

Piotr Tarka

Poznań University of Economics

MEASUREMENT SCALE FOR CONSUMERS' HEDONIC VALUES – A COMPARISON OF RELIABILITY TECHNIQUES

Summary: The main objective of this study is to explore a set of different value-items (indirectly observable variables, considered by young consumers) and transpose them according to statistical reliability rules into a single latent coherent scale dimension. Six items prepared and designed in the self-explanatory questionnaire were handed out to students at Polish and Dutch Universities across all classes. After collecting the answers from respondents and running in-depth reliability analysis based on various techniques of reliability testing, the author generated one scale being classified on the basis of items structure and their consistency. Different techniques of reliability assessment helped in this case to define the final resolution of the measurement and gave a description how small differences could exist among all of them.

Key words: reliability assessment techniques, scales, youth values and items configuration.

1. Introduction

Admittedly values are a sort of latent characteristics. Simply saying they are non-psychical and (on the basic level and rules of measurement) must be carefully treated in the same way as the measurement of some other human mental indicators. Moreover values are something that belongs to fundamental perceptual dimensions. In this sense, they are an interior part of a human “mental image” about the world, which helps people form (according to their mental characteristics) any evaluation or judgment about product, service based on values. Therefore one should have referred here to values as “latent dimensions” which can be turned into realistic objects [DeVellis 2003; Duncan 1984, Netemeyer et al. 2003]. These dimensions are usually a part of consumers’ perception of products or services. The objective of this article is to explore and compare different techniques of reliability one after another in order to fit a set of items into their best composition on projected scale. The author also wanted to prove how and when small differences existed among these techniques of reliability in the process of scale construction.

2. Methodology guidelines

In order to find out which particular reliability technique would yield the most preferred reliability score, contributing to the terminal value scale (being constructed on the basis of several proposed items), we have conducted three main reliability tests. The main objective obviously was to explore a set of value-items and transpose them into a single scale entitled initially “hedonistic youth’s consumption”.

The designed questionnaire was self-explanatory and took approximately 15 minutes to complete. It contained 6 items and all of them were prepared in a form of short statements accompanied by 5 – point Likert scale (ranging from 1 – *totally disagree*, and 5 – *totally agree*). Most of these statements were derived from Ro-keach’s [1973] or Kahle’s [1983] earlier studies and after some minor alterations in their word formation, they had been handed out to a number of individuals at universities in Poland and in the Netherlands for a final evaluation. Only a low percentage (less than 5%) of those contacted refused to participate in a study. We collected a sample (N = 232) of students enrolling in marketing, management and economics classes at universities in the Netherlands and Poland. The data were collected in both countries between May and June 2008.

The analysis was performed with the purpose of experiment. That means, the author interchangeably ran the analysis (either by combining both data sets including samples with Polish and Dutch respondents or by parting them into separate so called “subsamples”). Doing so we obtained an opportunity to check and compare the scores before a final conclusion according to each technique yielding the best and worst results for the scale.

3. Selected techniques of value reliability assessment

Overall Cronbach’s reliability technique

According to 6 items (short sentences) being interposed on questionnaire as a whole, the summary statistics indicated that there was clearly one type of scale referring to the hedonistic way of life. It was marked by initial descriptive statistics. Mean formula for particular i -th (item) was calculated as:

$$\bar{T}_i = \frac{T_i}{W}. \quad (1)$$

where: $T_i = \sum_{j=1}^N x_{ji} w_j$ – total score of the i -th (item),

$W = \sum_{j=1}^N w_j$ – sum of the weights, where w_j is the weight for the case j .

Also respectively mean scale:

$$M = \frac{G}{W}. \tag{2}$$

where: $G = \sum_{i=1}^k \sum_{j=1}^N w_{ji} w_j$ – grand sum of scores.

According to aggregate measurement and joint scores assessment (with both samples combined, including Polish and Dutch respondents), we identified a coherent scale within the proposed set of items. Having performed Alpha's Cronbach reliability test [Kelly 1958; Peter 1979; Ferguson, Takane 2009] for the items we obtained ratio $\alpha = 0.73$

$$\alpha = \frac{kr}{1 + (k - 1)r} = \frac{6 \cdot 0.33}{1 + (6 - 1) \cdot 0.33} = 0.73 \tag{3}$$

where: k – is the number of items in the scale,
 r – is the average correlation among the items in the scale.

Table 1. Total sample-items characteristics

	<i>i</i> -th item mean	Standard deviation
1	4.13	0.88
2	4.13	0.82
3	4.11	0.92
4	4.29	0.88
5	4.21	0.87
6	3.64	0.98

List of items: 1 – Freedom, 2 – Independence, 3 – Success, 4 – Enjoying life, 5 – Pleasure, 6 – Wealth

Source: own calculations based on Statistica software.

Later on when we decided to divide a total sample into two subsequent groups, we obtained α reliability ratio (accordingly for the Dutch sample), amazingly lower (down approximately four points) as compared to the total sample. And respectively in the Polish sample there was score reliability – about around two decimal points higher than at the level of measurement where the total sample was surveyed.

$$\alpha_{(\text{Dutch sample})} = 0.70$$

$$\alpha_{(\text{Polish sample})} = 0.75$$

In order to test differences in means (among items) we ran the variance analysis, parting total variance into the component due to true random error (i.e., within-group) and the components that were due to differences between the means of items.

Table 2. Parted samples-items characteristics

Dutch sample				Polish sample			
<i>i</i> -th item mean	Standard deviation	Summary statistics		<i>i</i> -th item mean	Standard deviation	Summary statistics	
4.06	0.94	Cronbach's alpha	0.70	4.21	0.83	Cronbach's alpha	0.75
4.04	0.81	Standard. alpha	0.70	4.22	0.82	Standard. alpha	0.76
3.92	0.92			4.30	0.89		
4.16	0.89			4.42	0.86		
4.15	0.92			4.28	0.82		
3.54	0.97			3.74	0.99		

List of items: 1 – Freedom, 2 – Independence, 3 – Success, 4 – Enjoying life, 5 – Pleasure, 6 – Wealth

Source: own calculations based on Statistica software.

Table 3. Total sample-analysis of variance on items

	SS Sum of Squares	df	MS Mean Square	<i>F</i>
Between Subjects	475.32	231	2.06	
Within Subjects	694.33	1160	0.60	
Between Items	59.98	5	12.00	21.84
Residua	634.35	1155	0.55	
Total	1169.66	1391		

Source: own calculations based on Statistica software.

Table 4. Parted samples-analysis of variance on items

	Dutch sample				Polish sample			
	SS Sum of Squares	df	MS Mean Square	<i>F</i>	SS Sum of Squares	df	MS Mean Square	<i>F</i>
Between Subjects	226.51	115	1.97		232.65	115	2.02	
Within Subjects	374.17	580	0.65		320.17	580	0.55	
Between Items	30.51	5	6.10	10.21	32.01	5	6.40	12.78
Residual	343.65	575	0.60		288.15	575	0.50	
Total	600.68	695			552.81	695		

Source: own calculations based on Statistica software.

Here for our scale, Anova, the significance value of the $F = 21.84$ test (for total sample) was nearly $p = 0.00$ and respectively it reached $F = 10.21$ (in the Dutch sample) and $F = 12.78$ (in the Polish sample). That was considerably below 0.05 value level. Therefore we must have rejected the hypothesis that the average assessment scores were equal across items. The items differed to some extent. Obviously the differences were smaller when the total sample was divided and analyzed in separate groups.

Reliability measure based on Anova was:

$$r_{\eta(\text{Scale})} = \frac{MS_A - MS_{AB}}{MS_A} = \frac{2.06 - 0.55}{2.06} = 0.73. \quad (4)$$

Although reliability analysis suggested that the six items were measuring the same underlying dimension, we observed an exception pertaining to *Wealth* item there. This might be a problematic variable to include in our summated scale for the following reasons. Numbers in **Table 5** in a column “alpha if item deleted” indicated that if this item was omitted from a scale, the coefficient alpha would increase considerably. Notice that the column “item total correlation” was 0.33 for *Wealth* while for the rest compared correlation levels the scores were: *Freedom* – 0.44, *Independence* – 0.48, *Success* – 0.47, *Enjoying life* – 0.56, *Pleasure* – 0.56. The examination of *Wealth* indicated that, although it loaded relatively high with the other items in the reliability analysis, it possessed some measurement error and was not as strongly linked theoretically to the underlying dimension as the other items were [Wilson 2005]¹. Consequently, in purifying this summated scale, *Wealth* should be excluded from further analysis. But this fact will be considered (if necessary) in a further part of the analysis.

Table 5. Total sample-correlation between items

	Item – Total Correlation	Alpha if Item deleted
1	0.44	0.70
2	0.48	0.69
3	0.47	0.70
4	<u>0.56</u>	0.67
5	<u>0.56</u>	0.67
6	0.33	0.74

List of items: 1 – Freedom, 2 – Independence, 3 – Success, 4 – Enjoying life, 5 – Pleasure, 6 – Wealth [item if deleted]

Source: own calculations based on Statistica software.

For the sake of parted subsamples it looks alike. Having deleted *wealth* item, we obtained higher item – correlation coefficients and also higher alpha ratio.

Hypothetically considering a higher level of reliability is easy to predict here if some additional items were added. For example if we had decided to add two new items in total sample analysis procedure, the reliability ratio would have risen to $r_{\eta} = 0.78$ and consequently for three more new items: $r_{\eta} = 0.80$ and for four items

¹ As a result α coefficient would increase its value up to 0.74. This level of reliability showing Alpha coefficient of 0.74 is very acceptable for most research. It is also widely known (according to practice rules) that increasingly large coefficient alphas beyond 0.80 may not significantly increase overall reliability.

Table 6. Parted samples-correlation between items

Dutch	Item – Total Correlation	Alpha if Item deleted
1	0.42	0.66
2	0.46	0.65
3	0.45	0.65
4	0.41	0.66
5	0.48	0.64
6	0.35	0.68
Polish	Item – Total Correlation	Alpha if Item deleted
1	0.44	0.73
2	0.49	0.72
3	0.45	0.73
4	0.69	0.66
5	0.64	0.68
6	0.30	0.78

List of items: 1 – Freedom, 2 – Independence, 3 – Success, 4 – Enjoying life, 5 – Pleasure, 6 – Wealth [item if deleted]

Source: own calculations based on Statistica software.

$r_{ii} = 0.82$. It clearly indicates to what extent a scale being examined, gives higher and improved value of its reliability. It also appears that having added four additional items (in questionnaire), the final reliability coefficient would have been on the level of 0.82 which is approximating 1.0 – a perfect reliability. Perhaps in our case, some additional, e.g. four statements would be perfect.

Table 7. Total sample – resultant alpha

Reliability level, if more added items	Total
2 items	0.786
3 items	0.805
4 items	0.821

Source: own calculations based on Statistica software.

The same performed hypothetically analysis with: two, three and four items (but this time for each separate group: Polish and Dutch samples), indicated one important fact. Namely, both samples differed much in their level of reliability coefficient, if more items were added. Higher coefficients were mainly associated with the Polish group, in which final reliability coefficient (at four items added along the six already existing on scale, attained the ratios $r_{ii(\text{with 4 new items})} = 0.83$ and respectively for the Dutch sample $r_{ii(\text{with 4 new items})} = 0.79$. Comparing these coefficients with divided samples and reliability coefficient of the total sample, one could have inferred that the score was considerably improved for the Polish group

but it got worsened for the Dutch one. These two scores were more consistent and clear (if two samples were examined separately). It also fulfilled a much better reliability condition (when running the whole analysis separately among the subsamples) – providing that one wants to attain unbiased reliability coefficient.

Table 8. Parted samples – resultant alpha

Reliability level, if more added items	Dutch sample	Polish sample
2 items	0.754	0.802
3 items	0.775	0.820
4 items	0.793	0.835

Source: own calculations based on Statistica software.

In the next part of reliability ratio analysis we focused on another technique called *first half and second half* and *odd even reliability*. This type of analysis differed from the previous one in the selection of particular items in the course process of measurement. **We wanted to know, if respondents' fatigue taken to complete a list containing first half of a scale and then second half, would determine a final reliability score.** Here we decided that in the first half there would be such items as: *Freedom, Independence, Success* and for the second half: *Enjoying life, Pleasure, Wealth*². Either the first half or the second was balanced with the equal number of items.

First half and second half reliability technique

Considering at first calculated scores if the total sample was analyzed, we might have concluded that there were a lot worse reliability coefficients in comparison to reliability coefficient obtained when all items were grouped together on the alpha Cronbach's analysis. As shown below, reliability ratios for split halves were more or less equal 0.60 – 0.62. Moreover the second half proved to be even less consistent and coherent than the first half (just to mention standard deviation and variance).

Table 9. Total sample

	Summary – 1st half	Summary – 2nd half
Reliability	0.61	0.63
ITEMS 1:	Freedom	Enjoying life
2:	Independence	Pleasure
3:	Success	Wealth

Source: own calculations based on Statistica software.

Joint split half reliability coefficient (0.679) and Guttman split half reliability were still lower than previous Cronbach's alpha (0.732). There was also one piece of mean-

² *Wealth* item was still considered, owing to different technique of reliability assessment.

ingful information as for both halves correlation (correlation between the first and second half was 0.51). It soon appeared that these halves were somehow and to some extent different to each other, although previous Cronbach's alpha (0.732) proved to combine them into the single scale. Undertaken (as a result) by respondent fatigue in this case it lowered the correlation between the halves because of the order in which the scale items were presented, not because of the quality of the scale items.

Later on we decided to retry analysis, with one little exception: total sample was again divided into two separate samples consisting of Polish and Dutch respondents. But still the process of partition did not help much either. Only the Polish sample in the second half successfully increased its reliability up to 0.68 value (at 0.62 reliability level when the total sample was analyzed).

The rest of the scores in the Dutch sample were as follows:

- correlation between first and second half: 0.46,
- split half reliability: 0.634,
- Guttman split-half reliability: 0.634.

And respectively in the Polish sample:

- correlation between the first and second half: 0.54,
- split half reliability: 0.702,
- Guttman split half reliability: 0.699.

Table 10. Parted samples

	Dutch sample		Polish sample	
	Summary – 1st half	Summary – 2nd half	Summary – 1st half	Summary – 2nd half
Reliability	0.61	0.55	0.59	0.68
ITEMS 1:	Freedom	Enjoying life	Freedom	Enjoying life
2:	Independence	Pleasure	Independence	Pleasure
3:	Success	Wealth	Success	Wealth

Source: own calculations based on Statistica software.

Odd – even reliability technique

Finally we arrived at odd – even reliability technique. For this purpose, we selected the subset of odd numbered items and compared them to the even numbered items. The halves for measurement were split so that each half contained properly ranked and sorted items. This was a process of items extraction proceeded in the following manner:

- A) first half = 1 – Freedom, 3 – Pleasure, 5 – Independence, 7 – Enjoying life,
- B) second half = 2 – *Success*, 4 – *Wealth*.

This split (on the total sample), as a comparison to the first half and the second half method improved Alpha of the first half of the scale (from 0.61 to 0.73 – first half) but decreased it in the second half (from 0.62 to 0.54). For the total sample, a correlation between the first and the second half was 0.40. The rest was as follows: split half reliability – 0.569 and Guttman split half reliability – 0.524.

Table 11. Total sample

	Summary – 1st half	Summary – 2nd half
Reliability	0.73	0.54
ITEMS 1:	Freedom	Success
2:	Independence	Wealth
3:	Enjoying life	
4:	Pleasure	

Source: own calculations based on Statistica software.

The results of odd – even reliability technique as for reliability's coefficients presented a similar configuration when parted samples and total sample were examined. Again, the first half was strongly reinforced and the second half was much more weakened. The correlation for the Dutch sample between the first and the second half was 0.41. The split half reliability was 0.581 and Guttman split-half reliability 0.539. Respectively in the Polish sample the correlation between the first and the second half was only 0.36 the split half reliability 0.526 and Guttman split-half reliability 0.477.

Table 12. Parted samples

	Dutch sample		Polish sample	
	Summary – 1st half	Summary – 2nd half	Summary – 1st half	Summary – 2nd half
Reliability	0.65	0.53	0.79	0.54
ITEMS 1:	Freedom	Success	Freedom	Success
2:	Independence	Wealth	Independence	Wealth
3:	Enjoying life		Enjoying life	
	Pleasure		Pleasure	

Source: own calculations based on Statistica software.

4. Conclusion

According to the examined set of items and different reliability techniques of comparison analysis (being applied in this work), we confirmed one type of scale. The reliability analysis measured the same underlying dimension/scale. This scale might be entitled *Hedonism* or *Consumerism* with items grouping: Enjoying life, Pleasure, Freedom Independence, Success. We decided to exclude the item called Wealth from our scale. These expressions are certainly a part of contemporary modern consumers' approach to the emerging trends in the market and the style of living and also consuming products. This scale assuredly stands for mass culture and mass products era (as it is nowadays) and certainly covers wider markets respectively.

In order to measure the consistency of the above mentioned items accurately, we applied in the first row “alpha's Cronbach's” technique, then “the first half and the second half” and eventually “odd – even” technique. Alpha's reliability test for the set of value items was performed on the total sample (including respondents

from Dutch and Polish universities). Later on, the same analysis was retried on the divided subsamples (in Dutch and Polish groups). As far as divided subsamples were taken into account, one could say that both groups (Polish and Dutch) differed much in their level of reliability coefficients. Higher coefficients were mainly associated with the Polish group. It also proved that running analysis separately among the subsamples could generate a better reliability goodness-of-fit of data set without a bias in reliability coefficient. Also a comparison between three different techniques of reliability assessment proved that the former technique (based on Cronbach's solution) worked the best and the most effectively. The second type of technique "the first half and the second half" and the third "odd – even reliability assessment" displayed a lot worse reliability coefficients.

Literature

- DeVellis R.F., *Scale Development – Theory and Applications*, Sage Publications, London 2003.
- Duncan O.D., *Notes On Social Measurement – Historical and Critical*, Russell Sage, New York 1984.
- Ferguson G.A., Takane Y., *Analiza statystyczna w psychologii i pedagogice*, PWN, Warszawa 2009.
- Kahle L.R., *Social Values and Social Change: Adaptation to Life in America*, Praeger, New York 1983.
- Kelly G.A., *The theory and technique of assessment*, "Annual Review of Psychology" February 1958, vol. 9, p. 323-352.
- Netemeyer R.G., Bearden W.O., Sharma S., *Scaling Procedures – Issues and Application*, Sage Publications, London 2003.
- Peter J.P., *Reliability: A review of psychometric basics and recent marketing practices*, "Journal of Marketing Research", February 1979, vol. XVI, p. 6-17.
- Rokeach M., *The Nature of Human Values*, The Free Press, New York 1973.
- Wilson M., *Constructing Measures: An Item Response Modeling Approach*, Lawrence Erlbaum Associates, New York 2005.

SKAŁA DO POMIARU WARTOŚCI HEDONISTYCZNYCH KONSUMENTÓW – ANALIZA PORÓWNAWCZA RÓŻNYCH TECHNIK OCENY RZETELNOŚCI

Streszczenie: W artykule autor dokonuje eksploracji wartości młodzieży na podstawie zbioru sześciu stwierdzeń (ukrytych zmiennych), w oparciu o które konstruuje jednowymiarową skalę do pomiaru wartości hedonistycznych konsumentów. Skala została opracowana w wyniku przeprowadzonych badań empirycznych, w ramach których studenci polskich i holenderskich uczelni oceniali rozpatrywane w analizie stwierdzenia. Po zakończeniu procesu gromadzenia danych przeprowadzono analizę rzetelności skonstruowanej skali.