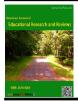
Research Article AJERR 2021,6:86



American Journal of Educational Research and Reviews (ISSN:2474-9265)



The Analysis of Frequency Terms Used in Rating Scales

Seock-Ho Kim¹, Hyo Jin Eom²

¹The University of Georgia, ²Korea University

ABSTRACT

The purpose of this paper is to provide a contemporary analy- *Correspondence to Author: sis of frequency terms commonly used in rating scales that are Seock-Ho Kim based on differing degrees of frequency with which events hap- The University of Georgia pen. A total of 20 frequency terms was analyzed using empirical data as well as employing a lexical database. Use of clearly separable frequency terms is imperative for the construction of a set How to cite this article: of rating scales as an effective measuring instrument.

Keywords: cluster analysis, frequency terms, rating scale, syn- in Rating Scales. American Journal set

Seock-Ho Kim, Hyo Jin Eom. The Analysis of Frequency Terms Used of Educational Research and Reviews, 2021,6:86.



Rating scales are frequently used to record the results of observations (cf. Berk, 1986; Guilford, 1954). References to the use of rating scales can be traced back to at least 150 BCE when the Greek astronomer Hipparchus used a six-point scale (see Footnote 1) to evaluate the brightness of stars (see Lodge, 1981). The great popularity of rating scales rests on the relative easiness with which they can be constructed and administered, and to their applicability for the many measurement settings (Pedhazur & Schmelkin, 1991).

Different formats of rating scales exist (for detailed explications, see Aaker & Day, 1983; Dawes, 1972; Dawes & Smith, 1985: Gable & Wolf, 1993; Guilford, 1954; Lemon, 1973; Lin, 1976; Nunnally, 1978; Saal, Downey, & Lahey, 1980). Three types of rating scales that are most frequently employed in educational psychological settings include numerical, graphic, and checklist (Payne, 1992). Sometimes frequency terms are accompanied with numerical values in rating scales. example, the following item with both numbers and frequency terms can be used to evaluate an aspect of instructor's teaching: The instructor's presentation was well organized. 1=Almost never, 2=Rarely, 3=Sometimes, 4=Usually, 5=Almost always. Another example can be found in the scale by Renzulli and Hartman (1981) to evaluate students' behavior with 1-Seldom or never. 2-Occasionally. Considerable degree, 4-Almost all of the time.

Sometimes frequency terms without numerical values are used in rating scales, but eventually summarized with certain numbers. The following is an example of such a format from a university course evaluation form (i.e., Mark Reflex) in the exact arrangement:

- 9. The instructor was willing to give individual assistance outside of class.
- Almost neverOften
- InfrequentlyAlmost always
- Occasionally

- 12. The instructor spent time on unimportant and irrelevant materials.
- Almost neverOften
- InfrequentlyAlmost always
- Occasionally

. . .

Note that the above item 12 is a reverse-scoring item. If an increasing set of numbers are listed together with options in the item 9, then one may ponder if a decreasing set of numbers should be used for the item 12 instead of an increasing set. It can be argued that, however, respondents might not select options based on accompanied numbers but mainly use the frequency terms select to the options. Quantification of frequency terms is considered in this study in conjunction with the selection of an ideal set of such terms to be used in rating scales.

Gable and Wolf (1993, pp. 53-54) presented example formats used for rating in terms of agreement, frequency, importance, quality, and likelihood without any numerical designations. There are seven different sets of frequency formats in Gable and Wolf (e.g., Always - Very often - Sometimes - Rarely - Never; Always -Usually - About half the time - Seldom - Never). Results from rating scales can be converted into a set of numbers and analyzed with fairly complicated psychometric models including polytomous item response theory models (e.g., Schriesheim Nerina & Ostini, 2010; Schriesheim, 1974) and item factor analysis (Wirth & Edwards, 2007). As indicated in Payne (1992, p. 196), however, the verbal definitions of rating and/or numbers on a numerical scale may lead to semantic confusion to the raters who are to record the results of observations. confusion may be exacerbated for items with frequency terms and numerical values when the items are reversely scored. In order to effectively construct the rating instrument, hence, it is necessary to use a proper, contemporary set of frequency words with some quantified interpretation based on empirical evidence.

. . .

One way to construct efficient rating scales with frequency terms can be accomplished by employing such terms with clear assignments of the numerical, probabilistic values. Many empirical investigations have been done to determine numerical or probability interpretation of the frequency terms. For a thorough review, the readers are referred to Mosteller and Youtz (1990; see also the comments by various scholars, and the rejoinder, as well as the references therein).

Obviously, frequency terms can be classified based on the dictionary definitions (e.g., Kipfer, 1992). The set of frequency words can also be analyzed with a current lexical database developed by researchers in the fields of artificial intelligence and linguistics. In order to assess the use of the frequency words, WordNet (http://wordnet.princeton.edu) (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990) was employed in this study. WordNet is a lexical ontology consisting of nouns, verbs, adjectives, and adverbs which are classified as sets of cognitive synonyms (i.e., synsets; synonym rings) based on human judgments.

Methods

Instruments

There exist many frequency terms that can be used in rating scales (e.g., Gable & Wolf, 1993, pp. 53-54; Kipfer, 1992, p. 960). In this study, two questionnaire forms based on Simpson (1944, 1963) were used to collect empirical evidence of 20 frequency terms. The first form was exactly the same as the one used in Simpson (1944). The second form is presented in Appendix (cf. Hakel, 1968). The 20 terms used were (1) Almost never, (2) Always, (3) About as often as not, (4) Frequently, (5) Generally, (6) Hardly ever, (7) Never, (8) Not often, (9) Now and then, (10) Occasionally, (11) Often, (12) Once in a while, (13) Rarely, (14) Rather often, (15) Seldom, (16) Sometimes, (17) Usually, (18) Usually not, (19) Very seldom, and (20) Very often.

Participants

Participants were 286 graduate students taking introductory research methods and applied statistics courses. There were 91 male and 195 female students. Each student provided answers to the two questionnaire forms. In the first form, participants were asked to provide a range (i.e., interval) of per cent of the time the term (e.g., Almost never: 0-5) indicates a thing would happen. In the second form, participants were asked to provide an exact number (i.e., point) of per cent of time (e.g., Almost never: 2).

Procedures

Many different approaches could be employed to analyze the empirical data from the two forms. The midpoint of the range for each term was used for the analysis of data from the first form. The value reported for each term was used for the analysis of data from the second form. Results from the two forms are compared in term of consistency across the forms. Clustering of the frequency terms was also performed to assess possible groupings of the frequency terms. All statistical analyses including the cluster analysis were done using the computer program SPSS.

The intent behind the cluster analysis was to report a study of possible grouping using the per cent measures of the frequency terms with an emphasis on the selection of the frequency terms for rating scales. For the purposes of the cluster analysis, the data set from the second form was employed. The data consisted of 286 rows of cases (i.e., participants) and 20 columns of variables (i.e., frequency terms). The data matrix was a 286 × 20 matrix, and the clustering was performed with regard to the 20 variables.

For the cluster analysis, the per cent measures of the 20 frequency terms were used without any transformation or standardization. The data matrix was complete without any missing. Euclidean squared distance was used as the index of dissimilarity. A Ward (1963) hierarchical analysis was used initially. The Ward method was chosen because of its overall cluster recovery capability and its consideration of the error sum of squares from the cluster centroid.

Because the popularity of the five-point rating scales, the number of clusters was specified to be five in a subsequent analysis. Note that there are other ways to determine the number of clusters.

Twelve frequency terms (e.g., Always, Frequently, Once in a while, etc.) have respective synsets (i.e., semantic relations) in WordNet, and portions of others (e.g., Almost

never, About as often as not, etc.) can also be found in WordNet. Synsets of the frequency terms are summarized and compared with dictionary definitions. Relations among the frequency terms can be analyzed based on other characteristics of WordNet (i.e., hypernymy, hyponymy, meronymy, metonymy, holonymy, etc.) in the context of the modern parlance.

Table 1 Median and Mean Values of the Frequency Terms from Form 1 and Form 2

	<u>Median</u>		Mean	
Term	Form 1	Form 2	Form 1	Form 2
Always	99 (99, 99)	100	97.62	98.79
Very often	90 (88, 88)	90	87.91	87.52
Usually	80 (85, 80)	80	77.71	78.30
Frequently	78 (73, 75)	80	77.33	77.12
Rather often	80 (65, 69)	80	76.72	75.50
Often	75 (78, 73)	75	75.51	74.63
Generally	75 (78, 77)	75	72.61	71.47
About AOAN	50 (50, 50)	50	43.37	44.45
Sometimes	43 (20, 27)	40	39.83	40.20
Occasionally	35 (20, 23)	33	34.38	35.40
Now and then	35 (20, 25)	33	33.85	34.89
Once in a while	23 (15, 16)	20	24.13	24.56
Usually not	15 (10, 14)	15	17.04	17.63
Not often	13 (13, 11)	15	15.38	16.77
Seldom	13 (10, 10)	15	14.19	15.35
Very seldom	8 (6, 6)	8	9.17	10.71
Rarely	8 (5, 5)	7	8.12	9.16
Hardly ever	7 (7, 6)	10	7.96	8.80
Almost never	3 (3, 3)	5	3.67	5.08
Never	0 (0, 1)	0	0.62	0.23

Note. Parentheses of Form 1 contain respective values from Simpson (1944) and Simpson (1963) separated with a comma. About AOAN = About as often as not.

Table 2 Third and First Quartiles, Interquartile Range, and Standard Deviation Values of the Frequency Terms from Form 1 and Form 2

	Q3-Q1=IQR		Standard Deviation	
Term	Form 1	Form 2	Form 1	Form 2
Always	100-98=2 (2, 2)	100-100=0	3.41	3.14
Very often	93-85=8 (13, 13)	95-85=10	8.17	8.00
Usually	86-72=14 (20, 14)	85-72=13	12.33	11.50
Frequently	85-75=10 (40, 20)	85-74=11	11.19	9.97
Rather often	85-70=15 (35, 23)	85-70=15	20.65	14.80
Often	85-68=17 (20, 30)	80-70=10	23.66	12.16
Generally	83-65=18 (22, 20)	80-60=20	12.80	13.21
About AOAN	50-48=2 (2, 3)	50-50=0	16.52	16.51
Sometimes	50-28=22 (22, 22)	50-30=20	14.53	14.29
Occasionally	45-23=22 (23, 19)	45-25=20	16.24	15.95
Now and then	50-23=27 (25, 28)	50-25=25	15.60	14.72
Once in a while	33-15=18 (19, 19)	31-15=16	14.85	13.93
Usually not	25-8=17 (12, 15)	25-10=15	10.81	10.48
Not often	20-8=12 (14, 14)	20-10=10	10.27	10.77
Seldom	18-8=10 (12, 11)	20-10=10	9.96	9.82
Very seldom	13-4=9 (7, 7)	10-5=5	8.67	12.45
Rarely	10-4=6 (7, 6)	10-5=5	5.82	6.87
Hardly ever	10-4=6 (9, 6)	10-5=5	5.78	6.26
Almost never	5-3=2 (2, 3)	5-2=3	2.34	6.61
Never	1-0=1 (2,1)	0-0=0	1.26	1.01

Note. Parentheses of Form 1 contain respective IQR values from Simpson (1944) and Simpson (1963) separated with a comma. About AOAN = About as often as not.

Results

Descriptive Statistics

The results from the two forms are presented in Table 1 in terms of the median and mean values. The median was reported as an integer. The order was from the highest to the lowest based

on the mean values obtained from the point values (see also Figure 1). This ordering of frequency terms is used in all other tables and figures. Form 1 values were based on the intervals and Form 2 values were based on the points (n.b., subscripts i and p in Figure 1

indicate that the values are from the intervals and the points, respectively).

The orderings of the means from the two forms are perfect. The maximum difference between two sets of means is 1.54 (i.e., 'Very seldom'). The orderings of the medians from the two forms are nearly the same. The maximum difference between two sets of medians is 3 (i.e.,

'Sometimes', 'Once in a while, and 'Hardly ever'). In comparison with results from Simpson (1944, 1963), the Form 1 column in Table 1 shows inflated per cent values for 'Rather often', 'Sometimes', 'Occasionally', 'Now and then', and 'Once in a while'. In Simpson (1944, 1963), the modifier 'Rather' reduced the per cent when used in 'Rather often', whereas it increased the value for the current sample.

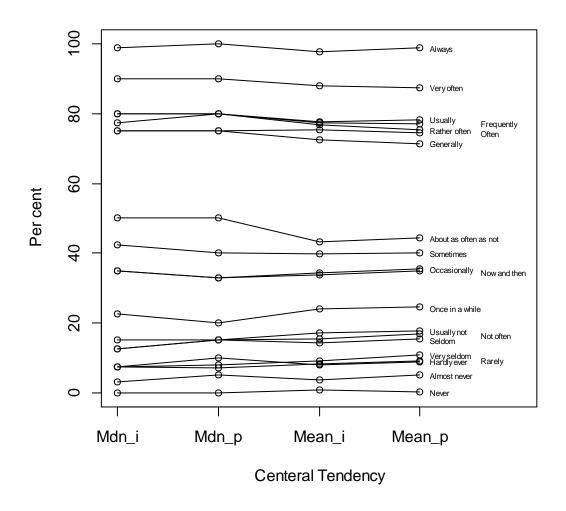


Figure 1. Median and mean per cent values of the frequency terms from Form 1 and Form 2.

Table 2 shows the variability measures for the values of the frequency terms. Interquartile range (IQR) values are shown in integers. The analysis of variability in term of IQR indicated that the term 'Now and than' yielded the largest IQR values (i.e., 27 from Form 1 and 25 from Form 2) and the term 'Never' yielded the smallest IQR values (i.e., 1 from Form 1 and 0 from Form 2). 'Always' and 'About as often as not' also yielded the smallest IQR value for Form 2. The largest difference in the IQR values from

the two forms was 7 from 'Often'. In comparison to Simpson (1944, 1963), the IQR values from 'Frequently', 'Rather often', and 'Often' were relatively smaller.

The standard deviation from 'Often' (23.66) was the largest for Form 1. The standard deviation from 'About as often as not' (16.51) was the largest for Form 2 (n.b., this seems to be a problematic term, and portions of participants indicated low per cent values). In both forms, 'Never' yielded the smallest standard deviations

(i.e., 1.26 for Form 1 and 1.01 for Form 2). The standard deviations from Form 1 and Form 2 are very similar except for 'Often' and 'Rather often' for which Form 1 yielded relatively larger

standard deviations. The largest difference in the standard deviation values from the two forms was 11.50 from 'Often'.

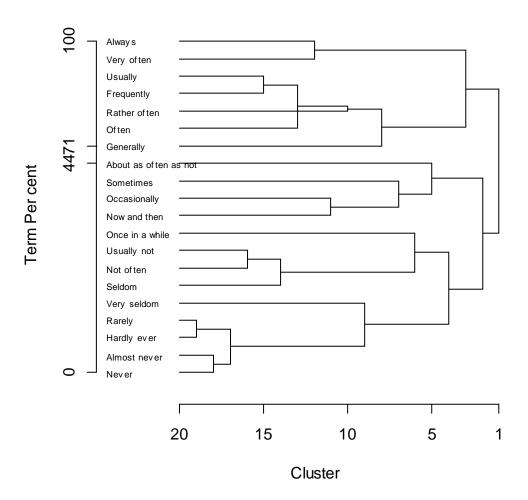


Figure 2. Dendrogram from the icicle plot of the frequency terms from cluster analysis.

Cluster Analysis

Figure 2 presents a modified dendrogram which is based on an icicle plot. If the number of clusters is not specified, the clustering algorithm yields the number of clusters from 20 to 1. The horizontal axis, abscissa, shows the number of clusters in the decreasing order from 20 to 1, while the vertical axis, ordinate, shows the ordered 20 frequency terms from 'Never' with the smallest per cent to 'Always' with the highest per cent. From the 20 starting frequency terms, 'Rarely' and 'Hardly ever' merged the first and formed the 19 clusters. From the 19 clusters, 'Almost never' and 'Never' then merged next and form the 18 clusters. Cluster separation of the

frequency terms can be inferred by drawing a vertical line on the number of clusters axis.

The dendrogram and the icicle plot from the computer didn't exhibit the frequency words in such an ordered arrangement shown in Figure 2. The dendrogram using Ward linkage, in fact, yielded initially only three clusters: {Always, Very often, Usually, Frequently, Rather often, Often, Generally}, {About as often as not, Sometimes, Occasionally, Now and then}, {Once in a while, Usually not, Not often, Seldom, Very seldom, Rarely, Hardly ever, Almost never, Never}. It then yielded the following two clusters: {Always, Very often, Usually, Frequently, Rather often, Often, Generally}, {About as often as not,

Sometimes, Occasionally, Now and then, Once in a while, Usually not, Not often, Seldom, Very seldom, Rarely, Hardly ever, Almost never, Never}.

A subsequent analysis was done with specifying the number of clusters. Based on the analysis for extracting five clusters (see Figure 2 also), the five clusters were: {Always, Very often}, {Usually, Frequently, Rather often, Often, Generally}, {About as often as not, Sometimes, Occasionally, Now and then}, {Once in a while, Usually not, Not often, Seldom}, {Very seldom, Rarely, Hardly ever, Almost never, Never}.

Table 3 WordNet Synsets of the Frequency Terms

Term	Synset	
Always	ever, e'ver, constantly, invariably, forever, perpetually, incessantly	
Very often		
Usually	normally, unremarkably, commonly, ordinarily	
Frequently	often, oftentimes, oft, offtimes	
Rather often		
Often	frequently, offentimes, oft, ofttimes, much, a great deal	
Generally	by and large, more often than not, mostly	
About AOAN		
Sometimes	on certain occasions, in certain cases but not always	
Occasionally	on occasion, once in a while, now and then, now and again, at times, from time to time	
Now and then	occasionally, on occasion, once in a while, now and again, at times, from time to time	
Once in a while	occasionally, on occasion, now and then, now and again, at times, from time to time	
Usually not		
Not often		
Seldom	rarely	
Very seldom		
Rarely	seldom	
Hardly ever		
Almost never		
Never	ne'er	

Note. About AOAN = About as often as not.

Synset Analysis

Table 3 presents a summary of the WordNet lexical analysis. The ordered frequency terms are listed in the far left column. The respective synsets, mostly different from the 20 frequency terms, are shown in Table 3. Among the 20

frequency terms, {Frequently, Often}, {Occasionally, Now and then, Once in a while}, and {Seldom, Rarely} are respectively perceived equivalently in the modern parlance.

Definitions from dictionaries were similar. Based on Sheffield (1981), synonym groups in terms of mostly the 20 frequency terms in this study are {Frequently, Often}, {Sometimes, Occasionally, Now and then}, and {Seldom, Rarely, Not often, Infrequently}. Synonym groups from Kipfer (1992) were broader and not mutually exclusive: {Frequently, Often, Usually, Generally, Now and then, Occasionally, Sometimes}, {Sometimes, Occasionally, Frequently, Now and then, Once in a while}, and {Seldom, Rarely, Not often, Hardly ever, Almost never, Sometimes, Now and then, Once in a while, Infrequently}. It is worth noting that the patterns from the synsets and dictionary definitions are not consistent with the empirical results from the cluster analysis.

Discussion

The results from the empirical data analyses as well as from the WordNet lexical analysis indicate a set of frequency terms can be perceived indistinguishably and semantically equivalent. Ideal use of equally spaced, separable frequency terms will enhance the quality of the measures from rating scales. Frequency terms in the same cluster or belong to the same synset may not be used together in a set of response options in a rating scale to prevent potential confusion by raters.

Probably the major reason why rating scales are vulnerable to misuse and misinterpretation stems from the fact that they reflect a perceptual process. It is, therefore, imperative for a researcher to design the rating scales as succinct and clear as possible. In this sense, a careful selection of frequency terms is essential in the use of rating scales. This study may provide a scheme that provides raters with a common, consistent frame of reference based on both empirical data and lexical analysis.

There seem to be some measurement instruments with frequency words for which not an ideal set is employed (see e.g., Bearden, Netemeyer, & Mobley, 1993; Bruner & Hensel, 1992). Considering the results from this study, a set of five frequency terms of {Never, Seldom, Sometimes, Often, Always}, with or without numbers 1 to 5, seems to be a good one to use in practice, provided that item scores without any

transformations will be used in further data analysis. Similar suggestions were made in other studies (e.g., Gable & Wolf, 1993; Payne, 1992; Pohl, 1981). Note that 'Sometimes' may be replaced with 'About half the time' if anchoring to the middle of the frequency scale is required. Although 'About as often as not' may function a similar way, this study showed the variability of 'About as often as not' was problematic.

In order to increase spreadness of the selection by encouraging to choose the two extreme options, a set of {Almost never, Seldom, Sometimes, Often, Almost always} can be a good alternative. The modifier, almost, has a diminishing or deflation function for the frequency terms in the extreme ends of the per cent continuum. Intensity would be reduced about five per cent if such a modifier is used with 'Never' and 'Always'.

Footnote

¹ The brightness of a star is referred to as the magnitude. The brightest stars are of first magnitude and those just barely visible with the unaided eye are of sixth magnitude (Chartrand, 1991, p. 24).

References

- [1]. Aaker, D. A., & Day, G. S. (1983). *Marketing research*. New York, NY: Wiley.
- [2]. Bearden, W. O., Netemeyer, R. G., & Mobley, M. F. (1993). Handbook of marketing scales: Multiitem measures for marketing and consumer behavior research. Newbury Park, CA: Sage.
- [3]. Berk, R. A. (1986). *Performance assessment: Methods & applications.* Baltimore, MD: The Johns Hopkins University Press.
- [4]. Bruner, G. C., & Hensel, P. J. (1992). *Marketing scales handbook: A compilation of multi-item measures*. Chicago, IL: American Marketing Association.
- [5]. Chartrand, M. R. (1991). National Audubon Society field guide to the night sky. New York, NY: Alfred A. Knopf.
- [6]. Dawes, R. M. (1972). Fundamentals of attitude measurement. New York, NY: Wiley.
- [7]. Dawes, R. M., & Smith, T. L. (1985). Attitude and opinion measurement. In G. Lindzey & E. Aronson (Eds.), Handbook of social psychology

- (Vol 1., 3rd ed., pp. 509-566). New York, NY: Random House.
- [8]. Gable, R. K., & Wolf, M. B. (1993). Instrument development in the affective domain (2nd ed.). Boston, MA: Kluwer.
- [9]. Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York, NY: McGraw-Hill.
- [10]. Hakel, M. D. (1968). How often is often? American Psychologist, 23, 533-534.
- [11]. Kipfer, B. A. (Ed.). (1992). Roget's 21st century thesaurus in dictionary form. New York, NY: Delacorte Press.
- [12]. Lemon, N. (1973). Attitudes and their measurement. New York, NY: Wiley.
- [13]. Lin, N. (1976). Foundations of social research. New York, NY: McGraw-Hill.
- [14]. Lodge, M. (1981). Magnitude scaling: Quantitative measurement of opinions. Newbury, CA: Sage.
- [15]. Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. International Journal of Lexicography, 3, 235-244. doi: 10.1092/ijl/3.4.235
- [16]. Mosteller, F., & Youtz, C. (1990). Quantifying probabilistic experiences. *Statistical Science*, 5, 2-12.
- [17]. Nering, M. L., & Ostini, R. (Eds.). (2010). Handbook of polytomous item response theory models. New York, NY: Routledge.
- [18]. Nunnally, J. (1978). Psychometric theory (2nd ed.). New York, NY: McGraw-Hill
- [19]. Payne, D. A. (1992). *Measuring and evaluating educational outcomes*. New York, NY: Merrill.
- [20]. Pedhazur, E. J., & Schmelkin, L. P. (1991). Measurement, design, and analysis: An integrated approach. Hillsdale, NJ: Erlbaum.

- [21]. Pohl, N. F. (1981). Scale considerations in using vague quantifiers. *Journal of Experimental Education*, 49, 235-240.
- [22]. Renzulli, J. S., & Hartman, R. K. (1981). Scale for rating the behavioral characteristics of superior students. In W. B. Barbe & J. S. Renzulli (Eds.), *Psychology and education of the gifted* (3rd ed., pp. 151-164). New York, NY: Irvington.
- [23]. Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*, 88, 413-428. doi: 10.1037/0033-2909.88.2.413
- [24]. Schriesheim, C. A., & Schriesheim, J. (1974). Development and empirical verification of new response categories to increase the validity of multiple response alternative questionnaires. Educational and Psychological Measurement, 34, 877-884.
- [25]. Sheffield, A. D. (1981). Soule's dictionary of English synonyms. New York, NY: Bantam Books.
- [26]. Simpson, R. H. (1944). The specific meanings of certain terms indicating different degrees of frequency. *Quarterly Journal of Speech, 30,* 328-330.
- [27]. Simpson, R. H. (1963). Stability in meanings for quantitative terms: A comparison over 20 years. *Quarterly Journal of Speech, 49*, 146-151.
- [28]. Ward, Jr., J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236-244.
- [29]. Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58-79. doi: 10.1037/1082-989X.12.1.58



Appendix

WHAT DO THESE WORDS MEAN TO YOU?

Below is a group of words which we use to indicate differing degrees of "frequency" with which events tend to happen. Obviously, some of the words mean different things to different people. We

wish to determine what each word means to you. For instance, if "fifty-fifty" indicates to you that a thing would happen about 50 times out of 100, you should mark 50 in the space before the expression: _____ fifty-fifty Simply indicate how many times out of 100 you think the word indicates an act has happened or is likely to happen. _____ almost never _____ often ____ always ____ once in a while _____ about as often as not _____ rarely _____ frequently _____ rather often _____ seldom _____ generally ____ hardly ever _____ sometimes ____ never ____ usually ____ not often _____ usually not ____ now and then _____ very seldom

< The End >

_____ very often

_____ occasionally

Gender: □ Female □ Male

Thank you very much for your participation.