Sensory Ratings of Commercial Plain Yogurts by Consumer and Descriptive Panels

STEVEN J. HARPER, DEBBIE L. BARNES, FLOYD W. BODYFELT, and MINA R. McDaniel
Department of Food Science and Technology
Oregon State University
Corvallis 97331

ABSTRACT

Ten trained panelists rated the intensity of sensory descriptors from 17 commercial plain yogurts, and 153 consumer panelists who consumed plain yogurt at least once a month evaluated the same yogurts for acceptance. Consumer responses were correlated with attribute ratings, and consumer responses and trained panel ratings were correlated with analytical measurements. There was a wide range in consumer hedonic ratings for all sensory factors with significant differences found for appearance and sweetness liking. Consumers found the vast majority of samples too sour and not sweet enough. Hedonic ratings for sweetness, sourness, appearance, and texture liking were positively correlated with overall liking. Samples rated highest in overall liking had sourness ratings closest to just right on the "just right" scale. Consumers responses for hedonic and just right factors differed by sex of respondent.

Ratings for all descriptors by the descriptive panel were significantly different. Sourness, with high intensity ratings, was most important in describing plain yogurt. Samples rated most favorably by consumers had lower intensity ratings for overall intensity, sourness, acetaldehyde, saltiness, and astringency and higher intensity ratings for sweetness, milkiness, and cooked milk. Titratable acidity and pH were correlated with many trained panel descriptors important to consumer acceptability. Better control of pH by yogurt processors would result in more favorable sourness levels, which should increase consumer acceptability.

(Key words: yogurt, sensory, consumer acceptability)

Abbreviation key: PC = principal component, PCA = principal component analysis.

INTRODUCTION

Culturing and processing techniques (4, 9), as well as skillful marketing (7), all play an important role in producing and selling quality yogurt. To reverse the current decline of 7.6% in the total consumption of cultured nonfrozen yogurt (20), it is equally important to understand better the acceptance and preference patterns of consumers regarding flavor and texture. This is not only the case with fruit-flavored yogurt, which comprises the largest share of the market, but with plain yogurt as well, with its approximate 11% share (13). Plain yogurt, although most often consumed with other foods or used as an ingredient, is the base for fruit-flavored yogurts. This product base, therefore, must be of the highest quality to ensure the optimal quality of the final fruit-flavored yogurt.

Yogurt score cards (3, 16) have been developed and used to provide direction to processors by evaluating yogurt quality based on criticisms or lack of criticisms judged by small expert panels. However, detailed descriptive analysis data, although previously reported for fruit-flavored yogurt (1), are not available for plain yogurt. Descriptive analysis results in a complete sensory profile of the product regardless of the quality level. Plain yogurt was therefore evaluated for flavor-by-mouth descriptors using a trained descriptive panel and for sensory attribute liking using a consumer panel to provide this descriptive information. Consumer panel hedonic ratings were...
TABLE 1. Reference standards used by the trained descriptive panel to define the flavor-by-mouth descriptors for plain yogurt.

<table>
<thead>
<tr>
<th>Descriptors</th>
<th>Reference standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetaldehyde</td>
<td>.66 ppm and 2 ppm acetaldehyde (IFF, Union Beach, NJ)/2% milk</td>
</tr>
<tr>
<td>Cooked milk</td>
<td>2% milk heated to 90°C for 30 s</td>
</tr>
<tr>
<td>Caramel</td>
<td>caramel candy (Kraft, Chicago, IL)</td>
</tr>
<tr>
<td>Milky</td>
<td>2% milk</td>
</tr>
<tr>
<td>Buttery</td>
<td>butter</td>
</tr>
<tr>
<td>Cheesy</td>
<td>Parmesan cheese (Frigo Cheese, Green Bay, WI)</td>
</tr>
<tr>
<td>Yeasty</td>
<td>.1% (wt/wt) baking yeast (Universal Foods, Milwaukee, WI)/H₂O solution</td>
</tr>
<tr>
<td>Salty</td>
<td>.2% (wt/wt) NaCl/H₂O solution</td>
</tr>
<tr>
<td>Sweet</td>
<td>5% (wt/wt) sucrose/H₂O solution</td>
</tr>
<tr>
<td>Sour</td>
<td>.32% (wt/wt) lactic acid (J. T. Baker Chemical, Phillipsburg, NJ)/H₂O solution</td>
</tr>
<tr>
<td>Astringent</td>
<td>.11% (wt/wt) alum (McCormack, Baltimore, MD)/H₂O solution</td>
</tr>
<tr>
<td>Bitter</td>
<td>.08% (wt/wt) caffeine (Fisher Scientific, Fair Lawn, NJ)/H₂O solution</td>
</tr>
</tbody>
</table>

correlated with trained panel ratings, and consumer and trained panel responses were correlated with analytical measurements.

MATERIALS AND METHODS

Samples

All commercial plain yogurts (17 brands) available during August 1989 in the Corvallis, OR market were evaluated. These included 7 nonfat, 9 lowfat, and 1 full-fat yogurt, all produced by different manufacturers. Because of the relatively large number of samples tested, it was not possible to obtain samples with the same expiration dates. However, samples were chosen on the basis of the latest expiration date available, which ranged from 7 to 28 d beyond the testing date. Samples were obtained no more than 3 d prior to testing. All containers (varying in size from 177.5 ml (.5 pt) to 946.7 ml (1 qt)) of each brand were stamped with the same expiration date. Yogurts from all containers of each sample were gently mixed together (20 folds) to ensure product consistency. After mixing, the samples were stored at 1.5°C in 2.3-kg polypropylene containers.

Consumer Panel

Screened panelists were at least 10 yr of age and consumed plain yogurt at least once a month. One hundred fifty-three consumers evaluated two samples each of plain yogurt while seated in portable booths at the August, 1989 Benton, OR County Fair. The consumer panelists were served 40 ml of sample in 60-ml odorless plastic cups. Sample temperature at time of serving was 3.3°C. Overall liking, appearance, sweetness, sourness, and texture liking were rated by using a nine-point hedonic scale (1 = dislike extremely, 5 = neither like nor dislike, 9 = like extremely). "Just right" ratings were also obtained for sweetness, sourness, and thickness using the seven-point just right scale (1 = way too little, 4 = just right, 7 = way too much). Demographic data were obtained regarding the age, sex, and consumption patterns of panelists. A randomized, balanced, incomplete block design was used (5), yielding 17 observations per sample.

Descriptive Analysis Panel

Ten volunteer students and staff from Oregon State University (Corvallis, OR) participated in training and subsequent evaluation of the yogurts. Six training sessions were conducted to determine the flavor-by-mouth descriptors of the plain yogurts. The panelists had previously been trained to evaluate fruit flavored yogurts; hence, they were well acquainted with the type of product being evaluated.

Descriptors agreed upon by the panel included overall intensity, acetaldehyde, cooked milk, caramel, milky, buttery, cheesy, fruity, yeasty, salty, sweet, sour, astringent, and bitter. Panelists were provided with references for all of the descriptors (Table 1) to aid in evaluation of the samples.

Testing was conducted at the Sensory Science Laboratory at Oregon State University (Corvallis, OR) in individual booths under red
lighting to mask color differences. Spring water (Aqua Cool, Portland, OR) was provided for rinsing between samples. Serving temperature and sample size were the same as for the consumer panel evaluations.

A randomized, balanced, complete block design was used (5) that resulted in two replications for all samples. Samples were tested on 4 consecutive d with the first replication being presented on the first 2 d and the second replication on subsequent days. Panelists evaluated three sets of three different samples on each day. Samples were counterbalanced throughout testing. Time breaks between sets were included to minimize panelist fatigue. One of the yogurt samples was presented throughout the evaluation as a test for possible day effects.

Analytical Measurements

Titratable acidity (determined by titration to pH 8.3 using a pH meter) and pH were measured at 22°C using the standard methods for dairy products (15). The pH meter was a Corning 125 (Corning, Medfield, MA) with a Sensorex epoxy body, sealed-reference combination electrode (S200C, Sensorex, Stanton, CA). The pH meter was calibrated with buffers of pH 3 and 9 (Microessential Laboratory, Brooklyn, NY), and 10N sodium hydroxide (J.T. Baker Chemical, Phillipsburg, NJ) was used for the titration analyses.

Statistical Analysis

Analysis of variance, principal component analysis (PCA), and correlations were conducted using a 1987 SAS statistical package (SAS Institute Inc., Cary, NC). The ANOVA and correlations of treatment effects were tested for consumer panel results. For the trained panel, correlations and PCA were applied, and panelist, replication, treatment, and interaction effects were tested using ANOVA. A fixed treatment, mixed effects F test model was used with the F value being determined by the treatment mean square divided by the panelist by treatment mean square. Trained panelists that judged a descriptor in a manner inconsistent with other panelists were deleted in the analysis for that particular descriptor.

RESULTS AND DISCUSSION

Trained Panel Results

Differences were significant for all of the descriptors tested except milky, cheesy, cara-

<table>
<thead>
<tr>
<th>TABLE 2. Mean ranges, F values, and statistical significances of plain yogurt descriptors from the trained panel evaluation.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plain yogurt descriptors</strong></td>
</tr>
<tr>
<td>-----------------------------</td>
</tr>
<tr>
<td>Overall Intensity</td>
</tr>
<tr>
<td>Sour</td>
</tr>
<tr>
<td>Astringent</td>
</tr>
<tr>
<td>Salty</td>
</tr>
<tr>
<td>Sweet</td>
</tr>
<tr>
<td>Cooked Milk</td>
</tr>
<tr>
<td>Buttery</td>
</tr>
<tr>
<td>Bitter</td>
</tr>
<tr>
<td>Yeasty</td>
</tr>
<tr>
<td>Fruity</td>
</tr>
<tr>
<td>Acetaldehyde</td>
</tr>
<tr>
<td>Cheesy</td>
</tr>
<tr>
<td>Milky</td>
</tr>
<tr>
<td>Caramel</td>
</tr>
</tbody>
</table>

¹Nonsignificant at P = .05.

*P ≤ .05.

**P ≤ .01.

***P ≤ .001.

****P ≤ .0001.
Overall intensity means for the samples were in the moderate to large range with sourness comprising the largest component. No other descriptor had means near those of sourness. This indicates the importance of sourness and the need to pay particular attention to the consumers' ratings.

Acetaldehyde, although not found to be significantly different among samples, had the next highest intensity rating after sourness. Its role in imparting a distinctive green (10, 11) or "green apple" (3) taste to plain yogurt has been widely noted. Trained panelists in this study also noted the green apple character both in the samples rated lower in acetaldehyde and in the 0.66-ppm acetaldehyde reference standard. However, the panelists characterized the 2-ppm acetaldehyde reference standard as being like white glue. This specific aroma was subsequently found in four of the yogurts and characterized as acetaldehyde.

Descriptors in the low to moderate range included astringent, salty, cheesy, sweet, milky, cooked milk, buttery, and bitter. These descriptors played a lesser role in describing the overall intensity of the plain yogurt. Finally, fruity, yeasty, and caramel descriptors probably did not play any important role in the description of the plain yogurt because of their low intensity ratings.

Multivariate tools such as PCA can be used to identify related, underlying characteristics of a food product such as yogurt (6, 14). The PCA of these data revealed three principal components (PC) predominantly explaining the relationship between the samples and the descriptors: 1) PC1, which explained 27% of the model, was characterized by salty, yeasty, sweet, buttery, and astringent; 2) PC2 was characterized by sourness and overall intensity and opposed by sweetness and cooked milk and explained 22% of the model; and 3) PC3 explained 12% of the model and was characterized by bitter and cooked milk and opposed by milky. Graphing of PC1 versus PC2 (Figure 1) serves to separate the yogurt samples into...
four distinct groups. Samples 6 and 16 were characterized by high bitterness and cooked milk and very low overall intensity and sour values. Samples 2, 3, 4, 10, 11, and 15 did not have high values for any of the components but exhibited moderate levels of milky, cooked milk, sweetness, and buttery and low levels of overall intensity, sourness, and astringency. Samples 1, 5, 7, 9, 12, and 14 also did not have high values in any of the components but showed moderate levels of overall intensity, sourness, and bitter and low levels of sweetness, cooked milk, buttery, and milky. Finally, samples 8, 13, and 17 had high values for both PC1 and PC2, which resulted in high levels of overall intensity, sour, astringency, salty, yeasty, sweet, and buttery.

No day effects were discovered in the analysis of the data.

**Consumer Panel Results**

Consumers found significantly \( P \leq .001 \) large differences between appearance and sweetness liking; mean ranges varied widely for the most liked and least liked samples (Table 3). These findings were somewhat surprising. The appearance liking of plain yogurt would not be expected to vary to a large extent because of the similarity of the ingredients and production techniques. However, in the mixing process used to create a uniform sample, the curd mass was disturbed, and this resulted in yogurt appearance that ranged from creamy smooth to curdy-lumpy. Consumers were probably detecting a difference based on consistency appearance. The wide range in sweetness liking values was not expected, because sweetness is not normally thought to be associated with plain yogurt. Consumers may have been responding more to the lack of sourness in the samples, because the samples rated highest in sweetness liking did not have the highest sweetness intensity ratings. Overall liking of the samples, despite wide variances in means (Table 3), was not significantly different \( P = .066 \). Most of the overall liking ratings of the samples were in the like slightly range, a few in the like moderately range, and none in the like very much range (Figure 2). These observations may suggest the need for a modification of the sensory quality of plain yogurt brands in the tested market and a better understanding by processors of consumers' perceptions of desirable plain yogurt attributes. Sourness liking was not significantly different among the samples. This may be partially explained by the lack of correlation with sourness intensity determined by the trained panel. The nonsignificant differences found for texture may have resulted from the blending process used to ensure homogeneity of each sample or from a difference among consumers in their perceptions of desirable texture.

Just right ratings for sweetness, sourness,
and thickness resulted in significant \( P < .05 \) differences among samples (Table 3). Consumers found all samples to be not quite sweet enough and, with the exception of one sample, slightly too sour. Thickness ratings ranged from slightly too thick to not quite thick enough. Five samples were rated quite close to just right, five samples were slightly too thick, and seven samples were not quite thick enough, which leads perhaps to the conclusion that consumers were fairly satisfied with the thickness, at least with respect to the way these samples were presented.

There were a number of factors rated by the consumers that were significantly \( P < .05 \) correlated with overall liking. These correlations (Figure 3), as well as Figures 4 to 6, have been graphed with the sample points connected for visual clarity. Samples were all independent of each other. Sweetness and sourness liking showed the highest correlation with samples rated the most liked in overall liking (Figure 3). Appearance liking was not as highly correlated with overall liking but was still highly significant (Figure 3). Texture liking was the least highly correlated but again showed a general increase in the texture liking as the overall liking increased (Figure 3). One question that is raised when correlating these hedonic ratings is whether consumers, when they assign a high rating to a sample for overall liking, are truly influenced and automatically give proportionally higher ratings to the other hedonic factors. When hedonic factors have been rated concurrently, individual factor ratings have not been shown to influence ratings of other factors (12). Close scrutiny of these data revealed that there was not a pattern in the data to suggest that the highest liked sample also had the highest ratings for all of the hedonic factors. It is interesting to note that sweetness liking was rated higher than sourness liking, and both were rated lower than overall liking. This may highlight the need to improve the flavor balance in plain yogurt. However, even if yogurt is in perfect balance, sourness liking usually rates lower than sweetness liking (2).
Figure 5. Nine-point hedonic mean ratings from the consumer panel evaluation for overall liking, sweetness liking, and sourness liking divided by sex. Overall liking (a), sweetness liking (b), and sourness liking (c).

If a comparison is made between the overall liking of the samples and the ratings of just right sourness (Figure 4), it is apparent that as the overall liking of the samples increased, the ratings of sourness moved closer to the just right level. Samples with lower levels of acceptance had too high levels of sourness. No other just right comparison with overall liking showed this trend.

When divided into male and female categories, analysis of the data revealed dramatically different response patterns, even though the mean response values of males and females across all the samples were not significantly different. These differences were especially evident for overall liking, sweetness liking, and sourness liking. For overall liking, there does not appear to be a pattern to the differences (Figure 5). However, in the case of sweetness liking, the samples rated the highest by the males were rated lower by the females, and the samples rated lower by the males were in general rated higher by the females (Figure 5). This pattern was also apparent for sourness (Figure 5). This difference in rating was probably a primary reason for nonsignificant differences among the samples for sourness liking. Unfortunately, no significant correlations were found between sweet and sour intensities as measured by the trained panel and liking as divided by sex.

Differences were also found in the ratings for just right sweetness and sourness. The just right scale results for sweetness revealed that females did not find the sweetness level as acceptable as did the males, and, in general, rated the samples not sweet enough and further from just right (Figure 6). For just right sourness (Figure 6), females found 10 samples to be farther from just right and five samples to be closer to just right when compared with...
ratings by the males. Females also did not find any of the samples to be not sour enough, but males found three samples to be not sour enough. These differences in flavor perception by sex have been observed by other researchers (18) and need to be considered by processors when developing yogurt products and targeting particular market segments. Some attention to these differences appears appropriate because Knutson et al. (8) reported that females comprise the larger share of the yogurt consumers. No correlation was found between any of the consumer rated factors and pH and titratable acidity.

Consumer Panel and Descriptive Panel Correlations

Although PCA did a good job of separating the yogurts into distinctive groups based on the quantification of the descriptors, it did not allow clear separation of the highest consumer rated samples from the lowest rated samples. Highest liked and lowest liked samples were scattered throughout all of the groups. The ratings data were carefully scrutinized, which allowed for broad characterization of the most liked and least liked samples. The most liked samples exhibited higher ratings for sweetness, milkiness, and cooked milk descriptors while manifesting lower ratings for overall intensity, sourness, acetaldehyde, saltiness, and astringency descriptors. The lowest rated samples showed a high level of saltiness and low levels of milkiness and sweetness.

Descriptive Panel and Analytical Measurement Correlations

Significant correlations were found for both pH and titratable acidity as related to a number of descriptive panel descriptors. Titratable acidity was positively correlated \((P \leq .01)\) with overall intensity, saltiness, astringency, and sourness and negatively correlated \((P \leq .01)\) with milkiness and sweetness. Sweetness and bitterness were positively correlated \((P \leq .01)\), and overall intensity, saltiness, astringency, and, most markedly, sourness were negatively correlated \((P \leq .01)\) with pH. Thus, samples with lower pH had lower levels of perceived sweetness and bitterness and higher levels of saltiness, astringency, and sourness compared with samples with higher pH. A number of studies (17, 19, 21) have focused on the possible correlation between low pH and high sourness.

The correlations involving pH and titratable acidity were extremely important, because the highest rated samples generally had lower intensity levels of overall intensity, sourness, saltiness, and astringency and higher intensity levels of sweetness. Because pH and titratable acidity were significantly negatively correlated, control of these process parameters—which markedly affect consumer acceptance—is accomplished most easily by strict control of product base pH during production. Maintenance of the desired pH range during the distribution is also critical for providing uniform yogurt product quality. Temperature abuse caused by stocking delays and improperly regulated display cases increases the production of acid by the live organisms in the yogurt, which results in decreased pH.

CONCLUSION

Although titratable acidity and pH did not correlate with consumer rated factors, they did correlate with many descriptive panel descriptors important to consumer acceptability. This relationship, along with the relationship of consumer acceptability to trained panel descriptor levels, should allow processors to predict the acceptability of plain yogurt better by controlling pH during production and distribution. In general, consumers indicated that all of the plain yogurts were too sour. Samples rated highest in overall liking were samples with sourness ratings closest to just right. Production of plain yogurt with higher pH and resultant decreased sourness should serve to increase consumer acceptability of plain yogurt.

Male and female consumers differed in their responses for overall liking, sweetness liking, sourness liking, and for ratings of sweetness just right and sourness just right. This difference is extremely important to yogurt manufacturers as consumption of yogurt by females is higher than for males.

This study provides the industry with important information regarding the relationship of flavor-by-mouth descriptors to consumer acceptability; however, descriptive panel analysis
of texture is needed to provide a complete understanding of consumer preference.

ACKNOWLEDGMENTS

The authors would like to thank the Western Dairy Foods Research Center for its financial support. We also wish to express our sincere appreciation for the panelists whose hard work and dedication made this research possible.

REFERENCES