STATISTICS IN TRANSITION-new series, July 2010
Vol. 11, No. 1, pp. 177—186

STATISTICAL CHOICE BETWEEN RATING AND RANKING METHOD OF SCALING CONSUMER VALUES

Piotr Tarka[1]

ABSTRACT

The article describes how many market researchers go wrong with statistical assumption as for the right choice in rating and ranking method application within consumers' values evaluation. Many mistakes still arise owing to ineffective comparison of one scale to another, which in fact constitutes a completely different method of analysis and usage. The author discusses advantages and disadvantages of both ranking and rating methods pertaining to further statistical possibilities in data processing against the theoretical background and later on given brief example. Next stage of description emphasizes the importance of both rating and ranking method, depending on type of consumer (subject) being interviewed, hypothetical type of question being asked, communication with subject or even data results to be obtained.

1. Key sources on the theory of value

Rokeach (1968) defined a value as 'a type of belief about how one ought or ought not to behave, or about some end-state of existence worth or not worth attaining'. The theory defines values as desirable, transsituational goals, varying in importance, that serve as guiding principles in people's lives. The crucial content aspect that distinguishes among values is the type of motivational goal they express. Different contents of values has three universal requirements of human existence: biological needs, requisites of coordinated social interaction, and demands of group survival and functioning. Groups and individuals represent these requirements cognitively as specific values about which they communicate in order to explain, coordinate, and rationalize behaviour. In fact, theoretical interest in values is spread across many disciplines including economics, sociology, psychology, political science and history. Values, then, are the most

[1] University of Economics in Poznań, Department of Strategy and Policy of International Competiveness Policy, al. Niepodległości, POLAND, e-mail: piotr.tarka@ue.poznan.pl.

abstract type of socio-economic cognition that people use to store and guide general responses to classes of stimuli.

Basically, the challenging questions belong to the branch of philosophical inquiry known as *axiology* or *the theory of value* (Hilliard, 1950; Taylor, 1961; Hartman, 1967). A crucial point is that one can understand a given type of value only by considering its relationship to other types of value. We can understand one type of value only by comparing it with other types of value to which it is closely or not-so-closely related.

Typical definition of consumer value in the market context can be taken in the following way [….]. *Value represents an interactive relativistic preference experience.* Typically, such consumer value refers to the evaluation of some *object* by some *subject.* Here, for our purposes, the "subject" in question is usually a consumer or other consumer, whereas the "object" of interest could be any product – a manufactured good, a service, a political candidate, a vacation destination, a musical concert, a social cause, etc.

2. Rating or ranking method for values measurement – what choices to make?

Researchers are still divided over the preference for rating or ranking methods applied for measuring values. However, the debate originates from the false premise that one method must be used to the exclusion of the other. A conceptualization of the value structure that uses characteristics of both rating and ranking systems opens up theory and research to a more complex understanding of values (Ovadia, 2004).

One way that research on values differs substantially from research on other science topics is the inexact definition of the object of analysis. While quantitative items such as wages and fertility are (largely) agreed upon in their definition and how they are to be measured, values do not have a single, unified definition nor a universal model of measurement. The choices that researchers make in defining what a value is not only determining the nature of the item of analysis (the value structure) but also the method of measurement (ratings or rankings). Therefore, a meta-analysis of values research must address two related questions. First, how are values structured in an individual's mind? Second, how does one measure a value?

In a ranking method, the respondent is simply asked to place a list of values in order of importance. This has been the method used by Rokeach in his Value Survey (1968), and many other studies. The other method of measuring values is a rating system, in which values are rated by the respondent independently of one another, typically on a Likert scale or some variant thereof. This approach has been used in the General Social Survey and modified versions of the Rokeach Values Survey and the Schwartz Values Survey. From the statistical point of view, rating scales are important instruments used to measure people's attitudes

towards a variety of stimuli including organizations, products, services, places, institutions, and advertisements. Like all measuring devices, the rating scale is only useful if it provides reliable and valid measurements. There is a considerable amount of research dealing with the problems of constructing questions and rating scales that are reasonably objective and relatively free of measurement errors.

3. Quantitative and qualitative variables differences

It is not a brand new discovery that quantitative variables assume numerical values, whereas qualitative variables assume nominal values indicated usually by some names. However, it sometimes happen that numbers are practiced in order to either denote the value of quality variable (e.g. shopping centre called M1 and measured as number 1) or mark and simply symbolize a name in itself "M1". In world of science, based on empirical research, undertaken analysis on different variables is becoming even more interesting, provided the increase in intensity level among variables can be measured. In this sense measurement is defined as a procedure and process of numbers assignment to different values stated on variable (in psychometrics literature – "constructs") with commonly accepted rules. These rules are set and remain usually arbitrary, which means all measurement units (e.g. length measured in: centimetres, kilometres or miles etc.) are imposed symbolically by community members. Therefore, when human values are considered, we have at our disposal the background of multiple **theoretical constructs**, or so called ideas (Francuz, Mackiewicz 2005). For them a set of rules of measurement and their comparison operation on all their levels as for selection of perfect unit should be initially proposed. Obviously, this measurement differs considerably from subjects or variables expression on natural or real units (e.g. length), where for example each unit falling on "length" variable is simply summed up according to assigned measurement unit (centimetres or inches). In case of constructs one needs to decide on selecting the observable rate or evaluation system to measure constructs in order to proceed further.

One way of considering constructs (subjects' values) as quantitative variables and conducting computations based on their structure or state of preference is to perceive constructs and process of their measurement according to some degree of intensity and their appearance on given numerical value. Construct (value) can be formed if there are higher (based on ascending intensity) numbers allocated (to construct) by subject (respondent). In other words, the greater allocation as for some construct, the better creation of final value and reference point from which any comparisons can be started (Sagan, 2004).

4. Statistical option in data analysis and field work (data collection) within Rokeach (RVS) and Schwartz scales

Despite its widespread use, there have been critiques of the RVS that are relevant to most ranking ones, even though Rokeach developed multiple methods for administering the survey in attempts to simplify the task for the ranking tool. First, a ranking question cannot be easily asked over the telephone, which has become the main mode of survey research in the last three decades. Secondly, requiring a respondent to place 18 items (on questionnaire) in an order of preference is a lengthy process when compared to asking for individual ratings (Szreder, 2004). Rokeach himself admitted that 'many respondents report the ranking task to be a very difficult one – one they have little confidence in having completed in reliable manner and one they are often sure they had completed more or less randomly'. Thirdly, rankings result in a data set that cannot be analyzed with standard statistical methods because of the interdependence of the ranks. If the rankings of 17 of the values in the RVS have been indicated by the respondent, the value of the final value is predetermined. This characteristic of the data (sometimes referred to as being 'ipsative') requires researchers to make complex adjustments in order to perform statistical analyses of the results.

In contrast, the rating system is advocated for its simplicity in execution and analysis, as the respondent can quickly and easily respond to each item without having to compare the value with others on the list of questionnaire. A number of researchers have also found that using a method in which each item is independently evaluated does not lead to a significantly different rank ordering of the items on the list. However, rating measurement have also been subject to considerable criticism. Researchers have found that rating tools allow low-and nonmotivated respondents to give largely uniform responses ('non-differentiation') and thereby understate the differences among values. Rating systems often yield statistical ties between values, which may be the result of indifference to the question, rather than a true equivalence of importance. Because the ranking procedure forces the respondent to give each value a different assignment, it prevents nondifferentiation, irrespective of motivation level. However, it is also possible that ranking procedures force individuals to overstate the differences between values and make comparisons of values that the respondent considers non-comparable or potentially make random responses to meet the requirements of the survey.

Generally respondents using a rating system often cluster their responses within a narrow subset of the range of options, but no such 'underdispersion' is possible in a ranking system where the respondent must use the entire range of the scale. Also rating systems cannot ensure that the scale is used consistently, either between or within respondents. It is possible that one respondent considers a score of 70 (out of 100) to be a high level of importance, while another respondent considers 70 to be a moderate level of importance. Although a rating system is clearly better than a ranking system for assessing the distance between values for

a single respondent, ratings cannot completely ensure that within-respondent ratings are consistent.

5. The pros and cons pertaining to either methods application

Since it is clear that both methods for measuring the consumers' values have strengths and weaknesses, one should ask what each method assumes about the object of analysis itself, the value system. In other words, what is the value system and how is it organized? By looking more closely at the methods, one can see how the conclusions about value systems are not only residuals of the method of data collection, but are expressions of the characteristics of the overall value system that are assumed to exist in the individual. Therefore, the question of ratings or rankings is not only the question of a method, but of the nature of values and their organization in the mind. In an ipsative system, such as the Rokeach Value Survey or any other ranking system, values are arranged in a zero-sum structure by definition. If the ranking of one value increases by one rank, another value must decline by one rank. For Rokeach and other advocates of the ranking method, values represent mutually exclusive choices. In situations that call multiple values into possible action, one must be prioritized over the other(s). This means that as value structures change over time or differ across groups, the higher importance of one value must come at the expense of the importance of another value.

In contrast, a rating method does not require changes in the importance of some values to be compensated for by changes in other values. Since the respondent has the ability to give each value any rating without regard to the ratings given to other items, the ratings of the values have no restrictions. The value system that is assumed in a rating system is one in which the importance of values are independent; there is no limit on the total amount of importance that is distributed among the values.

In general, when researchers are interested in measuring values belonging to subjects (respondents), they often use a rating rather than a ranking method because it is easier and faster to administer and yields data that are amenable to parametric statistical analyses. However, because subjects' values are inherently positive constructs, subjects often exhibit little differentiation among the values and end-pile their ratings toward the positive end of the scale. Such lack of differentiation may potentially affect the statistical properties of the values and the ability to detect relationships with other variables.

6. Brief example

An example of how these differences can exist were considered in a hypothetical value structure containing only four values: **accomplishment, happiness, pleasure,** and **salvation** (Tarka, 2008). Each of these values can be

associated with a specific amount of importance that is later scaled from 0 (no importance) to 100 (maximum importance). Now, consider two individuals (subjects) with the following distributions of importance.

Graph 1. Answers importance by two subjects making judgments

1 subject:	**Accomplishment - 90, Happiness - 80, Pleasure - 70, Salvation - 40**
2 subject:	**Accomplishment - 70, Happiness - 60, Pleasure - 50, Salvation - 30**

Source: own construction

In the very first view ranking survey would show that these two people have the same value structure. Each respondent would report that accomplishment is most important, followed by happiness, pleasure, and salvation, respectively. However, a rating method would show that there are substantial differences between subject (1) and subject (2) in their values judgement. Firstly, subject (1) places more importance on these values overall than subject (2). Second, (s.1) places more importance on any specific value than (s.2), even though they rank the value identically.

Depending on the question being asked in regard to the market values of these individuals, conclusions about the values of these individuals may differ. If the question is about which values will be acted upon in situations where choices must be made, the ranking and rating systems will both point us in the same direction. However, if the questions of interest are about the relative importance of a value for (s.1) as compared to (s.2), the method will lead to different conclusions: ranking will indicate no difference whereas rating will show a difference. In particular, we would conclude that while happiness has the same importance for each individual relative to the other values in the structure, the amount of importance that (s.1) assigns to happiness is greater than the value assigned to it by (s.2). Neither a ranking nor rating measure by itself would reveal this pattern. Both of these conclusions are equally valid and therefore, the 'most correct' answer about their values is one that reveals both facts.

A second advantage in this approach is shown when one introduces change in the value system of (s.1). Assume that over time or in response to some stimulus, the importance of accomplishment for (s.1) declines from 90 to 87, whereas the importance of happiness increases from 80 to 83. The ranking approach would conclude that there has been no change in the value system of (s.1) over time. This is true, as we would expect that in circumstances where accomplishment and happiness were to come into conflict, (s.1) would choose accomplishment at both times of measurement. However, the rating system also tells us that the importance of accomplishment has declined for (s.1) over time, while the importance of happiness has increased. If these observations were part of a study to determine whether (s.1) had been affected by some external stimulus, such as an buying a new car or clothes, **then ranking and rating approaches would**

lead us to two completely different conclusions. The best conclusion uses both pieces of information: the stimulus affected a change, but it was not large enough to change the ordering of the values of interest.

7. Still ongoing transformation dilemmas on ranking and rating scale measure

In market research, the assumption that many statistical (especially multivariate) procedures require at least is intervally scaled data (Green, Tull, 1978). This supposed relationship between measurement scales and statistical techniques can be traced to Stevens (1966) and represents the representational theory of measurement. In contradiction to this view, it is sometimes said that in mathematical statistics literature one will not find scale properties as a requirement for the use of various statistical procedures. The point being made here is that the assumptions for the use of a particular statistical procedure are based solely on the mathematical aspects underlying that procedure and nothing else.

Borgatta and Bohrnstedt (1980) argue that the question of measurement assumptions in statistical applications is often treated inappropriately by market researchers. This occurs because unobserved constructs like values are assumed to be continuous variables (approximately) normally distributed in the population of interest – by definition a metric (interval) variable. When such constructs are measured as discrete variables, they are interval scales with (sometimes considerable amounts of) error. This was a classical theory view of measurement where three primary types of ordinal data were identified (Kim, 1975): [(1) ordered categories, (2) ranking, (3) test and summated rating scales construction], and also a set of simple transformations used by many scientists: [(a) assign each category an evenly spaced numerical value that preserves the original order, (b) convert ranks into corresponding numbers, and (c) consider the given values as valid metric scores].

In literature, procedure of transformation from ranks into numbers is nothing else than just assigning meaning categories to definite numbers with expressions: *better, worse, worst*. A positive number is assigned and starting point falls on lowest number (for a rank) with number 1 – *better*, continuing through 2 – *worse* (next rank) 3, 4, 5 and so on. Problems associated with this sort of transformation appear on the gaps between numbers. In ranking scales we hardly differentiate accurate distances between numbers which are contrary to interval, rating scales where distances are easily measured and most of arithmetic computations can be performed. For that reason, in practice, subjects' characteristics are measured on rating scales where they can express their attitudes and opinion easily using numbers ranging from 1 to 7 or 1 to 5, at which subjects can point up human value as the most important for them (7) or as the least important (1). Because subjects' answers given on ratings (based on this type of scale) are perceived and treated as part of scale numbers, neighbouring numbers adjoining from both sides

to chosen number by subject are sometimes summed up and later single numerical value is retrieved. This numerical value makes up complementary and indicates importance of value (Walesiak, 1996), (Zaborski, 2001).

Virtually transformations make the unrealistic assumption of equal distances between each pair of scale categories and probably do not produce "true" interval scales under most conditions. To use this approach the researcher should have at least assumed that the errors and distortions introduced by the transformation are minor and that they do not affect conclusions drawn from the data analysis. Some researchers have already developed a wide variety of more complicated models which transform paired comparisons or rankings into unidimensional and multidimensional measures. These scales are thought to produce a better fit between the operationalized variable and the latent construct. The Thurston Case procedure – the normalized method of rank order, and similar method of successive categories – (Guilford, 1954) are examples of methods used to derive unidimensional (interval) scales from order-type comparisons. For example, let us assume the following experiment conducted on sample consisting of 250 respondents (men) ranking seven cars on two dimensions: 1). stated preference and 2). perceived quality of the car (Tarka, 2008). Guilford explained that a complex transformation of the rank data can be performed by using the method of successive categories to obtain an interval scale. In this sense each rank can be treated as a category of an underlying unidimensional continuum for which we derive an interval level value for the midpoint of each category. These values then replace the individual responses. A comparison of the original rank values and their rescaled equivalents can show that these methods effectively compress the scale (rank values = 1 to 7; above mentioned “preference” = 1.00, 1.86, 2.41, 2.84, 3.25, 3.72, 4.52; “quality” = 1.00, 1.86, 2.39, 2.81, 3.22, 3.68, 4.47. The new scale values differ by no more than 0.05 across the two dimensions. This implies that respondents are using comparable scaling categories to evaluate the set of quality elements associated with a car on each dimension. In order to prove it, one can further perform a goodness-of-fit test between the new scale values and the original data, indicating that the method of successive categories (or similar one: normalized method of rank order) cannot be rejected. The rest of nonmetric scaling methods (generating multidimensional metric scales from order data) can be found in works of Green and Rao (1972).

8. Conclusions

Comparison of ranking and rating scales and methods of evaluation in human values suggests that there are no perfect types of statistical instruments in evaluation genuine and actual state of human values. This is by reason of values, being hidden in subjects’ mind, which are deeply taking hold of subjective and mental expressions given by subjects (e.g. consumer’s evaluation of product or service). “Values” are still obscure and hard to express for people. For statistical

researchers they are sometimes inscrutable or impenetrable, no matter what type of scale is applied. In fact. this is subject to acceptance and we all have to leave with it. However, the truth is also that ranking and rating scales can be a valuable measurement instrument provided they are applied appropriately. Ultimately, each scale and method has its own unique way of values measurement (in approximate way), depending on research objective or subject being explored. Any transformation procedures from ranking to rating / interval scale (if ever undertaken) should be implemented very carefully. Finally, one can conclude that human values and their expressions are formed (or rather generated from value-items) not only under specific scale or techniques of statistical analysis but also under the influence of a number of factors such as: subject's determination to answer a posed question, time of interview or type of questionnaire, communication with them or even subjects' situational humour.

REFERENCES

BORGATTA, E.F., BOHRNSTEDT. G.W. (1980), *Level of measurement: Once over again,* Sociological Methods and Research, Vol. 9, No. 2, pp. 148–160.

FRANCUZ, P., MACKIEWICZ, R., (2005), *Liczby nie wiedzą skąd pochodzą.* Wyd. KUL Lublin

FRONDIZI, R. (1971), *What is value. An introduction to axiology,* 2nd edition, La Salle, IL, Open Court Publishing Co.

GREEN, P.E., TULL. D.S. (1978), *Research for marketing decisions,* Englewood Cliffs, Prentice – Hall.

GREEN, P.E., RAO. V.R. (1972), *Multidimensional scaling: a comparison of approaches and algorithms,* New York, Holt, Rinehart and Winston Publishing Co.

GUILFORD, J.P. (1954), *Psychometric methods,* New York, McGraw – Hill.

HARTMAN, R.S. (1967), *The structure of values,* Carbondale, IL, Southern Illinois University Press.

HILLIARD, A.L. (1950), *The forms of value: the extension of hedonistic axiology,* New York, Columbia University Press.

KIM, J. (1975), *Multivariate analysis of ordinal variables,* American Journal of Sociology, Vol. 8, No. 2, pp. 261–298

ROKEACH, M. (1968), *The nature of human values,* New York, The Free Press.

OVADIA, S. (2004), *Ratings and rankings: reconsidering the structure of values and their measurement*, Journal Social Research Methodology, Vol. 7, No. 5, pp. 403–414

SAGAN, A. (2004), *Badania marketingowe.* Wyd. AE Kraków

STEVENS, S.S. (1966), *Handbook of experimental psychology,* New York, John Wiley and Sons.

SZREDER, M. (2004), *Metody i techniki sondażowych badań opinii*, Warszawa, PWE

TAIWO, A., HERSHEY H. F., (200), *Overall evaluation rating scales: an assessment,* International Journal of Market Research Vol. 42, Issue 3, pp. 301–312

TARKA, P., (2008*), From ranking (Rokeach – RVS) to rating scales evaluation – some empirical observations on multidimensional scaling Polish and Dutch youth's values* – Innovative Management Journal, Vol. 1, No 2, pp. 24–42

TAYLOR, P.W. (1961), *Normative discourse,* Englewood Cliffs, New York, Prentice – Hall.

WALESIAK, M., (1996), *Metody analizy danych marketingowych*. PWN Warszawa

ZABORSKI, A., (1991), *Skalowanie wielowymiarowe,* Wyd. AE we Wrocławiu.