Development of a Scheme to Evaluate Consumer Apple Variety Preferences

Vicki A. McCracken, Bernd Maier, Terri Boylston, and Thomas Worley

Introduction

The opportunity for continued increases in exports of U.S. (mainly Washington State) apples is dependent on providing high quality fruit of the desired varieties to the international consumer. Southeast Asia, specifically Taiwan, Hong Kong, Singapore, Thailand, and Malaysia are currently the major markets in the Pacific Rim for apples exported from Washington State. China has recently agreed to import U.S. produced apples, and U.S. growers will likely gain access to the Japanese market in the near future.

The number of new varieties of apples introduced in Washington State and other states in the United States has increased dramatically in recent years, allowing apple producers to provide apples with a range of sensory characteristics to the consumer. Although Red Delicious has traditionally been the main apple produced and exported from the Pacific Northwest, several of the new varieties have gained popularity. There is a need to determine the sensory attributes which are important to consumers, and to develop effective marketing programs in which consumer preferences are matched with apple varieties.

There are large differences in the market shares of specific varieties in Washington State and their shares for the world (Figure 1). In the 1989/90 marketing year, Red Delicious and Golden Delicious apples constituted 70 percent and 17 percent, respectively, of total apple production in Washington State. For the same time period, Red Delicious and Golden Delicious apples constituted only 23 percent and 25 percent, respectively, of total world apple production. These shifts may be reflecting changing consumer preferences for sensory characteristics of apples. These differences, which have become larger in recent years,

illustrate the need for Washington producers to not become complacent with production and marketing of Red and Golden Delicious apples. With the introduction of the new apple varieties such as Fuji, Braeburn, and Gala, and the apparent rapid acceptance of these varieties in the international market place, a comparative evaluation of new and traditional varieties is essential.

Accordingly, the general objective of the preliminary research that is discussed in this report was to develop a methodology that can be applied on a general basis to evaluate consumer attitudes towards and acceptance of new varieties of a product. Specifically here, the research investigated the differences in the sensory quality of 11 apple varieties and integrated sensory and instrumental analysis to determine the impact of these parameters on consumer acceptability of a variety.

The remainder of this research report is organized as follows. First, the methods and procedures used in evaluating the apple varieties are identified. Second, the results of the empirical component of the study are presented. And in the last section of the report, limitations of this preliminary research are identified, and recommendations for future work are made.

Methods/Procedures

The apple varieties evaluated in this study were: Red Delicious, Golden Delicious (two different samples), Granny Smith, Fuji, Braeburn, Melrose, Gloster, Gala, Jonagold, Hawaii, and Mutsu. This array of varieties includes the traditionally important commercial varieties in Washington State, and some upcoming varieties such as Fuji, Gala, and Braeburn. Also included were some of the less familiar varieties that are being experimented with, in order to have a wide variation in apple characteristics in the sensory analysis.

The apples used in the study were purchased in November 1992 from selected commercial growers/packers in Washington and Oregon states. The apples were generally size 80/88s with the exception of Gala which were 138s. The apples were stored at 32°F in Washington State University storage facilities in Pullman, Washington.

Figure 1: Apple Variety Share
Washington State And The World
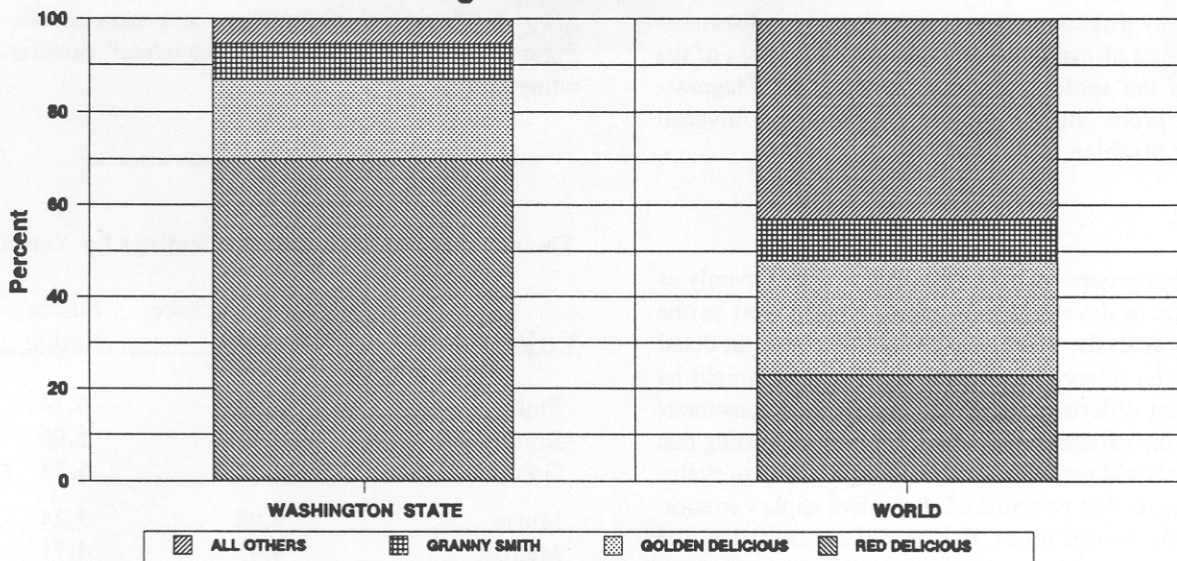
There were three components of the apple variety testing: focus group interviews; sensory (taste test) evaluations; and analytical (instrumental) evaluations. A review of the literature related to consumer preferences for apples provided only limited information on the attributes/bundle of attributes valued or even recognized in an apple by the typical consumer. Hence, the first stage of the research was to conduct focus group interviews in order to identify the attributes that consumers consider when determining whether they like or dislike a specific apple variety. Focus group interviews involve a small group and a moderator. The group is generally selected to include 10-12 individuals with common backgrounds or experiences regarding the subject of the interview. The moderator directs the discussion, but does not lead it. Focus groups have been used successfully, in marketing studies to provide information, for example, on consumers' frames of reference and to establish consumer vocabulary as a preliminary step in questionnaire development (Tull and Hawkins, 1984).

The focus group interviews in this study were conducted on or near the Washington State University campus in Pullman, Washington. Separate groups consisting of domestic consumers and selected international consumers (e.g., Korean, Taiwanese) were used for these depth interviews. The interviews were video taped for later review. The focus group results were used mainly to develop a questionnaire/instrument to use in the consumer sensory (taste test) evaluations.

The sensory evaluations were conducted at the Department of Food Science and Human Nutrition, Washington State University on two consecutive days. One hundred and twenty five untrained panelists participated in the sensory evaluations. Presentation of the samples was randomized with each sample identified with a random 3-digit code. Panelists evaluated apple slices under red lights in individualized booths. Soda crackers and distilled water was provided for rinsing between samples. Prior to evaluation of the samples, panelists were asked to complete a questionnaire regarding their purchasing habits for apples and demographic background. For each apple sample, the panelists evaluated firmness, sweetness, tartness, apple flavor, and overall acceptance. Firmness, sweetness, and tartness were evaluated using a 5-point "just-about-right" scale, with anchor points at 3 for "just-about-right," 1 for too little, and 5 for too much of the given attribute. Apple flavor and overall acceptance were evaluated using a 9-point hedonic scale, with 1 = uncharacteristic apple flavor or dislike apple extremely and 9 = characteristic apple flavor or like apple extremely, respectively.

The analytical (instrumental) evaluations of the apples involved the same characteristics evaluated by the sensory panelists: flavor compounds, soluble solids (sweetness), titratable acidity (tartness), and firmness. Volatile flavors were isolated using headspace analysis techniques and quantified using gas chromatography. Gas chromatography-mass spectrometry were used for

the identification of the volatile flavor compounds. Soluble solids and titratable acidity were measured from juice expressed from the fruit. Soluble solids were determined using a refractometer, and titratable acidity, by titration with sodium hydroxide to determine the amount of malic acid equivalents. Firmness of the flesh of the apples were measured using a Magness-Taylor probe attachment for the Instron Universal Testing Machine.

Results

The focus groups interviews were conducted mainly as guidance in developing the questionnaire used in the sensory analysis, and hence are not discussed in detail here. The interviews did indicate that there might be important differences in the preferences of consumers among the various Asian groups, hence suggesting that Asians should not be treated as a single group in evaluating the market potential of alternative apple varieties. Given the limited funds, this market segmentation was not pursued in this research.

In terms of demographic composition of the sensory panel, the gender distribution was almost equal, with 45.9 percent female and 54.1 percent male panelists. The panelists ranged in age from 19 to 74 years, and averaged about 32 years. Most of the panelists attended or worked at the university, being undergraduate or graduate students (65.5%), staff members (18.9%), or faculty members (13.9%). Most of the panelists indicate that they like to eat apples (98.4%), and most often ate them as fresh apples (96.7%). Before evaluating the different apple varieties, the panelists indicated that their favorite varieties were Red Delicious, Golden Delicious, and Granny Smith, closely followed by Fuji.

The panelists were asked to rank the importance of selected attributes in affecting their purchase of apples. These attributes were divided into two categories: attributes obvious to the consumer before purchase (pre-purchase attributes) and attributes known by the consumer only after purchase (post-purchase attributes). The apple's price and its color were ranked as the most important pre-purchase attributes by 38 and 41 percent of the panelists, respectively. Size and shape of the apple were not viewed by most to be the most important pre-purchase factors. Many panelists deemed the texture of the apple (39%) and its flavor (35%) as the most important factors affecting their post-purchase evaluation. An apple's sweetness and tartness, while important, were not the most important post-purchase attributes for this group of consumers.

The panelists tasted and evaluated the 11 different varieties (2 samples of Golden Delicious apples were included) on firmness, sweetness, and tartness using a 5-point scale (3 being the highest rating) and flavor and overall preference using a 9-point scale (9 being the highest rating). The descriptive results for the three highest and three lowest rated varieties for firmness, sweetness and tartness are summarized in Figures 2-4 and for flavor and the overall preference rating in Table 1.

Table 1
Flavor and Overall Preference Ratings by Variety

| Variety	Overall Preference Rating	Flavor Rating
Fuji	6.22	6.14
Braeburn	5.83	5.96
Golden Delicious 2	5.13	5.38
Mutsu	4.98	5.24
Melrose	4.47	4.71
Granny Smith	4.33	4.64
Jonagold	4.12	4.50
Gloster	4.11	4.71
Gala	4.02	4.07
Golden Delicious 1	3.69	4.51
Red Delicious	3.16	3.61
Hawaii	3.15	3.49

The three varieties that the panelists rated highest on firmness were Fuji, Braeburn, and Granny Smith, rating just right (scale = 3) for over 60 percent of the panelists (Figure 2). The Red Delicious, Golden Delicious 1, Jonagold, and Hawaii varieties received the lowest ratings on firmness, with the general pattern that these apples were not firm enough. Similarly for tartness (Figure 3), Fuji, Golden Delicious 2, and Braeburn received the highest ratings, with approximately 60 percent indicating that the tartness of these apples were just right. The lowest rated apples, the Golden Delicious 1, Hawaii, and Red Delicious varieties, were all judged to be not tart enough. Figure 3 also includes the tartness rankings for Granny Smiths, as this variety is known for its tartness. Grannies followed the pattern of the top ranked varieties, peaking on just right, but was also rated as too tart by a significant number of the panelists. The sweetness of the Fuji apple was rated just right by almost three-fourths of the panelists, followed by Braeburn, Melrose, and Golden Delicious 2 (Figure 4). A significant number of the panelists thought that the Red

Figure 2: SENSORY FIRMNESS
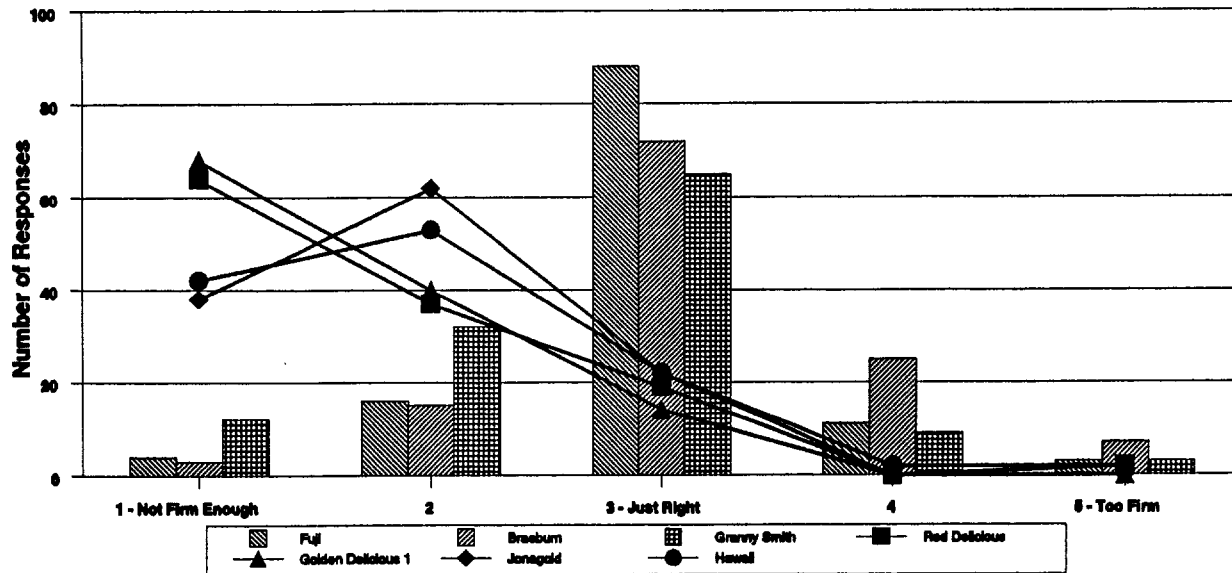
Footnote: Bars denote 3 highest "Just Right" varieties; Lines denote 4 lowest "Just Right" varities.

Figure 3: Sensory Tartness
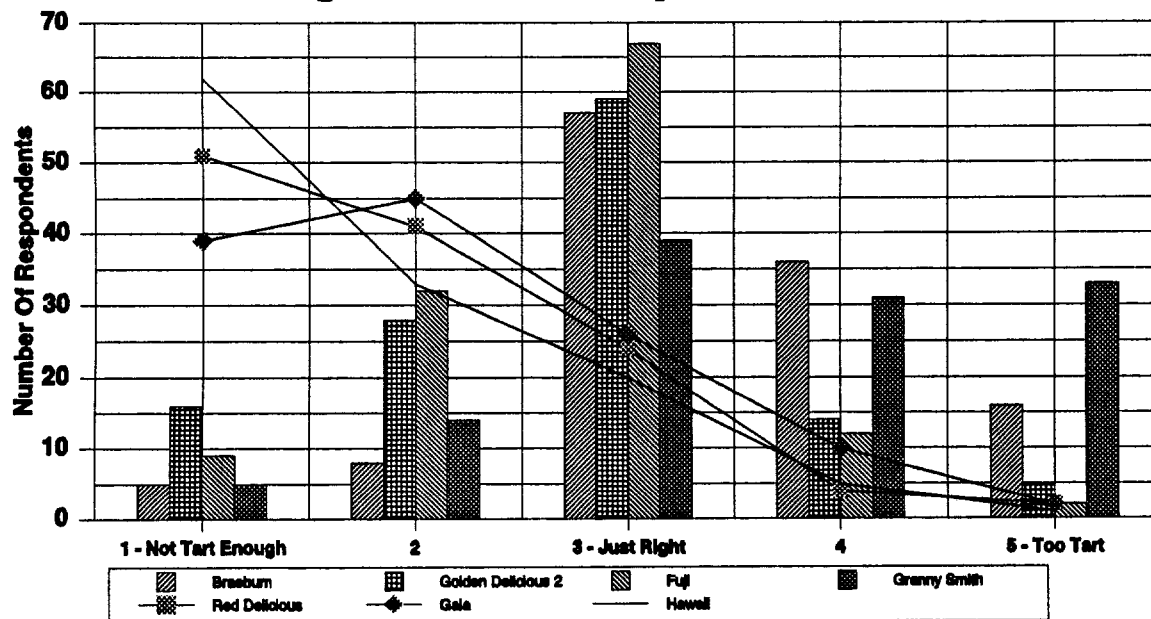
Footnote: Bars denote 3 highest "Just Right" varieties plus Granny Smith; Lines denote 3 lowest "Just Right " varieties.

Delicious, Hawaii, and Granny Smith apples were not sweet enough.

There were large differences in the average flavor rating for the various varieties, ranging from a high of 6.14 for the Fuji apple to a low of 3.49 for the Hawaii apple (Table 1). The Braeburn, Golden Delicious 2, and the Mutsu apples all received relatively high flavor ratings, while the Red Delicious and Gala received relatively low flavor ratings. Panelists were also asking to give each apple an overall preference rating (Table 1). Fuji received the highest score, on average, followed by Braeburn and the Golden Delicious 2. The apples receiving the lowest overall preference ratings were the Hawaii, Red Delicious, and Golden Delicious 1. In summary there was considerable consistency in the sensory ranking of the apple varieties for the various attributes. Fuji consistently ranked highest, followed by Braeburn and Golden Delicious 2. And Red Delicious, Hawaii, and Golden Delicious 1 varieties consistently ranked lowest.

The sensory data were also analyzed using multivariate statistical techniques in order to identify statistically significant relationships among selected variables. A goodness of fit test based on the χ^2 statistic was used to test for significant relationships for the nominally scaled qualitative variables. A simple analysis of variance procedure based on the F ratio was used to test for significant relationships for quantitatively measured variables. Only selected results are mentioned in the following paragraphs.

The relationship between preferred pre-purchase and post-purchase attributes and the age (i.e., younger = 25 years or less; older = over 25 years) and gender of the panelist were analyzed using χ^2 tests. The color of the apples, while ranked high by most panelists, was found to be a more important pre-purchase attribute for younger panelists than for older panelists. And the shape of the apples, while ranked low in importance by most panelists, was ranked less important for older panelists. The post-purchase attribute ratings did not differ by age, but did differ by gender of the panelist. Females were more likely than males to rate texture of the apple as an important attribute, while males tended to rate flavor as more important than females.

Goodness of fit tests were also used to identify relationships between the sensory ratings for firmness, sweetness, and tartness and other characteristics of the panelist or apple. The results indicate that panelists rated firmness differently depending on the specific variety of apple. Surprisingly, the firmness ratings did not depend upon the panelist's ranked importance of firmness on their post-purchase decisions. The firmness ratings differed by gender of the panelist--females were more likely to rate the apples as being too soft.

Not surprisingly, the tartness ratings differed across the specific varieties considered as well as by the panelist's stated preference for tartness (prior to tasting the apples). The interactive effect of preference for tartness by variety on the tartness sensory rating was significant for a number of the varieties including Braeburn, Melrose, Red Delicious, Fuji, Gloster, and Hawaii. The tartness sensory rating also differed depending upon the panelist's stated preference for the importance of sweetness. Panelists for whom sweetness was a more important post-purchase attribute were more likely to rate the tartness of an apple as okay or too tart; and panelists for whom sweetness was not so important were more likely to rate the apples as not tart enough. In addition, the tartness sensory ratings differed by gender of the panelist, with the differences depending upon variety. For example, females were more likely than males to indicate that the Melrose apple was too tart, and more likely to indicate that the Hawaii apple was not tart enough.

The sweetness sensory ratings of the panelists also differed by variety, as well as by stated importance of sweetness in post-purchase apple decisions. Panelists rating sweetness as relatively unimportant were more likely to indicate that the sweetness of an apple sample was just right, as compared to panelists who rated sweetness as an important post-purchase factor. Gender appeared to affect sweetness sensory ratings, as females were more likely than males to rate an apple as not sweet enough.

The analysis of variances for the flavor sensory ratings documented differences in ratings across varieties. These ratings also differed depending upon the panelist's stated importance of the pre-purchase attributes of color, shape, and price and the post-purchase attributes of flavor and tartness. For example, panelists who rated flavor as an important post-purchase attribute were more likely to give lower sensory flavor ratings to the apples. There were significant differences by age but not gender in the flavor ratings, with younger panelists tending to give higher flavor ratings relative to the older panelists.

The overall preference ratings that the panelists assigned to each apple variety differed significantly by apple variety, as well as by all pre-purchase and post-purchase attributes except for size and sweetness. Overall preference ratings also varied by gender and age of the panelist. Males tended to have higher ratings than females, and older panelists tended to have lower ratings than younger panelists. These results suggest that there are potentially definable market segments that have stronger or weaker preferences for the various apple varieties.

Figure 4: Sensory Sweetness
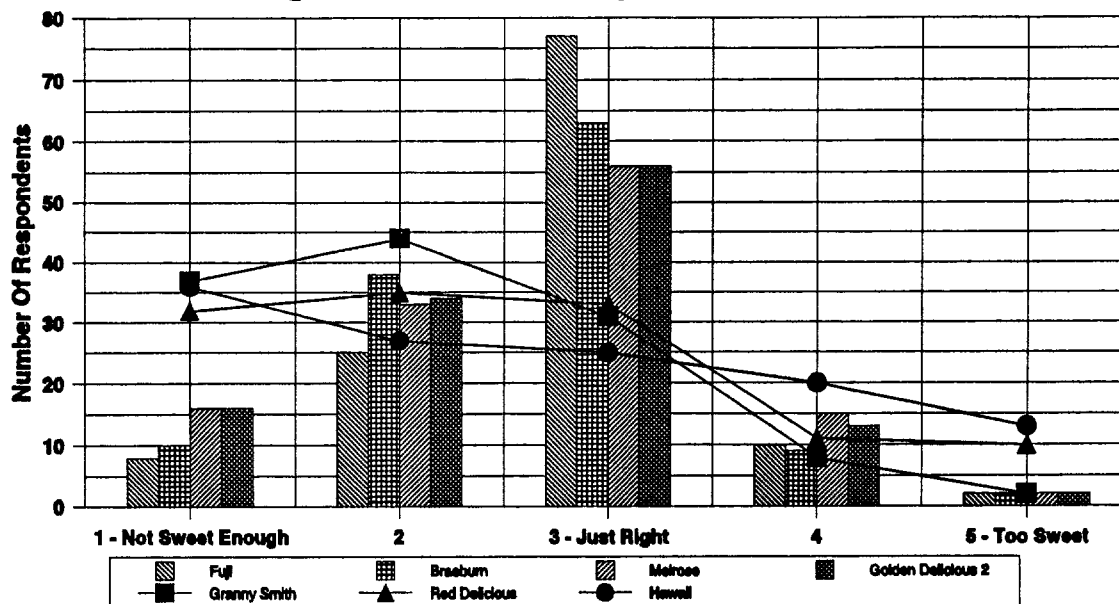
Footnote: Bars denote 4 highest "Just Right" varieties; Lines denote 3 lowest "Just Right" varieties.

Analytical testing was also used to measure the attributes of the apple varieties. The averages for the replications for the analytic measures and the ascending ranks for each attribute are listed in Table 2. These ranks are based solely on the level of each attribute, and do not consider consumer acceptability of the level of the attribute and the level of other attributes in the same apple. Combining the consumer sensory ratings of the attributes with the analytic measures will shed some light on the desired level of the attribute in an apple.

The average sensory ratings for firmness, tartness, and sweetness were plotted against the corresponding analytic measures for each variety (not shown here, for sake of space)[1]. The six varieties with the lowest analytic values for firmness (i.e., Jonagold, Melrose, Golden Delicious 1, Gloster, Red Delicious, and Hawaii) were also rated significantly lower by consumers with respect to the desirability of their texture. The four varieties with (intermediate) analytic firmness values between 45 and 55.5 Newton Force (i.e., Golden Delicious 2, Granny Smith, Mutsu, and Gala) all had sensory ratings for texture of approximately 2.4 (where 3 indicates that the consumer perceived the texture of the apple as "just right"). Fuji had the second highest analytic firmness value (i.e., 61.6 Newton Force), and was ranked by consumers as having the most acceptable texture (i.e., sensory rating of 2.7). Braeburn had the highest analytic firmness value (i.e., 67.5 Newton Force), but was ranked by

consumers as the second most acceptable apple in terms of texture (sensory rating of 2.5). Hence combining the sensory and analytic measures for firmness suggests that firmness beyond a certain level (i.e., for the apples used here that level was about 62 Newton Force, the level for Fuji) is not necessarily more desirable.

The results for tartness and sweetness are not quite so clear cut, possibly because what is considered to be an acceptable level of each attribute depends upon the level of the other attributes in an apple variety. This appears to be particularly true for high values of the attribute (i.e., tartness or sweetness).

Considering tartness only, the varieties with the smallest amounts of titratable acidity were rated lowest by the sensory panel. For example, Hawaii, Red Delicious, Golden Delicious 1, Gala, Jonagold, and Mutsu had less than .31 grams of malic acid per 100 grams of apple and were given average ratings of less than 2.1 by the sensory panelists. Fuji was rated highest by the sensory panel with regards to tartness, and had an intermediate analytic value for tartness (.34 grams of malic acid per 100 grams of apple). Increases in acidity beyond this level appeared to be less acceptable to the sensory panelists (e.g., Melrose, Golden Delicious 2, Gloster, and Granny Smith). An interesting exception was Braeburn: Braeburn had the highest acidity (i.e., .58 grams of malic acid per 100 grams of apple) yet was still acceptable to the panelists (2.30) likely because of its high level of soluble solids (sweetness).

Table 2. Analytical Measures For Tartness, Sweetness, and Firmness

Variety	Tartness (a)	(b)	Sweetness (c)		Firmness (d)	
Braeburn	.5798	1	14.98	3	67.470	1
Fuji	.3405	6	15.43	1	61.578	2
Gala	.2160	9	12.33	9	55.502	3
Golden Delicious 1	.2044	10	11.80	11	36.152	10
Golden Delicious 2	.4090	4	13.35	5	45.185	6
Gloster	.4940	3	12.93	8	40.045	9
Granny Smith	.5346	2	11.83	10	49.016	5
Hawaii	.1522	12	11.70	12	40.672	7
Jonagold	.2741	8	14.38	4	33.849	12
Melrose	.3955	5	13.32	6	34.745	11
Mutsu	.3071	7	15.20	2	52.874	4
Red Delicious	.1822	11	13.10	7	40.113	8

(a) Titratable Acidity is measured as grams of malic acid per 100 grams of apple.
(b) Numbers are the variety rank for the specific attribute. The ranks were assigned in ascending ordering of the values for the attribute, where 1 was assigned to the largest value.
(c) Sweetness is measured as soluble solid percentage.
(d) Firmness was tested with an Instrom Universal Testing Machine using a Magness-Taylor probe (11 mm diameter). The measurement units are Newton Force.

Sweeter apples (based on the analytic measures) were generally rated as more acceptable with regards to sweetness than less sweeter apples. Fuji, for example, had the highest soluble solids level (15.4%) and was rated highest by the sensory panelists (2.6). Mutsu and Braeburn also had high soluble solids levels (15.2% and 15.0%, respectively) and were rated high by the sensory panelists (2.3 and 2.4, respectively). Hawaii had the lowest soluble solids level (11.7%) and received the lowest sensory rating (1.8). Interestingly, Red Delicious had an intermediate level of soluble solids (13.1%) yet was rated low by the panelists (1.9), likely because of its low level of titratable acidity (tartness).

The overall preference ratings by the sensory panelists were plotted against the analytic measures of firmness, tartness, sweetness, and the ratio of sweetness to tartness. These graphs show some interesting patterns, but should be interpreted with caution since the overall preference rating is affected by other than the single attribute included in the graph.

There was a general tendency for firmer apples to be more preferred by the sensory panelists than less firm apples. Fuji and Braeburn, the varieties with the

highest firmness values (61.6 and 67.5 Newton Force, respectively), were also the most preferred varieties (6.2 and 5.8, respectively). The least preferred varieties, namely Red Delicious (3.2) and Hawaii (3.1), had intermediate analytic firmness values (40.1 and 40.7 Newton Force, respectively). And the less firm apples, Jonagold (33.8 Newton Force) and Melrose (34.7 Newton Force), received intermediate preference ratings (4.1 and 4.5, respectively). Obviously texture is not the only attribute driving the sensory preference rating, but it does appear to be an important attribute.

Consistent with a previous reported result for tartness, increases in the titratable acidity in an apple were desirable up to a level and then further increases had a negative impact on the desirability of an apple. The least tart apples, Hawaii (.15 grams of malic acid per 100 grams of apples) and Red Delicious (.18 grams of malic acid per 100 grams of apples), were also the least preferred apples (3.12 and 3.16, respectively). The most preferred apple, Fuji (6.2) had an intermediate level of titratable acidity (.34 grams of malic acid per 100 grams of apples). Varieties with higher levels of titratable acidity (e.g., Granny Smith) received low sensory preference ratings.

There was a tendency for sweeter apples to receive higher overall preference ratings by the sensory panelists. Fuji had the highest level of soluble solids (15.4%) and was the most preferred (6.2), and Hawaii had the lowest level of soluble solids (11.7%) and was the least preferred (3.1). There were some deviations from this general tendency as Red Delicious had a moderate level of soluble solids (13.1%) but was the second least preferred variety (3.2).

Another issue concerns how each of the individual attribute (firmness, tartness, sweetness, and flavor) ratings correlate with overall preference rating for each variety in the sensory test. The sensory ratings for each attribute were plotted against the overall varietal preference rating. These graphs indicate a fairly strong positive correlation between the average sensory ratings for tartness, sweetness, and flavor and the preference rating for the variety. This suggests that the panelists were fairly consistent in their ratings of the individual attributes for a variety and their overall preference for the variety, and that there was a general tendency for an apple variety to have either a satisfactory or an unsatisfactory level of all of these attributes. This relationship was not so strong for the texture attributes.

Implication/Recommendations

The procedure developed in this research appears to have potential for evaluating new apple varieties. In evaluating the potential of new varieties, there is a need to determine the sensory quality attributes which are important to consumers. This research investigated the differences in the sensory quality of eleven different apple varieties and integrated sensory and analytical/instrumental to determine the impact these parameters have on consumer acceptability.

The specific findings for the apple varieties should be interpreted as preliminary as they are based on limited observations at one point in a single season, but they are suggestive of the type of information generated by the procedure developed in this study. Future applications of this procedure should involve sensory panels conducted and analytical measurements taken at different times throughout the season, as apple varieties have different storage capabilities, and during additional seasons. The number of sensory panelists should be increased, and should include people from different ethnic groups. This additional data will be useful to assess which new varieties have market potential, and to identify market niches for these varieties and to measure the magnitude of these niches.

Endnote

[1] The sensory ratings for firmness, tartness, and sweetness, originally measured as 1, 2, 3, 4, 5, were recorded as 1, 2, 3, 2, 1 in order to calculate the average sensory ratings. A higher average, therefore, can be interpreted as the apple containing a more desirable level of the attribute. A lower average indicates that the apple contains a less desirable level of the attribute (i.e., either too little or too much).

Reference

Donald S. Tull and Del I. Hawkins. 1984. *Marketing Research: Measurement and Method*, 3rd Edition New York: Macmillan Publishing Co.