



## A Comparison of Consumer Responses Using Paper and Digital Ballots for Eating Quality Assessment of Beef Steaks

Kyle T. Mahagan, Andrea J. Garmyn\*, Jerrad F. Legako, and Mark F. Miller

Animal and Food Sciences Department, Texas Tech University, Lubbock, TX 79409, USA

\*Corresponding author. Email: [garmynan@msu.edu](mailto:garmynan@msu.edu) (Andrea J. Garmyn)

**Abstract:** A consumer study was conducted to determine whether consumers scored beef palatability traits differently on paper versus digital ballots. Beef subprimals representing 4 treatments with inherent variation in eating quality were collected: USDA Select eye of round aged 7 d postmortem, USDA Select strip loin aged 7 d postmortem, USDA Choice tenderloin aged 21 d postmortem, and USDA Prime strip loin aged 21 d postmortem. Accessory muscles, external fat, and connective tissue were removed from subprimals. Muscles were fabricated into 2.5-cm steaks and further divided into 2 equal halves for consumer testing. Consumers ( $n = 360$ ) evaluated 8 samples divided into 2 blocks representing the 2 ballot types. Within each ballot block, Select *longissimus lumborum* samples were always served in the first and fifth position, followed by the remaining 3 treatments served in a randomized order among the latter 3 positions. Consumers rated each steak sample for tenderness, juiciness, flavor liking, and overall liking from 0 to 100 on either a paper or a digital ballot and then rated the paired steak halves on the opposite ballot during the second block of sample testing. Ballot type influenced ( $P < 0.02$ ) all traits, as consumers scored traits greater ( $P < 0.05$ ) on paper compared with digital ballots, regardless of treatment. The magnitude of differences between ballot types was much smaller than the magnitude of differences between cut treatments, which also differed ( $P < 0.01$ ). The smallest margin between ballot type was observed for tenderness (1.8 points); juiciness, flavor liking, and overall liking all differed by 3.4 points. Independent studies could and have utilized digital ballots without concern, as consumers sorted samples by treatment in the current study similarly, regardless of ballot type. However, researchers should consider ballot type for their sensory studies, especially if data will be added to a collective data set.

**Key words:** beef, consumer, digital ballot, eating quality, paper ballot

*Meat and Muscle Biology* 5(1): 48, 1–12 (2021)

doi:10.22175/mmb.12611

Submitted 31 May 2021

Accepted 1 October 2021

## Introduction

Meat Standards Australia (MSA) has a very robust system in place that has been used to evaluate thousands of consumer responses across many different countries dating back to the late 1990s (Bonny et al., 2018). When collecting consumer data using MSA protocols, 7 individual sheets of paper are required per consumer panelist for evaluation of tenderness, juiciness, flavor liking, and overall liking (Watson et al., 2008). Samples are rated for each trait on 100-mm lines (Watson et al., 2008). Once ballots

are completed, ballots are scored digitally or by hand using various measuring tools before consumer responses are entered and compiled. This process can be very time-consuming, depending on the measuring tool. The use of electronic devices equipped with sensory software or digital ballots has emerged as a more efficient means of collecting and compiling data from sensory trials. Several researchers have used digital ballots through Qualtrics Software Survey (Qualtrics, Provo, UT) to collect sensory data of cooked meat samples from trained panelists (McKillip et al., 2018; Vierck et al., 2018; Ponce et al.,

2019) and untrained consumer panelists (Morrow et al., 2019; Sepulveda et al., 2019; Fletcher et al., 2021).

Tenderness, juiciness, and flavor all contribute to palatability and ultimately drive consumer satisfaction (Felderhoff et al., 2020). Muscle function as supportive or locomotive can influence palpability, namely tenderness (Ramsbottom et al., 1944; McKeith et al., 1985; Belew et al., 2003; Hunt et al., 2014). Additionally, greater quality grade (marbling score) will lead to greater palatability scores and a better overall eating experience for the *longissimus lumborum* (Smith et al., 1987; O'Quinn et al., 2012; Corbin et al., 2015). By deliberating choosing different muscles and quality grades for inclusion in this study, differences are to be expected in consumer palatability scores.

The objective of this study was to determine whether consumers score beef samples differently on paper versus digital ballots. To achieve this, samples representing a wide range of tenderness and flavor were deliberately selected and served to consumers for evaluation using both ballots independently. We hypothesized that samples would be scored similarly, regardless of ballot format. Another objective was to determine whether variability, and therefore the detectable difference or power, was different between paper and digital ballots. Again, we hypothesized that variability in responses would be similar, regardless of ballot format. Lastly, we aimed to determine whether consumers spend similar amounts of time evaluating samples using paper and digital ballots. We believed consumers would initially take longer to complete digital ballots, but as they became familiar with the software, completion time would not differ between ballot types.

## Materials and Methods

### Production selection and collection

Carcasses ( $n = 39$ ) of beef (USDA, 2017) were selected from a commercial beef processing facility in Friona, Texas. Thirteen carcasses per quality grade category (USDA Prime, USDA Choice [lower 1/3], and USDA Select) were selected. Paired strip loins (Institutional Meat Purchase Specification [IMPS] 180) and paired eye of rounds (IMPS 171C) were collected from each USDA Select carcass (North American Meat Institute [NAMI], 2014). Paired tenderloins (IMPS 189A) were collected from each USDA Choice carcass (NAMI, 2014). Paired strip loins (IMPS 180) were collected from each USDA Prime carcass (NAMI, 2014).

Carcasses were selected and verified by trained Texas Tech University personnel through visual appraisal of marbling and maturity (USDA, 2017) of the carcass at the time of selection. Carcass data, including beef carcass yield and quality grade traits and *longissimus* muscle pH (TPS Model WP-90 with pH sensor part #111227; TPS Pty Ltd., Brendale, QLD, Australia), were collected and recorded by trained personnel. All subprimals were transported by refrigerated truck to the Gordon W. Davis Meat Science Laboratory in Lubbock, Texas. Subprimals were stored in vacuum bags at 2°C to 4°C in the absence of light until steak fabrication on day 7 postmortem.

### Steak fabrication

All external fat and connective tissue were removed from all subprimals prior to steak fabrication. In addition, the *gluteus medius* was removed from the strip loin, leaving only the *longissimus lumborum*. The *psaos minor* and *iliacus* were removed from the tenderloin, leaving only the *psaos major*. Subprimals were then fabricated into 2.5-cm steaks, maintaining position number for each steak. Steaks were further portioned into 2 equal pieces for consumer testing, and the 2 pieces were assigned randomly to half A and half B. Steak halves were individually vacuum packaged, maintaining the identity of the subprimal, steak position, and half (A or B). All samples were held at 2°C to 4°C and frozen on the respective postmortem aging day: Select *longissimus lumborum* steaks–7 d (SE-LL); Select *semitendinosus* steaks–7 d (SE-ST); Choice *psaos major* steaks–21 d (CH-PM); and Prime *longissimus lumborum* steaks–21 d (PR-LL). Variation in postmortem aging was deliberate to promote variation in eating quality, specifically tenderness.

### Compositional analysis

Proximate analysis was conducted using the anterior most steak from each subprimal. External fat and connective tissue were trimmed from the steak, and the remaining muscle was ground (ProSeries DC Meat Grinder, Cabela's L.L.C., Sidney, NE) in triplicate through a 4-mm plate for proximate analysis. Proximate analysis of raw steaks was conducted by an AOAC official method (Anderson, 2007) using a near infrared spectrophotometer (FoodScan; FOSS NIRsystems, Inc., Laurel, MD). Chemical percentages of fat, moisture, and protein were determined for each subprimal.

## Consumer panels

Consumer panels were conducted in the Texas Tech University Animal and Food Sciences Building. Consumer panelists ( $n = 360$ ) were recruited from Lubbock, Texas, and the surrounding local communities. Each consumer was monetarily compensated for being a participant and was only allowed to participate one time. Fifteen sessions, each consisting of 24 people, were conducted over 8 evenings. Each session lasted approximately 60 min.

Cooking procedures followed MSA protocols (Watson et al., 2008) with modification for number of samples cooked per round. The grill was preheated 45 min prior to cooking, and 12 steak pieces (unrelated to the trial) were cooked to condition the grill and ensure stable temperatures throughout all heating elements (Gee, 2006). An exact time schedule was followed to ensure all steaks were prepared identically and facilitate continued consistency of the heating elements. Each cooking round consisted of 12 samples (as opposed to the traditional 10 samples for standard MSA sessions) that were cooked on the grill at one time. Twelve samples were cooked to accommodate 24 panelists per session, rather than the traditional 20 panelists. Each session consisted of 8 cooking/serving rounds, corresponding to the 8 samples that would be served to each consumer. For each round of sample evaluation, steak samples were cooked following a precise time schedule. Each cooking round lasted 5 min and 45 s, to target a medium degree of doneness. Samples were rested for 3 min and cut into 2 equal halves and served to 2 separate predetermined consumer panelists. Panelists received their samples 7 min apart. By employing this fabrication and cooking method, 2 consumers were able to evaluate the same sample using both ballot formats and limit the variation in responses between the 2 ballot formats that can be attributed to the samples themselves.

Each consumer evaluated 8 samples, with 4 samples per section or block. USDA SE-LL steaks aged 7 d were included in the cooking order and served in the first position of each block of 4 samples—one for the paper ballot and one for the digital ballot block. These warm-up samples were always and only served in the first and fifth rounds as described by Watson et al. (2008), followed by 3 test samples served in a predetermined, balanced order.

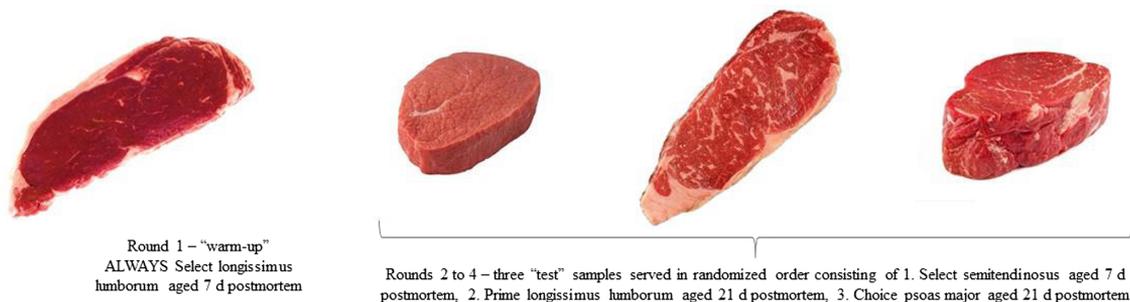
Six test samples were served representing muscle/quality grade combinations described earlier and were evaluated using 2 ballot formats (paper and digital). Following the warm-up samples (round 1 and round 5), muscle/quality grade presentation order were

randomized within each ballot format (in advance). Half of the consumers in a session (consumers 1–12) scored their first 4 samples using the digital ballot, while the other half (consumers 13–24) scored their first four samples using paper ballots. In adherence to MSA cooking style, 2 consumers evaluated 1 steak half. For samples 6 to 8, muscle/quality grade presentation order was re-randomized (in advance), and consumers used the alternative ballot format to evaluate the opposite half of the steaks that were evaluated in rounds 1 to 4, as shown by Figure 1.

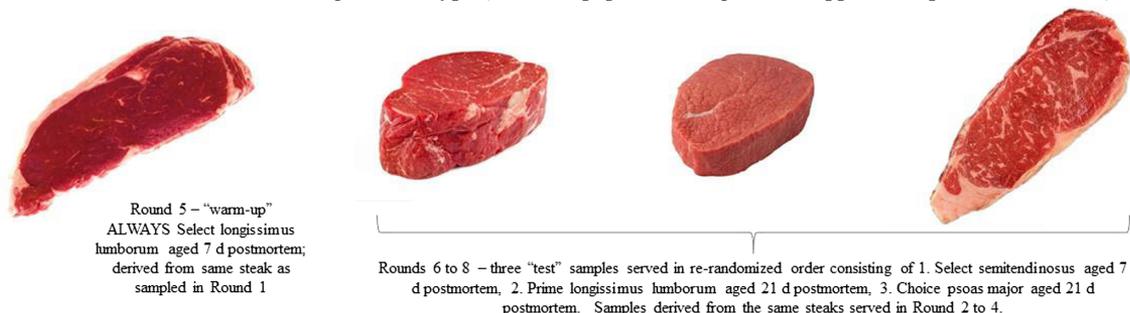
Consumers rated tenderness, juiciness, flavor liking, and overall liking on 100-mm line scales following MSA protocol (Watson et al., 2008) on a paper or digital ballot. Specifically, consumers would mark the line on the paper ballot by drawing a single vertical mark using a pencil on the horizontal line scale for each attribute (Figure 2a). Consumer responses recorded on paper ballots were digitized on a GTCO Calcomp Drawing Board connected to a corded click tip pen (The Logic Group, Austin, TX). The Logic Group's digitizing software was used to set up parameters for data collection and export all data into an electronic format. Each paper ballot was digitized independently in duplicate. Data were checked for accuracy before averaging the 2 responses. When starting the digital evaluation for each sample, the slider was always aligned to the left and set at zero by default (Figure 2b). Participants had to press and drag the marker to the desired location on the scale. Numerical scores were not displayed on the tablet during evaluation. That feature was disabled for this study. The zero anchors were labeled as not tender, not juicy, and dislike extremely of flavor and overall. The 100 anchors were labeled as very tender, very juicy, and like extremely of flavor and overall. Vertical hash marks were present on the lines, regardless of ballot format, at 20, 40, 60, and 80 mm. In addition, a composite score to predict meat quality using 4 traits, referred to as MQ4, was calculated as follows:  $\text{Tenderness}(0.3) + \text{Juiciness}(0.1) + \text{Flavor Liking}(0.3) + \text{Overall Liking}(0.3) = \text{MQ4}$ . Consumers were also asked to check one of 4 boxes for eating quality level to indicate whether they considered each sample "unsatisfactory," "good everyday quality," "better than everyday quality," or "premium quality," as is customary with MSA consumer testing protocols.

Each panelist was seated at a numbered booth and was provided with an electronic tablet, 4 paper ballots, plastic utensils, a toothpick, unsalted crackers, a napkin, an empty cup, a water cup, and a cup with diluted apple juice (10% apple juice and 90% water). Each

Rounds 1 to 4 – all evaluated using 1 ballot type (either all paper or all digital)



Rounds 5 to 8 – all evaluated using 1 ballot type (either all paper or all digital, but opposite of previous 4 rounds)



**Figure 1.** Serving order for samples.

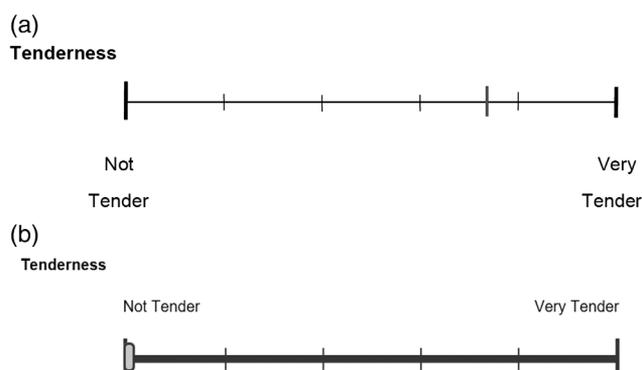
ballot consisted of a demographic questionnaire (digital), 4 paper ballots, 1 iPad (Apple Inc., Cupertino, CA) preloaded with 4 corresponding digital ballots, and a post-panel survey regarding beef purchasing habits (digital). For samples evaluated digitally, samples were rated on digital ballots designed through the Qualtrics survey software (Qualtrics, Provo, UT). Before beginning each session, consumers were given verbal instructions by Texas Tech personnel about the ballot and the process of testing samples. Panels were conducted in a large classroom under fluorescent lighting with tables that were divided into individual consumer booths.

For each session, panelists were selected at random (10 per session) and were timed to determine the length of evaluation time for a sample from the time they received the sample to the time the evaluation of the sample was complete (i.e., pencil down or advanced to the next ballot on the tablet). Times were recorded by 2 individuals who were observing each session.

### Statistical analysis

All data were analyzed using the GLIMMIX procedure (unless noted otherwise) of SAS (version 9.4; SAS Institute Inc., Cary, NC). Carcass and compositional data were analyzed with treatment as the fixed effect. Homogeneity of variance was tested for the

ballot type treatment means for the consumer data using Levene’s Test in PROC GLM of SAS. From the analysis, the  $P$  values were 0.26, 0.07, 0.69, and 0.22 for tenderness, juiciness, flavor liking, and overall liking, respectively. Therefore, homogeneity of variance assumptions were met ( $P > 0.05$ ). Consumer data were analyzed as a  $2 \times 4$  factorial design using ballot format (paper, digital), treatment (SE-LL, SE-ST, CH-PM, PR-LL), and their interaction as fixed effects. Consumer nested within testing day was included as a



**Figure 2.** (a) The top line illustrates the Tenderness line scale on the paper ballot. The (gray) mark shows how a consumer would score the tenderness trait for a given sample—with a single vertical mark. (b) The bottom line represents the Tenderness line scale on the digital ballot. When starting the digital evaluation for each sample, the slider is always aligned to the left at zero. Participants must press and slide the marker to the desired location on the scale. No numbers appear when sliding the marker.

**Table 1.** Least-squares means for carcass traits based on quality grade

Quality Grade	Fat Thickness, mm	Ribeye Area (cm <sup>2</sup> )	Marbling <sup>1</sup>	Lean Maturity <sup>2</sup>	Ossification <sup>2</sup>	pH
Prime	11.9	91.1	736 <sup>a</sup>	187	162 <sup>a</sup>	5.49 <sup>b</sup>
Choice	11.6	96.8	457 <sup>b</sup>	202	135 <sup>b</sup>	5.58 <sup>a</sup>
Select	8.1	94.0	346 <sup>c</sup>	155	138 <sup>b</sup>	5.57 <sup>a</sup>
<i>P</i> value	0.09	0.36	<0.01	0.09	0.03	<0.01
SEM <sup>3</sup>	1.3	2.88	10.4	14.85	7.9	0.01

<sup>1</sup>Marbling scores: slight-00 = 300; small-00 = 400; modest-00 = 500; moderate-00 = 600; slightly abundant-00 = 700; moderately abundant-00 = 800.

<sup>2</sup>Lean maturity/ossification scores: A-00 = 100, B-00 = 200.

<sup>3</sup>Pooled (largest) standard error of the least-squares means (SEM).

<sup>a-c</sup>Within a column, least-squares means without a common superscript differ ( $P < 0.05$ ).

random effect. Ballot completion time data were analyzed as a factorial design using ballot format (paper, digital), round (1–8), and their interaction as fixed effects. A subsequent analysis of consumer data was performed to investigate any interactive effects between ballot type and demographic traits (age, gender, consumption level, education, income, and heritage) and preferred ballot type. The interaction between ballot type and the aforementioned traits were all considered fixed effects. Treatment least-squares means were separated with the PDIF option of SAS using a significance level of  $P < 0.05$ . A chi-square analysis was conducted to determine whether the distribution of responses into the 4 eating quality categories differed between paper and digital ballots ( $P < 0.05$ ). Demographic data were summarized using PROC FREQ.

## Results and Discussion

### Carcass data and compositional analysis

Table 1 shows the differences in carcass traits between the 3 different quality grades used in this study. Quality grade influenced ( $P < 0.05$ ) marbling, ossification, and pH of the beef carcasses. Fat thickness, ribeye area, and lean maturity were similar ( $P > 0.05$ ) between quality grades. As expected, Prime had the most ( $P < 0.05$ ) marbling followed by Choice and Select, with a significant difference between each grade where Prime > Choice > Select. Ossification scores were greater ( $P < 0.05$ ) for Prime carcasses compared with Choice and Select carcasses, which were similar ( $P > 0.05$ ). However, it should be noted that all carcasses were considered “A” maturity. Choice and Select carcasses had similar ( $P > 0.05$ ) pH values in the *longissimus* muscle, but both had greater ( $P < 0.05$ ) pH values than Prime carcasses. However, it is important to note that the average pH value for all

carcasses was below 5.6. Despite statistical differences in pH between quality grades, this would likely not translate to relevant differences in eating quality, especially because muscles other than the *longissimus* were used for half of the treatments.

Results for the compositional analysis are shown in Table 2. As expected, treatment (USDA quality grade and muscle) influenced ( $P < 0.01$ ) fat, moisture, and protein percentage. PR-LL had the greatest percentage of fat, followed by CH-PM, SE-LL, and SE-ST, with each treatment differing ( $P < 0.05$ ). SE-LL had greater ( $P < 0.05$ ) protein percentage than all other treatments except CH-PM; CH-PM had similar protein as SE-ST, and PR-LL had lower protein percentage than all other treatments. Moisture percentage was greatest ( $P < 0.05$ ) in SE-ST, followed by SE-LL and CH-PM, which were similar ( $P > 0.05$ ) but were greater ( $P < 0.05$ ) than PR-LL, which had the lowest moisture percentage of all the treatments. Chemical percentages of the PR-LL were similar to previous percentages (Corbin et al., 2015; Legako et al., 2015; McKillip et al., 2018). The CH-PM and the SE-LL had similar

**Table 2.** Least-squares means by treatment (quality grade × muscle combination) for percentage of chemical fat, moisture, and protein for beef subprimals ( $n = 104$ ) used in consumer sensory panels

Treatment <sup>1</sup>	Fat, %	Protein, %	Moisture, %
CH-PM	4.7 <sup>b</sup>	26.1 <sup>ab</sup>	72.1 <sup>b</sup>
PR-LL	10.4 <sup>a</sup>	25.0 <sup>c</sup>	66.8 <sup>c</sup>
SE-ST	2.2 <sup>d</sup>	25.8 <sup>b</sup>	73.9 <sup>a</sup>
SE-LL	3.4 <sup>c</sup>	26.3 <sup>a</sup>	72.2 <sup>b</sup>
<i>P</i> value	<0.01	<0.01	<0.01
SEM <sup>2</sup>	0.3	0.2	0.3

<sup>1</sup>CH-PM = Choice *psaos major*; PR-LL = Prime *longissimus lumborum*; SE-LL = Select *longissimus lumborum*; SE-ST = Select *semitendinosus*.

<sup>2</sup>Pooled (largest) standard error of the least-squares means (SEM).

<sup>a-d</sup>Within a column, least-squares means without a common superscript differ ( $P < 0.05$ ).

**Table 3.** Demographic characteristics of consumers ( $n = 360$ ) who participated in consumer sensory panels

Characteristic	Response	Percentage of Consumers, %
Age Group, y	<20	8.6
	20–29	23.6
	30–39	24.4
	40–49	19.4
	50–59	11.1
	>60	12.8
Gender	Male	45.0
	Female	55.0
Occupation	Tradesperson	16.9
	Professional	23.1
	Administration	12.8
	Sales & Service	14.2
	Laborer	8.9
	Homemaker	2.8
	Student	10.0
	Unemployed/retired	11.4
Beef Consumption	Daily	15.0
	4–5 times per week	26.1
	2–3 times per week	40.6
	Weekly	12.2
	Biweekly	4.7
	Monthly	0.8
	Rarely/never	0.6
Degree of Doneness	Blue	0.3
	Rare	3.6
	Medium rare	34.4
	Medium	33.3
	Medium well done	18.6
	Well done	9.7
Household Size, No. of Adults	0	0.6
	1	13.1
	2	55.3
	4	17.8
	5	8.1
	6	3.6
Household Size, No. of Children	0	46.4
	1	20.0
	2	15.8
	3	11.1
	4	4.4
	5+	2.2
Income	<\$20,000	12.5
	\$20,000–\$50,000	28.9
	\$50,001–\$75,000	23.6
	\$75,001–\$100,000	14.4
	>\$100,000	20.6
Education	Non-high school graduate	4.4
	High school graduate	25.6

**Table 3.** (Continued)

Characteristic	Response	Percentage of Consumers, %
	Some college/technical school	39.7
	College graduate	20.3
	Post-college graduate	10.0
Heritage	African American	1.9
	Asian	1.1
	Caucasian/white	47.2
	Hispanic	45.8
	Native American	1.4
	Mixed race	2.2
Ballot Preference	Other	0.3
	Digital	79.9
	Paper	20.1

chemical percentages to previously reported literature (Legako et al., 2015).

### Demographic profile

Table 3 outlines the panelist demographics for this study. Most of the panelists (67.5%) were aged 20 to 49 years old. There were slightly more female than male participants (55.0% vs. 45.0%, respectively). Most participants were either employed full-time (79.6%) or were students (10.0%). A majority (55.3%) of panelists lived in a 2-adult household, with no children being the most common (46.4%) number of children in the household. Over half of consumers had an annual income ranging from \$20,000 to \$75,000. A vast majority (95.6%) of panelists were at least a high school or college graduate. Caucasian/white was the primary ethnic group followed closely by Hispanic. When asked about ballot preference between the digital ballot and the paper ballot, a clear majority (79.9%) preferred digital over paper. A large proportion (81.7%) of consumers ate beef multiple times a week (2–3 times or more). When asked about how they like their beef prepared, a majority (86.4%) of consumers preferred their beef cooked to medium-rare to a medium-well done degree of doneness.

### Ballot type and treatment effects on palatability ratings

As shown in Table 4, ballot type and treatment influenced ( $P \leq 0.02$ ) tenderness, juiciness, flavor liking, overall liking, and MQ4. Despite significant differences for both ballot type and treatment, the magnitude of difference between ballot types, was generally much smaller than the magnitude of differences between muscle  $\times$  quality grade treatments. Eating

**Table 4.** The effects of ballot and treatment<sup>1</sup> (quality grade × muscle combination) on consumer scores ( $n = 360$ ) for palatability traits

	Tenderness <sup>2</sup>	Juiciness <sup>2</sup>	Flavor Liking <sup>2</sup>	Overall Liking <sup>2</sup>	MQ4 <sup>3</sup>	Quality Level <sup>4</sup>
<b>Ballot</b>						
Digital	56.8 <sup>b</sup>	56.3 <sup>b</sup>	54.8 <sup>b</sup>	55.7 <sup>b</sup>	55.8 <sup>b</sup>	3.2
Paper	58.6 <sup>a</sup>	59.7 <sup>a</sup>	58.2 <sup>a</sup>	59.1 <sup>a</sup>	58.7 <sup>a</sup>	3.3
<i>P</i> value <sup>5</sup>	0.02	<0.01	<0.01	<0.01	<0.01	0.35
SEM <sup>6</sup>	0.7	0.8	0.9	0.8	0.7	0.03
<b>Treatment</b>						
CH-PM	84.8 <sup>a</sup>	68.0 <sup>b</sup>	70.4 <sup>a</sup>	75.2 <sup>a</sup>	75.9 <sup>a</sup>	3.9 <sup>a</sup>
PR-LL	73.0 <sup>b</sup>	75.0 <sup>a</sup>	69.0 <sup>a</sup>	71.9 <sup>b</sup>	71.6 <sup>b</sup>	3.8 <sup>b</sup>
SE-ST	30.9 <sup>d</sup>	37.6 <sup>d</sup>	39.1 <sup>c</sup>	36.3 <sup>d</sup>	35.6 <sup>d</sup>	2.5 <sup>d</sup>
SE-LL	42.1 <sup>c</sup>	51.4 <sup>c</sup>	47.5 <sup>b</sup>	46.3 <sup>c</sup>	45.9 <sup>c</sup>	2.8 <sup>c</sup>
<i>P</i> value <sup>7</sup>	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
SEM <sup>6</sup>	0.9	1.0	1.0	0.9	0.8	0.03
Ballot × Treatment <i>P</i> value <sup>8</sup>	0.75	0.75	0.37	0.79	0.58	0.45

<sup>1</sup>CH-PM = Choice *psaos major*; PR-LL = Prime *longissimus lumborum*; SE-LL = Select *longissimus lumborum*; SE-ST = Select *semiteminosus*.

<sup>2</sup>Consumer tenderness, juiciness, and flavor liking recorded on anchored 100-mm line scale: 0 = very tough, very dry, and dislike extremely of flavor and overall; 100 = very tender, very juicy, and like extremely of flavor and overall.

<sup>3</sup>MQ4 scores were calculated using the formula Tenderness(0.3) + Juiciness(0.1) + Flavor liking(0.3) + Overall Liking(0.3).

<sup>4</sup>2 = unsatisfactory, 3 = good every day, 4 = better than every day, and 5 = premium.

<sup>5</sup>Observed significance levels for consumer scores based on ballot type.

<sup>6</sup>Pooled (largest) standard error of the least-squares means (SEM).

<sup>7</sup>Observed significance levels for consumer scores based on treatment.

<sup>8</sup>Observed significance levels for consumer scores based on the interaction between ballot type and treatment.

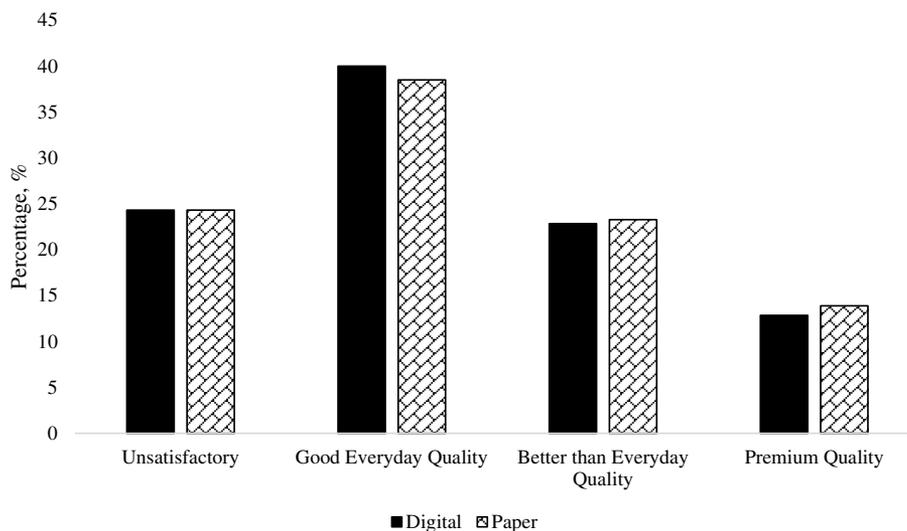
<sup>a-d</sup>Within a column and treatment (ballot type or treatment), least-squares means without a common superscript differ ( $P < 0.05$ ).

quality classification was affected by treatment ( $P < 0.01$ ) but was not impacted ( $P > 0.05$ ) by ballot type. No interactions between ballot and treatment were observed for any palatability traits ( $P > 0.05$ ). Consumers scored tenderness, juiciness, flavor liking, and overall liking greater ( $P < 0.05$ ) on paper ballots compared with digital ballots, regardless of treatment. Because MQ4 is a composite trait of 4 four traits, it followed a similar trend. The smallest margin between paper and digital ballots was observed for tenderness, where scores only differed by 1.8 points when traits were scored on 100-mm lines. However, juiciness, flavor liking, and overall liking all differed by 3.4 points in favor of paper ballots. When weightings were applied to determine the MQ4 score, the composite score differed by 2.9 points. Despite differences in all palatability traits, eating quality classification was similar between ballots. This is further illustrated in Figure 3, which shows the distribution of eating quality classification by ballot type. A chi-square analysis revealed that the frequency of distribution of the various categories did not differ between paper and digital ballots ( $P > 0.05$ ). It is of particular note, however, that despite similar percentages in each category, there were instances when consumers would designate the same

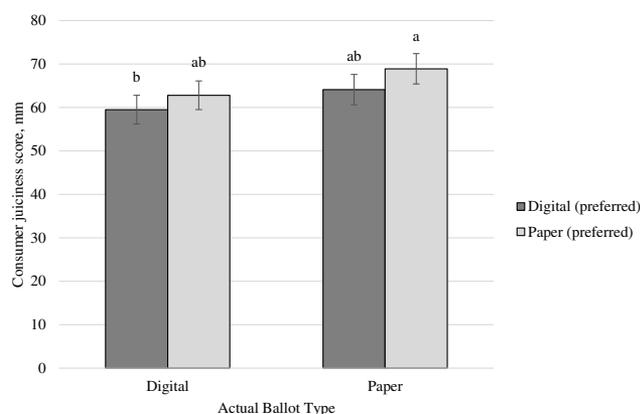
steak sample as a different category on the 2 ballot types. As an example, a consumer might indicate that Steak A was “unsatisfactory” on the digital ballot, but they would classify the corresponding Steak B as “good everyday quality” on the paper ballot. Specifically, from the 1,440 paired comparisons (4 comparisons per consumer), 51% were classified the same on both ballots, 23% were placed in a higher category on the digital ballot, and 26% were designated as a higher category on the paper ballot (data not shown in tabular form).

Tenderness, overall liking, MQ4, and quality level were greatest ( $P < 0.05$ ) for CH-PM compared with all other treatments, with a significant difference between the remaining treatments where CH-PM > PR-LL > SE-LL > SE-ST. Consumers scored juiciness greater ( $P < 0.05$ ) in PR-LL compared with all other treatments, with a significant difference between the remaining treatments where PR-LL > CH-PM > SE-LL > SE-ST. Flavor liking was similar ( $P > 0.05$ ) and greater ( $P < 0.05$ ) for CH-PM and PR-LL than SE-LL, which was intermediate, and SE-ST had the lowest flavor liking scores.

Muscles that were considered support muscles (PR-LL, CH-PM, SE-LL) were scored as more tender



**Figure 3.** Distribution of eating quality classification by ballot type ( $n = 360$  consumers; 4 samples/ballot type). Chi-square analysis indicated the frequency of distribution of the 4 eating quality categories did not differ between paper and digital ballots ( $P = 0.79$ ). A similar percentage appeared in each category.



**Figure 4.** Influence of ballot preference on juiciness scores. Consumer ( $n = 360$ ) juiciness recorded on anchored 100-mm line scale: 0 = very dry; 100 = very juicy. <sup>a,b</sup>Bars without a common superscript differ ( $P < 0.05$ ).

than the locomotive muscle (SE-ST). The CH-PM were scored most tender, and the SE-ST were scored as least tender, which was consistent with reports by Rhee et al. (2004), who evaluated the palatability traits of 11 beef muscles. Juiciness scores were related to fat percentages, with consumers scoring the PR-LL as most juicy and scoring the SE-ST as least juicy. Overall liking mirrored that of tenderness scores; this finding was not surprising given the documented importance of tenderness to overall palatability (Boleman et al., 1997; Miller et al., 2001; O’Quinn et al., 2018).

Consumers were able to identify differences in treatments and rate accordingly, regardless of ballot. However, on average consumers scored samples higher on paper ballots compared with digital ballots. The

magnitude of difference between ballot types, however, was generally much smaller than the magnitude of difference between cut treatments. The reason for the difference between the 2 ballots is unclear, as it was expected that samples would be scored similarly on both ballots. In part, the adequate sample size ( $n = 360$  consumers) allows for easier detection of differences because it reduces the standard error, but the adequate sample size also improves the validity of the results because it increases power. It is possible that, in a smaller test sample (i.e., fewer consumer panelists), the difference in tenderness between ballot type could be inconsequential. Even so, despite what seems to be a relatively small difference (1.8- to 3.4-point difference scored on 100-mm line scales) could have significant bearing on a consumer’s final classification of overall eating quality. A 1-point difference on this scale can equate to a difference between “good everyday quality” and “premium quality.” It only takes a 1-point shift, but it depends where on the scale that 1-point shift occurs.

On the digital ballots, the marker is always initially placed at zero by default for each trait, and consumers must drag the marker to the desired score on the line. They cannot simply point to the precise spot on the line where they want to score the trait but must drag the marker to that point. Perhaps consumers become fatigued with this dragging exercise and just get to a point that is “close enough” in their mind. Conversely, on the paper ballot, the consumers simply mark the exact spot where they wanted to score a particular trait for a given sample. Essentially, the consumer could be suffering from what we have coined as “lazy finger”

when they score on a digital ballot as opposed to the paper ballot.

It was also noted that, when comparing the 4 palatability traits on a paper ballot versus a digital ballot, tenderness scores had the lowest difference at 1.8 points. One reason could be that this was the first trait listed and scored on the ballot, and perhaps consumers were more critical of this trait and then relaxed as they observed the next 3 traits. However, there were no clear reasons as to why these discrepancies existed.

### Ballot completion times

Tables 5 and 6 show that ballot type and ballot round affected ( $P < 0.01$ ) time of sample evaluation. No interaction was observed between ballot type and round ( $P = 0.11$ ). Consumers used more time to record responses ( $P < 0.05$ ) on digital ballots compared with paper ballots regardless of ballot round. Consumers used the most time ( $P < 0.05$ ) for ballot completion in round 1, followed by round 5, compared with all other rounds, regardless of ballot type. Consumers were likely taking longer to evaluate those samples as they were getting acclimated to the ballot. However, once the consumers became accustomed to the ballot they were using, consumers completed their assessments more rapidly for the remaining samples in that ballot block.

### Demographic traits and ballot preference effects on palatability ratings

To determine whether demographic characteristics (age, gender, consumption, income, education, and heritage) or preferred ballot type influenced palatability scores, we evaluated the interactive effects of those traits with the actual ballot type. As seen in Table 7, preferred ballot influenced ( $P = 0.01$ ) juiciness scores when comparing paper and digital ballots. Gender contributed to variation in flavor and overall liking scores ( $P < 0.01$ ), and education level resulted in variation between ballot type for flavor liking scores ( $P <$

**Table 5.** Average completion time of digital ballot versus paper ballot per round

Ballot Type	Completion Time (s)
Digital Ballot	120.2 <sup>a</sup>
Paper Ballot	109.2 <sup>b</sup>
<i>P</i> Value	<0.01
SEM <sup>1</sup>	1.62

<sup>1</sup> Pooled (largest) standard error of the least-squares means (SEM).

<sup>a-b</sup>Least-squares means without a common superscript differ ( $P < 0.05$ ).

**Table 6.** Average completion time of each panel round

Round	Completion Time (s)
1	153.4 <sup>a</sup>
2	114.5 <sup>c</sup>
3	104.5 <sup>d</sup>
4	102.7 <sup>d</sup>
5	126.6 <sup>b</sup>
6	106.1 <sup>cd</sup>
7	102.1 <sup>d</sup>
8	107.7 <sup>cd</sup>
<i>P</i> Value	<0.01
SEM <sup>1</sup>	3.23

<sup>1</sup> Pooled (largest) standard error of the least-squares means (SEM).

<sup>a-d</sup>Least-squares means without a common superscript differ ( $P < 0.05$ ).

0.01). Age, consumption level, income level, and heritage did not contribute to the variation between ballot type for any of the palatability traits ( $P > 0.05$ ). When considering the demographic characteristics, we suspected that age could influence palatability ratings between ballot types, as younger generations might be more familiar with technology. However, according to Kakulla (2020), tablet adoption among adults aged 50+ years has risen from 30% in 2014 to 52% in 2019, actually surpassing tablet adoption among 18- to 49-year-olds (49%). As such, age is not one of the demographic characteristics that influenced scores between ballot types.

As previously noted in Table 3, when asked about ballot preference between the digital ballot and the paper ballot, a clear majority (79.9%) preferred digital over paper. Figure 4 illustrates the interaction between actual and preferred ballot type for juiciness scores. This was the only palatability trait for which this

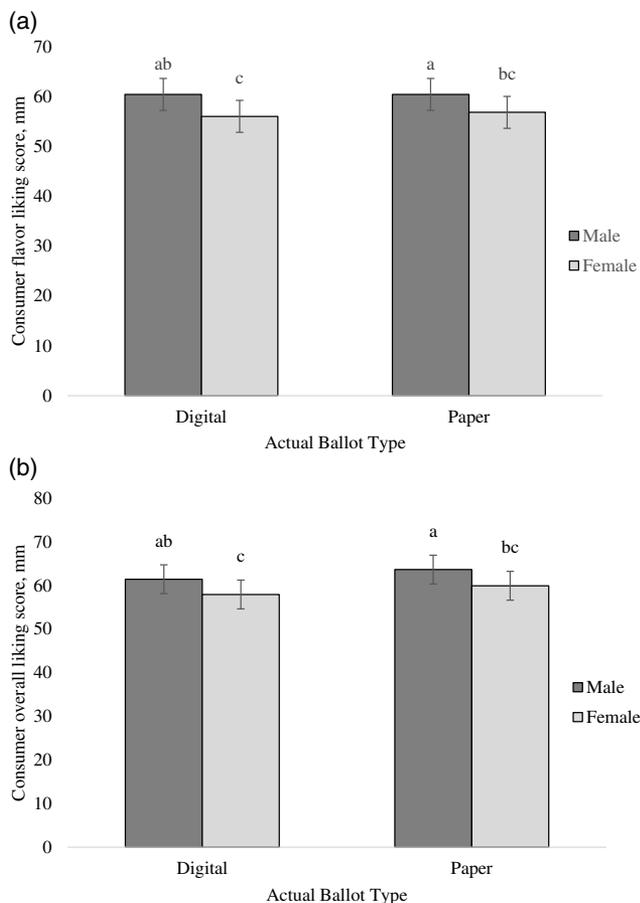
**Table 7.** Examination of the interactive effects of actual ballot type with demographic characteristics and preferred ballot type on consumer sensory scores ( $n = 360$ ) for palatability traits

Interaction	Tenderness	Juiciness	Flavor Liking	Overall Liking
Ballot × Preferred Ballot	0.16	0.01	0.30	0.20
Ballot × Age	0.82	0.54	0.60	0.69
Ballot × Gender	0.29	0.13	<0.01	<0.01
Ballot × Consumption	0.41	0.16	0.14	0.40
Ballot × Income	0.58	0.92	0.27	0.19
Ballot × Education	0.87	0.39	<0.01	0.53
Ballot × Heritage	0.42	0.19	0.09	0.10

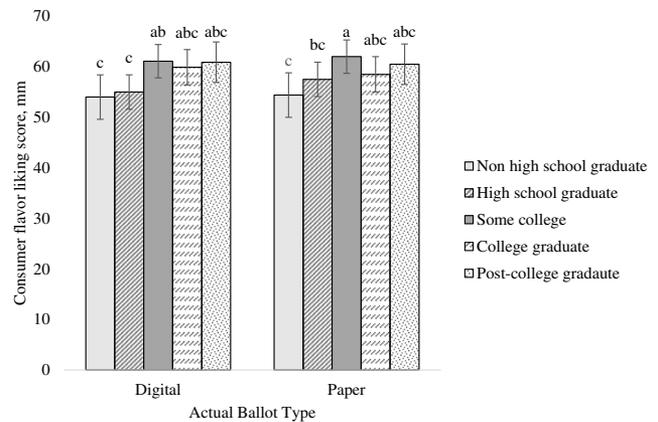
Values appear as observed significance levels (*P* values) for consumer scores based on the interaction between ballot type and the respective traits.

interaction was observed. When consumers were using a digital ballot for evaluation and preferred the digital ballot, juiciness scores were lower ( $P < 0.05$ ) than when scored on a paper ballot with a preference for a paper ballot. However, within each tested ballot type, there were no differences in juiciness scores owing to the preferred ballot type ( $P > 0.05$ ). Overall, there does not appear to be any consistent pattern or advantage in scores among all palatability traits when using the preferred ballot as indicated by the consumer, but we did want to investigate and report any differences that were observed.

Figure 5 illustrates the interaction between gender and ballot type for flavor liking (Figure 5a) and overall liking (Figure 5b). Regardless of the ballot type used, males scored flavor liking and overall liking greater ( $P < 0.05$ ) than females. Additionally, when scoring samples on paper ballots, males scored traits greater than females on both paper and digital ballots.



**Figure 5.** (a) Influence of gender on flavor liking scores. (b) Influence of gender on overall liking scores. Consumer ( $n = 360$ ) flavor and overall liking recorded on anchored 100-mm line scale: 0 = dislike extremely of flavor and overall; 100 = like extremely of flavor and overall. <sup>a,b</sup>Bars without a common superscript differ ( $P < 0.05$ ).



**Figure 6.** Influence of education level on flavor liking scores. Consumer ( $n = 360$ ) flavor liking recorded on anchored 100-mm line scale: 0 = dislike flavor extremely; 100 = like flavor extremely. <sup>a-c</sup>Bars without a common superscript differ ( $P < 0.05$ ).

Finally, Figure 6 shows the variation in flavor liking scores due to education level across ballot type ( $P < 0.01$ ). Participants with at least some college education similarly liked the flavor more than non-high school and high school graduates, regardless of ballot type. Participants within the 3 higher levels of education scored flavor liking similarly ( $P > 0.05$ ), regardless of the ballot type. When assessing the potential influence of these demographic factors, there does not appear to be any consistent pattern or advantage in scores across all palatability traits, but we did want to investigate and report any differences that were observed.

## Conclusions

Most of the consumers in this study indicated that they preferred the use of a digital ballot over a paper ballot. Despite consumer preference for the digital ballot, the results of this study show that consumers scored palatability traits greater for samples when using the paper ballot as opposed to the digital ballot; however, the use of their preferred ballot type does not seem to be a major contributing factor. Scores were greater by 1.8 points in favor of the paper ballot for tenderness and nearly 3.5 points greater for paper in the other 3 palatability traits when traits were scored on 100-mm line scales. There were no obvious reasons for discrepancies in scores between ballot types. Additionally, the variability in responses (standard deviation) was deemed similar between paper and digital ballots. The magnitude of difference between ballot types, however, was much smaller than the magnitude of difference

between muscle  $\times$  quality grade treatments. In light of these results, researchers should consider ballot type for their sensory studies if data will be added to a collective data set, such as the consumer data for MSA testing purposes, when studies are conducted over a period of time and/or in multiple locations. If data will be compiled from multiple locations and over a period of time, an adjustment factor may be warranted for data collected using digital ballots. However, the results from this study would need to be validated in other locations with more diverse test samples (i.e., more muscles, different cattle diets, other species, different cook methods) and using different cook methods to ensure that no interactions occur between ballot type and treatments. Independent studies could and have utilized digital ballots without any issues, as consumers sorted samples by treatment in the current study similarly, regardless of ballot type.

## Literature Cited

- Anderson, S. 2007. Determination of fat, moisture, and protein in meat and meat products using the FOSS FoodScan near-infrared spectrophotometer with FOSS Artificial Neural Network Calibration Model and Associated Database: Collaborative study. *J. AOAC Int.* 90:1073–1083. <https://doi.org/10.1093/jaoac/90.4.1073>.
- Belew, J. B., J. C. Brooks, D. R. McKenna, and J. W. Savell. 2003. Warner–Bratzler shear evaluations of 40 bovine muscles. *Meat Sci.* 64:507–512. [https://doi.org/10.1016/S0309-1740\(02\)00242-5](https://doi.org/10.1016/S0309-1740(02)00242-5).
- Boleman, S. J., S. L. Boleman, R. K. Miller, J. F. Taylor, H. R. Cross, T. L. Wheeler, M. Koochmarai, S. D. Shackelford, M. F. Miller, R. L. West, D. D. Johnson, and J. W. Savell. 1997. Consumer evaluation of beef of known categories of tenderness. *J. Anim. Sci.* 75:1521–1524. <https://doi.org/10.2527/1997.7561521x>.
- Bonny, S. P. F., R. A. O'Reilly, D. W. Pethick, G. E. Gardner, J. F. Hocquette, and L. Pannier. 2018. Update of Meat Standards Australia and the cuts based grading scheme for beef and sheepmeat. *J. Integr. Agr.* 17:1641–1654. [https://doi.org/10.1016/S2095-3119\(18\)61924-0](https://doi.org/10.1016/S2095-3119(18)61924-0).
- Corbin, C. H., T. G. O'Quinn, A. J. Garmyn, J. F. Legako, M. R. Hunt, T. T. N. Dinh, R. J. Rathmann, J. C. Brooks, and M. F. Miller. 2015. Sensory evaluation of tender beef strip loin steaks of varying marbling levels and quality treatments. *Meat Sci.* 100:24–31. <https://doi.org/10.1016/j.meatsci.2014.09.009>.
- Felderhoff, C., C. Lyford, J. Malaga, R. Polkinghorne, C. Brooks, A. Garmyn, and M. Miller. 2020. Beef quality preferences: Factors driving consumer satisfaction. *Foods.* 9:289. <https://doi.org/10.3390/foods9030289>.
- Fletcher, W. T., A. J. Garmyn, J. F. Legako, D. R. Woerner, and M. F. Miller. 2021. Investigation of smoked beef brisket palatability from three USDA quality grades. *Meat Muscle Biol.* 5:1,1–12. <https://doi.org/10.22175/mmb.10963>.
- Gee, A. 2006. Protocol Book 4: For the thawing preparation, cooking and serving of beef for MSA [Meat Standards Australia] pathway trials. Meat and Livestock Australia, North Sydney.
- Hunt, M. R., A. J. Garmyn, T. G. O'Quinn, C. H. Corbin, J. F. Legako, R. J. Rathmann, J. C. Brooks, and M. F. Miller. 2014. Consumer assessment of beef palatability from four beef muscles from USDA Choice and Select graded carcasses. *Meat Sci.* 98:1–8. <https://doi.org/10.1016/j.meatsci.2014.04.004>.
- Kakulla BN. 2020. Older adults keep pace on tech usage. AARP. <https://www.aarp.org/research/topics/technology/info-2019/2020-technology-trends-older-americans.html>. (Accessed 7 September 2021).
- Legako, J. F., J. C. Brooks, T. G. O'Quinn, T. D. J. Hagan, R. Polkinghorne, L. J. Farmer, and M. F. Miller. 2015. Consumer palatability scores and volatile beef flavor compounds of five USDA quality grades and four muscles. *Meat Sci.* 100:291–300. <https://doi.org/10.1016/j.meatsci.2014.10.026>.
- McKeith, F. K., D. L. De Vol, R. S. Miles, P. J. Bechtel, and T. R. Carr. 1985. Chemical and sensory properties of thirteen major beef muscles. *J. Food Sci.* 50:869–872. <https://doi.org/10.1111/j.1365-2621.1985.tb12968.x>.
- McKillip, K. V., A. K. Wilfong, J. M. Gonzalez, T. A. Houser, J. A. Unruh, E. A. E. Boyle, and T. G. O'Quinn. 2018. Sensory evaluation of enhanced beef strip loin steaks cooked to 3 degrees of doneness. *Meat Muscle Biol.* 1:227–241. <https://doi.org/10.22175/mmb2017.06.003>.
- Miller, M. F., M. A. Carr, C. B. Ramsey, K. L. Crockett, and L. C. Hoover. 2001. Consumer thresholds for establishing the value of beef tenderness. *J. Anim. Sci.* 79:3062–3068. <https://doi.org/10.2527/2001.79123062x>.
- Morrow, S. J., A. J. Garmyn, N. C. Hardcastle, J. C. Brooks, and M. F. Miller. 2019. The effects of enhancement strategies of beef flanks on composition and consumer palatability characteristics. *Meat Muscle Biol.* 3:457–466. <https://doi.org/10.22175/mmb2019.07.0030>.
- NAMI. 2014. The meat buyer's guide. 8th ed. North American Meat Institute, Washington, DC 20036.
- O'Quinn, T. G., J. C. Brooks, R. J. Polkinghorne, A. J. Garmyn, B. J. Johnson, J. D. Starkey, R. J. Rathmann, and M. F. Miller. 2012. Consumer assessment of beef strip loin steaks of varying fat levels. *J. Anim. Sci.* 90:626–634. <https://doi.org/10.2527/jas.2011-4282>.
- O'Quinn, T. G., J. F. Legako, J. C. Brooks, and M. F. Miller. 2018. Evaluation of the contribution of tenderness, juiciness, and flavor to the overall consumer beef eating experience. *Trans. Anim. Sci.* 2:26–36. <https://doi.org/10.1093/tas/txx008>.
- Ponce, J., J. C. Brooks, and J. F. Legako. 2019. Consumer liking and descriptive flavor attributes of *M. longissimus lumborum* and *M. gluteus medius* beef steaks held in varied packaging systems. *Meat Muscle Biol.* 3:158–170. <https://doi.org/10.22175/mmb2018.12.0041>.
- Ramsbottom, J. M., J. Strandine, and C. H. Koonz. 1944. Comparative tenderness of representative beef muscles. *J. Food Sci.* 10:497–509.

- Rhee, M. S., T. L. Wheeler, S. D. Shackelford, and M. Koohmaraie. 2004. Variation in palatability and biochemical traits within and among eleven beef muscles. *J. Anim. Sci.* 82:534–550. <https://doi.org/10.2527/2004.822534x>.
- Sepulveda, C. A., A. J. Garmyn, J. F. Legako, and M. F. Miller. 2019. Cooking method and USDA quality grade affect consumer palatability and flavor of beef strip loin. *Meat Muscle Biol.* 3:375–388. <https://doi.org/10.22175/mmb2019.07.0031>.
- Smith, G. C., J. W. Savell, H. R. Cross, Z. L. Carpenter, C. E. Murphey, G. W. Davis, H. C. Abraham, F. C. Parrish, and B. W. Berry. 1987. Relationship of USDA quality grades to palatability of cooked beef. *J. Food Quality.* 10:269–286. <https://doi.org/10.1111/j.1745-4557.1987.tb00819.x>.
- USDA. 2017. Official United States Standards for Grades of Carcass Beef. Agricultural Marketing Service, USDA, Washington, DC.
- Vierck, K. R., J. M. Gonzalez, T. A. Houser, E. A. E. Boyle, and T. G. O’Quinn. 2018. Marbling texture’s effects on beef palatability. *Meat Muscle Biol.* 2:142–153. <https://doi.org/10.22175/mmb2017.10.0052>.
- Watson, R., A. Gee, R. Polkinghorne, and M. Porter. 2008. Consumer assessment of eating quality development of protocols for Meat Standards Australia (MSA) testing. *Aust. J. Exp. Agr.* 48:1360–1367. <https://doi.org/10.1071/EA07176>.