

stantly looking for cheaper and more efficient methods to transport their products from grower to consumer in the best possible condition, with minimum losses.

The tests made by Reynolds strongly indicate that foil-lined containers for packaging and storing citrus fruit without refrigeration may be a practical approach to this problem. Let me stress, however, that all tests conducted thus far have been made with California and Arizona citrus and may not necessarily hold true with Florida citrus. A research program

for Florida citrus shipments is planned for this season and we are hopeful that results will prove as encouraging as the California-Arizona tests.

These are only the highlights of an unlimited field. There are many more possibilities which could be explored if time permitted, however, I can assure you that we are all looking forward with a great deal of anticipation to the continued development and growth of aluminum foil in the citrus industry.

CONSUMER PREFERENCES FOR CANNED ORANGE JUICES OF 12 TO 22 BRIX-ACID RATIOS¹

RICHARD L. D. MORSE

Florida State University

Tallahassee

Consumer preferences for canned orange juices ranging from 8.2 to 14.8 Brix-acid ratio at 12.6° Brix were reported at the 1952 proceedings of the Florida State Horticultural Society. Likewise reported were preferences for juices ranging from 8° to 15° Brix at both the 12 and the 16 Brix-acid ratios, (Morse, 1952). Juices of higher Brix-acid ratios were more favorably scored although the increase in preference ratings diminished in the higher Brix-acid ratios. The purpose of the present study was to ascertain consumer reaction to juices in even higher Brix-acid ratios.

Secondary to this overall objective was interest in furthering knowledge as to methods for measuring preferences. A complete critical analysis of the methods will not be presented in this paper. However, the results will be reported by test methods so that contrasts will be permitted. The three methods employed were first, the Single Juice Score method; second, the "Blind Rank," so called because the consumer did not know the differences in the juices and ranked the six juices in order of preference; and third, the "Known Rank," in which the respondent was informed as to the significance of code markings on the cans, 12, 14, 16, 18, 20, 22.

All juices were administered to a panel of 69 families with persons over ten years of age participating, in all, 160 persons. The fami-

lies were essentially those members of the 1951 and 1952 panels remaining at the Florida State University. A careful review of those families who formerly participated, but were not a part of this study, indicates no apparent bias. Although, of course, the results of this and previously reported studies are strictly applicable only to the panel families concerned since they were not selected at random from a clearly defined population for them to represent. Technically, therefore, the title of this paper is misleading, since the results are not appropriate for all consumers. A similar study to this is being reported by the United States Department of Agriculture which used the same juices and score sheet in a probability sample of the families of Indianapolis. Any decided biases in this panel of Florida faculty families should be evident from a comparison of these two reports.

The juices were delivered, in the fall and winter of 1953-1954, in person to the office of the faculty member who took the juices (plus score sheet) home for testing, and the score sheets were returned via campus mail. Reminder notes and phone calls were necessary to stimulate returns. However, a genuine spirit of cooperation prevailed among the participants who viewed their role as experimental subjects. The record of cooperation is beyond expectation: of the 70 families (162 persons) contacted, only one failed to score the juices at all, while the other 69 (160 persons) cooperated 100% in all of the four test runs of the Single Score method. Of these, 67 (154 persons) followed through on both the "Blind Rank" and "Known Rank" tests.

¹In cooperation with the Florida Citrus Commission and the Florida Citrus Experiment Station.

The juices used were of 12, 14, 16, 18, 20 and 22 Brix-acid ratios at 12.6° Brix. In the "Single Score method," juices 14 and 20 were repeated.

SINGLE JUICE SCORE METHOD

This test is referred to as the "Single Juice Score" method because only one juice was scored at a time. Each family participating was given a package of four 10 ounce cans; two of one juice labeled with masking tape and marked as "O," and two of another similarly labeled as "K." The family was instructed to refrigerate the juices immediately and at their next *normal* juice consumption time to open the cans labeled "O" for scoring. At their next normal juice consumption time, at least a day later, they were asked to open and score juice "K." The score sheets labeled "O" and "K" were returned via campus mail to the investigator who then scheduled another delivery of juices. This was repeated four times and each time the juices were labeled "O" and "K," although for each family and each test run these represented a different pair of juices. The test was popularized as the "O-K" test to emphasize the importance of order and to give identification to the juices without revealing to the participants the nature of the juices. The masking tape device concealed the raised code numbers in the tops of the cans which revealed the Brix-acid ratio.

At the same time that the "K" score sheet was filled out the "O" score sheet was available for comparison, thus the two scores were not entirely independent. Each pair of juices was given in both orders an equal number of times to avoid order bias. Subsequent analysis of the data did not reveal a significant bias. They had been paired to facilitate distribution because by pairing, the number of contacts with the participants was reduced to four. Also, because such tests extend over a period of time during which there is possible change in juice from storage conditions and change in taste of the participants from climatological and other seasonal influences, all the juices, and even the same juices (ratios 14 and 20), were paired to permit separate analyses of scores for any one of the four stages (test runs) of the experiment.

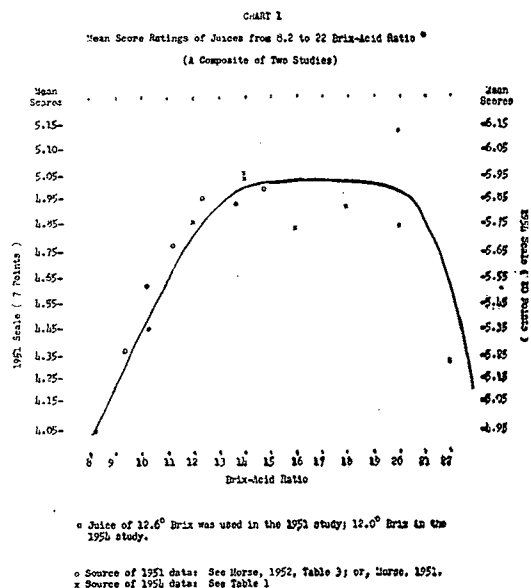
The experimental design employed resulted from extended correspondence with Dr. H. O. Hartley, visiting Professor of Statistics at the Iowa Statistical Laboratory. It consisted of

the assignment an random of code numbers 1 through 8 to the ratios 12 through 22 with 14 and 20 repeated. It provided for a panel of 28 families, assigned at random by number. To these panel families were assigned four pairs of juices to be given over four test runs. The order of these test runs was also randomized. Dr. Hartley described the design as having the following properties: "All juices are tasted by two families in the order '1 then 2' and by two families in the order '2 then 1'. All juices are tasted 14 times first and 14 times second within a run." An additional panel of 28 families was added to increase the reliability of the estimates of juice scores; thus there were 28 sets of two families.

After the experiment was under way the distribution of the 28 sets of families by sex and (representation in the) age classes of 21-40 and 41-60 revealed that only 14 families were needed to give representation to each age and sex class for each of the 28 sets. Representation for 13 of these 14 positions was attained. A comparison of the results from these 13 families with those from the other families revealed no significant differences, so their reports are included in the 160 persons who constitute the population reported on in this paper.

The score sheet was cooperatively developed with Dr. Forrest Clements, then Head, Division of Special Surveys, Bureau of Agricultural Economics, who was directing the Indianapolis study. He conducted a pretest using a 100 point scale, a 7 point scale, and a 10 point scale. The latter proved to be the most sensitive and was adopted. It was highly unstructured and consisted of 10 squares arranged vertically, below which was written "Very Poor" and above, "Excellent." The individual was asked to place a check in that square between the extremes which best expressed his opinion about the juice. (These were later coded to read as scores from 1 through 10 for computational purposes.)

The second half of the score sheet consisted of statements classified by the essential juice characteristics of sweetness, tartness, body, and overall flavor and concluded with a statement regarding disposition to buy. The first three followed a pattern of: (1) not observed, (2) not — enough, (3) just right in —, and (4) too —. The exact wording appears in Table 1 which summarizes the responses. Mr. Lucian Martinez, Research Assistant, who had



laboriously coded the open end "why" questions in the 1952 panel study was well qualified to word this phase of the schedule.

Results

The mean scores increase from 5.76 for juice of 12 Brix-acid ratio to an average of 5.94 for the two estimates (5.93 and 5.95) for ratio 14. Ratios 16 and 18 rated below ratio 14 with mean scores of 5.74 and 5.82, respectively. The two estimates for ratio 20 average the same as ratio 14, but their individual values, 6.12 and 5.74 are disconcerting. They bracket all previous estimates. The estimate for the highest ratio, 22, is sufficiently low at 5.21 for this juice to be classified as inferior. These means and their standard errors are presented in Table 1. The standard errors are so large that none of the means by the usual statistical tests is significantly different. These mean scores are shown in Chart 1 with the 1951 panel mean scores for the juices of ratios 8.2 to 14.8. Since these were based on a 7 point scale the two sets of data were married by equating in the chart the average of the 1951 12-14 ratio juices (4.95) with the average (5.85) of the 12-14 ratios for this study.

A free-hand curve was fitted to the plotted points showing a rise from ratio 8.2 to a plateau from 14 to 20, and a drop to 22. If it were not for the high estimate for ratio 20, it

could be argued that the peak of preference is around ratio 14, with gradual reduction in preference with increase in ratio until 22, for which there is a sharp drop. The legitimacy of the marriage of the 1951 and the 1954 data may also be questioned particularly since juices of 12.6° rather than 12.0° Brix were used in the '51 study, the panels were not identical, and two years had elapsed.

Another method of representing the relative position of these six juices is to sum the differences in the scores of each pair of juices. For example, if a family were to score the "O" juices as 5 and the "K" juices as 7, then the difference is -2. If the other family which had the same juices, but in the reverse order scored them the same, their difference would be -2. (O-K) - (K-O) would equal -4. That is, "O" is 4 points preferred to "K." The same result would be obtained if the scores had been 1 and 3 or 7 and 9, etc. Thus, concentration on differences in scores between paired juices avoids giving undue influence to those who generally give high scores. These scores are presented in Table 2. Juices of ratios 12 and 22 definitely are not favored with 22 least in favor. Ratios 14 and 16 are most favored with some slight reduction in favor for ratios 18 and 20. Interesting is the closer agreement of the two estimates of ratio 20 and the greater disparity of ratios 14, the opposite of the mean scores. Noteworthy, however, is the agreement in general of the two methods of analysis.

A third analytical method would be simply to record whether "O" was rated higher, equal to, or lower than "K," regardless of how large the difference. This resembles the paired comparison test with ties permitted. The same pattern prevails: juices of ratio 12 and particularly 22 found less favor than the other juices with which they were paired.

The evidence, while not entirely clear from the mean scores, but when supported by other analytical methods, indicates ratio 12 to be inferior and 22 decidedly inferior to the intermediate juices. The peak of preference lies in the range of from 14 to 20 ratio.

The second part of the score sheet which provided for a descriptive evaluation of the juices is quite revealing. The data presented in Table 1 gives the frequency with which the 160 persons checked each of the statements. It must be remembered that the participants were completely ignorant of the facts with

TABLE 1

Composite Summary of Scores and Comments by Juices
Single Juice Score Method

	Brix-Acid Ratios							
	12	14	14	16	18	20	20	22
Mean Scores	5.76	5.93	5.95	5.74	5.82	6.12	5.74	5.21
Sweetness								
Not observed	7	12	9	8	6	4	7	5
Too sweet	14	18	23	26	33	46	42	70
Just the right sweetness	80	92	93	92	97	94	96	64
Not sweet enough	59	38	35	34	24	16	15	21
Tartness or Sourness								
Not observed	8	19	17	13	16	18	20	19
Too tart or sour	64	38	35	39	22	19	23	25
Just the right tartness	64	85	90	76	87	82	66	58
Not tart or sour enough	24	18	18	32	35	41	51	58
Body or Consistency								
Not observed	2	10	11	6	9	10	12	8
Too thin or watery	62	61	62	53	65	71	56	70
Just the right body or consistency	92	86	84	96	82	76	83	78
Too thick	4	3	3	5	4	3	9	4
Flavor								
Tastes like fresh orange juice	14	6	12	7	9	12	7	12
Does not taste like fresh orange juice, but still pretty good	68	72	73	69	72	73	56	47
Tastes tinny	34	29	27	29	23	19	27	24
Tastes artificial	37	46	40	43	46	43	57	62
Other taste characteristics	7	7	8	12	10	13	13	15
If this juice were available on the market, would you like it served in your home?								
Yes	48	55	56	48	54	55	54	48
No	112	105	104	112	106	105	106	112

TABLE 2
Sum of Differences in Scores of Paired Juices
By Age Groups and by Juice Brix-Acid Ratio

Age	Brix-Acid Ratio (at 12.0° Brix)								Check Sum
	12	14	14	16	18	20	20	22	
11-20	-11	1	23	4	-7	13	-9	-14	0
21-40	-5	20	2	11	5	23	45	-101	0
41-60	-52	44	4	33	24	2	-13	-43	0
Total	-67	65	29	48	22	38	33	-158	0

respect to the relative sweetness or tartness of the juices. It is gratifying that these frequencies fall into a pattern which conforms to the physical differences in the juices and which

validate the scores. For example, only 14 checked ratio 12 as too sweet, yet 70 claimed this for ratio 22. Likewise 59 checked ratio 12 as not sweet enough, but only 21 checked this for ratio 22. The frequencies for the intermediate juices, while not very different, do move consistently in the direction of the physical characteristics of the juices. Interesting also is the lower frequency of those who considered ratio 12 and 22 to be "just the right sweetness." This same pattern prevails with respect to comments on tartness or sourness. There is a definite tendency to check ratio 22 and one of the 20 ratios as tasting artificial or

having other taste characteristics, and not possessing the attributes of fresh juice. There was a slight tendency to classify juice 12 as "tinny."

The frequency with which the statements under "Body or Consistency" were checked shows no pattern. Likewise there is no pattern to be found in the willingness-to-buy commitment. Since most of the participants normally consumed reconstituted frozen orange concentrate, the buying decision question was not in keeping with their customary choices that they make in the market. Many of those who said they would buy qualified their position with such a statement as "would buy only if concentrate were not available". Secondly, to a concentrate consumer, body or consistency would be in terms of pulp content which, of course, is absent in canned orange juice. So this terminology probably was not particularly well suited to this population.

All the evidence secured by the Single Juice Score method indicates a preference for juices of ratios 14 through 20 over ratio 12, with ratio 22 least preferred. Juice of ratio 22 was most frequently cited as too sweet, not tart enough, and having an artificial or other taste characteristic.

BLIND RANK TEST

Direct quotation from the form letter sent to the panel families best describes the method employed. The format of the letter cannot be preserved, of course. Liberty has been taken to insert explanatory remarks.

"Now that the O-K test is finished we can reveal that there were only six different juices. To these six juices we have assigned the letters: M L R P O E. After the cans have been refrigerated *open all of them at one time*. Taste a little from each, and, as you do so, arrange the cans in the order of their acceptability. Taste and re-taste until you are reasonably certain you have them in the order of your choice. Transcribe this order in the boxes below by writing into the boxes the letters on the cans." Presented horizontally were six boxes labeled from left to right, "Best," "Second Choice," "Third Choice," "Fourth Choice," "Fifth Choice," and "Poorest." A footnote read as follows: "Suggestion: This is difficult. But if you will first select the 2 better and the 2 poorer, and then choose between the 2 juices in each pair, the test may be less difficult."

In order to get not only the relative preference position of each juice, but also a threshold level, the following device was created:

"*The Bonus Juice*—We will attempt to reward you by giving you two cans of the juice you like best. But since our stock is limited, we may have to substitute for your first choice, your second or third choices. If there are any of the juices you would rather not have, even as a bonus, let us know, so those juices can be given to those families who do not consider them objectionable.

"Which juice(s) do you consider to be so objectionable that you would not want as a bonus?" Listed were the letters for the individual to circle and a check box for the individual to mark if he considered none objectionable.

Only one of the 69 families and two individuals failed to perform this test, leaving 156 participants. The failures were the result of moving and apparent loss in the mails.

Results

The frequency with which each juice was assigned the rank position of 1 (Best), 2 (Second Choice), 3 (Third Choice), 4 (Fourth Choice), 5 (Fifth Choice), and 6 (Poorest) is presented in Table 3. Also shown is the

TABLE 3
"Blind Rank" Preference Position of Six Juices
Ranging from 12 to 22 Brix-Acid Ratios

Brix-Acid Ratio	Frequency by "Blind Rank"						Median Rank	Objectionable (Frequency)
	Best	1	2	3	4	5		
12	21	19	16	21	30	19	4.5	43
14	24	25	32	29	28	18	3.4	20
16	24	19	34	36	26	17	3.5	23
18	30	34	28	24	18	22	3.0	20
20	41	28	29	23	21	14	2.8	17
22	14	29	21	27	31	34	4.0	29

median rank of each juice. The juices most favored (of lowest median rank) are ratios 20 and 18. Juices of ratios 14 and 16 are next in favor, followed by an equal difference in median rank by ratio 22 and finally by the least preferred juice, ratio 12. Also of interest in Table 3 is the frequency with which each juice was marked as so objectionable it would not be wanted even as a bonus. Here again the extreme juices are most frequently cited, although contrary to the results of the Single Juice Score method, yet consistent with

the median "Blind Rank" values, ratio 12 was in greater disfavor than ratio 22. Furthermore, if attention is focused on the juice most frequently cited as "Best," it is ratio 20, and if attention is focused on the juice most frequently cited as "Poorest" and most objectionable, it is ratio 12, with 22 a close second.

"KNOWN RANK" TEST METHOD

The bonus juice promised in the "Blind Rank" test method was not awarded and an additional test was inserted. Only one additional family failed to participate in this extra test, leaving 154 persons participating. Again the cause was of a natural sort, unrelated to the test.

The method of this test and the cooperation solicited can again best be presented by quoting at length from the letter-schedule form employed:

"Many of you have expressed doubt as to the difference between the best juices. And others have expressed interest in knowing what the juices are. For these reasons and for the purpose of noting any significant difference in your reactions when you know the characteristics of the test juices, I am asking you to re-take the test with the blind-folds off.

"Juice Characteristics: You will note that each can is stamped in code, and that the first line is different for each can. The code reads: 012, 014, 016, 018, 020, 022. This technically speaking, is the Brix-acid ratio of the juice. And since all juices are of 12° Brix (degree of sweetness), the higher numbers reflect decrease in acidity. . . . In language I understand, all these juices came from fruit from the same grove. The only difference is that the fruit was picked from early to late in the season. The lower numbers are from the early picked Valencia fruit; the higher, from the late picked Valencia fruit.

"Now that you know what to look for in the juices, how do you rank them?" And again were presented the six blank boxes for insertion of the code numbers. Following this was the question: "Which of these juice(s) do you find so objectionable that you would not want as a bonus?" The code numbers were listed for circling and a box provided for checking if none was objectionable.

A post script on this form read: "This is the last test! Many thanks for your assistance in making this study a success. As soon as possible we will send the bonus juices and a

summary statement of your scores on all the tests." A summary sheet was later devised for recording each individual's scores on the four test runs of the Single Juice method and their rankings on the "Blind" and "Known" methods. This form, a letter of explanation reviewing the multistage study, and a sack of the bonus juice was given each participant.

Results

The frequency with which each juice was assigned the rank position of 1 (Best), . . . 5 (Fifth Choice), and 6 (Poorest) is shown in Table 4. Also presented are the calculated

TABLE 4
"Known Rank" Preference Position of Six Juices
Ranging from 12 to 22 Brix-Acid Ratios

Brix-Acid Ratio	Frequency by "Known Rank"						Median Rank	Objectionable (frequency)
	Best	1	2	3	4	Poorest		
12	19	11	10	12	16	86	5.6	60
14	19	23	23	19	65	5	4.1	22
16	31	23	24	54	16	6	3.5	17
18	36	30	43	33	8	4	2.8	13
20	28	37	32	23	25	9	2.9	15
22	21	30	22	13	74	14	3.8	34

median rank values. The favored juices, as in the "Blind Rank" test, were ratios 18 and 20. Likewise ratio 16 placed third. In contradiction with the "Blind Rank" results and contrary to the Single Juice Score placement ratio 22 placed fourth. Ratio 14 ranked in fifth position with ratio 12 trailing.

The "Known Rank" test is marked by a greater number of declared "objectionable" juices at the extremes. Under the "Blind Rank" test the extremes were not known to the participant as they were in this test. Note, for example, the distribution of the 6th rank position by juices for these two tests and the evidence in support of "extreme avoidance" of bias is clear.

The juice most frequently ranked as "Best" is ratio 18 with the other middle juice next highest. The juice most frequently classed as poorest was ratio 12 with ratio 22 next in order.

REPEATED "BLIND RANK" TEST

Prior to the administration of the "Known Rank" test and following the "Blind Rank" test, the "Blind Rank" test was repeated with a random sample of ten families. While the

test was made prior to the "Known Rank" test, the discussion appears here because it is appropriate to question at this point whether such differences as appear between the "Blind" and "Known" tests are a result of differences in test methods or of variations to be expected in the repetition of tests.

The results from the ten families (24 individuals) for each of the three tests: the original "Blind Rank," the repeated "Blind Rank," and the "Known Rank," have been summarized. The median rank values for each test for each juice are presented in Table 5.

TABLE 5
Median Rank by Juices by Test Method
Based on Reports from 10 Families (24 Individuals)

Rank Method	Ratio					
	12	14	16	18	20	22
Original-Blind	4.4	2.8	4.7	2.3	3.2	4.7
Repeat-Blind	4.1	2.4	3.8	3.4	3.2	4.6
Known	5.6	3.8	3.2	3.1	3.1	4.5

Three facts seem evident: (1) The families as a group were able to repeat the test with fairly uniform results. (2) The difference in median values was greater between the two test methods than between the original and repeated "Blind Rank" test. (3) There was little change in median value for the juices of higher ratios and greater change for those in the lower ratios. A study especially designed to include a repeated ranking test for the entire population, rather than a small sample such as was used in this study, would yield more conclusive evidence.

CONCLUSIONS

Three different methods were employed to estimate consumer preferences for six canned orange juices of high Brix-acid ratios. The several interpretations of the results from these different methods makes clearly evident: (1) that each method and interpretation may yield a somewhat different estimate, and (2) that in spite of the obvious lack of complete agreement in the estimates, there is a satisfactory degree of conformity in the pattern of estimates that evolves.

The resulting pattern of preference estimates shows clearly that juices of ratios 12 and 22 are in relative disfavor; that within the range of ratios of 14 through 20 lies the juice of greatest favor; that the pin-pointing of the

most favored juice requires clarification and interpretation of the meaning of preference, for each of the four juices could be defended as the best.

The importance of method used to secure the data from the participants and the methods of analysis of such data (which permit different interpretations of the data) becomes increasingly evident. Space will not permit a complete review of the methods and analyses, but since this point is becoming increasingly evident to this investigator who is essentially concerned with evaluation of the appropriate design for such studies, a few observations will be made to give meaning to the position taken. Essentially two methods were employed, one which secured the reaction of the participant as he "normally" consumed juice, while the other created an artificial and forced contrast of one juice with others. It may be observed that the tart juice lost in favor when it was tasted alongside other juices, while the least tart juice gained in favor. And when the participant learned that the high ratio juice was from the riper fruit, it gained in number of first choices (which tended to counteract the bias inherent in extremes). The analysis and subsequent interpretation of the comparison-type data can be and was focused on the median rank values, on the modal frequency of first choices or of last choices, or on the frequency with which the juice was classed objectionable. The appropriateness of the alternatives depends in part on whether one wants to know the one juice that will have greatest popularity, or the juice(s) not extremely liked or disliked, but widely acceptable. The Single Juice Score method likewise lends itself to several analyses of the actual scores with focus on the mean or median score values; on the difference-in-score values between the juices, thereby reducing all participants to the same over-all level; or simply on the direction of preference between pairs of juices. Furthermore, attention could be focused simply on an analysis of the comments.

Further critical review of the data from this and previous studies is contemplated.

REFERENCES

- . Preferences of a Consumer Panel for Canned Orange Juice of Specified Brix-acid Ratios (8.2-14.8). Unpublished mimeographed report, F.S.U. Library, July, 1951. P. 12.
- Morse, Richard L. D. Selected Studies of Consumer Preferences for Canned Orange Juices. Florida State Hort. Soc., Proc. 65:230-234, 1952.