Comparison of Mean Scores and R-indices for Consumer Preferences of Apple Cultivars

M.A. Cliff, M.C. King, and C. Hampson

Agriculture and Agri-Food Canada, Pacific Agri-Food Research Centre, Summerland, B.C., Canada V0H 1Z0

Additional index words. preference testing, consumer testing, R-index, fresh apples, flavor, visual appearance, *Malus ×domestica*

Abstract. Conventional (analysis of variance, mean preference scores) and novel (R-index) methodologies for hedonic assessments of 'Silken' and 'Creston' apple (*Malus ×domestica* Borkh.) cultivars from the breeding program at Summerland, B.C., were compared with the standard cultivars Royal Gala, Jonagold, and Golden Delicious. Visual and flavor preferences were evaluated for either three or five cultivars by panels of 50 to 200 consumers. Consumers were successfully able to evaluate five apple samples at a given session. Significant differences in mean preference scores and R-indices were found among cultivars. Both 'Silken' and 'Creston' had higher flavor and lower visual preference ratings than did 'Royal Gala'. Results were consistent for both methodologies when panels consisted of 100 or more consumers. R-index, however, expressed the results as a probability rather than a mean score, and was a more understandable and interpretable measure of consumer preference than were preference ratings.

Obtaining reliable predictions of consumer preference is an important stage in the breeding and commercialization of new apple cultivars. Traditionally, consumer preferences or hedonic measurements are evaluated by pair-comparison, ranking, and rating procedures. Typically, 5-, 7-, and 9-point category scales and uni- or bipolar magnitude estimation scales have been used (Pearce et al., 1986). The 9-point hedonic scale, developed originally to assess acceptability of military foods (Jones et al., 1955), has become an industry benchmark. Results from this scale are remarkably "stable" regardless of panel size and region (Stone and Sidel, 1985).

Recently, signal detection procedures have been adapted for difference (O'Mahony, 1988; O'Mahony et al., 1979) and hedonic testing (Vié et al., 1991). These new R-index methodologies have the sensitivity of difference testing with the added benefit of being able to quantify the difference, like scaling procedures. Vié et al. (1991) proposed calculating an R-index for hedonic responses that expressed consumer preference as a probability of preferring a test product over the control.

The R-index takes on values between 0 and 1 with the control designated as 0.5. Samples with R-indices >0.5 are preferred over the control; the reverse is true for those with R-indices <0.5. The R-index can also be thought of as a "hypothetical paired-comparison" (n =

100) between the sample and the control. An R-index of 0.8 indicates that 80% of consumers would prefer the experimental sample and 20% would prefer the control.

This new methodology has been used to evaluate consumer preferences of honey (Swanson and Lewis, 1991/92) and apple juice (Cliff et al., 1996), but has yet to be widely accepted or understood by the industry. Only recently have Bi and O'Mahony (1995) published statistical tables to determine the critical values for the R-indices, i.e., the amount by which the R-index must differ from 0.5 to be statistically significant.

To date, the R-index preference methodology has not been widely used; thus, little is known about its effectiveness and stability across varying numbers of consumers. Therefore, the purposes of this research were to: 1) compare mean preference scores and R-index values for visual and flavor preferences of two new apple cultivars with those of three standard cultivars; 2) evaluate the effect of panel size on R-index estimates; and 3) assess whether consumers can successfully evaluate five apple samples in a given session.

Materials and Methods

Apples. Five apple cultivars grown at Pacific Agri-Food Research Centre, Summerland, B.C., were evaluated in this research: three commercial cultivars ('Golden Delicious', 'Jonagold', 'Royal Gala') and two new cultivars ('Creston', 'Silken') from the breeding program at Summerland. 'Creston' is a cross of 'Golden Delicious' × 'NJ381049'. 'Silken' is a cross of 'Honeygold' × 'Sunrise' (Quamme et al., 1997). Both cultivars have shown promise in preliminary sensory evaluation panels, but have not been evaluated

using a consumer panel. Fruit were selected to be defect- and disease-free. 'Royal Gala' was selected as the control and was presented in all sample sets.

For the visual evaluation, six fruit of each cultivar were polished and placed on white polyethylene trays labeled with three-digit random numbers. For the flavor evaluation, either three or five apple samples (2-cm wedges) were placed on white paper plates, labeled with three-digit random numbers. For the five-sample evaluation, all cultivars were evaluated. For the three-sample evaluation, either set 1 ('Silken', 'Golden Delicious', 'Royal Gala') or set 2 ('Creston', 'Jonagold', 'Royal Gala') was evaluated. Sample orders were assigned using a completely randomized design.

Consumer evaluation. Panelists were recruited from visitors to the Living Landscape Festival (Kelowna, B.C., 1996) who liked apples and were interested in evaluating new apple cultivars. Evaluations were conducted between 1000 and 1500 hr on 2 d, in portable white cardboard "taste panel booths" placed on tables covered with white paper.

For the visual evaluation, panelists (n = 201) evaluated all five apple cultivars according to a completely randomized design. For each sample, they placed a mark on the line scale (8 cm) according to their degree of liking. Scales were labeled at 1, 3, 5, and 7 cm with the terms "dislike," "neutral," "like moderately," and "like very much," respectively.

For the flavor evaluation, consumers evaluated either five cultivars (n = 163) or three cultivars [set 1 (n = 151) or set 2 (n = 162)]. The particular set was assigned at random to the panelists. All samples were evaluated according to a completely randomized design. Consumers evaluated the samples using the previously described scale.

Statistical analysis. Panelists' evaluations were quantified using a digitizing pad, by measuring the distance between their mark, in centimeters, and the origin on the 8-cm line scale. The effects of cultivar and gender on scores were analyzed using analysis of variance. Fisher's least significant difference (LSD) was used for multiple comparisons. Calculations were conducted using the SAS GLM procedure (SAS Institute, 1990).

For calculation of R-indices, the responses "dislike," "neutral," "moderately like," and "like very much" were given values of 1, 2, 3, and 4, respectively. Scores that lay between these categories were assigned to the nearest category. Frequency counts of the number of scores per category were determined and R-indices were calculated using the formula outlined by O'Mahony (1986). 'Royal Gala' was designated as the reference and assigned an R-index of 0.5. Significant differences between R-index values were determined using critical values published by Bi and O'Mahony (1995).

Results and Discussion

Comparison of methodologies and cultivars. Analysis of variance of the visual and flavor scores indicated highly significant dif-

Received for publication 2 Mar. 1998. Accepted for publication 3 June 1998. Contribution no. 1070. We thank Kimberly Dever, Rod King, and Paul Ferguson for assistance with the evaluations. The cost of publishing this paper was defrayed in part by the payment of page charges. Under postal regulations, this paper therefore must be hereby marked *advertisement* solely to indicate this fact.

ferences among cultivars ($P < 0.0001$), but no effect of panelists' gender ($P > 0.40$). Therefore, only cultivar means are reported in Table 1.

For the visual evaluation ($n = 201$), 'Royal Gala' had the highest ($\bar{x} = 6.6$) and 'Creston' and 'Silken' the lowest mean scores ($\bar{x} = 5.4$) ($n = 201$). These results parallel those reported as R-indices (Table 1) ($n = 201$), where 'Royal Gala' had the highest (0.5) and 'Creston' and 'Silken' the lowest (0.35 and 0.37) R-indices. In other words, 'Creston' or 'Silken' would be visually preferred over 'Royal Gala' only 35% of the time.

The rather low visual preferences may be attributed to the low familiarity or visual recognition of both these cultivars in the marketplace. 'Creston' is greenish yellow with red stripes, with a total over-color of 65%, while 'Silken' is a transparent cream color (Quamme et al., 1997) that is uncommon among apple cultivars. The visual acceptance of these cultivars may improve as consumers gain familiarity/confidence with their other quality attributes (flavor/texture).

'Creston' had the highest mean preference score (6.7) and R-index value (0.60), for flavor ($n = 163$) (Table 1). Its flavor would be preferred 60% of the time when compared with 'Royal Gala', probably because of its crispness, firmness, juiciness, and aroma (Quamme et al., 1997). Similarly, the high preference score for 'Silken' ($\bar{x} = 6.2$) may be attributed to its sweetness, acidity, and good textural attributes (Quamme et al., 1997).

Sample number. Flavor preferences for set 2 ('Creston', 'Royal Gala', and 'Jonagold') were consistent whether tasted as a set of three or five apples (Table 1). Apparently, consumers were not only willing, but able, to successfully evaluate five apple samples in a given session.

However, results from set 1 ('Silken', 'Jonagold', and 'Golden Delicious') did vary between the three- and five-sample evaluations. 'Golden Delicious' had the lowest preference score ($\bar{x} = 5.6$; $R = 0.47$) in the five-apple evaluation, but its score did not differ from the control ($\bar{x} = 6.2$; $R = 0.50$), in the three-sample evaluation. This discrepancy may be due in part to paradigm differences between the three- and five-sample sets. Although the consumer is asked to make an independent evaluation of his/her liking for each apple, human perception is relative (Parducci, 1963). McBride (1982) found that the perception of the sweetness of a flavored milk varied with the sweetness range pre-

Table 1. Preferences for apple cultivars expressed as means and as R-indices for visual and flavor ratings.

Cultivar	Means ²	R-index ³				
		Visual rating				
No. of consumers:	201	50	100	150	201	
Golden Delicious	5.9 bc	0.34*	0.42	0.45	0.42	
Creston	5.4 c	0.40	0.39*	0.39*	0.35*	
Royal Gala (control)	6.6 a	0.50	0.50	0.50	0.50	
Silken	5.4 c	0.38	0.40*	0.41*	0.37*	
Jonagold	6.3 ab	0.42	0.47	0.48	0.47	
Flavor rating ⁴						
No. of consumers:	163	Set 1		Set 2		
		151	162	50	100	163
Golden Delicious	5.6 c	6.2 a		0.50	0.50	0.47
Creston	6.7 a		6.8 a	0.59	0.63*	0.60*
Royal Gala (control)	5.9 bc	6.2 a	6.0 b	0.50	0.50	0.50
Silken	6.2 b	6.3 a		0.51	0.56	0.56
Jonagold	6.0 bc		6.3 b	0.61	0.52	0.52

²Means separation within columns and parameters by LSD, $P \leq 0.05$.

³Within a column, R-indices followed by an asterisk (*) are different from 'Royal Gala' by critical values (Bi and O'Mahony, 1995), $P \leq 0.05$.

⁴Each evaluation consisted of five samples (cultivars), except sets 1 and 2, in which three samples were used.

sented to the panelists. In this study, consumers apparently were better able to differentiate among samples when presented with a wider range of product.

Panel size. Results for both the visual and flavor preferences were consistent for panels of all sizes, except $n = 50$ (Table 1). While a sample size of 50 has been used in previous studies, results obtained here indicated that 50 panelists do not give reliable estimates of consumer preference. This finding is consistent with the recommendations of Basker (1988), who suggested that at least 100 panelists are necessary for consumer research.

In this research, both methodologies (mean scores, R-indices) were successful in elucidating differences in consumer preference among apple cultivars. While the results were consistent between the methodologies, the usefulness and understandability differed. The mean preference scores expressed results in relative terms, which are somewhat difficult to interpret without familiarity with typical consumer responses. In contrast, the R-index expressed results as probabilities, which are easily interpreted in meaningful terms with little or no prior experience.

Literature Cited

- Basker, D. 1988. Critical values of differences among rank sums for multiple comparisons. *J. Food Technol.* 42:79-84.
- Bi, J. and M. O'Mahony. 1995. Table for testing the significance of the R-index. *J. Sensory Studies* 10:431-437.
- Cliff, M.A., M.C. Dever, C. Scaman, and B.J. Edwards. 1996. Evaluation of R-indices for preference testing of apple juices. *Food Qual. Prefer.* 8:241-246.
- Jones, L.V., D.R. Peryman, and L.L. Thurston. 1955. Development of a scale for measuring soldier's food preferences. *Food Res.* 20:512-520.
- McBride, R.L. 1982. Range bias in sensory evaluation. *J. Food Technol.* 17:405-410.
- O'Mahony, M. 1986. Sensory evaluation of food: Statistical methods and procedures. Marcel Dekker, New York.
- O'Mahony, M. 1988. Sensory difference and preference testing: the use of signal detection measures, p. 145-175. In: H. Moskowitz. *Applied sensory analysis of foods*. CRC Press, Boca Raton, Fla.
- O'Mahony, M., J. Kulp, and L. Wheeler. 1979. Sensory detection of off-flavors in milk incorporating short-cut signal detection measures. *Dairy Sci.* 62:1857-1864.
- Parducci, A. 1963. Range-frequency compromise in judgement. *Psychol. Monogr.* 77:1-50.
- Pearce, J.H., B. Korth, and C.B. Warren. 1986. Evaluation of three scaling methods for hedonics. *J. Sensory Studies* 1:27-46.
- Quamme, H.A., R.A. MacDonald, C. Hampson, and W. Couriard. 1997. Apple catalogue-1997. Agriculture and Agri-Food Canada, Pacific Agri-Food Research Centre, Summerland, B.C.
- SAS Institute. 1990. User's guide: Statistics. vers. 5 ed. SAS Inst., Cary, N.C.
- Stone, H. and J.L. Sidel. 1985. Sensory evaluation practices. Academic, San Diego.
- Swanson, R.B. and C.E. Lewis. 1991/1992. Premium honeys: Response of sensory panelists. *Food Qual. Prefer.* 3:215-221.
- Vié, A., D. Gulli, and M. O'Mahony. 1991. Alternative hedonic measures. *J. Food Sci.* 56:1-46.