# Sensitivity, reliability and neutral tendency of hedonic ratings as affected by scale types and lengths and overall product impression

Wannita Jirangrat[1], Jing Wang[2], Jose Andres Herrera Corredor[3], Sujinda Sriwattana[4] and Witoon Prinyawiwatkul[1]

[1] Department of Food Science, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, USA

[2] Department of Experimental Statistic, Louisiana State University, Baton Rouge, LA 70803-4200

[3] Colegio de Postgraduados, Campus Córdoba, Veracruz, México, C.P. 94946

[4] Sensory Evaluation and Consumer Testing Unit, Faculty of Agro-Industry, Chiang Mai University, Chiang Mai, Thailand 50100

## ABSTRACT

Several scales are available for assessing degrees of liking/disliking. Characteristics of hedonic ratings as affected by scale types/lengths and overall product impression are limited in the literature. The objectives of this study were to compare a discriminatory power, sensitivity, reliability and neutral responses' behavior of hedonic ratings collected from 9-point categorical (CAT), line (LIN) and labeled affective magnitude (LAM) scales; each with two scale lengths (100-mm vs. 300-mm length).

Results showed that the scale length effects were slightly different. The longer scale tended to have better discrimination for LAM scale. Hedonic scores tended to be rated lower with the longer scale. The Cronbach's alpha values of all three scale types and two lengths were comparable ranging from 0.777-0.919. The higher correlation among attribute questions and overall preference score was observed when presenting disliked products. The LAM scale seemed to be slightly advantageous over the categorical and LIN scales with the coefficient ranging from 0.46-0.72. The hedonic categorical scale weighed the lowest neutral responses' percentages (1.69-10.17) following by LIN scale (7.63-14.41) and LAM scale (4.66-16.10). Regarding the permitted area in LAM scale, the longer line accounted for 28.82% comparing with that of the shorter scale (17.80%); this was due to more flexible responses above extremely like/dislike categories.

**Practical Application:** This study demonstrated that characteristics of various hedonic ratings were affected by scale types, lengths and overall product impression. Using proper scales to assess liking/disliking would help increase power of the experiment. It was found that categorical scale has a high discriminative power, sensitivity and a low rate of neutral responses. Besides the LAM scale has similar Cronbach's alpha value with higher correlation coefficient when testing with identical samples.

**Keywords:** hedonic scales; scale types, scale lengths

## INTRODUCTION

Several scales have been proposed to improve reliability of hedonic methods to measure degree of liking/disliking. Comparing to several scales and procedures, hedonic rating provides unique features such as general applicability, not requiring trained panelists, getting meaningful results, and data that can be analyzed by various statistical methods (Cordonnier and Delwiche, 2008)

A few years after the 9-point hedonic scale had been created. Jones and Thurstone developed a balanced 9-point hedonic scale with a neutral point. However, the highest and lowest scale categories were frequently unused. The frequent use of the midpoint was also questionable. Forty-five years later, Schutz and Cardello (2001) introduced one of the most popular scale, a Labeled Affective Magnitude (LAM) scale. It was more sensitive with better discriminative power than a 9-point hedonic scale in well-liked product (Greene et al., 2006) and (2) permitted the use of areas above the like extremely category (Lawless *et al.*, 2010) giving consumers more flexibility. However, some the disadvantages were found on consumers performing a category behavior on a continuous LAM scale (Cardello *et al.*, 2008; Lawless *et al*, 2010).

Beside the 9-point hedonic categorical scale, some preferred to use a line scale. The scale was used alternatively so as to (1) provide "a zone of psychological comfort" for participants (Lawless and Heymann, 1999), (2) lessen a categorical behavior, (3) better perform by reducing a contextual effect, (4) give more freedom/flexibility (Giovanni and Pangborn, 1983), (5) lower deviation from normality (McPhearson and Randall, 1985), (6) maximize a discriminative power comparing between the best and worst samples (Villanueva *et al*., 2005) and (7) present a good correlation between rating of acceptability and preferences (Resano *et al*., 2009). However, Lawless and Malone (1986) had mentioned some disadvantages in terms of time consuming and complication for user.

## OBJECTIVES

The objectives were to compare the scale types and scale lengths in terms of (1) the ability to differentiation products and/or the sensitivity, (2) the reliability of scales, (3) the consistency of scales when testing two identical samples, and (4) the neutral responses' behavior.

## MATERIALS AND METHOD

**Materials and methods:** Three commercial grape juices were classified into 3 categories: poor (50% Welch's light dilution), moderate (Welch's light), and good quality (Welch'100% juices) using 7 expert panels and 30 consumers (using a ranking method).

**Procedure:** Four juice products were served and each participant using one of the 4 possible random serving orders: AB and AC, BA and CA, AC and AB or CA and BA. In each session a duplicated sample A was served to test consistency of scales. A total of 6 all possible permutations (6 independent sessions) derived from three different scale types (9-point categorical scale, 9-point line scale and LAM scale) and two scale lengths (100 and 300 mm).

**(1) Discriminative power/Sensitivity test:** ANOVA table was used to describe a discriminative power and sensitivity. If the ANOVA F test was significant, a Tukey's procedure was conducted to calculate the confidence interval (CI) estimating for hedonic mean score of each treatment. These CI estimates were then compared to assess sensitivity affected by treatments, scale types and lengths on a hedonic mean score of each product. The variances of each factor were estimated by Proc Mixed.

**(2) Reliability:** consistency of responses based on the different scale types and lengths among three attribute questions was evaluated by the Cronbach's alpha.

**(3) Consistency**: the Pearson correlation coefficients were used to measure a score consistency of duplicate samples.

**(4) Neutral responses' behavior:** the response percentages of a neutral response, which were define as "5" on CAT scale were calculated. To prevent an unfair count from LIN and LAM scale, the measure of neural response include all scores placed between "4.5-5.5" and "45-55" for LIN and LAM scale, respectively (Schutz and Cardello, 2001).

**Statistical analysis:** The analysis was carried out with a mixed procedure analysis of variance and once the significant difference ($\alpha < 0.05$) was detected, the follow up analysis would be carried out to compare mean pairwisely using multiple comparison test, Tukey' s Studentized Range (HSD) test (SAS, 2003).

## RESULTS AND DISCUSSION

(1) Discriminative power: Hein et al. (2008) suggested that the higher the F value, the better the discriminative power of the scale. Considering an OL, CAT yielded the highest discriminative power (Table 1). The power was ranked from CAT, LIN and LAM scale for 100 mm scale length; however, the order changed to LIN, CAT and LAM scale for 300 mm scale length. Cordonnier and Delwiche (2008) found that although using a hedonic categorical scale seemed to yield better discrimination, the difference on scales was not obvious. Hein et al. (2008) also found that CAT occupied the highest F-ratio. Overall, length effects were slightly different. The longer scale tended to be slightly better in discrimination and the effect was more obvious for LAM scale. Sensitivity test: The sensitivity was defined as a number of differentiation of a pair of mean (Greene et al., 2006; Lawless et al., 2010) (Table 2). More significant pairs indicated higher power to differentiate among products. CAT held the highest sensitivity yielding 8 pairs out of 12 pairs accounting for 66.67% comparing to that of for LAM (8.33%) and LIN scale (16.67%).

(2) Reliability was defined as a responses' consistency evaluated by a Cronbach's alpha value (Table 3). The score ranged from 0.777-0.919 which implied reliable use of CAT, LIN or LAM scale on a consumer acceptance test. There was no pattern observed between two lengths. However, somewhat higher (0.891-0.919) Cronbach's alpha values were observed for disliked product (less discrepancy) than for a well-liked and/or moderate-liked samples.

**Table 1 Discrimination power of scale**

| Scale | Length (mm.) | F value | | |
|---|---|---|---|---|
| | | OL | OC | OT |
| Category | 100 | 141.35 | 71.08 | 132.48 |
| | 300 | 99.387 | 107.84 | 101.52 |
| LIN | 100 | 80.67 | 113.79 | 85.22 |
| | 300 | 109.87 | 86.22 | 98.71 |
| LAM | 100 | 92.93 | 50.91 | 79.62 |
| | 300 | 94.97 | 98.28 | 94.30 |

**Table 2 Sensitivity test**

| Attributes | Scale | P value | | | |
|---|---|---|---|---|---|
| | | Good | Moderate | Good | Bad |
| OL | Category | 0.0319 | 0.0311 | 0.0423 | ns |
| | LIN | ns | ns | ns | 0.0024 |
| | LAM | 0.0397 | ns | ns | ns |
| OC | Category | 0.0118 | 0.0285 | ns | 0.0114 |
| | LIN | ns | ns | ns | ns |
| | LAM | ns | ns | ns | ns |
| OT | Category | 0.0486 | ns | 0.0136 | ns |
| | LIN | ns | ns | ns | 0.0127 |
| | LAM | ns | ns | ns | ns |

1) P value < 0.05 in the column indicated that at alpha=.05 their corresponding juices differ significantly (between 100 vs. 300 mm.) in their mean scores.
2) "ns" indicated that at alpha=.05, the corresponding juices were not significantly different (between 100 vs. 300 mm.) in their mean scores.

**Table 3 Reliability test**

| Lengths (mm.) | Cronbach's alpha | | | | |
|---|---|---|---|---|---|
| | Scale | Good | Moderate | Good | Bad |
| 100 | Category | 0.879 | 0.827 | 0.845 | 0.888 |
| | LIN | 0.777 | 0.833 | 0.855 | 0.909 |
| | LAM | 0.859 | 0.843 | 0.816 | 0.896 |
| 300 | Category | 0.853 | 0.892 | 0.849 | 0.914 |
| | LIN | 0.785 | 0.856 | 0.850 | 0.891 |
| | LAM | 0.880 | 0.898 | 0.781 | 0.919 |

**Table 4 Pearson correlation coefficient**

| Scale | Lengths (mm.) | OL | OC | OT |
|---|---|---|---|---|
| Category | 100 | 0.504* | 0.569* | 0.745* |
| | 300 | 0.424* | 0.655* | 0.580* |
| LIN | 100 | 0.485* | 0.398* | 0.580* |
| | 300 | 0.377* | 0.742* | 0.323** |
| LAM | 100 | 0.451* | 0.747* | 0.390* |
| | 300 | 0.558* | 0.694* | 0.527* |

**Table 5 Neutral tendency percentage**

| Lengths(mm.) | Scale | Overall Liking (OL) | | | |
|---|---|---|---|---|---|
| | | Good | Moderate | Good | Bad |
| 100 | Category | 1.69 | 8.47 | 3.39 | 11.86 |
| | LIN | 6.78 | 8.47 | 11.86 | 23.73 |
| | LAM | 1.69 | 15.25 | 5.08 | 18.64 |
| 300 | Category | 1.69 | 11.86 | 6.78 | 15.25 |
| | LIN | 8.47 | 20.34 | 11.86 | 23.73 |
| | LAM | 10.17 | 16.95 | 1.69 | 16.95 |

(3) Consistency was defined as the consistency of hedonic responses between the two identical samples (Lawless et al., 2010) assessed with Pearson correlation coefficients. The length effects were observed with an unclear explanation; however, less complicated attributes or surface attributes (OC) was less affected by the scale types (higher value). This study agreed that LAM scale was slightly better than CAT and LIN in terms of reliability of testing two identical products (Schutz and Cardello, 2001).

(4) The quality of sample showed an immense influence on a chance to get a mid response. The better the quality the lower the percentage of neutral tendency. The length of the scale showed a minor influence while the scale types inserted more effects. CAT weighed the lowest neutral responses' percentages (1.69-10.17) followed by LIN (7.63-14.41) and LAM (4.66-16.10) when the data were averaged among products ignoring the length effect for each attribute.

## CONCLUSION

CAT has a high discriminative power, sensitivity and low rate of neutral response. LIN has moderated discriminative power, sensitivity, reliability, correlation and neutral tendency. LAM has similar Cronbach's alpha value with higher correlation coefficient when testing two identical samples. The potential benefit from LAM was also observed with the high ratings in the permitted area above like/dislike extremely for well-liked and disliked products. The length effects exhibited differences in discriminative power and sensitivity but with no pattern. This study suggested the use of the 9-point hedonic categorical scale; however, using LAM scale is undoubtedly beneficial from the permitted area above extremely like/dislike.

## SELECTED REFERENCES

1. Cardello, A., Lawless, H.T. and Schutz, H.G. 2008. Effect of extreme anchors and interior Label spacing on labeled affective magnitude scales. ***Food Quality and Preference*** 19: 473-480.
2. Cordonnier, S.M. and Delwiche, J.F. 2008. An alternative method for assessing liking: positional relative rating versus the 9-point hedonic scale. ***J. Sensory Studies 23: 284-292.***
3. Giovanni, M.E. and Pangborn, R.M. 1983. Measurement of taste intensity and degree of liking of beverages by graphic scaling and magnitude estimation. J. Food Sci. 48: 1175-1182.
4. Greene, J.L., Bratka, K.J., Drake, M.A. and Sanders, T.H. 2006. Effective of category and line scales to characterize consumer perception of fruit fermented flavors in peanuts. J. Sensory Studies 21: 146-154
5. Hein, K.A., Jaeger, S.R., Carr, B.T. and Delahunty, C.M. 2008. Comparison of five common acceptance and preference methods. Food Quality and Preference 19: 651-661.
6. Lawless, H.T. and Malone, G.J. 1986. A comparison of scaling methods: Sensitivity, replicates and relative measurement. J. Sensory Studies 1: 155-174.
7.Lawless, H.T., & Heymann, H. (1999). ***Sensory Evaluation of Food: Principal and Practices***. Maryland: Aspen publisher.
8. Lawless, H.T., Sinopoli, D. and Chapman, K.W. 2010. A comparison of the labeled affective magnitude scale and the 9-point hedonic scale and examination of categorical behavior. J. Sensory Studies 25: 54-66.
9. McPherson, R.S. and Randall, E. 1985. Line length measurement as a tool for food preference research. Ecology of Food and Nutrition, 17: 149-156
10. Resano, H., Sanjuan, A.I. and Alibisu, L.M. 2009. Consumer acceptability and actual choice. An exploratory research on cured ham in Spain. Food Quality and Preference 20: 391-398.
11. Schutz, H.G. and Cardello, A.V. 2001. A labeled affective magnitude (LAM) scale for assessing food liking/disliking. J. Sensory Studied. 16: 117-159.
12. Villanueva, N.D.M., Petenate, A.J., and Da Silva, M.A.A.P. 2005. Performance of the hybrid hedonic scale as compared to the traditional hedonic, self-adjusting and ranking scales. Food Quality and Preference, 16: 691-703.
13. SAS 2003. SAS/IML software: Usage and reference, Version 9. SAS Institute Inc., Cary, NC, USA.

# Sensitivity reliability and neutral tendency of hedonic ratings as affected by scale types and lengths and overall product impression

Jirangrat, W

2012-06-26