Use of Preference Testing to Identify Tolerance Limits for Fruit Visual Attributes in Apple Breeding

C.R. Hampson¹ and H.A. Quamme

Agriculture and Agri-Food Canada, Pacific Agri-Food Research Centre, Summerland, BC V0H 1Z0, Canada

Additional index words. consumer, fruit shape, fruit size, lenticels, *Malus × domestica*, russet, sensory evaluation

Abstract. Sensory evaluation methods were used to establish tentative guidelines for screening apple (*Malus × domestica* Borkh.) breeding selections for four visual attributes. A panel of 42 regional consumers rated sample selections for fruit size on the 7-point “Just Right” (JR) scale, for fruit shape on a 7-point hedonic (liking) scale, and for the appearance of lenticels and stem bowl russet (SBR) on a 7-point affective (acceptability) scale. The panel most preferred a fruit about 7.5 cm in diameter. No evidence was found for range bias or for differences between yellow and red apples in size preference. Women and panelists over 55 years of age tended to prefer a slightly smaller apple. Panelists liked all the most common apple shapes. Lenticels generally became unacceptable when they exceeded 1.0 mm in diameter, but lenticel density was not related to acceptability. For red or yellow apples, SBR was acceptable on average, provided its maximum extent did not exceed about 55% of the fruit diameter. The panel’s tolerance to SBR resembled that of a larger regional population, and their fruit size preferences resembled those reported elsewhere for European consumers. Similar methods could be used by other breeders to assess the preferences of their target consumer population.

Attractive fruit appearance is commonly considered to be important to the commercial success of an apple cultivar. Consumers use appearance to assess fruit quality at the point of purchase, although repeat sales also depend on internal fruit quality. Quality defects such as bruises or stem punctures are clearly unacceptable. Color is widely deemed to be a cardinal attribute of appearance (Canadian Horticultural Council, 1979; Goedegebure et al., 1990). Fruit size is also important, but populations differ in the size preferred (Goedegebure et al., 1990; Redalen, 1988), and the preferences of Canadian consumers do not appear to have been characterized.

Little quantitative information is available regarding consumer preferences for visual characters such as fruit shape, while tolerance limits for lenticels or russet have not been documented. Acceptability guidelines would be useful to the apple breeder when selecting seedlings in the field. For example, if a selection has excellent internal quality but a tendency for stem bowl russet (SBR), is it worthy of further evaluation or not?

Kappel et al. (1995, 1996) recently used sensory evaluation techniques to establish breeding guidelines for sweet cherry (*Prunus avium* L.) and pear (*Pyrus communis* L.), based on consumer preferences. The goal of this investigation was to establish initial guidelines for screening apple seedlings for fruit size and shape, appearance of lenticels, and the presence of SBR.

Materials and Methods

Orchard management and fruit storage. All fruit were harvested from trees at the Summerland research facility. Most of the samples (6–10 fruit per genotype) were gathered from first-stage, numbered breeding selections growing on M.26 rootstocks in double rows, supported by post-and-wire trellises, and planted at 1.8 × 0.6 m (alleys 3.1 m). Fruit of named cultivars came from a parent block of freestanding trees growing on semidwarf rootstocks (4.6 × 3.7 m). All trees were irrigated and managed according to commercial recommendations (British Columbia Ministry of Agriculture, Fisheries and Food, 1996), except that no calcium-containing fertilizers were applied to the breeding selections. Fruit were selected to exhibit each trait beyond the range typically seen in commerce and beyond the expected range of acceptance.

All fruit were harvested in early Oct. 1996, placed in perforated plastic bags inside wooden apple boxes, and stored in a 1 °C common air storage. Apples were washed and buffed with a soft cloth prior to evaluation to remove dirt.

Consumer panel recruitment and panel protocols. Volunteers were solicited by advertising in two local newspapers. Forty-two consumers (21 male, 21 female) were obtained, all from the southern Okanagan Valley of British Columbia. Twenty-seven were 18 to 55 years old, and 15 were over 55. The relatively high proportion over 55 is typical of the regional population.

Panel sessions were conducted over six evenings in mid-Mar. 1997 at the Summerland research facility. Each person attended two evening sessions 1 week apart. The consumers were familiarized with the rating systems being used, but were not otherwise trained. Each person independently rated the coded samples in the order on their scorecard, and each scorecard was differently randomized. The scores were based on whole samples (three to five apples), not on individual apples within a sample. The panel was reminded to focus only on the trait being assessed. Fruit were displayed under artificial light in covered white plastic trays.

Fruit size study. Three red and three yellow sets of apples were rated on the 7-point “Just Right” (JR) scale (Meilgaard et al., 1991), where 1 = not nearly large enough, 2 = not large enough, 3 = not quite large enough, 4 = just right, 5 = a little too large, 6 = too large, and 7 = much too large. Each set consisted of five size categories, with each category represented by three apples. The sets covered the full size range, only the large end of the range, or only the small end of the range, in order to permit detection of range bias, the phenomenon by which the JR point changes, depending on the range of samples presented (McBride, 1982). The yellow apples were ‘Golden Delicious’. The red apples were all globose, blushed-red apples with little or no calyx end lobing (‘Co-op 9’, ‘Jonagored’, ‘Empire’, ‘Spartan’, ‘BB-122-21’, ‘Idared’, ‘BH-31021’, ‘Canada Baldwin’, and ‘Haralson’).

After panel sessions were complete, each apple was weighed and its diameter (two measurements perpendicular to each other) and maximum length were measured to the nearest 0.01 cm with digital calipers (Mitutoyo Digimatic, Tokyo).

Fruit shape. The consumers used a 7-point hedonic scale (Meilgaard et al., 1991; Peryam and Pilgrim, 1957) to rate fruit shapes, where 1 = dislike extremely, 2 = dislike moderately, 3 = dislike slightly, 4 = neither like nor dislike, 5 = like slightly, 6 = like moderately, 7 = like extremely. Sample apples were presented on their sides. Color varied among the 26 samples, but none was completely yellow or green. Breeding selections were used where possible to obtain the shape range desired and to avoid potential cultivar prejudices. The shape categories were taken from Watkins and Smith (1982), except that cylindrical and cordate were added to describe the shapes of two genotypes more accurately. Shapes were rated in a crossover design (Cox, 1957), where half the consumers rated the first 13 genotypes in Session I and the second 13 in Session II. The other half

Received for publication 17 Feb. 1999. Accepted for publication 5 Dec. 1999. Agriculture and Agri-Food Canada Contribution No. 1083. We thank John Hall for his statistical advice, Warren Walters for his technical assistance, and the consumer panel for its input. 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.

¹To whom requests for reprints should be addressed. E-mail address: hampsonc@em.agr.ca

rated the second set of 13 in Session I and the first set in Session II. After all sessions, the authors measured maximum and minimum diameter (from which taper was calculated) and maximum length, all to the nearest 0.01 cm, and weight (g) for each fruit in each sample. Calyx end lobing was rated on a 0–9 scale where 0 = absent and 9 = extreme.

Lenticels. Samples of 20 genotypes were collected, representing a range for the density, color, and size of lenticels. All genotypes were red, and all but one ('Jonamac') were breeding selections. Three fruit per genotype were placed on their sides for the evaluations. A crossover design was used, with judges evaluating 10 selections per session. The fruit were rated on a 7-point affective scale (1 = highly unacceptable, 2 = moderately unacceptable, 3 = slightly unacceptable, 4 = neither acceptable nor unacceptable, 5 = slightly acceptable, 6 = moderately acceptable, 7 = highly acceptable).

After all sessions, the number of lenticels in a 2×2 -cm area was counted on two mid-cheek areas per fruit. The diameter of three midcheek lenticels per apple was measured with digital calipers to the nearest 0.01 cm. Lenticel color was recorded (white, beige, brown).

Stem bowl russet. SBR was rated on a scale similar to that used for lenticels, except that intent to buy was included (Meilgaard et al., 1991), where 1 = highly unacceptable, certainly would not buy; 2 = moderately unacceptable, probably would not buy; 3 = slightly unacceptable, might not buy; 4 = neither acceptable nor unacceptable; 5 = slightly acceptable, might buy; 6 = moderately acceptable, probably would buy; and 7 = highly acceptable, certainly would buy. All judges rated six yellow selections in the first session and 17 red selections in the second. Apples were presented stem end up (three per genotype).

After the sessions, fruit diameter, maximum extent of SBR as a percentage of apple diameter, and "average" extent of SBR as a percentage of area in top view were recorded for each apple. For the "average" extent, three to seven measurements of the diameter of the russet were taken per fruit. The higher number of measurements was taken if the russet was more uneven. Area was estimated using the average radius of the russet and the average fruit diameter, assuming both were circular.

Additional consumers were solicited at a convention on 21 Feb. 1997. The attendees were from all regions of British Columbia. Of the 101 respondents, 50 were female, 49 were male, and two did not indicate their gender. Seventy-three were between 18 and 55 years of age, 26 were over 55, and two did not state their age category. A subset of nine of the red selections was rated for SBR acceptability. A separate analysis for this subset of nine, as rated by the panel, was also run and compared with the convention attendees' scores.

Statistical analysis. Previous work with the scales used has demonstrated that their properties approximate those of an interval scale and parametric statistics can be used (Land and Shepherd, 1988; Meilgaard et al.,

1991; Peryam and Pilgrim, 1957). All data were analyzed with the SAS statistical software package (SAS Institute, 1990). Split-plot analysis of variance (ANOVA) was conducted (SAS procedure general linear model) to check for differences in scores attributable to consumer age category, sex, age $\times$ sex (tested with error a), and effects of sample and remaining interaction terms tested against residual (error) variance. For shape and lenticels, the model also included a term for session (i.e., differences in mean scores of samples from one session to the next), tested against error variance. Means were separated with the Waller–Duncan k ratio t test (k ratio = 100). Panel mean scores were regressed on objective measurements for the visual attributes (e.g., acceptability was regressed on lenticel density and diameter). Multiple linear regression used the SAS regression procedure with forward stepwise inclusion of independent variables. Only variables that contribute significantly to the model ($P < 0.05$) remain after this procedure.

Results and Discussion

Fruit size. The ANOVA showed that age and sex affected fruit size preference in some of the data sets, with women and judges over 55 years of age tending to prefer slightly smaller apples (data not shown). The JR rating was a linear function of apple diameter (Fig. 1 a and c). Fruit weight and length did not contribute significantly to the multiple linear model, probably because they represented redundant measures of size. Weight was curvilinearly related to JR score (Fig. 1 b and d).

Interpolating from the equations obtained, a JR score of 4 (ideal size) corresponded to a fruit diameter of 7.5 cm (yellow) or 7.7 cm

(red). The slope of the line indicated a low tolerance to departures from the ideal size, with fruit diameters only 1 cm different from the ideal being considered too large (JR = 6) or too small (JR = 2) on average. Separate regressions for each gender and age category found the ideal diameter to be between 7.4 and 7.7 cm (≈ 3 inches) for all subgroups, or about box size 100. This size is similar to the preferred fruit size of consumers in Germany and the Netherlands (7.0–8.5 cm), and the United Kingdom (7.0–7.5 cm) as reported by Goedegeure et al. (1990). Redalen (1988) noted that 90% of the Norwegian consumers in his survey found an apple of 90 g to be "about right." A much larger apple was preferred by the consumers here, ≈ 190 g (Fig. 1 b and d).

The JR point was virtually the same for the three size ranges for both yellow (Fig. 2) and red apples (data not shown). Either range bias was absent or the size ranges did not differ enough to detect it. However, the fruit presented considerably exceeded the size range found in most retail outlets.

Shape. The ANOVA for fruit shape showed no significant session effect, indicating that any difference in the experimental environment or fruit appearance from one session to the next made no difference to scoring (data not shown). Consumer age category and gender did not influence scores. The r^2 of the ANOVA was low ($r^2 = 0.29$, data not shown), indicating that much variation in score remained unexplained. The range in hedonic scores for genotypes of similar shape (Fig. 3) suggests that other aspects of appearance were affecting the evaluation. Only cordate, oblong, ovate, and cylindrical apple shapes were disliked (score < 4). All other shapes scored 4.0 to 5.5 (neutral to "like slightly"). When hedonic scores were regressed

Fig. 1. "Just Right" scores for size of red (▲) and yellow (△) apples as related to (a, c) fruit diameter and (b, d) fruit weight. The 95% confidence intervals are depicted by dotted lines.

Fig. 2. "Just Right" scores for three size ranges of yellow apples.

Fig. 3. Mean panel scores for acceptability of different fruit shapes.

on the objective measurements, the best model obtained explained $\approx 80\%$ of the variation in scores, and was a complex function of length, diameter, lobing, and taper (Table 1).

In summary, no clear-cut shape preferences emerged from the data. Judges rated most of the shapes in the middle of the scale (neutral to "like slightly"), suggesting that they did not feel strongly about shape. All the most common apple fruit shapes were liked similarly.

Lenticels. When scores were regressed onto objective measurements, acceptability was linearly related to lenticel diameter (Fig. 4a) and unrelated to lenticel density (Fig. 4b). Lenticels larger than 1.0 mm generally were rated below neutral (<4). This work suggested that genotypes with lenticels larger than 1 mm could be screened out during field selection. No data could be found in the literature for comparative purposes. Disagreement among consumers was higher for selections with large lenticels, producing a wider confidence interval at that end of the linear function (Fig. 4a).

Lenticel color affected acceptability ($P \leq 0.01$). The mean scores were: 6.4 ± 0.08 for white ($n = 252$), 5.3 ± 0.11 for beige ($n = 252$), and 4.0 ± 0.10 for brown lenticels ($n = 336$). White lenticels were usually smaller and unobtrusive, so color may have been confounded with diameter.

Stem bowl russet. Consumer age and gender did not significantly influence scores. Only six samples of yellow apples were scored, but plots suggested that tolerance for SBR was similar for red and yellow apples, and the data were pooled for regressions. Scores were related to the maximum (Fig. 5a) and the average extent of SBR (Fig. 5b). The upturned end of the curve on the latter reflects that totally russeted fruit were more acceptable to some consumers than heavy streaks of russet.

A subset of genotypes was also rated by 101 consumers at a convention. Again, scores were unaffected by age or gender. Acceptability scores for the subset were plotted to compare the responses of the convention attendees and the recruited consumer panel

Table 1. Regression of shape preference on objective measures of shape.

Variable	Estimate	Standard error	Partial r^2
Intercept	3.73	1.83	NS
Length	-0.88	0.17	0.28***
Diameter	1.21	0.21	0.19***
Calyx end lobing	-0.30	0.06	0.24***
Taper	-1.72	0.60	0.10*
Total			0.81

NS, *, ** Nonsignificant or significant at $P \leq 0.05$ or 0.001.

(Fig. 6). The scores showed similar trends, but the convention attendees tended not to use the low end of the scale. The difference could arise from reluctance to use the negative end of the scale (Land and Shepherd, 1988), an effect of sample size, or a true difference in preference. The recruited consumer panel may have had a greater than ordinary interest in apples, and consequently could have been more discriminating.

If a rating of 4 (neutral) is chosen as the minimum desirable score, then SBR became unacceptable if its maximum extent exceeded $\approx 55\%$ (panel) or 71% (convention) of the fruit diameter. Usually only a few streaks of russet reached past the "average" extent, indicating that such streaks may have been important to acceptance. The linear relationship indicates that the less SBR, the more acceptable the appearance, but for breeding purposes, a guideline of $\approx 50\%$ could be used for field screening. Similarly, selections with russet covering more than $\approx 20\%$ of the area in top view could be rejected (Fig. 6a). No studies of other populations could be found in the literature for comparison.

Consumer panels provided useful information on fruit size preferences and tentative tolerance limits for lenticels and SBR. The information could be used to form initial guidelines for screening breeding selections. Additional samples of consumers can be surveyed over time to increase confidence in the guidelines. The consumer panel's preference for fruit size closely resembled that of a larger population of European consumers (Goedegebure et al., 1990), and their tolerance of SBR was similar to, or lower than, that of a large regional population sample.

Apple selections whose visual attributes lie outside the guidelines may still appeal to some consumers, and may have niche markets. However, they may lack the widespread acceptance or "mass appeal" important to large retail operations. Consumer subgroups reportedly differ in taste preferences for apples (Dailliant-Spinnler et al., 1996; Stebbins 1994). Populations may also differ in preferences for visual attributes. Tastes may also change over time. The methods described here could be used by other fruit breeders to assess the preferences of target consumer groups of interest.

Fig. 4. Acceptability of lenticels as related to (a) diameter and (b) density. The 95% confidence intervals are depicted by dotted lines.

Fig. 5. Consumer panel acceptability of stem bowl russet (SBR) as related to (a) maximum extent of SBR and (b) average extent of SBR. The 95% confidence intervals are depicted by dotted lines.

Fig. 6. Subset of nine samples rated for acceptability of stem bowl russet (SBR) by consumer panel (○) and convention attendees (▲) relative to (a) maximum extent of SBR and (b) average extent of SBR.

Literature Cited

- British Columbia Ministry of Agriculture, Fisheries and Food. 1996. Tree fruit production guide for commercial growers, interior districts. Province of British Columbia, Victoria, B.C., Canada.
- Canadian Horticultural Council. 1979. A survey of Canadian consumers concerning apples and other fresh fruit. Ronalds-Reynolds and Co., Toronto.
- Cox, D.R. 1958. Planning of experiments. Wiley, New York.
- Dailliant-Spinnler, B., H.J.H. MacFie, P.K. Beyts, and D. Hedderley. 1996. Relationships between perceived sensory properties and major preference directions of 12 varieties of apples from the Southern Hemisphere. *Food Quality and Preference* 7:113-126.
- Goedegebure, J., H. Barendse, and F. Pladdet. 1990. For the sake of the consumer. Dutch Fruit Growers' Organization, The Hague, Netherlands.
- Kappel, F., R. Fisher-Fleming, and E.J. Hogue. 1995. Ideal pear sensory attributes and fruit characteristics. *HortScience* 30:988-993.
- Kappel, F., R. Fisher-Fleming, and E.J. Hogue. 1996. Fruit characteristics and sensory attributes of an ideal sweet cherry. *HortScience* 31:443-446.
- Land, D.G. and R. Shepherd. 1988. Scaling and ranking methods, p. 155-185. In: J.R. Piggott (ed.). *Sensory analysis of foods*. Elsevier, London.
- McBride, R.L. 1982. Range bias in sensory evaluation. *J. Food Technol.* 17:405-410.
- Meilgaard, M., G.V. Civille, and B.T. Carr. 1991. *Sensory evaluation techniques*, 2nd ed. CRC Press, Boca Raton, Fla.
- Peryam, D.R. and F.J. Pilgrim. 1957. Hedonic scale method of measuring food preferences, p. 9-14. In: *Studies in food science and technology. 1. The methodology of sensory testing: Contributions to the IFT symposium in Pittsburgh*. In: *Food Technol.* 11(9). (Insert with separate paging, between p. 472 and 473.)
- Redalen, G. 1988. Quality assessment of apple cultivars and selections. *Acta Hort.* 224:441-447.
- SAS Institute. 1990. *SAS/STAT user's guide*, version 6 ed. vol. 2. SAS Inst., Cary, N.C.
- Stebbins, R.L. 1994. Consumer's tastes in apple varieties profiled. *Good Fruit Grower* 45(15):32-33.
- Watkins, R. and R.A. Smith (eds.). 1982. *Apple descriptors*. Intl. Board for Plant Germplasm Resources Secretariat and Commission of European Communities Secretariat, Rome.