In the study of the concrete subject, moreover, the interval of two scores can be measured more easily than the interval of those in the study of the abstract subject. When the mass of two tennis balls is measured, for example, the mass of those balls can clearly measured; the ball A is 56 grams, and the ball B is 57 grams. It clearly indicate that the ball B is slightly heavier than the ball A. However, it may be difficult for measuring which ball is heavier only by holding those balls in our hands.[1,6]

The same problem, furthermore, also occurs in measuring abstract phenomenon or an activity or an object with a questionnaire as the measurement tool. For those reasons, in order to obtain continuous interval or score variations, the determined measurement tool is needed to be on a trial by using certain methods as the following.

**Equal interval method**

The objective of the equal interval method is to assess the options of the answers. For example, respondent is asked to assess an object with criteria; Good, Average, and Poor. Each of criteria then will be on a trial with a 100 respondents even though 30 respondents. However, the sample for the trial can also be used later in analysis of the study so that it will not be useless.[7]

The first step is the respondents must determine criteria Good, Average, and Poor into a continuum divided into nine or eleven intervals.

A B C D E F G H I

The letter E is the central point and the letter A is the lowest score. Thus, the quality of the score is getting high as to the right side. The quantitative interval among letters is still not known. However, the interval qualitatively increas as to letter I. It is indicated that the order of those letters is on ordinal scales. The interval among those letters then must be transformed on interval scales (equivalence). For instance, if the respondents choose the category Good, this option must be divided more specifically into, Good Enough, Good, or Very Good. Therefore, Box F to I can be choosed.

Moreover, if the respondents choose the option Poor, Box D to A can be choosed. If the respondents choose the option Average, Box D to F can be choosed. Nevertheless, there will be possibilities of overlapping during .

Thus, the respondents must be clearly instructed that if they want to choose Good, they must choose boxes the right side; if they want to choose Poor, they must choose boxes the left side; and if they want to choose Average, they must choose boxes on the center. Moreover, in each form, the respondents must write down Good, Average, or Poor on the top right of the form since the respondents are sometimes not consistent with their. Therefore, it can be useful in classifying the data. Finally, after all the respondents (100) have done their, for example, there are 35 respondents Good, 35 respondents choosed Average, and the rest respondents choosed Poor, the data tabulate as the following.

Table 1 showed that out of 35 respondents, the letter f means frequency that means the number of respondents choosing 'Good' category. For letter F the frequency is 10: it means that there are 10 respondents assessing the object with option Good, or for letter G the frequency is 15 meaning that there are 15 respondents assessing the object with 'Really Good' option. Moreover, p stand for proportion is a comparison between the frequency of each letter and the number of all respondents. Thus, p = f/n. For example, for letter F, p = 10/35 = 0.286.

The 'pk' symbol means cumulative proportion is proportion number on interval of certain interval or a

**Table1**. The result data of the respondents' towards the category "Good"

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| F | 0 | 0 | 0 | 0 | 0 | 10 | 15 | 6 | 4 |
| p | 0 | 0 | 0 | 0 | 0 | .286 | .428 | .171 | .114 |
| pk | 0 | 0 | 0 | 0 | 0 | .286 | .714 | .885 | 1.0 |

certain number added with all the proportions less than that number. For instance, for number 7 or G letter the cumulative proportion is: 0.286 + 0.428 = 0.714. Finally, the calculation of discriminal modal value as the final step is a process of evaluating a value representing rating or judgment of a measurement group towards an object. The value is estimated based on the median value, and symbolized with S. The formulation is as the following:

S = bb + I [(0.50 – pkb)/p]
bb = Minimal limit of the number category with median value inside
pkb = cumulative proportion below the number category with median value inside
p = proportion of the number category with median value inside
i = the width of interval (equal with 1)

In statistics, median is a number that limits 0.50 of proportion or 50% of frequency – a number smaller than the median itself. In order to determine the position of median, could be seen in the column or the category in the table which there is 0.50% of cumulative proportion inside. For instance, in Table 1 the cumulative proportion of the measurement result towards the criteria "Good" is in letter G's column or in number 7. Its minimal limit (bb), is about 6.5. This minimal limit is between the $6^{th}$ and $7^{th}$ box. The cumulative proportion below the number category with median inside (pkb) is about 0.286, and the proportion of the number 7 (p) is about 428. Thus, the calculation of the score (on the scale) is as the following:

S = 6.5 + 1 [(0.50 – 0.286)/0.428]
S = 6.44

The total score of the criterion "Good" is 7
The next example is the measurement towards the criterion "Average".

Before, the median of the data must be determined. Based on Table 2, the median is on column E with pk = 0.6, so that:

bb = 4.50
pkb = 0.314
p = 0.286

Thus, the total of the criterion "Average" is:
S = 4.50 + 1 [(0.50 – 0.314)/0.286]
= 5.150

Moreover, the result data of the towards the criterion "Poor" is the following:
The result of the is as the following:

S = 2.5 + 1 [(0.50 – 0.263)/0.333]
S = 3.21

The total of the criterion "Poor" = 3.21
In conclusion, the total of the criterion "Good": 7

"Average": 5.15
"Poor" : 3.21

The above result is an example of the towards an object. However, the result can also become the for a research about the index status of tooth cleaning (OHI'S) with 100 respondents.

The problem, moreover, is that the from those 100 respondents is divided into three distributions. Thus, if the scoring can be put in order as the example above, the continuum of the can be analyzed well. Nevertheless, if the is coincided, the interval of the continuum can not be analyzed. Therefore, in order to solve the problem, another method is needed for the ordinal data.

**Rating summative method**

This method is used to solve overlapping results of distribution from and respondents since the of the respondents can be too homogenous or too heterogeneous. The respondents, thus, are put into one distribution and are asked to assess an object with 5 options. The options of the answer are usually Totally Disagree (STS), Disagree (TS),

**Table 2**. The result data of the respondents' towards the category "Average"

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| F | 0 | 0 | 5 | 6 | 10 | 7 | 7 | 0 | 0 |
| p | 0 | 0 | .142 | .172 | .286 | .2 | .2 | 0 | 0 |
| pk | 0 | 0 | .142 | .314 | .6 | .8 | 1.0 | 1.0 | 1.0 |

**Table 3**. The result data of the respondents' towards the category "Poor"

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| F | 1 | 7 | 10 | 10 | 2 | 0 | 0 | 0 | 0 |
| P | .03 | .233 | .333 | .333 | .06 | 0 | 0 | 0 | 0 |
| Pk | 03 | .263 | .596 | .929 | .989 | 1.0 | 1.0 | 1.0 | 1.0 |

Abstain (TM), Agree (S), and totally Agree (SS).

The category of those options of the answers, moreover, is clearly on ordinal scale, which then is assessed so that it can be changed onto interval scale (Figure 3 & 4).[8–10]

The following table is an example of measuring the Dental Services at clinic of Faculty of Dentistry Airlangga University from and respondents. The question is: was are Dental Services at Faculty of Dentistry Airlangga University "Satisfying".

**Table 4**. Respons distribution from or 100 respondents as a object (service) **Category of answer**

| | STS | TS | TM | S | SS |
|---|---|---|---|---|---|
| F | 4 | 49 | 22 | 17 | 8 |
| P = f/n | 04 | 49 | 22 | .17 | .08 |
| Pk | 04 | 53 | 75 | .92 | 1.00 |
| Pk-t | 02 | 285 | 640 | .835 | .96 |
| Z | –2.054 | –.568 | .358 | .974 | 1.751 |

STS = Totally Disagree
TS = Disagree
TM = Abstain
S = Agree
SS = Totally Agree

The first of the table is the frequencies of answers (f) for each responcategory. The total of all frequencies is the same as the total of respondents (n), in this case 100 respondents.

Proportion (p) is obtained by dividing each frequency with number of respondents. For instance, the proportion of TM responds is 22/100 = 0.22

Cumulative proportion (pk) is a proportion of one respon category added to a proportion of all categories on the left side. For instance, pk for S category is obtained by adding (0.17 + 0.22 + 0.49 + 0.04) = 0.92.

Pk-t, is a median of cumulative proportion formulated as a half proportion of one category is added to cumulative proportion of another category on the left side, as the following formula:

pk-t = ½ p + pkb
p = a proportion of one category
pkb = cumulative proportion of another category on the left side.

For example, pk-t for answer category TS is: 0.49 /2 + 0.04 = 0.285.

The score of Z is a median of each respon category for one continuum with interval scale. The interval among respond categories is stated by the interval of score Z. Deviation score for each pk-t is based on the table of normal deviation (Appendix A). Thus, the normal standard score in the curve can be determined by pk-t score. This process is ordinal data into an interval one or the semi quantitative data by using Table Z (Appendix). For instance, SS category with pk-t = 0.96 has Z score (see Table) = 1.751

Moreover, STS category with pk-t = .020 has Z score= –2.054, and so do the other categories. If all Z scores of each respon category are put into one continuum line, it will be as the following:

-3 -2 -1 0 1 2
…..… ..,…..…,……..,……....,……,

STS TS TM S SS
(–2.054) (–.568) (.358) (.974) (1.751)

Those respon scores are now in interval measurement ranks so that the lowest score can be changed into 0 by taking linier transformation, as the following:

Y = 2.054 + (1) X

The result of the transformation is as the following:

| | | STS | TS | TM | S | SS |
|---|---|---|---|---|---|---|
| X | = | –2.054 | –.568 | .358 | . 974 | 1.751 |
| Y | = | 0 | 1.486 | 2.412 | 3.028 | 3.805 |

## DISCUSSION

Transformation of ordinal scale as the measurement scale must be done before its validity and reliability are measured parametric statistics is required in analysis process. However, this scale transformation needs a long process. First, a trial must be taken in preparing scale transformation. Second after the transformation, the validity and reliability process of measurement scale must be on another trial.

In equal interval method the measurement is simple and easier. Nevertheless, this method has some weaknesses. In measuring an object with scale, for example, there will be possibility of overlapping values since its categories are more than three categories (Good, Average, Poor). If the respondents are homogenous, furthermore, there will also be possibility of coinciding values or even closely coinciding values among categories; Good, Average, Poor. Therefore, this equal interval method in measuring an object on index scales (withorank) so that the options of the answer; A, B, and C, can become B, A, and C after , or can also become another combination depending on the result of its equal interval.

For example, in the measur of Oral Hygiene Index Simplified the mean score (code) of the data can not be estimated directly since the data still on ordinal scales. Thus, in measuring this OHI'S data, a researcher must equal. In this case, scores 0, 1, 2, and 3 must be equal so that the exact interval can be obtained. For this reason, the OHI'S scores are not always 0, 1, 2, and 3. The scores can possibly become 0, 1.5, 2.3, and 2.9. For OHI'S tabulation, then, can be analyzed with parametric statistics by measuring its mean score. Therefore, the score of Debris Index can not be the same as that of Calculus Index since the trial is separately taken in the equivalence of the measurement scales.

For Rating Summative Method, measuring approach,

on the other side, is longer and uses table Z (table of normal deviation) in making the ordinal data become interval one like equivalent interval method, this method an object. This method, thus, is more appropriate to be used in measuring an object with ranking scale and with unnecessary homogenous respondents.

For instance, the questionnaire with Likert scale usually has options of answers divided into five grades. However, if those grades must be transformed into interval scales, scales 1 to 5 must be equal a trial, not as a separate one like on index scales, since those grades are based on ranks. Thus, as a unit, scale 1 can impossibly be changed into scale 2. Nevertheless, only if the respondents are very homogenous, it will be possible that scale 1 and 2 will closely be coincided, so will scale 3, 4, and 5.

As a conclusion, both of those methods can be applied as alternatives in conducting questionnaire, especially in determining values or scores from the options of the answers wh distribution approach is closed to normal distribution based on interval measurement scale. They, finally, can also be analyzed by parametric statistics.

## REFERENCES


1. Azwar, Saifuddin. Dasar-dasar psikometri. Edisi ke-1. Yogyakarta: Pustaka Pelajar Offset; 1999. p. 112, 117.
2. Ferguson GA. Statistical analysis in psychology and education. Auckland: McGrawHill; 1981. p. 34.
3. Howell DC. Statistical methodes for psychology. Boston: Duxbury Press; 1982. p. 23.
4. Zimmerman DW. A simplified probability model of error measurement. Psychological Reports 1969; 25:175–86.
5. McCrae RR, Costa PT. Validation of the five factor model of personality across instruments and observers. Journal of Personality and Social Psychology1987; 52:81–90.
6. Allen MJ, Yen WM. Intoduction to measurement theory. Monterey: Brooks/Cole Publishing Company; 1979. p. 7.
7. Cliff N, Keats JA. Ordinal measurement in the behavioral sciences. Mahwah: NJ Lawrence Erlburn; 2003. p. 23.
8. Michell J. Measurement scales and statistics : a clash of para Digsm psychological Bulletin 1986; 3:398–407.
9. Babbie E. The Practice of Social Research. 10th edition. Wadsworth: Thomson Learning Inc; 2004. p. 14.
10. Velleman PF, Wilkinson L. Nominal, ordinal, interval,and ratio typologies are misleading. The American Statistician 1993; 47(1):65–72. Available at: http://www.spss.com/ research/Wilkinson/ Publications/Stevens.pdf. Accessed January 9, 2009.