Garden To Café Taste Test Report

**An analysis of 'components of taste' and 'willingness to try new foods' results
from the May 14, 2019 taste test of a green salad with vegetables
using the Mid-reflective Taste Test Survey
conducted at a public elementary school in NYC**

Robert Abrams, Ph.D.
1/13/2022

Executive Summary

1. Eighty students taste tested a green salad with vegetables on seven components of taste (salty, sweet, bitter, sour, spicy, flavorful and temperature), and on their overall judgment of the salad (deliciousness).

2. The students quantified their willingness to try foods they hadn't eaten before (WTT new foods): today's dish, fruit, vegetables and the sum of the three variables.

3. Students self-reported their demographics (grade level and prior participation in taste tests and Garden To Café events).

4. While there was often consensus among the students about each taste component, the consensus was not total. This window into students' perceptions of food may help improve taste education, and through that, find additional ways to increase school meal participation and students' enjoyment of healthy foods.

5. Using four willingness to try new foods variables crossed with two prior participation in taste tests variables, with a small set of variable weightings, 48 Univariate ANOVAs were run.

6. As described in detail in this report, taken together the analyses provide evidence that:

    a. Garden To Café has been achieving its program goal of increasing students' willingness to try new foods.

    b. Taste testing as an instructional and programmatic activity is associated with, and may directly contribute to causing, increased willingness to try new foods, and that this is a hypothesis worth confirming in future studies.

7. A power analysis showed that this willingness to try new foods hypothesis could be confirmed with a sample of 345 to 700 students in a next study.

Table of Contents

Introduction

In May 2019, I had been working for the New York City Department of Education's Office of School Support Services (NYC DOE OSSS) for about five years[1], where my work was primarily focused upon school food research and program evaluation. In this capacity, I was the program evaluator for the Garden To Café program (GTC)[2].

This program was and is led by a chef (Chef George Edwards). GTC's chefs obtain fresh, seasonal food (raw ingredients) from local farms and the schools' own gardens and then prepare the available ingredients into one or more dishes. The students at participating schools try those dishes at special tasting events, often during lunch in the school cafeteria.

The chefs had a pretty good sense of how students were responding to the dishes. The GTC coordinator also wanted formal feedback on the dishes. Over a period of four to five years, he and I worked together to develop on-site taste testing methodology that would efficiently deliver quality feedback to the program. We also wanted to make sure the schools benefited from the time they gave us for the taste testing.

On May 14, 2019, I conducted a taste test for the Garden To Café program (GTC). This taste test was held at a public elementary school in NYC, with the warm and generous support of the school's principal and science teacher. This was a pilot test of the latest evolution of our taste test data collection instruments: the Mid-reflective Taste Test Survey.

This taste test turned out to be my last for GTC. This report delivers analysis of that taste test. For various reasons, including but not limited to the pandemic, the data from this taste test has not been analyzed until now.

---

[1] Around January 2019, the Office was reorganized. I ended up working within the Office of Food and Nutrition Services (OFNS), but the work of the organization remained the same.

[2] GTC is still operating, although the name has been changed to 'The New York City Farm To School Program'.

I felt I had an obligation to my colleagues at the Garden To Café program, to the staff and students at the school at which the taste test was held, to the field of school food research and to the taxpayers of New York City to complete the work. Part of this report was written as my final project for the Teachers College, Columbia University class *Introduction to R*, which felt like an ideal opportunity to bring the work to completion, while simultaneously expanding my analysis skills. The final part of this report was written after the class concluded.

As this report will show, the taste test data from May 14, 2019 provides evidence that the Garden To Café program has been increasing students' willingness to try new foods (foods they haven't eaten before) and supports the hypothesis that taste testing as an instructional and programmatic activity more broadly increases students' willingness to try new foods.

Motivating questions

1. How did students respond to the green salad being taste tested?
2. How did assessments of taste components relate to overall assessments of the dish?
3. How did student responses to the dish being taste tested vary by demographic characteristics (age and previous experience with taste tests)?
4. To what extent did Garden To Café achieve the program goal of increasing students' willingness to try new foods?

This report is primarily concerned with Motivating Questions #1, part of #3 and #4.

The dish that was taste tested

A Garden To Café chef prepared a salad with arugula, spinach and sliced carrots. The salad was lightly dressed with an apple-based dressing. Three photos of the salad are shown below: with flash, without flash and as served in sample cups plus that day's school food lunch (hamburger, fries, onion rings and a pear or apple). The last photo also shows the ingredient handouts that were available to the teacher and students who wished to take a set. (Photography by the author.)
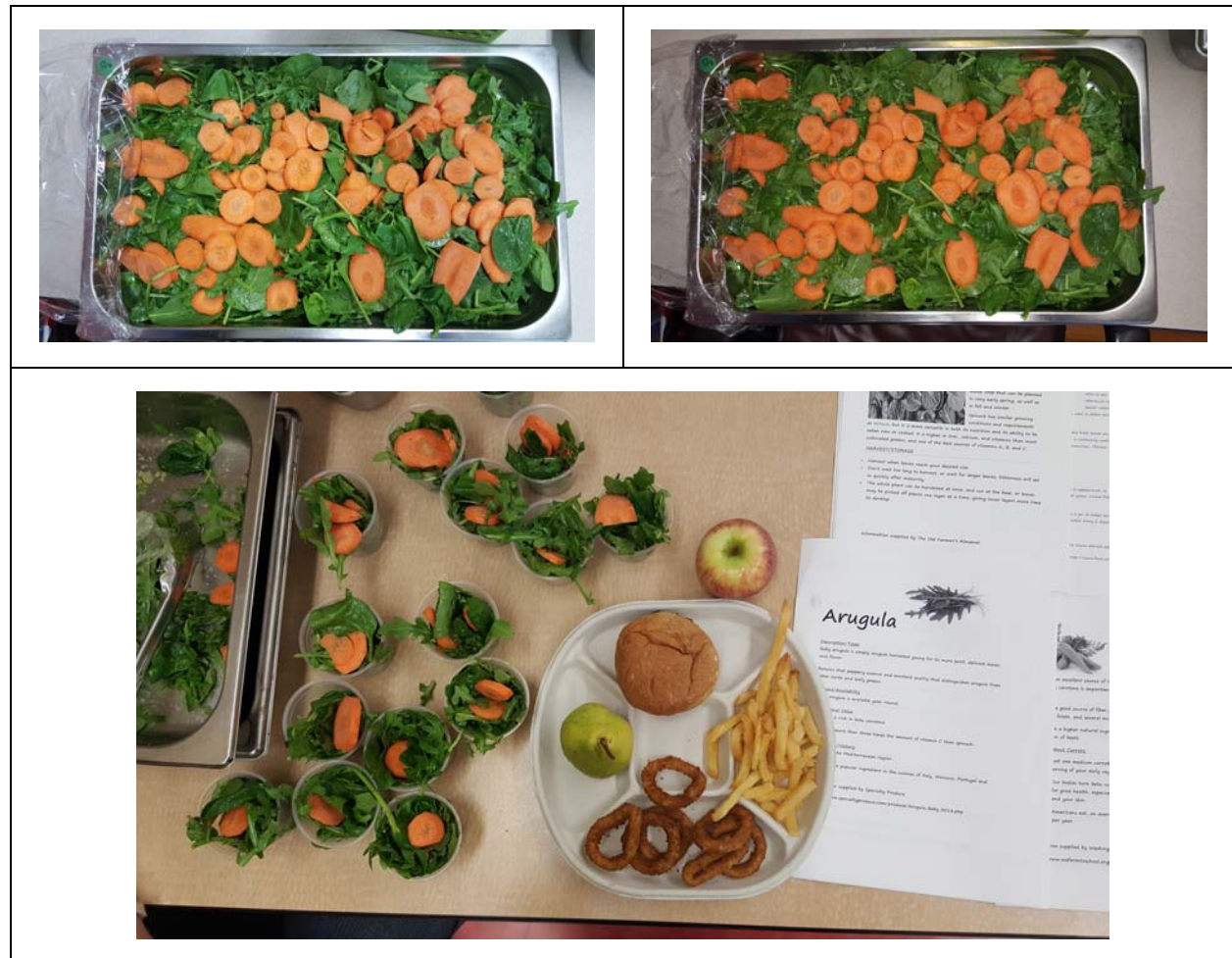
Results

Assessments of taste components

The survey asked students to assess seven components of taste prior to rendering an overall assessment of the dish. This gives the student time and multiple taste points to consider the dish before rendering judgment. This process of tasting is expected to result in more thoughtful and accurate taste judgments compared to typical one taste point taste tests. The components of taste on the survey are: salty, sweet, bitter, sour, spicy, flavorful and temperature.

Salty

In the case of Salty, there was a substantial consensus that the salad was low for Salty (45%). However, 30% of students didn't know how salty the salad was or left it blank. This suggests an opportunity for salt education. Given that it is known that perception of salt is relative and can shift if salt intake is adjusted slowly[3], the variability makes some sense.

Mode = 45.0%[4]
Other 2 levels = 25.0%
Mode/O2 = 1.80
M – O2 = +20.0%

```
## Tell us how today's dish tastes: Salty? :
##              Frequency Percent Cum. percent
## Low                36     45.0         45.0
## Medium             12     15.0         60.0
## High                8     10.0         70.0
## I don't know       21     26.2         96.2
## LeftBlank           3      3.8        100.0
##    Total           80    100.0        100.0
```


Tell us how today's dish tastes: Salty?
(percentages)

---
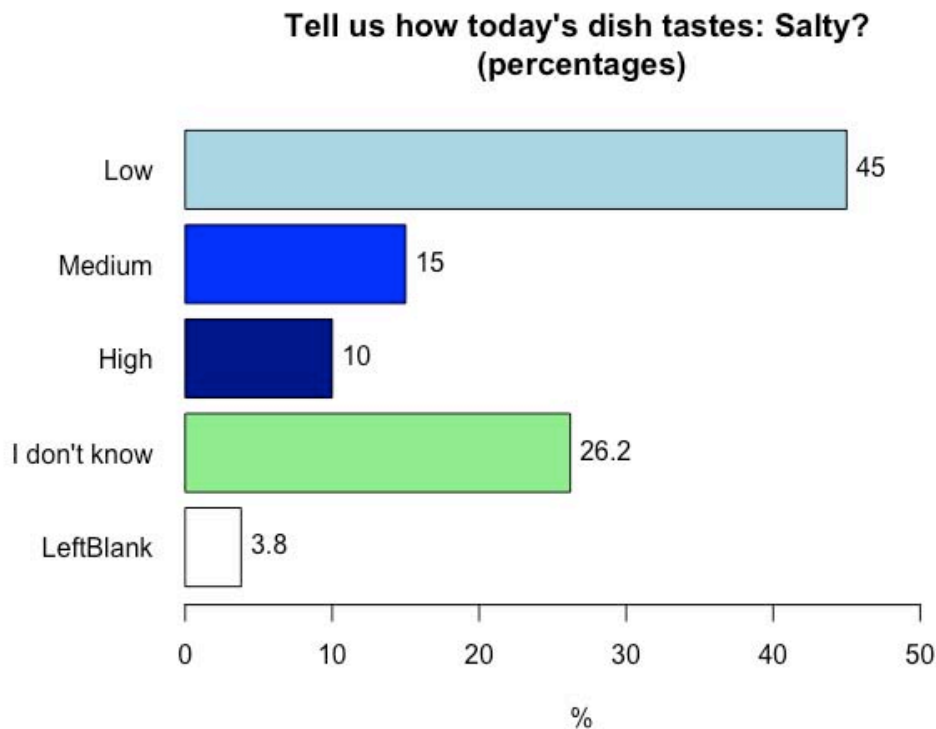
[3] *Find references for an academic paper version of this report.*

[4] I am experimenting with calculations to represent the variability in taste component responses. Note that "O2" is "Other 2 levels". There are three possible response options, so the analysis identifies the mode (M), and then calculates the percentage for the other two levels combined.

Sweet

For Sweet, the mode response was Low, which is consistent with the dish (a green salad with vegetables). However, using Mode/O2 as a measure of variability (high #s = high consensus, low variability and low #s = low consensus, high variability), there was quite a lot of variability in how students rated the sweetness of the salad. 23.7% of students didn't know how sweet the salad was (includes blanks).

Mode = 35%
O2 = 41.2%
M/O2 = 0.85
M – O2 = -6.2%

```
## Tell us how today's dish tastes: Sweet? :
##              Frequency Percent Cum. percent
## Low                28    35.0          35.0
## Medium             20    25.0          60.0
## High               13    16.2          76.2
## I don't know       13    16.2          92.5
## LeftBlank           6     7.5         100.0
##   Total            80   100.0         100.0
```

**Tell us how today's dish tastes: Sweet?
(percentages)**
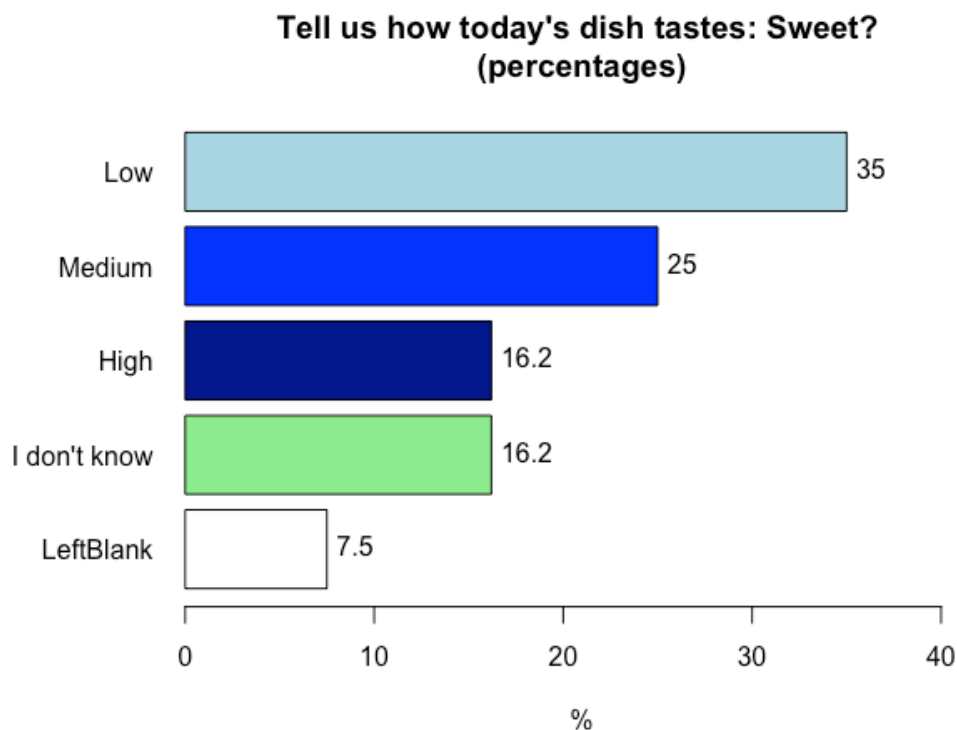
Bitter

For Bitter, the mode of High makes sense given the kind of salad. That said, the consensus was low, with a M/O2 of 0.59. 32.5% of students couldn't rate the bitterness of the salad (includes blanks).

Mode = 25.0%
O2 = 42.5%
M/O2 = 0.59
M – O2 = -17.5%

```
## Tell us how today's dish tastes: Bitter? :
##               Frequency Percent Cum. percent
## Low                  16    20.0         20.0
## Medium               18    22.5         42.5
## High                 20    25.0         67.5
## I don't know         20    25.0         92.5
## LeftBlank             6     7.5        100.0
##   Total              80   100.0        100.0
```


Tell us how today's dish tastes: Bitter?
(percentages)
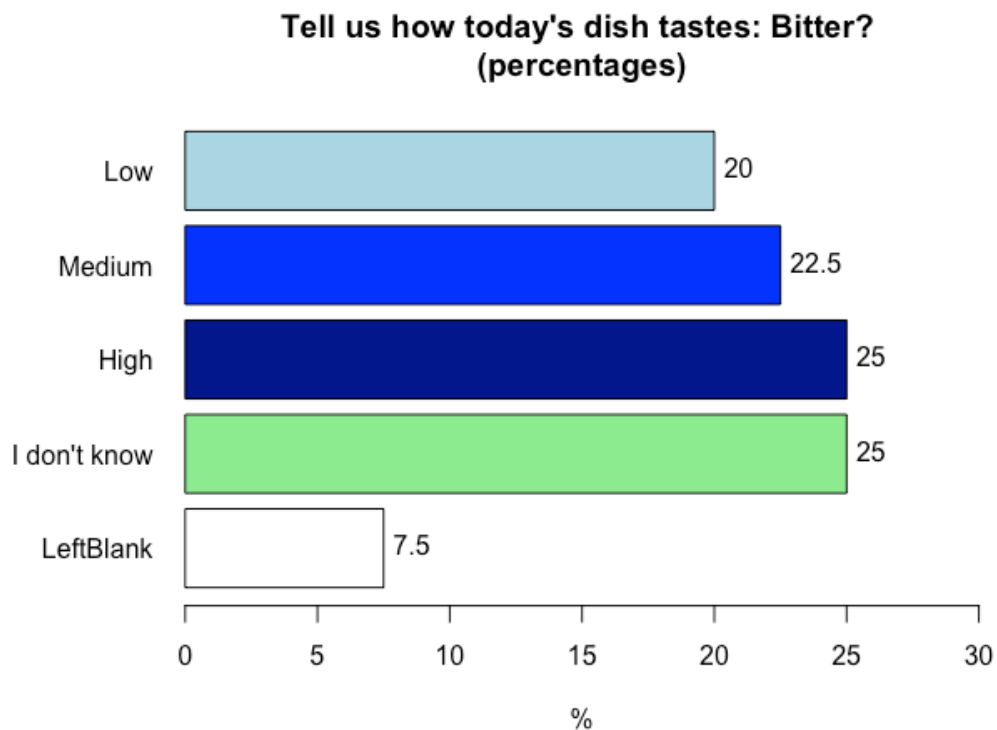
Sour

Unlike the previous taste component, which were unimodal, Sour was bimodal. M/O2 = 0.76. Don't know (plus blanks) was 27.4%.

    Mode = 31.2%
    O2 = 41.2%
    M/O2 = 0.75
    M – O2 = -10.0%

```
## Tell us how today's dish tastes: Sour? :
##               Frequency Percent Cum. percent
## Low                  25    31.2         31.2
## Medium               12    15.0         46.2
## High                 21    26.2         72.5
## I don't know         13    16.2         88.8
## LeftBlank             9    11.2        100.0
##   Total              80   100.0        100.0
```


Tell us how today's dish tastes: Sour?
(percentages)
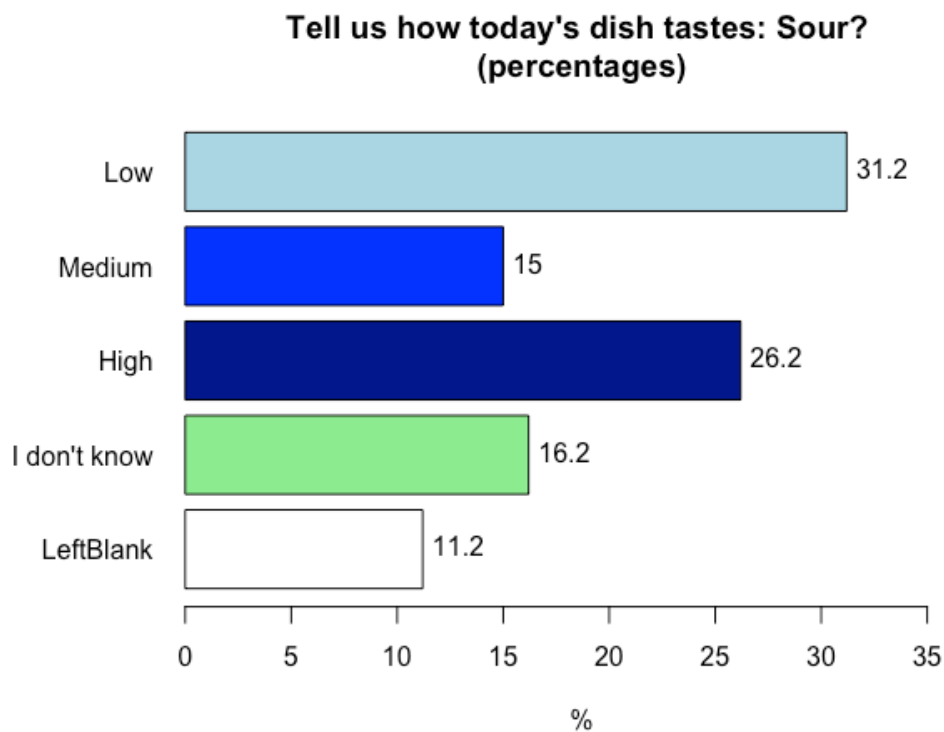
Spicy

Spicy had high consensus. Low spicy (40% of students) makes sense for this salad. Still, 15% of students though it was High spicy. 37.6% of students didn't know the spiciness of the dish (includes blanks).

Mode = 40%
O2 = 22.5%
M/O2 = 1.78
M − O2 = +17.5%

```
## Tell us how today's dish tastes: Spicy? :
##             Frequency Percent Cum. percent
## Low                32    40.0         40.0
## Medium              6     7.5         47.5
## High               12    15.0         62.5
## I don't know       19    23.8         86.2
## LeftBlank          11    13.8        100.0
##   Total            80   100.0        100.0
```

Tell us how today's dish tastes: Spicy? (percentages)
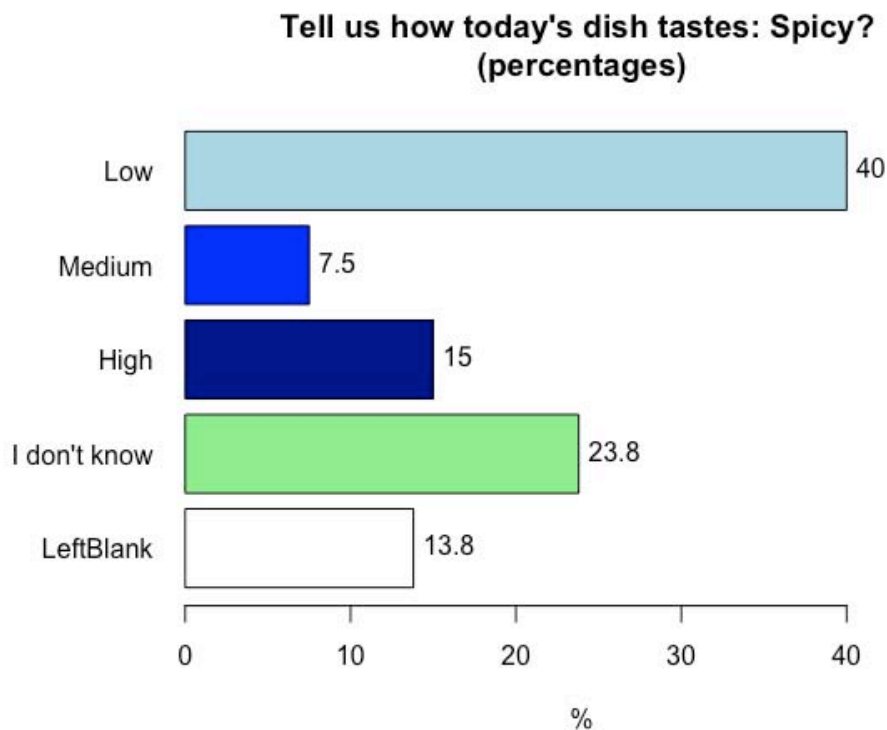
Flavorful

The mode for Flavorful was Low (22.5%). M/O2 = 0.62. Don't know (including blanks) = 41.2%.

Mode = 22.5%
O2 = 36.3%
M/O2 = 0.62
M – O2 = -13.8%

```
## Tell us how today's dish tastes: Flavorful? :
##              Frequency Percent Cum. percent
## Low                18    22.5         22.5
## Medium             15    18.8         41.2
## High               14    17.5         58.8
## I don't know       23    28.7         87.5
## LeftBlank          10    12.5        100.0
##   Total            80   100.0        100.0
```


Tell us how today's dish tastes: Flavorful?
(percentages)
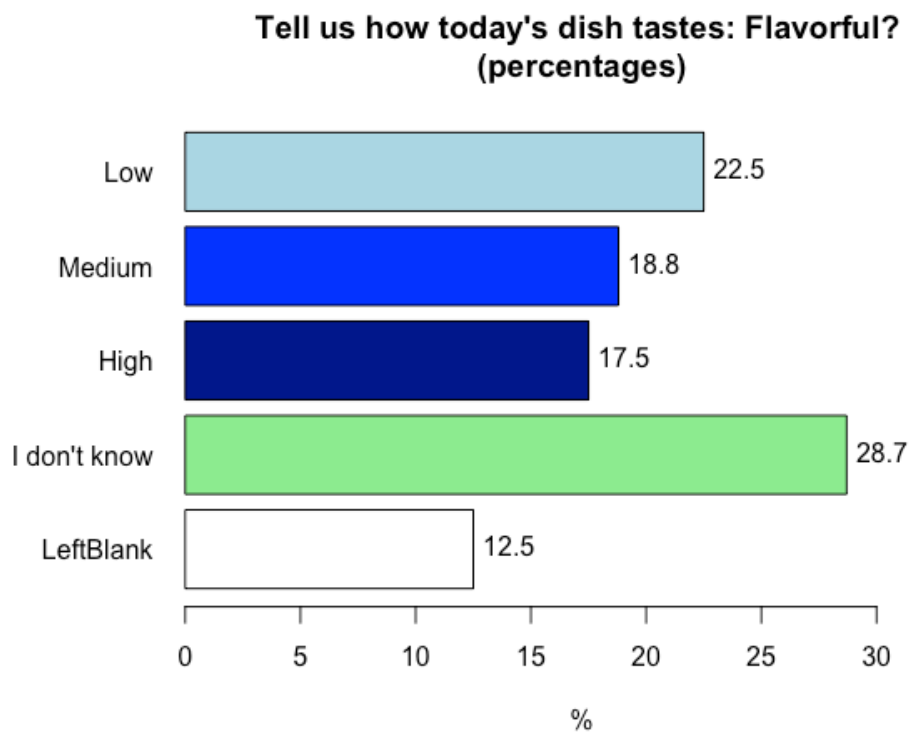
Temperature

The mode for Temperature was as expected: Cold (50%), although Warm could also be a fair description for room temperature, which is generally, but not always, the expectation for a green salad. Only 17.5% of students didn't know the temperature of the dish.

```
## What is the Temperature of today's dish? :
##              Frequency Percent Cum. percent
## Frozen               6     7.5          7.5
## Cold                40    50.0         57.5
## Warm                17    21.2         78.8
## Hot                  2     2.5         81.2
## Very Hot             1     1.2         82.5
## I don't know         6     7.5         90.0
## LeftBlank            8    10.0        100.0
##   Total             80   100.0        100.0
```


What is the temperature of today's dish?
(percentages)
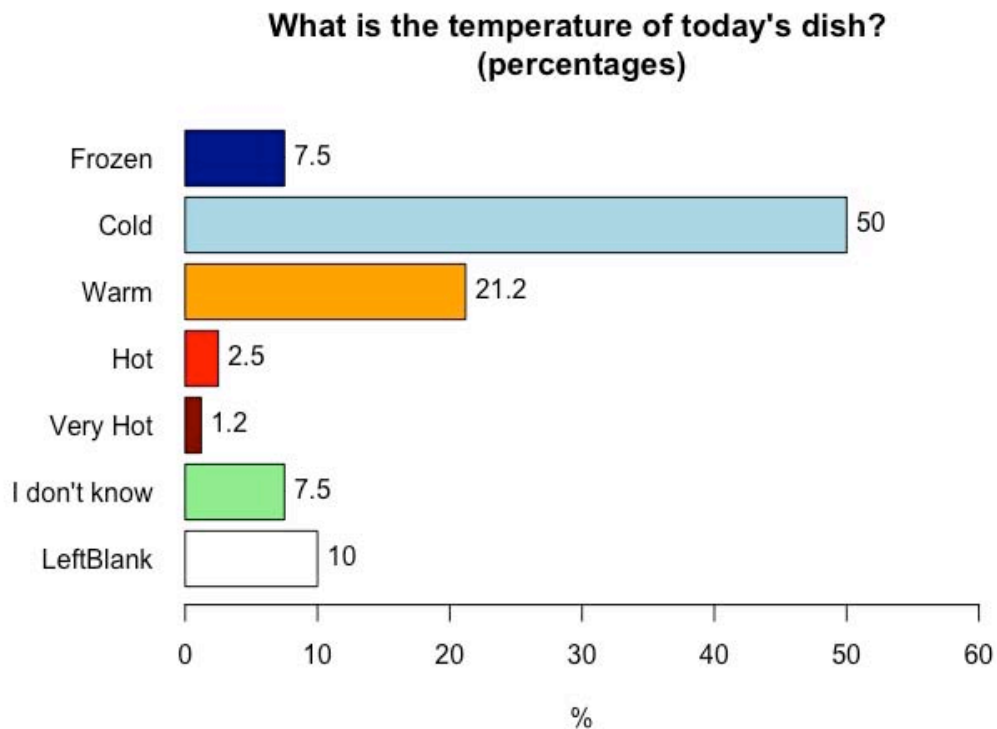
Overall assessment of the green salad

The overall assessment of the salad was fairly evenly distributed among Delicious (30%), Okay (27.5%) and Unsatisfying (32.5%). Don't know plus Blanks was 6.2%. Didn't try it was 3.8%.

I need to go back and check how students who answered Didn't try it to this question answered the taste component and other questions. There were only three such students, so it won't impact the results much either way.

```
## Overall, I think today's dish tastes… :
##                          Frequency Percent Cum. percent
## Delicious (Smile Face)          24    30.0         30.0
## Okay (Flat Line Face)           22    27.5         57.5
## Unsatisfying (Frown Face)       26    32.5         90.0
## I don't know (\\)                1     1.2         91.2
## I didn't try it (\\)             3     3.8         95.0
## LeftBlank                        4     5.0        100.0
##   Total                         80   100.0        100.0
```


Overall, I think today's dish tastes...
(percentages)
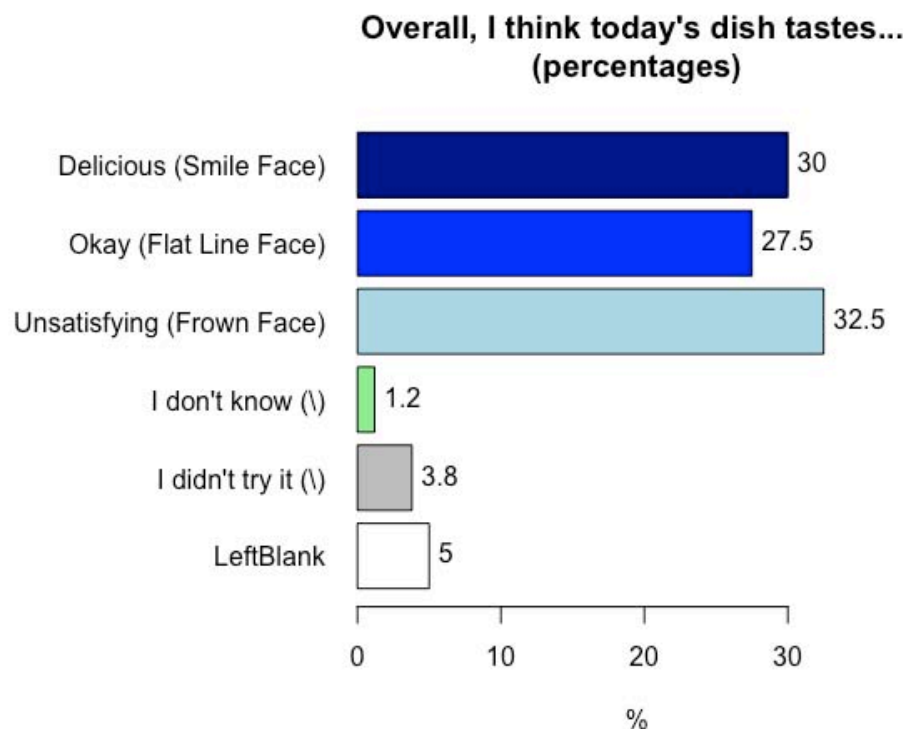
Willingness to try new foods

In terms of willingness to try new foods (WTT new foods), in this case different kinds of today's dish, 36.2% of students were strongly willing to try (3+ times), 27.5% (1-2 times) were willing to try, and 20% were not willing to try. 16.3% didn't know (includes blanks).

```
## I would like to try different kinds of today's dish … :
##              Frequency Percent Cum. percent
## 3+ times          29    36.2        36.2
## 2 times            4     5.0        41.2
## 1 time            18    22.5        63.8
## Never             16    20.0        83.8
## I don't know       6     7.5        91.2
## LeftBlank          7     8.8       100.0
##   Total           80   100.0       100.0
```


I would like to try different kinds of *today's dish*...
(percentages)
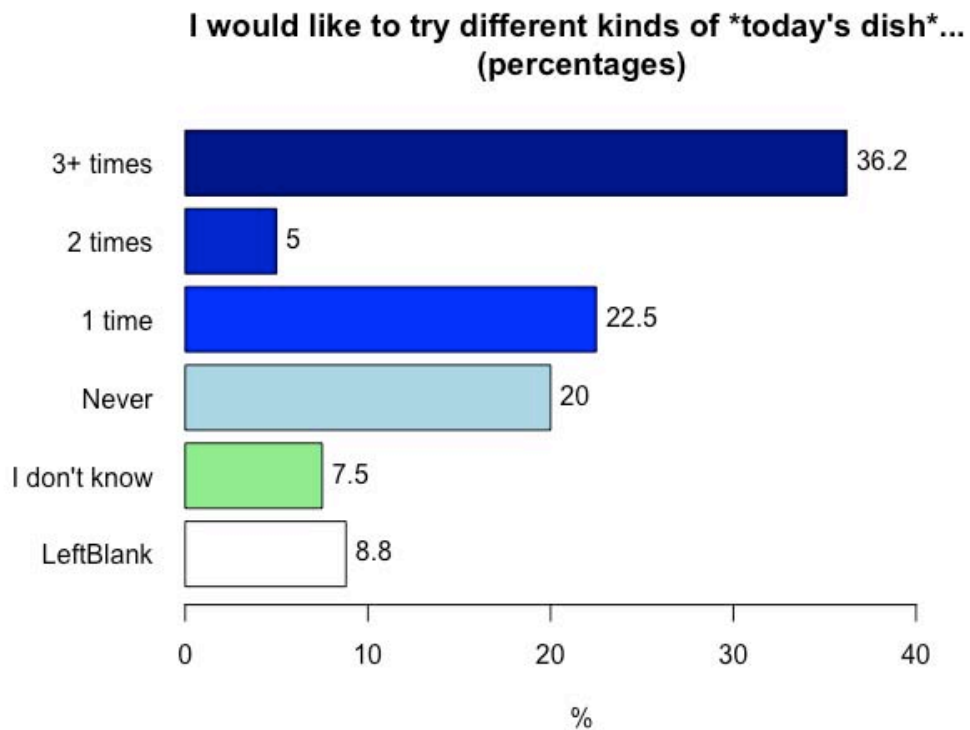
Willingness to try new fruits (that is, fruits the student hadn't eaten before), which is expected to be high, or at least higher than for vegetables, was 55% for 3+ times, 17.5% for 1-2 times, and 7.5% for never. 20% didn't know (including blanks).

```
## I would like to try fruit I haven't eaten before… :
##               Frequency Percent Cum. percent
## 3+ times            44    55.0         55.0
## 2 times             10    12.5         67.5
## 1 time               4     5.0         72.5
## Never                6     7.5         80.0
## I don't know         4     5.0         85.0
## LeftBlank           12    15.0        100.0
##    Total            80   100.0        100.0
```


I would like to try *fruit* that I haven't eaten before... (percentages)
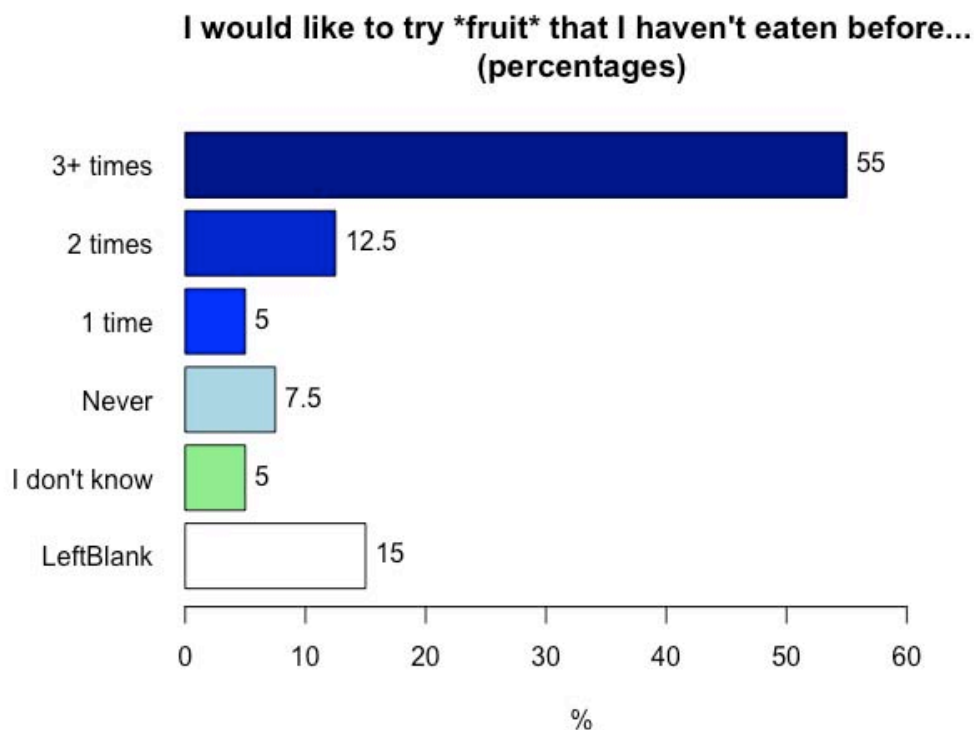
Willingness to try new vegetables (that the student hadn't eaten before), which is expected to be low, or lower than fruits, was 26.2% for 3+ times, 36.2% for 1-2 times, and 15% for never. I don't know (including blanks) was 22.4%.

```
## I would like to try vegetables I haven't eaten before… :
##              Frequency Percent Cum. percent
## 3+ times          21    26.2          26.2
## 2 times           13    16.2          42.5
## 1 time            16    20.0          62.5
## Never             12    15.0          77.5
## I don't know       5     6.2          83.8
## LeftBlank         13    16.2         100.0
##   Total           80   100.0         100.0
```


**I would like to try *vegetables* that I haven't eaten before...
(percentages)**
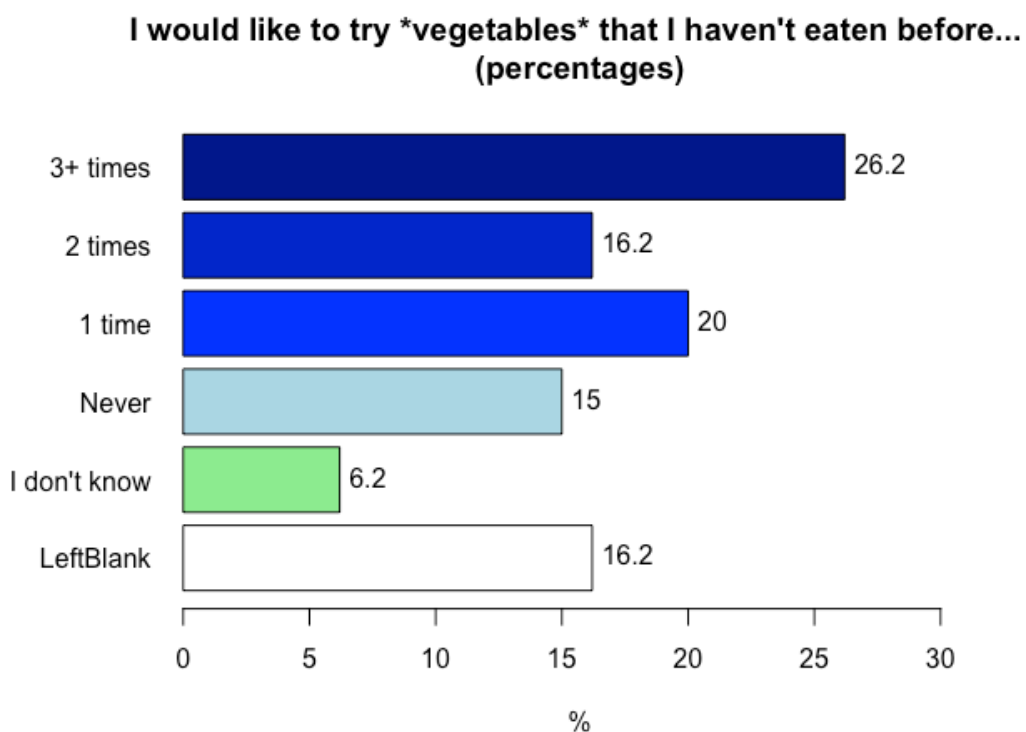
Willingness to try new fruit 3+ times was 55.0% and 7.5% for Never, versus willingness to try new vegetables 3+ times was 26.2% and 15.0% for Never, so WTT new fruit was greater than WTT new vegetables, as predicted.

A concept behind these willingness to try questions was to measure the Garden To Café's effectiveness: if willingness to try new foods increased over time, this would be one indicator of GTC's meeting its goals, keeping in mind that most schools participating in GTC have only a small number of GTC events, so expectations for change should be set appropriately – unless the school staff reinforced GTC messages such as "You don't have to like the new food, you just have to try it" in between GTC events.

Demographics

Grade levels

Students were largely in grades 1 and 2, with smaller, but still large numbers in grades 3 and 5. These were all self-reports on the survey. One student self-reported being in Kindergarten. Only 2.5% left the question blank. The participating students were classes selected by the school staff based on availability and scheduling. Thus, while the selection of classes (groups of students) was not entirely random on the part of the school staff, I as the GTC program evaluator had no hand in the selection. Since entire classes were selected, within classes we had the whole population of students attending school in each such class on that day.

```
## What grade are you in?
(If this is the Summer, what grade did you just complete?) :
##           Frequency Percent Cum. percent
## 0 (K)            1     1.2          1.2
## 1              27    33.8         35.0
## 2              22    27.5         62.5
## 3              13    16.2         78.8
## 4               0     0.0         78.8
## 5              15    18.8         97.5
## LeftBlank       2     2.5        100.0
##    Total       80   100.0        100.0
```

What grade are you in?
(If this is the Summer, what grade did you just complete?)
(percentages)
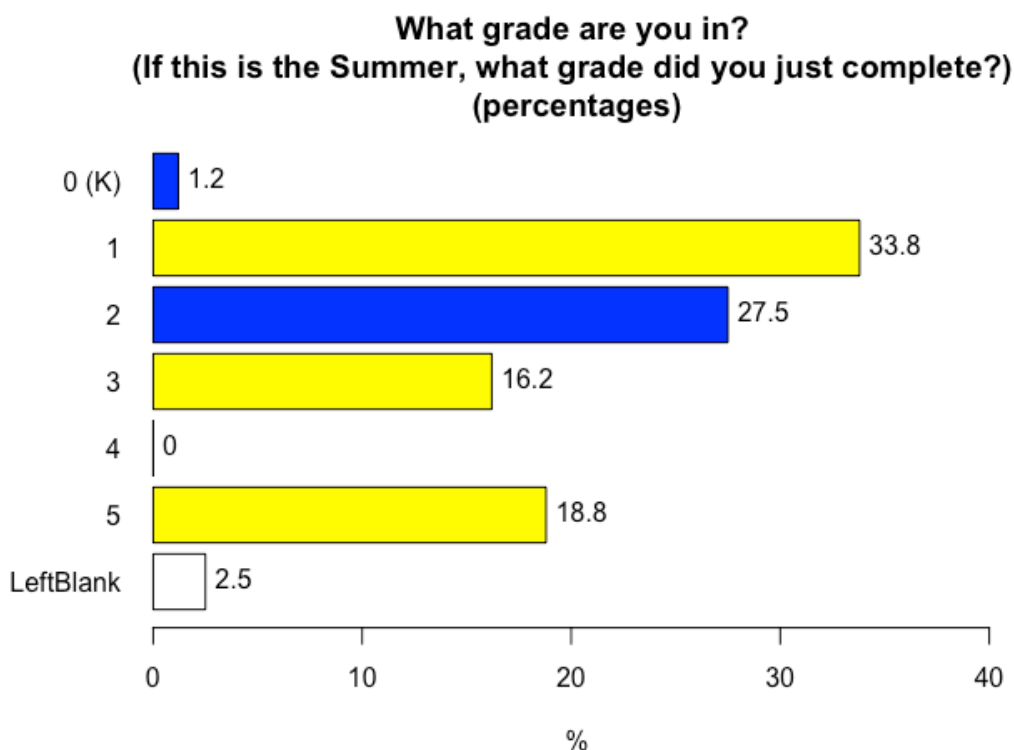
Prior taste test and Garden To Café event participation

Half of students (50%) reported participating in a taste test before this event. The other half said No or were uncertain (Maybe or blanks).

25% of students reported having participated in a GTC event before. The other 75% either said No (50%) or were uncertain (Maybe or blank).

```
## Have you taken part in a taste test before? :
##           Frequency Percent Cum. percent
## Yes              40    50.0         50.0
## Maybe             8    10.0         60.0
## No               30    37.5         97.5
## LeftBlank         2     2.5        100.0
##   Total          80   100.0        100.0
```

Have you taken part in a taste test before?
(percentages)
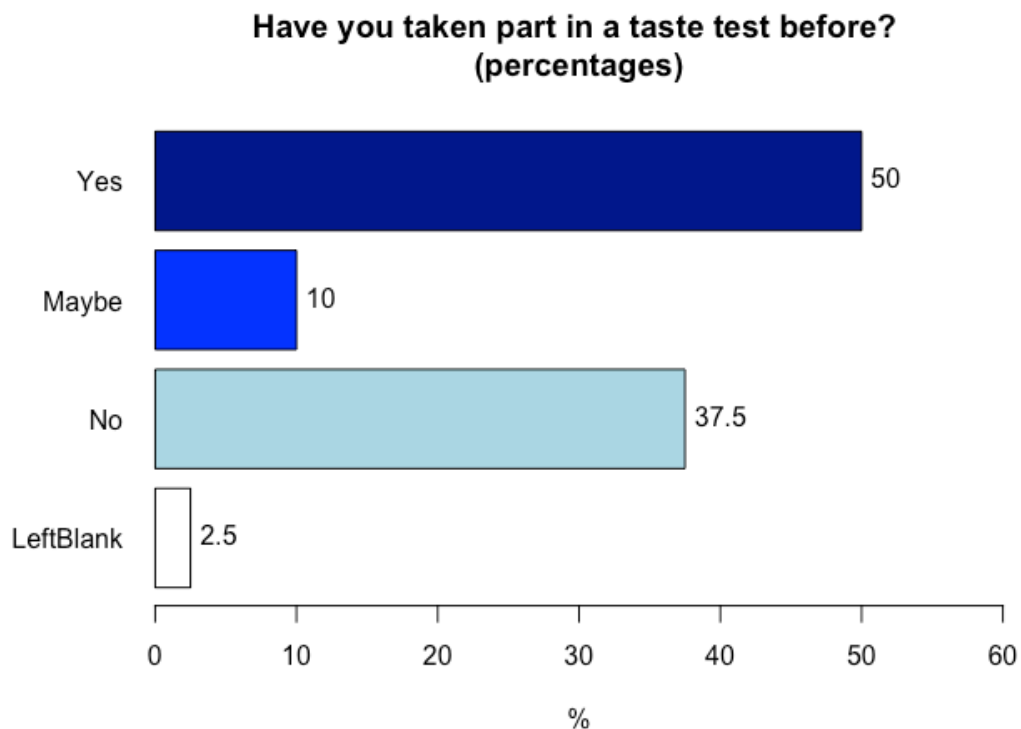
```
## Have you taken part in a Garden To Café event before? :
##           Frequency Percent Cum. percent
## Yes              20    25.0         25.0
## Maybe            15    18.8         43.8
## No               40    50.0         93.8
## LeftBlank         5     6.2        100.0
##    Total         80   100.0        100.0
```


Have you taken part in a Garden To Café event before?
(percentages)
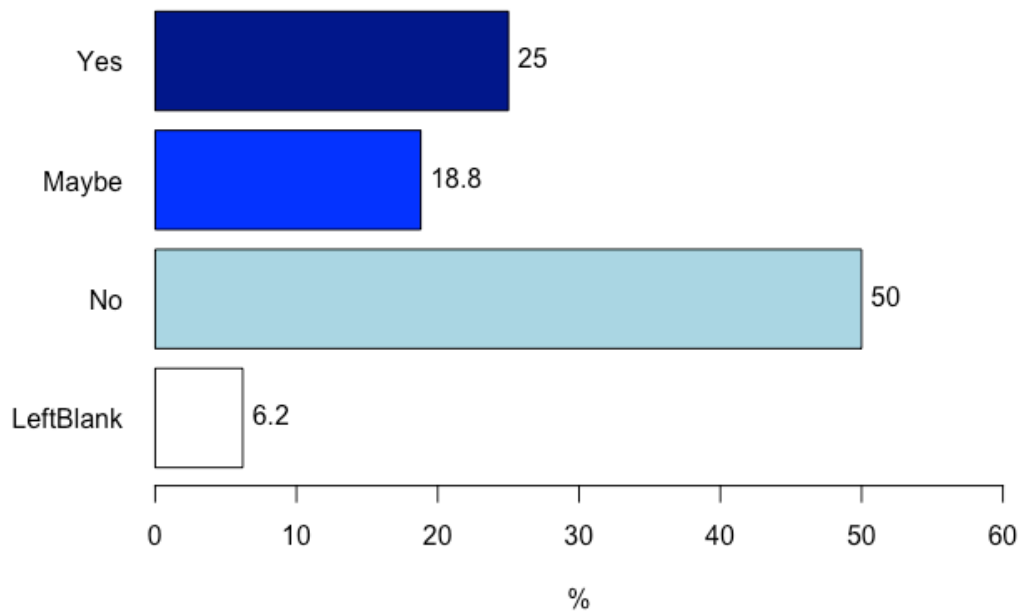
Class period

Responses by class period ranged from a low of 13 students in period 1 to a high of 21 students in period 2.

```
## Class Period :
##          Frequency Percent Cum. percent
## 1              13    16.2         16.2
## 2              21    26.2         42.5
## 3              17    21.2         63.8
## 6              15    18.8         82.5
## 7              14    17.5        100.0
##    Total       80   100.0        100.0
```

Class Period
(percentages)
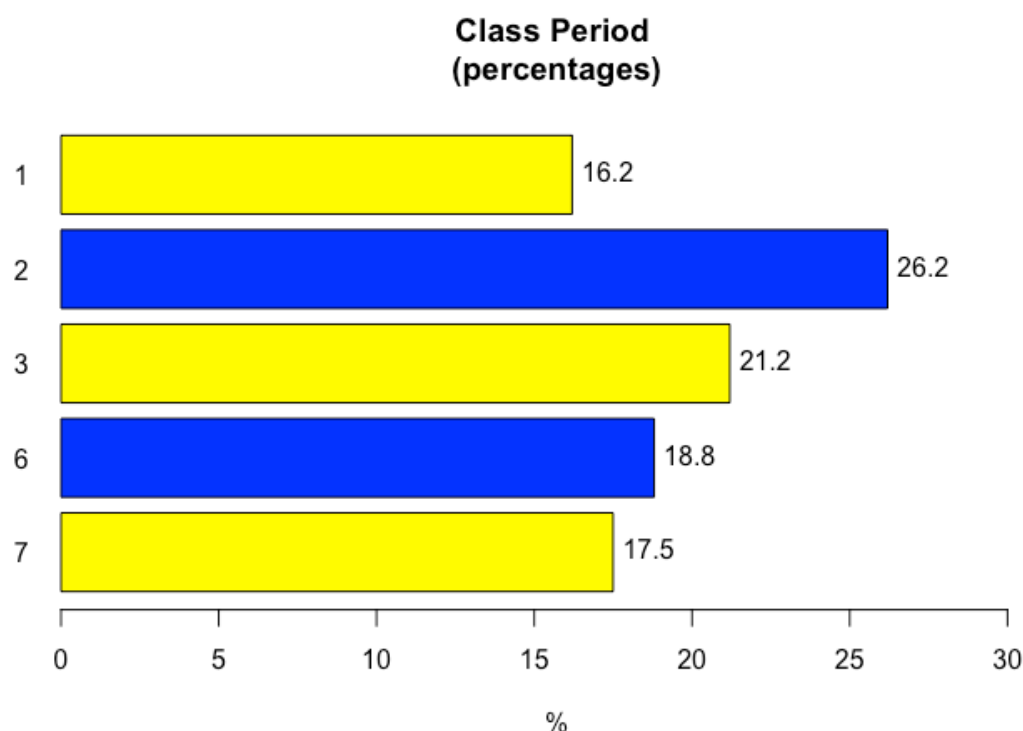
Other demographics

We intentionally did not include other demographic questions, such as gender, race or ethnicity on the survey. This was partly because these were not a priority for the Garden To Café program, and also because there was no way to make any additional questions fit onto one letter size page. In the case of gender, the response options were and probably still are in flux in the research community at large, and there was no official guidance from the NYC DOE on what gender response options ought to be used on surveys given to NYC DOE students. Additionally, we knew that a one-day data collection would not produce a large enough sample to make a large number of sub-group comparisons statistically meaningful.

The survey can be modified for future studies. For instance, prior cooking experience or languages spoken at home could make for interesting demographic questions.

Analysis of missing data

Response rate

Eighty out of 82 students present completed a survey, for an overall response rate of 97.6%. One of the students who did not complete a survey could not participate in the taste test due to an allergy to an ingredient in the dressing (apples).

Missing values (I don't know and blank responses)

To fully understand a dataset, it is important to understand the distribution of missing values. The following bar charts and tables provide insight into missing values in this dataset.

The surveys in this dataset had some missing values in response to some questions, but overall missing values were tolerable.

Number of "I don't know" or Blank responses per student: The distribution of I don't knows and blanks for each student, across all questions (14 sub-questions total), can be found in the bar chart titled "Count of IDK-s and Blanks for each student". The maximum number of IDKs and blanks for any one student was 11. Eight out of 80 students (10%) left half or more of the questions blank or answered I don't know. While questions left blank are missing data, they also potentially provide the program with an indication of where taste education might be targeted.

```
## Count of LeftBlanks for each student :
##          Frequency Percent Cum. percent
## 0              44    55.0         55.0
## 1              15    18.8         73.8
## 2               6     7.5         81.2
## 3               5     6.2         87.5
## 4               3     3.8         91.2
## 5               2     2.5         93.8
## 6               2     2.5         96.2
## 7               2     2.5         98.8
## 8               1     1.2        100.0
##    Total       80   100.0        100.0
```

Count of LeftBlanks for each student (percentages)
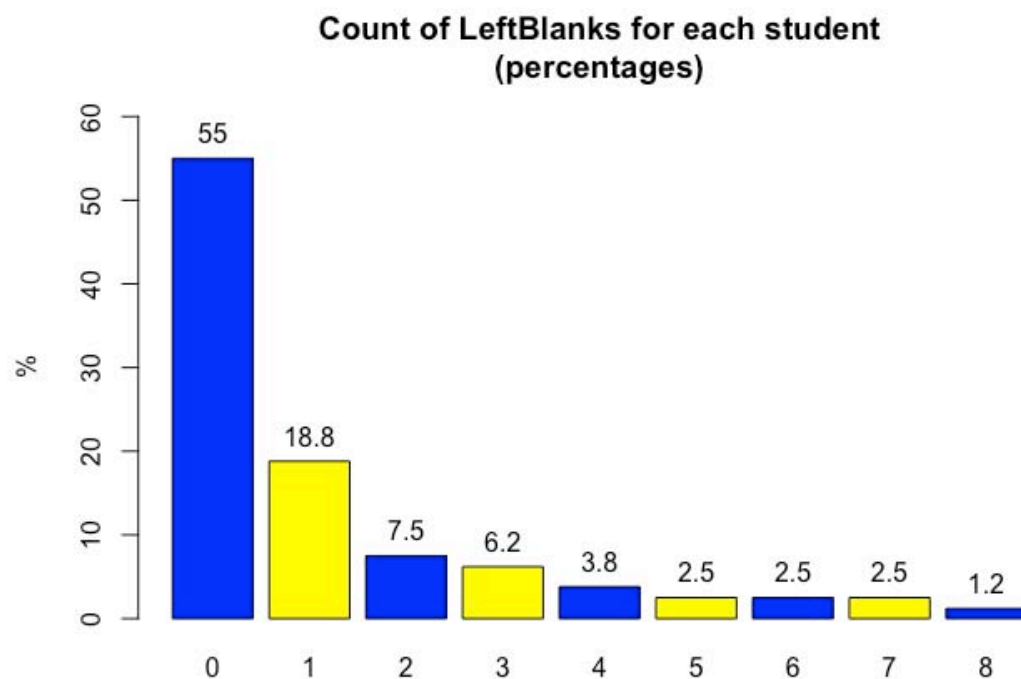
```
## Count of I don't knows for each student :
##          Frequency Percent Cum. percent
## 0               24    30.0         30.0
## 1               24    30.0         60.0
## 2               14    17.5         77.5
## 3               11    13.8         91.2
## 4                3     3.8         95.0
## 6                1     1.2         96.2
## 7                1     1.2         97.5
## 10               1     1.2         98.8
## 11               1     1.2        100.0
##   Total         80   100.0        100.0
```

**Count of I Don't Know-s for each student
(percentages)**
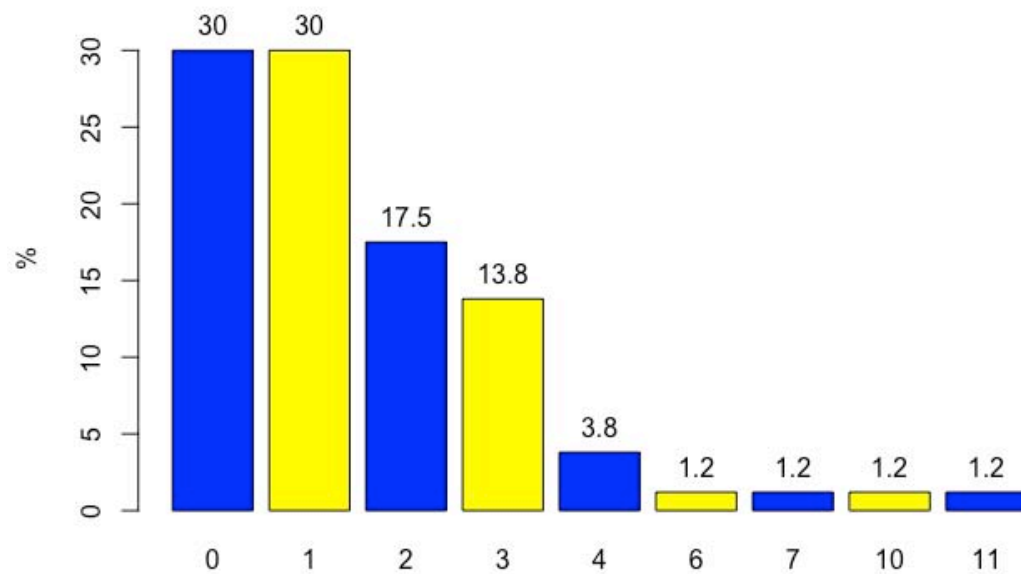
```
## Count of I don't knows and Blanks for each student :
##          Frequency Percent Cum. percent
## 0               9    11.2        11.2
## 1              19    23.8        35.0
## 2              15    18.8        53.8
## 3              13    16.2        70.0
## 4               9    11.2        81.2
## 5               3     3.8        85.0
## 6               4     5.0        90.0
## 7               4     5.0        95.0
## 8               1     1.2        96.2
## 9               1     1.2        97.5
## 10              1     1.2        98.8
## 11              1     1.2       100.0
##    Total       80   100.0       100.0
```

Count of IDK-s and Blanks for each student (percentages)
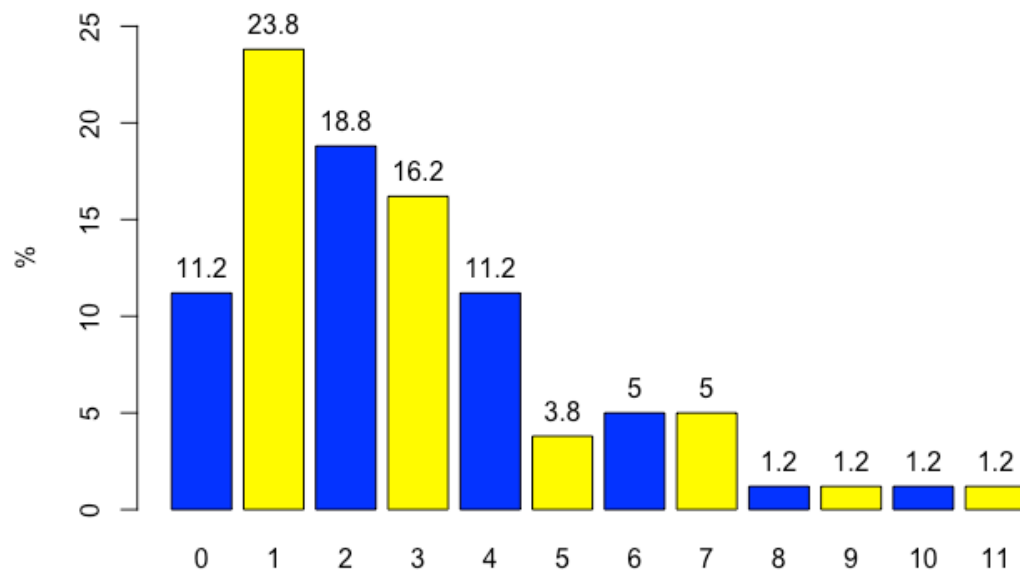
Measuring association of taste testing on change in willingness to try new foods

The Mid-Reflective Taste Test Survey contains three questions related to willingness to try new foods (WTT). It contains two questions related to prior participation in taste tests.

Analysis of the results of the pilot test of the Mid-Reflective Taste Test Survey suggests that these questions can be used demonstrate that taste tests in general and Garden To Café events in particular can create increased willingness to try new foods.

In my analysis, there were eight base comparisons. With variations in the ways the variables can be weighted, there were 48 comparisons. Keep in mind that the 48 comparisons were not a search for significance. They are different ways of presenting the data, both to report current results and to decide on adjustments to the survey for the next study. In any case, with a total sample of only 80 students, it would have been unlikely to find statistical significance at any p value suitable for broad generalization. For hypothesis generation, though, the sample was of sufficient size. (And 80 students was a very successful one day data collection, so the chance of obtaining a larger sample from one school on one day is small.)

Students were asked:

"I would like to try different kinds of today's dish…"

"I would like to try fruit that I haven't eaten before…"

"I would like to try vegetables that I haven't eaten before…"

The response options were:

3+ times, 2 times, 1 time, Never, and I don't know (plus Left Blank).

There are two basic ways to present results from these questions.

The first option is to ask, How many students wanted to try new foods? To find this number from the data, I combined 3+ times, 2 times and 1 time into one category "Willing to try new food" and grouped Never, I don't know and Left Blank into "Not willing to try new food". It is true that Left Blank (and/or I don't know) would often be coded as Missing, but in this case the student could have responded in the affirmative, but didn't, and in an instructional context, the teacher or chef could and should follow up with those students to find out more about the why of their responses. Or to put it another way, the data shows which students said they wanted to try new foods, and which students didn't say they wanted to try new foods. The latter is not exactly the same thing as students saying they don't want to try new foods (except for those in the group who responded Never), but it is close. In this case, I recoded the 1 to 3+ response options as "100" and the others as "0". This produces results that are the same as those produced by a cross-tab showing column percentages. This is a convenient and succinct presentation of the data, but sometimes one

might want to look at the results in more detail, so for that I also ran and report the corresponding cross-tabs. As shorthand, this variable weighting is indicated as "v100".

The second option is to treat the response options as a numeric scale. In this option, I treat the responses as an accurate representation of the intensity of a student's willingness to try new foods. Willingness to try new fruit one time is more than willingness to try new fruit never. Willingness to try new fruit two times is more than willingness to try new fruit one time. Three times is more than two times, and so on. "3+ times" was coded as 3, "2 times" as 2, 1 time as 1, "Never" as 0 and "I don't know" and "Left Blank" as Missing. This variable weighting is indicated as "vNum". In the case of this survey, if, for example, a student had never had a pear before, two times could be two trials of the same kind of pear, or one trial each of two different kinds of pears. I would regard those as equal levels of WTT new foods.

In a variation of the second option, I also ran analyses where I recoded "3 times" to "5". The survey specified the "3" response option as "3+ times" to account for more than three. Five was tried as a weighting because it is larger than three, and not so much larger that it would be implausible. This weighting is indicated as "vPl5" for "Version Plus is 5."

I also created a 4th way of measuring WTT by summing the three WTT variables within each of the three weightings. The three variables were evenly weighted in this computation. This results in a scale of 0 to 300 for the sum of v100 variables, a scale of 0 to 9 for the sum of vNum variables and a scale of 0 to 15 for the sum of vPl5 variables. WTT for vegetables (or its converse, vegetable neophobia) is generally regarded as lower than that for fruit, so I could have weighted WTT for vegetables more than the others, but in this case I didn't. I might run such a weighting later, to help decide on primary intended analysis for the next study.

For the second option of WTT questions, the variation of the second option and the computed WTT option of those two, I recoded I don't know and Left Blank as Missing.

Students were asked two prior participation questions:

> "Have you taken part in a taste test before?"

> "Have you taken part in a Garden To Café event before?"

The response options were:

> Yes, Maybe, No and Left Blank

Originally, I thought of this response option set as a scale from complete certainty in the positive (Yes) to some uncertainty (Maybe) to complete certainty in the negative (No). In this sense, I coded the responses as Yes = 1, Maybe = 0.5 and No = 0. Left Blank was coded as missing, which was only the case for two students. This weighting is indicated as "vFull" for the full range of response options.

That said, Maybe could also be uncertainty equivalent to "I don't know". Maybe could also be described as an un-collapsed probability function, where the student isn't quite sure if his/her past experience matches what the question asks, and so flickers back and forth between answering Yes and answering No. If this is the case, it might be best addressed by a branch of research called Chaos Theory. To address this potential problem, I recoded a second version of the two prior participation questions where Yes = 1, No = 0 and both Maybe and Left Blank were recoded as Missing. This weighting is indicated by "vEnds" in that I am using the ends of the response option scale. This also causes some students to drop out of the analyses, so analyses using this version of the prior participation questions will have a smaller sample, and will thus be that much less likely to find significance. This is normal in research, and is a reason why hypothesis generation can afford a relatively small sample (because significance isn't the first concern), but an attempt to find generalizable results needs a sample large enough that some students can drop out due to missing data, and still have a large enough sample left to have enough statistical power so that a significant result can be found if it exists.

Initial analysis: Univariate ANOVA of WTT fruit (v100) vs prior taste test participation (vFull)

Between the four WTT questions (the three questions plus the computed scale) times the two prior participation questions times the three WTT weightings times the two prior participation weightings, there were 48 Univariate ANOVAs run. Below is one set of results.

Descriptive Statistics for Dependent Variable: I would like to try <u>fruit</u> that I haven't eaten before... (v100)

Have you taken part in a taste test before? (vFull)	Mean	Std. Deviation	N
No	60.0000	49.82729	30
Maybe	75.0000	46.29100	8
Yes	80.0000	40.50957	40
Total	71.7949	45.29108	78

In the table above, we see that among the 30 students who said they had not taken part in a taste test before, 60% of those students said they would like to try fruit they hadn't eaten before at least once. Among the 8 students who said maybe they had taken part in a taste test before, 75% said they would like to try fruit they hadn't eaten before at least once. Among the 40 students who said they had taken part in a taste test before, 80% said they would like to try fruit they hadn't eaten before at least once. These results clearly show an association between prior taste test participation and higher likelihood of being willing to try new fruits, which strongly suggests that participation in taste tests increases students' willingness to try new foods, increased WTT new foods being a major goal of the Garden To Café program. Huzzah!

Now the caveats.

The result is not statistically significant. P = 0.185, which means that there is an 18.5% chance that the results which look different really are not different. The usual rule of thumb in research is that P should be less than 0.05 to be regarded as significant, so this result is too high. That said, the P value tells us the risk of being wrong, and what level of risk is acceptable depends on the circumstances. Here, I might say that 18.5% is an acceptable risk if I am only considering these specific students in the sample, or only generalizing to the roughly 200 students in the school, but 18.5% would be an excessive risk if I wanted to generalize to a much broader population. In fact, if I wanted to generalize to a much broader population, I would want to see a P value <= 0.01, to give me only a 1% chance of being wrong. For hypothesis generation, P = 0.185 looks pretty good, because with a larger sample, the odds are good that the results would be solidly significant.

Significance is not the only statistical measure to examine. Effect size, which measures the extent to which the result is meaningful, and is less influenced by sample size, is also very important. Here, the effect size, partial eta-squared, was 0.044, which is larger than the oft-cited small effective size rule of thumb of 0.01 for partial eta-squared, but smaller than the medium effect size rule of thumb of 0.06[5]. For the effect size to be considered meaningful, partial eta-squared should be 0.14 or greater. On the other hand, effect size has a practical dimension, so, in this case, if a 20% difference between the prior participation No group and the Yes group is a meaningful result to the program, then the result would be meaningful.

Another caveat is that the Maybe group only has eight students. Any sub-group with less than 10 should be treated with caution, and I would prefer sub-groups to all have at least 20 for analysis to withstand the weight of practical use.

Finally, not all of the 48 analyses produced such clear results. I will discuss the implications of the full set of analyses in the next section.

---

[5] Such as, for example, Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Second analysis: WTT fruit (vNum) vs prior taste test participation (vFull)

If I treat this same analysis of WTT for new fruit as a numeric scale from 0 to 3, I get the following results. Now, the same general trend is seen, in that students with prior taste test participation wanted to try new fruit a mean 2.54 times, while students with no prior taste test experience wanted to try fruit a mean 2.14 times. Huzzah!

Descriptive Statistics for Dependent Variable: I would like to try <u>fruit</u> that I haven't eaten before... (vNum)

Mean	Std. Deviation	N
2.1429	1.15264	21
2.8333	.40825	6
2.5429	.91853	35
2.4355	.98548	62

Except, the Maybe group now wanted to try new fruits 2.83 times, which is more than the Yes group, which makes the pattern less clear. The Maybe group now only has six students. P is a little higher at 0.199, which is worse than the previous analysis. Partial eta-squared is a little better at 0.053, but that is still below the medium effect size rule of thumb for partial eta-squared of 0.06. The total N for the analysis is now down to 62 students because those who responded to the WTT question with I don't know or Left blank were removed as missing data.

Still, both versions of the analysis are heading in the hoped for direction, so I think that increases the chance that the trend will hold with a larger sample size and would have a good chance of being significant and meaningful.

Now for the more difficult WTT question: vegetables!

Students with prior taste test participation were 10.8 percentage points more likely to want to try vegetables they hadn't eaten before compared to students with no prior taste test participation. Speaking from my experience as both a researcher and a parent, that's a difference I will take as a win. (Huzzah!)

The Maybe group deviates from the trend, but only has eight students. P = 0.519 and partial eta-squared = 0.017, which aren't so good, but for hypothesis generation, given that they fit the same trend as WTT for new fruit, and how difficult vegetables can be, I would be willing to put my own skin in the game and roll the dice on the next study, meaning I will buy $100 of vegetables with my own money for the next study if Chef George is willing to prepare them.

Descriptive Statistics for Dependent Variable: I would like to try <u>vegetables</u> that I haven't eaten before... (v100)

Have you taken part in a taste test before? (vFull)	Mean	Std. Deviation	N
No	56.6667	50.40069	30
Maybe	50.0000	53.45225	8
Yes	67.5000	47.43416	40
Total	61.5385	48.96532	78

Fourth analysis: WTT vegetables (vNum) vs prior taste test participation (vFull)

When the WTT for vegetables analysis is run with a numeric scale, the trend doesn't hold. In this case, the No prior participation group wanted to try new vegetables 1.86 times, while the Yes group wanted to try new vegetables 1.68 times. P = 0.519, which is the same as before. Partial eta-squared has now gone up to 0.023. I used to play backgammon a lot, which is a game of both skill and chance. Played full out as a gambling game, with the doubling die, backgammon requires the player to have blood like ice water running through one's veins. Research is not so different. These results tell me I have slightly less than even odds of finding the hoped for trend if I had a larger sample. Playing against an opponent as formidable as Vegetables, I like those odds. I am willing to double my wager to $200 worth of vegetables for the next study.

Descriptive Statistics for Dependent Variable: I would like to try vegetables that I haven't eaten before... (vNum)

Have you taken part in a taste test before? (vFull)	Mean	Std. Deviation	N
No	1.8571	1.23635	21
Maybe	1.2000	1.09545	5
Yes	1.6765	1.12062	34
Total	1.7000	1.15421	60

Fifth analysis: WTT today's dish (v100) vs prior taste test participation (vFull)

Now let's look at the results for today's dish, which was a green salad with vegetables.

Of the Yes prior participation group, 72.5% of students wanted to try different kinds of salads, compared to only 56.7% of students in the No prior participation group (a 15.8% difference in favor of prior taste test experience). Not bad for a green salad. (P = 0.275, partial eta-squared = 0.034) Also keep in mind that as a program evaluator, I don't necessarily "want" the No group to have a WTT result a lot lower than the Yes group, because the program should ideally elicit WTT from all groups of students.

Descriptive Statistics for Dependent Variable: I would like to try different kinds of <u>today's dish</u> ... (v100)

Have you taken part in a taste test before? (vFull)	Mean	Std. Deviation	N
No	56.6667	50.40069	30
Maybe	50.0000	53.45225	8
Yes	72.5000	45.22026	40
Total	64.1026	48.28045	78

Sixth analysis: WTT today's dish (vNum) vs prior taste test participation (vFull)

With the numeric scale analysis of the same question, the Yes prior participation group wanted to try different kinds of salads 1.75 times while the No group wanted to try the same 1.86 times. (P = 0.416, partial eta-squared = 0.028) I would argue that this salad was roughly equivalent to vegetables, so these results are consistent with the WTT vegetables results. I would also hypothesize that a salad with more fruit would have a WTT result that resembled WTT fruit more than WTT vegetables. I am also wondering if students tend to have a relatively low WTT vegetables because vegetables contain more roughage (which occurred to me because I was looking at the phrase "roughly equivalent" while needing to stop for lunch).

I am going to lump the WTT salad with the WTT vegetables, since these results are so similar, and hold my WTT backgammon-style wager at $200.

Descriptive Statistics for Dependent Variable: I would like to try different kinds of <u>today's dish</u> ... (vNum)

Have you taken part in a taste test before? (vFull)	Mean	Std. Deviation	N
No	1.8636	1.28343	22
Maybe	1.1429	1.34519	7
Yes	1.7500	1.22766	36
Total	1.7231	1.25633	65

Seventh analysis: Sum of three WTT variables (v100) vs prior taste test participation (vFull)

Finally, we come to the computed measure of combined WTT for new salad plus new fruit plus new vegetables.

In this version, where the component variables are coded such that expressing a willingness to try at least once is 100 and not expressing that willingness is 0, a score of 300 means that all students in the sub-group wanted to try all three types of new foods at least once. A score of 0 would mean no students in the sub-group wanted to try any type of new food. A score of 100 could be thought of as the equivalent of all students wanted to try one type of new food at least once. A score of 200 could be thought of as the equivalent of all students wanted to try two types of new food at least once.

The combined results show that, for these 78 students, those in the No prior participation group wanted to try the equivalent of 1.73 types of new foods at least once. Those in the Maybe group wanted to try the equivalent of 1.75 types of new foods at least once. Those in the Yes prior participation group wanted to try the equivalent of 2.20 types of new foods at least once. As with the WTT fruit results using the same weighting, the pattern is clear and in the hoped for direction, but this time with all three types of food taken into account. P =

0.181, partial eta-squared = 0.045, which for hypothesis generation is pretty good. With these odds factored in, I am willing to double my wager to $400 of farm and garden fresh ingredients. (The backgammon reference is a metaphor, but the money is actual cash out of my wallet towards the budget for a proposed grant proposal for a next study.)

Descriptive Statistics for Dependent Variable: Sum of three WTT variables (v100)

Have you taken part in a taste test before? (vFull)	Mean	Std. Deviation	N
No	173.3333	117.24814	30
Maybe	175.0000	116.49647	8
Yes	220.0000	101.77905	40
Total	197.4359	110.45874	78

Eighth analysis: Sum of three WTT variables (vNum) vs prior taste test participation (vFull)
Almost finally, the numeric scale version of the analysis shows that students in the No prior participation group wanted to try new foods 5.79 times (on a scale of 0 to 9), while students in the Yes prior participation group wanted to try new foods 6.13 times. (P = 0.541, partial eta-squared = 0.023) Since in this analysis, a missing value in any one question will cause the student to have a missing value for the computed variable, the N is only 55. Since this is in the hoped for direction for both weightings, and this analysis incorporates all three types of new foods, I will double my wager to $800 of fresh ingredients.

Descriptive Statistics for Dependent Variable: Sum of three WTT variables (vNum)

Have you taken part in a taste test before? (vFull)	Mean	Std. Deviation	N
No	5.7895	2.34708	19
Maybe	5.0000	1.58114	5
Yes	6.1290	2.14075	31
Total	5.9091	2.16258	55

Bonus analysis: WTT fruit (vPl5) vs prior GTC event participation (vEnds)
To show one final analysis, there was one analysis with a statistically significant result. When using the taken part in Garden To Café events before question with the Maybe option

set to Missing, instead of the taken part in taste test before question used above with the Maybe option always included, and when using the modified numeric scale, where the "3+ times" option was recoded to 5 instead of 3, the hoped for direction was seen, with the No prior participation group wanting to try new fruit 3.33 times compared to the Yes prior participation group wanting to try new fruit 4.56 times. P = 0.027 and partial eta-squared = 0.101, which is between a medium and a large effect size.

This result by itself doesn't prove that prior participation in taste tests and GTC events is correlated with, let alone causes, increased willingness to try new fruits, but it does provide evidence that with a properly sized sample, statistical significance and meaningful effect size could be found if it exists as a generalized phenomenon.

Descriptive Statistics for Dependent Variable: I would like to try fruit that I haven't eaten before... (vPI5)

Have you taken part in a Garden To Cafe event before? (vEnds)	Mean	Std. Deviation	N
.00 No	3.3333	2.02286	30
1.00 Yes	4.5556	1.33823	18
Total	3.7917	1.87887	48

Tests of Between-Subjects Effects

Dependent Variable: I would like to try fruit that I haven't eaten before... (vPI5)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	16.806[a]	1	16.806	5.184	.027	.101	5.184	.606
Intercept	700.139	1	700.139	215.989	<.001	.824	215.989	1.000
Have you taken part in a Garden To Cafe event before? (vEnds)	16.806	1	16.806	5.184	.027	.101	5.184	.606
Error	149.111	46	3.242					
Total	856.000	48						
Corrected Total	165.917	47						

a. R Squared = .101 (Adjusted R Squared = .082)
b. Computed using alpha = .05

Summation of all 48 Univariate ANOVA analyses

Out of the 48 analyses run, 28 analyses had results in the hoped for direction, where the Yes group had higher WTT than the No group, while 20 analyses had results running in the contrary direction, although as seen above, some of those contrary results were close to even.

Of the 28 analyses with results in the Yes > No direction, there were four analyses that were almost significant (P < 0.100), and no almost significant results in the No > Yes direction. These were:

Sum of 3 WTTs, v100; Taste tests, vEnds; P = 0.080, partial eta-squared = 0.044
Yes = 220.00 > No = 173.33

Fruit, v100; Taste tests, vEnds; P = 0.068, partial eta-squared = 0.048
Yes = 80.00 > No = 60.00

Fruit, vNum; GTC events, vEnds; P = 0.053, partial eta-squared = 0.079
Yes = 2.78 > No = 2.20

Fruit, vPl5; GTC events, vFull; P = 0.082, partial eta-squared = 0.081
Yes = 4.56 > No = 3.33

It is also useful to note that among the significant and almost significant results, there were two using the v100 WTT weighting (the 100 = try new food at least once recoding), one using the vNum weighting (the 0 to 3 scale) and two using the vPl5 weighting (the same as the 0 to 3 scale, but with 3 recoded to 5). Thus, no one of these weightings is necessarily "better" for inferential analysis. For the next study, a decision should be made ahead of time as to which weighting to use for the primary analysis.

With 58.3% (28 out of 48) of the analyses running in the hoped for direction of Yes prior participation in taste tests and GTC events across three kinds of new foods, one significant result in the hoped for direction with medium to large effect size, four almost significant results in the hoped for direction, and no significant nor almost significant results in the contrary direction, I am willing to double my wager to $1600 of farm fresh vegetables, paid from my own wallet, for the next study.

Cross-tabulations

This section reports cross-tabs for the WTT questions against prior taste test participation using the full scale. The tables show the v100 weighting of the WTT variable (students who said they wanted to try a new food at least once versus those who did not) and the vNum weighting (the 0 to 3 scale). Each of the results is reported in two tables: first a table of the counts of students, and second a table of the column percentages. Often counts and percentages are presented in one cross-tab table, but that style of presentation can get confusing so for now I have chosen to use two tables for each analysis. The other cross-tab results, such as WTT versus prior participation in GTC events, is available on request. Since prior taste test participation and prior GTC event participation are non-identical but

somewhat overlapping constructs, the results of the one set of analyses will be similar to the other.

The rest of the cross tabs and ANOVAs available on request.

NOTE: For each cross-tab, each subscript letter denotes a subset of the "Taken Part In" variable's categories whose column proportions do not differ significantly from each other at the .05 level.

I would like to try different kinds of <u>today's dish</u> ... (v100) *
Have you taken part in a taste test before? (vFull) Crosstabulation

Count		Have you taken part in a taste test before?			Total
		No	Maybe	Yes	
I would like to try different kinds of <u>today's dish</u> ...	Never, I don't know, Left Blank	13_a	4_a	11_a	28
	1 to 3+ times	17_a	4_a	29_a	50
Total		30	8	40	78

I would like to try different kinds of <u>today's dish</u> ... (v100) *
Have you taken part in a taste test before? (vFull) Crosstabulation

Column %		Have you taken part in a taste test before?			Total
		No	Maybe	Yes	
I would like to try different kinds of <u>today's dish</u> ...	Never, I don't know, Left Blank	$43.3\%_a$	$50.0\%_a$	$27.5\%_a$	35.9%
	1 to 3+ times	$56.7\%_a$	$50.0\%_a$	$72.5\%_a$	64.1%
Total		100.0%	100.0%	100.0%	100.0%

I would like to try <u>fruit</u> that I haven't eaten before... (v100) *
Have you taken part in a taste test before? (vFull) Crosstabulation

Count		Have you taken part in a taste test before?			Total
		No	Maybe	Yes	
I would like to try <u>fruit</u> that I haven't eaten before...	Never, I don't know, Left Blank	12$_a$	2$_a$	8$_a$	22
	1 to 3+ times	18$_a$	6$_a$	32$_a$	56
Total		30	8	40	78

I would like to try <u>fruit</u> that I haven't eaten before... (v100) *
Have you taken part in a taste test before? (vFull) Crosstabulation

Column %		Have you taken part in a taste test before?			Total
		No	Maybe	Yes	
I would like to try <u>fruit</u> that I haven't eaten before...	Never, I don't know, Left Blank	40.0%$_a$	25.0%$_a$	20.0%$_a$	28.2%
	1 to 3+ times	60.0%$_a$	75.0%$_a$	80.0%$_a$	71.8%
Total		100.0%	100.0%	100.0%	100.0%

I would like to try <u>vegetables</u> that I haven't eaten before... (v100) *
Have you taken part in a taste test before? (vFull) Crosstabulation

Count		Have you taken part in a taste test before?			Total
		No	Maybe	Yes	
I would like to try <u>vegetables</u> that I haven't eaten before...	Never, I don't know, Left Blank	13$_a$	4$_a$	13$_a$	30
	1 to 3+ times	17$_a$	4$_a$	27$_a$	48
Total		30	8	40	78

I would like to try <u>vegetables</u> that I haven't eaten before... (v100) *
Have you taken part in a taste test before? (vFull) Crosstabulation

Column %		Have you taken part in a taste test before?			Total
		No	Maybe	Yes	
I would like to try <u>vegetables</u> that I haven't eaten before...	Never, I don't know, Left Blank	43.3%$_a$	50.0%$_a$	32.5%$_a$	38.5%
	1 to 3+ times	56.7%$_a$	50.0%$_a$	67.5%$_a$	61.5%
Total		100.0%	100.0%	100.0%	100.0%

Sum of three WTT variables (v100) *
Have you taken part in a taste test before? (vFull) Crosstabulation

Count		Have you taken part in a taste test before?			Total
		No	Maybe	Yes	
Sum of three WTT variables (v100)	.00	7[a]	1[a]	4[a]	12
	100.00	4[a]	3[a]	5[a]	12
	200.00	9[a]	1[a]	10[a]	20
	300.00	10[a]	3[a]	21[a]	34
Total		30	8	40	78

Sum of three WTT variables (v100) *
Have you taken part in a taste test before? (vFull) Crosstabulation

Column %		Have you taken part in a taste test before?			Total
		No	Maybe	Yes	
Sum of three WTT variables (v100)	.00	23.3%[a]	12.5%[a]	10.0%[a]	15.4%
	100.00	13.3%[a]	37.5%[a]	12.5%[a]	15.4%
	200.00	30.0%[a]	12.5%[a]	25.0%[a]	25.6%
	300.00	33.3%[a]	37.5%[a]	52.5%[a]	43.6%
Total		100.0%	100.0%	100.0%	100.0%

I would like to try different kinds of today's dish ... (vNum) *
Have you taken part in a taste test before? (vFull) Crosstabulation

Count		Have you taken part in a taste test before?			Total
		No	Maybe	Yes	
I would like to try different kinds of today's dish ...	.00	5$_a$	3$_a$	7$_a$	15
	1.00	4$_a$	2$_a$	11$_a$	17
	2.00	2$_a$	0$_a$	2$_a$	4
	3.00	11$_a$	2$_a$	16$_a$	29
Total		22	7	36	65

I would like to try different kinds of today's dish ... (vNum) *
Have you taken part in a taste test before? (vFull) Crosstabulation

Column %		Have you taken part in a taste test before?			Total
		No	Maybe	Yes	
I would like to try different kinds of today's dish ...	.00	22.7%$_a$	42.9%$_a$	19.4%$_a$	23.1%
	1.00	18.2%$_a$	28.6%$_a$	30.6%$_a$	26.2%
	2.00	9.1%$_a$		5.6%$_a$	6.2%
	3.00	50.0%$_a$	28.6%$_a$	44.4%$_a$	44.6%
Total		100.0%	100.0%	100.0%	100.0%

I would like to try <u>fruit</u> that I haven't eaten before... (vNum) *
Have you taken part in a taste test before? (vFull) Crosstabulation

Count		Have you taken part in a taste test before?			Total
		No	Maybe	Yes	
I would like to try <u>fruit</u> that I haven't eaten before...	.00	3$_a$	0$_a$	3$_a$	6
	1.00	3$_a$	0$_a$	1$_a$	4
	2.00	3$_a$	1$_a$	5$_a$	9
	3.00	12$_a$	5$_a$	26$_a$	43
Total		21	6	35	62

I would like to try <u>fruit</u> that I haven't eaten before... (vNum) *
Have you taken part in a taste test before? (vFull) Crosstabulation

Column %		Have you taken part in a taste test before?			Total
		No	Maybe	Yes	
I would like to try <u>fruit</u> that I haven't eaten before...	.00	14.3%$_a$		8.6%$_a$	9.7%
	1.00	14.3%$_a$		2.9%$_a$	6.5%
	2.00	14.3%$_a$	16.7%$_a$	14.3%$_a$	14.5%
	3.00	57.1%$_a$	83.3%$_a$	74.3%$_a$	69.4%
Total		100.0%	100.0%	100.0%	100.0%

I would like to try <u>vegetables</u> that I haven't eaten before... (vNum) *
Have you taken part in a taste test before? (vFull) Crosstabulation

		Have you taken part in a taste test before?			
Count		No	Maybe	Yes	Total
I would like to try vegetables that I haven't eaten before...	.00	4_a	1_a	7_a	12
	1.00	5_a	3_a	7_a	15
	2.00	2_a	0_a	10_a	12
	3.00	10_a	1_a	10_a	21
Total		21	5	34	60

I would like to try <u>vegetables</u> that I haven't eaten before... (vNum) *
Have you taken part in a taste test before? (vFull) Crosstabulation

		Have you taken part in a taste test before?			
Column %		No	Maybe	Yes	Total
I would like to try vegetables that I haven't eaten before...	.00	$19.0\%_a$	$20.0\%_a$	$20.6\%_a$	20.0%
	1.00	$23.8\%_a$	$60.0\%_a$	$20.6\%_a$	25.0%
	2.00	$9.5\%_a$		$29.4\%_a$	20.0%
	3.00	$47.6\%_a$	$20.0\%_a$	$29.4\%_a$	35.0%
Total		100.0%	100.0%	100.0%	100.0%

Sum of three WTT variables (vNum) *
Have you taken part in a taste test before? (vFull) Crosstabulation

Count		Have you taken part in a taste test before?			
		No	Maybe	Yes	Total
Sum of three WTT variables (vNum)	.00	1_a	0_a	1_a	2
	3.00	1_a	1_a	4_a	6
	4.00	4_a	1_a	0_b	5
	5.00	2_a	1_a	3_a	6
	6.00	4_a	1_a	12_a	17
	7.00	2_a	1_a	2_a	5
	8.00	2_a	0_a	4_a	6
	9.00	3_a	0_a	5_a	8
Total		19	5	31	55

Sum of three WTT variables (vNum) * Have you taken part in a taste test before? (vFull) Crosstabulation

Column %		Have you taken part in a taste test before?			
		No	Maybe	Yes	Total
Sum of three WTT variables (vNum)	.00	$5.3\%_a$		$3.2\%_a$	3.6%
	3.00	$5.3\%_a$	$20.0\%_a$	$12.9\%_a$	10.9%
	4.00	$21.1\%_a$	$20.0\%_a$		9.1%
	5.00	$10.5\%_a$	$20.0\%_a$	$9.7\%_a$	10.9%
	6.00	$21.1\%_a$	$20.0\%_a$	$38.7\%_a$	30.9%
	7.00	$10.5\%_a$	$20.0\%_a$	$6.5\%_a$	9.1%
	8.00	$10.5\%_a$		$12.9\%_a$	10.9%
	9.00	$15.8\%_a$		$16.1\%_a$	14.5%
Total		100.0%	100.0%	100.0%	100.0%

Power analysis for estimating needed sample size for next study

I used G*Power version 3.1.9.6 for the MacOS to estimate the sample size needed for a next study.

The original sample had a total of 80 responses, but there were also two students who did not complete the survey (one of those students couldn't complete the survey due to an allergy to one of the ingredients in the salad). This was a very good response rate, but one not guaranteed to be achieved all the time. To account for this, I added an 11.17% of the original sample additional non-response, which works out to a little more than 9 students. The rationale for 11.17% was that I wanted at least 10% additional non-response, and then I added a little more so that the final number was a whole number. Then, I looked for the analysis with the most missing data, and thus the lowest N. This was 56 students. The adjusted sample, 91.1594 students / 56 students results in a sample size multiplier of 1.6278.

In G*Power, I selected "t tests", "Means: Difference between two independent means (two groups)", and "A priori: Compute required sample size – given alpha, power, and effect size".

The G*Power defaults for these selections are:

Tail(s) = Two

Effect size d = 0.5 (medium, note that this method of calculating effect size has a different rule of thumb than partial eta-squared)

Alpha err prob = 0.05

Power (1-beta err prob) = 0.95

Allocation ratio N2/N1 = 1

With these default parameter values, G*Power calculates a total sample size needed to achieve the desired level of significance and effect size of 210 students. Multiplied by the sample size multiplier to account for a lower response rate, 341.8 students would be needed to be enrolled in the study.

However, I don't want P (alpha) = 0.05. I want P < 0.05. So, I then set alpha to 0.049. Now 212 students are needed, or 345.1 with the multiplier.

Next, I need to account for the fact that the allocation ratio in the next study is not likely to be 1, since more students will likely be in one sub-group than the other. A fair allocation ratio feels like three. This increases the needed sample to 282 (459.1 with multiplier).

P = 0.049 would be generally acceptable for this research, but I need to allow for multiple sub-groups and multiple comparisons. To give myself plenty of room for this, I set P = 0.0049. This results in a total needed sample size of 430, which when multiplied by 1.6278 to account for extra non-response, recommends a total sample size of exactly 700 students.

Thus, the next study needs to target a minimum sample of 345 students, and an ideal sample of 700 students. These are ambitious but achievable sample size targets.

This means that if 91 students participate in taste tests per day, the study needs a minimum of 3.79 data collection school days and an ideal of 7.68 days. If data collection goes less than perfectly efficiently, that might be eight to 16 school days. The school for the current study had a pre-pandemic enrollment of about 200 students. Thus, to achieve these sample size targets, we need to plan for a minimum of two similar schools (if we had almost all students in each school participate) to an ideal of 16 similar schools (to enroll 700 students, assuming 25% of the students in each school participate, which is about half the rate achieved in the current study). Again, all of these targets are ambitious yet achievable, especially if we can give something of instructional value back to the schools as partial recompense for giving the study instructional time, access to their students and the assistance of school staff – which we did in the current study and have ideas for expanding the instructional benefit of the research further.

Discussion

The theory behind the Mid-Reflective Taste Test Survey is that students would provide more accurate overall taste-response assessments of a dish if they had time to reflect on that dish for multiple taste points, compared to the single taste point of many taste tests (and certainly of previous taste test surveys that led to this survey). Another purpose of this taste test design is to provide the client with a more detailed understanding of how students perceived the dish. An extension of this purpose is the theory that students often have an underdeveloped vocabulary of taste.

Where students in a class either have a high number of "I don't know" or blank responses to a taste component question, or have large variability in their perception of a taste component of the same dish, this may indicate an opportunity for the client and/or the school to invest in additional taste education on that taste component.

In terms of the relationship between prior taste test and GTC event participation on the one hand and willingness to try new foods on the other, this study provided evidence that taste test participation does increase willingness to try new foods. A future taste test study with 345 to 700 students would be able to confirm this hypothesis.

Next steps

One next step would be to conduct further exploratory analysis on the relationships between taste components and overall taste assessment.

A key next step would be to obtain input from the client on reporting needs, especially given programming changes due to the pandemic. This would include discussing the current results with the client.

A third next step would be to share the results of the taste test data analysis with the school at which the data was collected, in a much shorter form that could also function as a math or science lesson for a teacher to try with students. I might also print up small "reward

tokens" or stickers to give to the students who participated in the taste test, especially since the school already uses similar small reward items to motivate students to learn.

A series of pre-, immediate-post- and later-post- willingness to try assessments could be warranted, if feasible within budget constraints.

While there was often consensus among the students about each taste component, the consensus was not total. This window into students' perceptions of food may help improve taste education, and through that, find additional ways to increase school meal participation and students' enjoyment of a wide variety of healthy foods.

Conclusions

In the cross-tabs for fruit versus demographics, we saw that for the 80 students who took part in this taste test, students who had taken part in a taste test before were more likely to want to try new fruits they hadn't eaten before. Students who had taken part in a Garden To Café event before were also more likely to want to try new fruits they hadn't eaten before. There was also a general trend favoring prior taste test participation seen in the 48 Univariate ANOVAs. **These analyses provide evidence that *Garden To Café* has been meeting its overarching program goal of increasing students' willingness to try new foods.**

The analyses also provide evidence that taste testing, as a programmatic and instructional activity, increase students' willingness to try new foods. The evidence from this study is strong enough to be a confirmable hypothesis. Larger samples sizes of 345 to 700 students, at two to 16 schools, over four to 16 data collection days, would be needed to confirm this hypothesis.

Methodology

Software used

The analysis for this report started with a small amount of work in Excel to combine data files, continued with data processing and analysis in R using R Markdown and finished with additional data processing and analysis in SPSS. The R Markdown file was knitted to Word. Selected SPSS output was copied to Word and combined with the R Markdown output.

The survey itself is a special kind of Scantron form. It could be adapted to other formats as needed.

EpiDisplay and tab1()

The EpiDisplay package and its tab1() function for R was written by Virasakdi Chongsuvivatwong.

Color schemes for the `tab1()` bar charts were chosen with sequences of blues, in the case of temperature also reds, to represent sequences of response options. "I don't know" is always light green so that it can be recognized consistently across questions. "LeftBlank" is white, also for consistency across questions, and to reflect that this is missing data. Where the responses options are not a logical sequence, or where using a sequence wouldn't add

useful information, such as for Grade Level, alternating blue and yellow bars are used to make adjacent bars more easily distinguishable.

I like the `tab1()` bar charts because they use a reasonable amount of Tuftean data-ink minimization, while not being overly minimalist. I also like that the tab1() function generates the frequency table at the same time as the bar chart.

The `tab1()` output isn't always perfect, but I have been able to tweak it so that it works well enough, and then some. One tweak in particular was needed to obtain bar charts and tables where the sequence of the categories displays in the same order. (Many software packages display reversed order lists in the chart versus the legend by default. Why, I don't know.) To get this to work, I had to recode the taste component variables twice, and then factorize each pair of variables so that one has the categories listed in one direction, and othe other variable has the categories listed in the opposite direction. Then, I call `tab1()` twice, and keep the chart from one and the table from the other. Unfortunately, `tab1()` has an option to suppress the chart while displaying the table, but not for reverse. This means that I have to manually delete one table I don't need. And this means that this procedure only works when knitting the R Markdown file to Word, not when knitting to HTML or a PDF (unless one were then to manually edit out the extra table from the HTML code, which would go against the point of knitting to HTML for automated and updatable reporting in the first place). There may be a solution to this, but I haven't had time to find it yet.

The above particularly applies to variables where the categories have long-ish names. In that case, you may have no choice but to use horizontal bar charts because some long category labels may not be rendered correctly or at all in vertical bar charts but will be fine in horizontal bar charts.

Also, if you knit your file to HTML, where the charts display fine, but then you knit to Word and the charts get cut off, the problem may be in R Studio's setting for the dimensions of figures. In that case, look for the cog icon with the downward triangle next to it (in the icon bar at the top of your R Markdown code window). Click that, and a menu will appear. Select Output Options… at the bottom of the menu. At the top of the dialog box is a drop down menu called Output Format. Set that menu to HTML and then click Figures in the button bar below. Write down the default figure width and height. Now switch the Output Format menu to Word. Click Figures again. Set the default figure width and height for Word so that they match the settings for HTML. Click OK. Now, if you knit to Word, your charts should display as intended. If you have no idea what I am talking about, enroll in Teachers College's HUDM 5026 and it will all start to make sense.

The Mid-reflective Taste Test Survey

The Mid-reflective Taste Test Survey is the result of about five years of work at the New York City Department of Education's Office of School Support Services and School Food (NYC DOE OSSS/SchoolFood, which later because the Office of Food and Nutrition Services (OFNS)). It was part of my work as the program evaluator for NYC DOE's Garden To Café project. Shortly after the data presented in this report was collected, organizational issues arose that prevented my completing the data analysis at the time. (Those issues are still being addressed. I am happy to discuss them separately.)

Current and past surveys

Mid-reflective taste test survey used in the current study

Garden To Cafe Taste Test Survey

Tell us your opinion of today's Garden to Cafe dish. * There are no wrong answers * We want to know what you really think.

Dish # ☐☐☐☐   Your Code # ☐☐☐☐☐

1. Tell us how today's dish tastes, for each component of taste.

	Low	Medium	High	I don't know
Salty?	○	○	○	○
Sweet?	○	○	○	○
Bitter?	○	○	○	○
Sour?	○	○	○	○
Spicy?	○	○	○	○
Flavorful?	○	○	○	○

2. What is the temperature of today's dish?

Frozen	Cold	Warm	Hot	Very Hot	I don't know
Ⓕ	Ⓒ	Ⓦ	Ⓗ	Ⓥ	⑦

3. Overall, I think today's dish tastes...

Delicious	Okay	Unsatisfying	I don't know	I didn't try it
Ⓓ	Ⓞ	Ⓤ	⑦	○

4. Describe the dish in one or two words ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

5. Tell us how many times you would like to try different kinds of food.

	3+ times	2 times	1 time	Never	I don't know
I would like to try different kinds of today's dish...	③	②	①	⓪	⑦
I would like to try fruit that I haven't eaten before...	③	②	①	⓪	⑦
I would like to try vegetables that I haven't eaten before...	③	②	①	⓪	⑦

6. What grade are you in? (If this is the Summer, what grade did you just complete?)

K	1	2	3	4	5	6	7	8	9	10	11	12	Adult
○	○	○	○	○	○	○	○	○	○	○	○	○	○

7. Have you taken part in a taste test before?   Ⓨ Yes   Ⓜ Maybe   Ⓝ No

8. Have you taken part in a Garden To Cafe event before?   Ⓨ Yes   Ⓜ Maybe   Ⓝ No

0054513003157

Mid-reflective taste test survey used in the current study

Taste test survey used in the previous study

Garden To Cafe Survey

Tell us your opinion of today's Garden To Cafe dishes.

There are no wrong answers. We want to know what you really think.

	Delicious	Okay	Nasty	Didn't try it
I think these **apples** taste...	○	○	○	○
I think these **pears** taste...	○	○	○	○

	Three times or more	Two times	One time	Never
I would like to try different kinds of **apples**...	③	②	①	⓪
I would like to try different kinds of **pears**...	③	②	①	⓪
I would like to try other kinds of **fruit** that I haven't eaten before....	③	②	①	⓪
I would like to try **vegetables** that I haven't eaten before....	③	②	①	⓪

1. Are you a boy or girl? Ⓑ Boy Ⓖ Girl

2. What grade are you in? K 1 2 3 4 5 6 7 8 9 10 11 12 Adult
○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Ⓐ

3. Have you taken part in a taste test before? Ⓨ Yes Ⓜ Maybe Ⓝ No

4. Have you taken part in a Garden To Cafe event before? Ⓨ Yes Ⓜ Maybe Ⓝ No

5054471262417

Taste test survey used in the previous study

References

Abrams, R. (2019). *Taste test data collected at a NYC public school in May 2019 using the Mid-reflective Taste Test Scannable Survey* (Unpublished) [Data set]. Contact the author for more information: rha2121@tc.columbia.edu or Robert@AbramsData.com.

Initial analysis was modeled on past Garden To Café reports, which are available from the author on request. They were publicly published on the web, but are not currently available on the web. Citations below.

Abrams, R., Arnold, N., Sorkin, H., Sedito, V., Edwards, G. (2017). Pilot test of a Garden To Café scannable taste test survey for snack fruit administered in classrooms at PSABX on 12/14/2017 *New York City Department of Education, Office of School Food* (This report was dated 12/21/2017 and was made publicly available on the web. It is currently not available on the web, so please contact Robert Abrams for a copy of the report.)

Abrams, R. (2018). Supplemental results from a Garden To Café scannable taste test survey for snack fruit administered in classrooms at PSABX on 12/14/2017 *New York City Department of Education, Office of School Food* (This report was dated 2/14/2018 and was made publicly available on the web. It is currently not available on the web, so please contact Robert Abrams for a copy of the report.)

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum. [See especially, Chap. 7 for the computation of effect sizes.]

EpiDisplay package and tab1() function for R was written by Virasakdi Chongsuvivatwong.

Tufte refers primarily to *The Visual Display of Quantitative Information, Second Edition* by Edward R. Tufte. Graphics Press. (2001)., as well as to his work in general.

WG = *R for Data Science: Visualize, model, transform, tidy and import data* by Hadley Wickham & Garrett Grolemund. O'Reilly. r4ds.had.co.nz/index.html *accessed 12/21/2021*.

Thank you

Thank you to Dr. Bryan Keller, who taught the Teachers College HUDM 5026 *Introduction to R* class with aplomb (or possibly + `aplomb()`), not just for what I learned, but also for what I now know I need to learn next.

Thank you to Rachel Lee, the course TA, for her support and responsiveness.

Thank you to my classmates, who provided engaging discussion and invaluable collegiality.

Thank you to Vincent Sedito, for his help with scannable survey design.

A big thank you to the students and staff of the school at which the data for this study was collected.

Thank you to Dr. Matthijs Koopmans, who has provided sound statistical advice over the years, and who turned me into a disciple of effect size.

Thank you to Dr. Joseph Novak, without whom I would not have become an educational researcher.

Thank you to Gail Weinstein of Fried Frank, Brian Meegan of Evergent Law, Rosa Aliberti of Berke-Weiss Law, and Jeanne Hamburg of Norris McLaughlin.

Thank you to Robert Deschak, my former NYC DOE OSSS supervisor, who had the foresight to approve my intrepreneurial request to become the Garden To Café program evaluator, and for his support of my point of sale meal participation data analysis and other senior analyst/marketing manager duties.

Thank you to Dr. Pam Koch, who provided technical and subject matter expert advice (and moral support) for my Garden To Café and school food work.

Last but not least, thank you to Chef George Edwards. A chance meeting for lunch at our common office table during an NYC DOE OSSS employee appreciation day changed my life for the better.