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). However, consumers performing a category behavior on a continuous LAM scale (Cardello et al., 2008; Lawless et al., 2010) giving consumers more flexibility. Forty-five years later, the importance of such a scale increased; it was developed by Jones and Thurstone in 1965. The 9-point hedonic scale became a well-liked product in psychology and consumer science. However, it has been criticized because it is linear. The length of the scale showed a minor influence while the scale types inserted more effects. 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%). Reliability was defined as a responses’ consistency evaluated by a Cronbach’s alpha value. The score ranged from 0.770-0.915 with reliable usable cat, LIN or LAM scale on a consumer acceptance test. There was no pattern observed between two lengths. However, CAT yielded the highest (0.891-0.919) Cronbach’s alpha values were observed for disliked product (less discrepancy) than for a well-liked and/or moderately-liked samples.

Objectives

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

Materials and Methods

Materials and methods: Three commercial grape juices were classified into 3 categories: poor (50% Welch's light dilution), moderate (Welch's 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. 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 determine 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 defined as “5” on CAT scale were calculated. To prevent an unfair count from LIN and LAM scale, the measure of neutral 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 (p < 0.05) was detected, the follow up analysis would be carried out to compare mean painlessly using multiple comparison test, Tukey’s Studentized Range (HSD) test (SAS, 2003).

Results and Discussion

Conclusions

CAT has a high discriminative power, sensitivity and low rate of neutral response. LIN has moderate 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 well-liked and disliked products. The length effects exhibited a different pattern, the longer scale tended to be slightly better in discrimination and the effect was more obvious for CAT scale. However, using LAM scale is undoubtedly beneficial from the permitted area above extremely dislike/dislike.

SELECTED REFERENCES


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

Jirangrat, W

2012-06-26