Point of Scale Differences in the Rating Scales of the Liking Score as the Determinant of Efficiency in Penalty Analysis of Liking Score and Other Sensory Attributes of Plantain Chips on Just About Right Scale

Olugbemi A. Olujimi
Data Management Dept., FactInfo, Lagos Nigeria
1 Asabi Cole Stree, Opposite Daily Times, Ikeja Lagos, Nigeria
Email: olugbemi.olujimi@factinfo.com.ng , olugbemi.olujimi@roehampton-online.ac.uk , olugbemijimi@yahoo.com

Abstract

Important factors in product innovation is product's attributes optimisation. Here, consumer's acceptance or rejection of the product is determined for the product formulation. For this exercise, Penalty Analysis a statistical technique used by food scientists and consumer product development experts to evaluate the consumer behaviour on the product in either penalising it or not due to these consumers' responses on the tested product. Usually, researchers often uses 5-point Just About Right scale but known to use different levels of rating scale on overall liking score, but there exist controversies on this. This study evaluates the effect of different points of scale for overall liking score and the sensory attributes with just about right (JAR) scale on the Penalty Analysis. In this study, three different points of scales on rating scales of Overall liking (dependent variable) of the plantain chips were selected: 5-points scale, 7-points scale and 9-points scale against the sensory attributes that are independent variables of this plantain chips: Saltiness, Sweetness, and Crunchiness on JAR scale. These independent variables were evaluated on the each of overall liking scores separately. After running the Penalty Analysis on each of the liking scores with different point scales, it was found out that at $\alpha = 0.05$ the result of the Penalty Analysis run on both 9-points scale and 7-points scale of overall liking score were the same but these two were different from the one run of 5-Point scale of the liking score. Here, Penalty analysis performs well on a liking score of 7 and 9-points of scale which is in line with the recommendation of a well-known computer software for this analysis.

Keywords: Penalty Analysis, JAR (Just About Right) scale, Rating Scales, mean drop

1.0. Introduction

1.1 Scale Measurements

Measurement form an integral part of any research and evaluation work in quantitative research studies in as much as the experimenter is concern about the magnitude of some phenomena of interest Harpe (2015). Measuring research response based on the magnitude cannot be assessed without the rating scale. Several research studies in psychology were undertaken to quantify the non-physical phenomena and to put numerical strength to them (Harpe, 2015). Part of these works is the research work done by Steven (1946) where he classified scale of measurement to Nominal, Ordinal, Interval, and Ratio scales, and his scaling method was to simply assigned the number to objects using some rules.

However, in the early part of 20th century, several researchers developed various scaling methods to measure the relative strength of the psychological phenomena such as attitude and mental ability (Singleton & Straits, 1999). Through these research works there is development of several methods of measuring attitude which resulted in noteworthy scaling methods including Thurstone Scales, Guttman Scales, and Likert Scales (Cohen *et al.*, 2011).

1.2 Rating Scales - Likert scales

Likert scales can be classified under the group of larger measures that are oftentimes referred to as summated or aggregated rating scales, since they are based on some underlying principle that phenomenon can be measured by aggregating an individual's ratings of his/her feelings, attitude, or perception related to a series of individual statements or items (Harpe, 2015).

In the original version of his work, Likert used five options (Five Points of Scale), which included a neutral option (Likert, 1932). The response format has since been expanded to varying numbers of response options including removing the neutral category (Smith & Albaum, 2005; Cohen *et al.,* 2011). Furthermore on his original article, Likert proposed an equidistance response number between the numbers in the response set. Similarly, the distances between the response anchors (e.g., "Strongly Agree" to "Agree") were equal (Likert, 1932). As a result of this, from a statistical stand point, this suggested that Likert Scales is an interval level of measurement.

1.3 Just-about-right (JAR) scale

As reported by Lawless & Heymann (1998), the just-about-right (JAR) scale measures the desirability of a specific attribute at the same time determine the optimum level of the product. This scale of quantitative measurement typically consist of five or seven points where the middle value is the labelled as "Just Right" but the two opposite ends of the scale is either "too little" or "too much" (Xiong & Meullenet, 2006).

This implies that this scale runs from too little to too much while the just about right is the mid-point value of this scale for a particular characteristics. If one end of a characteristic is labelled too little, the other end will be labelled too much. JAR scale and the hedonic scale have been used in several consumer product evaluation studies as the diagnostic tools in determining the optimal product formulation direction (Johnson & Vickers, 1987; Meullenet *et al.*, 2002).

1.4 Penalty Analysis

Penalty analysis is widely used in the market research industry and academic communities over the past decade (Xiong & Meullenet 2006). It is the also known as mean drop analysis and has been the tool used by the market researchers and product developers to gain the understanding of the product attributes that is most affecting the liking, purchase intent or any other product-related measures (Plaehn & Horne, 2008). Apart from the overall liking or purchase intent which are on Likert scale, the other product attributes used in penalty analysis are measured on a JAR scale.

This analysis provides the product developers and market research scientists with the important list of critical product characteristics that are most-penalizing product performance. It is a graphical technique to reveal the possible penalty paid by the product as a result of the reduction in the overall liking by not being ''just about right'' on a characteristic (Xiong & Meullenet, 2006).

However, it is pertinent to note here that the penalty analysis is not a regression-based method, ignores correlations among product characteristics and cannot be used to predict consumer overall liking from JAR data (Xiong & Meullenet, 2006). Furthermore, the mean drop estimated by penalty analysis for a specific attribute is not the estimate of the ''true'' mean drop on overall liking.

According to Pagès *et al.* (2014) the principle of Penalty Analysis are summarised in five points below:

1. A set of consumers is asked to evaluate several products using a battery of variables called JAR variables (just about right). Also, they are asked to express their overall liking of each product.

2. A JAR variable is a bipolar labelled attribute scale; it measures levels of a product attribute relatives to a respondent's theoretical ideal level. These scales have an anchored midpoint of ''just about right'' (Rothman & Parker, 2009).
3. For a product p, a first indicator is the frequency of each non-JAR category, for example, the percentage of consumers who score product p as too sweet.
4. To assess the influence of a non-JAR category m on the liking, the difference between the average liking score for the consumers who selected the category m and the average liking score for the consumers who selected the corresponding JAR category is calculated. This difference is called ''mean drop'' or ''penalty'' (associated with m).
5. These results are summarised by scatter plotting each non-JAR category with respect to its frequency (on the x-axis) and its penalty (on the y-axis).

Although Xiong & Meullenet (2006) in their work noted that penalty analysis is not a regression-based method. This may be due to the measurement scale of the JAR data and the liking score. But the way the two variables works together is somehow similar to regression analysis as penalty analysis assess the impact of the consumer rating of Overall liking of a product, relative to this product assessment on the sensory attributes on a JAR scale.

This was further supported by Plaehn & Horne (2008) in their work, as they showed that traditional penalty analysis also known as mean drop analysis is equivalent to Ordinary Least Square (OLS) regression model where product reference variable is the response variable and regressors derived from the given JAR variables transformed to dummy variables with the JAR category removed (and a column of 1s added).

$$y_i = \beta_0 + x\beta + \varepsilon \dots\dots\dots (0)$$

Let c be as above, y_i the reference variable score for respondent *i*, *i* = 1, 2, 3, . . ., I, μ_{JAR} the reference variable mean for those respondents giving the JAR response, and x^i the given JAR variable response for respondent *i* after transforming to dummy variables, then the above OLS model is given by

$$\beta_0 = \mu_{JAR}$$

$$y_i = \mu_{JAR} + \sum_{j \neq (c+1)/2} x_j^i \beta_j + \varepsilon_i \quad \dots\dots\dots (1)$$

where ε_i is the model error for respondent *i*. If, for the given product, μ is the reference variable grand mean, μ_j is the reference level mean for JAR level *j* and p_j is the associated sample size proportion, *j* = 1, 2, . . ., c, (number of levels in JAR scale), then Plaehn & Horne (2008) showed

$$\beta_j = \mu_j - \mu_{JAR}\ , \qquad j \neq \frac{(c+1)}{2}$$

$$\mu = \mu_{JAR} + \sum_{j \neq (c+1)/2} p_j \beta_j$$

$$\mu = \mu_{JAR} + \sum_{j \neq (c+1)/2} w_j \dots\dots\dots\dots\dots\dots \quad (2)$$

where w_j is the so-called weighted penalty associated with JAR category *j*, $w_{(c+1)/2} = 0$. The β_js are the (unweighted) penalties and are typically negative. Together the weights, w_j, sum to the difference

between the grand mean and the JAR mean (Plaehn, 2013). According to Plaehn (2013) where the equation two was well examined, the major problem associated with the weighted penalty is the penalty inflation. In this phenomenon, for each of the product and each of the JAR attribute, the weighted penalty sum to the different between grand mean and the mean for JAR ($\mu - \mu_{JAR}$). As a result of this, it is possible that as the number JAR variables increases, the sum of the penalties for the given product increases as well. This phenomenon is called the Penalty Inflation. As we are having the penalty inflation, the grand mean which reflect the individual penalty remain constant.

1.5 Research motivation

Different controversies associated with the Overall Linking Score on Likert scales and JAR Scale prompt the researcher to evaluate the effect of different levels of Overall Liking on a fixed level of JAR scale run on the Penalty analysis. In this research work, Likert scales at three different points of scales (5, 7 and 9 points of scales) will be examined with only 5-points Just-about-right scale in a Penalty Analysis.

2.0 Materials and Methods

The data used in this study was randomly collected from tasters who are regular consumers of plantain chips for a sensory evaluation test of a known brand of plantain chips. This is in a product test carried out on 124 randomly selected tasters. The randomly selected regular plantain chips consumers were given this product to evaluate. Each consumer gave his/her opinion on JAR scale (1 to 5) for three attributes (Saltiness, Sweetness, and Crunchiness), and overall liking score on experimented points of scales; 5-Points, 7-Points and 9-Points Likert scales.

Our goal is to identify if there exist changes in the result of this analysis with respect to various levels of rating scales used in overall liking, so that while identifying possible directions for the development of a new product, researcher will not be misled in the inference of the analysis due to the choice of the point of scale used for the overall liking score of the product.

At the end of this assessment by each consumer, score sheets were given to the tasters to rate the tested plantain chips on Saltiness, Sweetness, Crunchiness and the overall liking (at three different levels). The data collected from these respondents were analysed for Penalty Analysis through XLSTAT-MX.

3.0 Result and Discussion

The result of the Penalty on tables 3, 4 and 5, for the fixed sensory attributes against Overall liking at 5-Point scale, 7-Point scale and 9-point scale shows that the at 5% confident level, the results of the Penalty Analysis run based on 7-Point of scale and 9-point of scales are same but this was different from the result of same analysis run on overall liking that was on 5-point scale. It is possible that as the point of scale increases, the respondents (consumers of the tested plantain chips) were able to rate the tested product appropriately.

In addition to the tables of Penalty Analysis discussed above, the charts of scatter plot of the Mean Drop and the percentage of non-JAR as shown on charts 1, 2 and 3 shows the same result as the Penalty Analysis. Although the distribution of the mean drops across the charts (charts 2 and 3) showing similar behaviours that might not be totally be identical, but a critical look at these charts shows that there is almost similar distribution of mean drops for Penalty analysis that produced them.

Analysing charts for the Penalty on charts 4, 5 and 6, a similar result as above was found there. Here, charts 5 and 6 shows a similar distribution of penalty across the two Points of scales (7 and 9 scale levels) examined.

Let us recall that the experimental materials (plantain chips) administered to the tasters is a known and well acceptable brand in Nigerian Snacks markets, therefore, there is no way its overall liking will be penalised by any of its attributes by not been just right, which support the result of the penalty analysis on the product using the Overall Liking score rated on 7 and 9 Points of scale as shown in tables 4 and 5, and charts 2, 3 and charts 5 and 6 respectively.

But this is against the result of this analysis run with overall Liking score of 5-Points scale, which shows that there is significant penalty on the overall liking of the product due to the consumers assessment of the product for not been just right on sweetness and crunchiness which might not be true as the tested product is a well-known and accepted brand in the market.

4.0 Conclusion

The result of this investigation showed that the higher the points of scale of overall liking, the better the result of the penalty analysis. It further showed that Penalty analysis performs well on a liking score of 7 and 9 points of scale. This support the recommendation of the well-known computer software for this analysis XLSTAT, which recommend the analysis could be run on either 9 or 10 points scale (XLSTAT, 2016).

References:

Carifio, J & Perla, R. (2008) Resolving the 50-year debate around using and misusing Likert scales. *Med Educ.* 42(12), 1150–1152.

Cohen, L., Manion, L. & Morrison, K. (2011) *Research Methods in Education.* (Seventh Edition ed.), London, UK: Routledge.

Harpe, S. E. (2015) How to Analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*. 7(1), 836–850.

Jamieson, S. (2004) Likert scales: how to (ab)use them. *Medical Education*. 38(12), 1217–1218.

Johnson, J. R., & Vickers, Z. (1987). Avoiding the centering biases or range effect when determining an optimum level of sweetness in lemonade. *Journal of Sensory Studies*. 2(1), 283–292.

Lawless, H. T., & Heymann, H. H. (1998). *Sensory evaluation of food*. New York: Chapman & Hall.

Likert R. (1932) A technique for the measurement of attitudes. *Archive of Psychology*. 22(140), 5–55.

Meullenet, J.-F., Xiong, R., Monsoor, M., Bellman-Horner, T., Zivanovic, S., Dias, P., Fromm, H., & Liu, Z. (2002). Preference mapping of commercial toasted white corn tortilla chips. *Journal of Food Science*. 67(1), 1950–1957.

Pagès, J., Berthelo, S., Brossier, M., & Gourret, D. (2014) Statistical penalty analysis *Food Quality and Preference*. 32(1), 16–23

Plaehn, D. (2009). A variation on external preference mapping. *Food Quality and Preference* 20(1), 427–439.

Plaehn, D. (2013). What's The Real Penalty in Penalty Analysis. *Food Quality and Preference* 28(1), 456–469.

Plaehn, D., & Horne, J. (2008). A regression-based approach for testing significance of ''just-about-right'' variable penalties. *Food Quality and Preference* 19(1), 21–32.

Rothman, L., & Parker, M. J. (2009). *Just-about-right (JAR) scales: Design, usage, benefits and risks*. ASTM International.

Singleton, R. A. Jr, Straits, B. C (1999). *Approaches to Social Research*. (Third Edition ed.), New York, NY: Oxford University Press.

Smith, S. M. & Albaum, G. S. (2005) *Fundamentals of Marketing Research.* Thousand Oaks, CA: SAGE Publications.

Xiong, R., & Meullenet, J. F. (2006). A PLS dummy variable approach to assess the impact of jar attributes on liking. *Food Quality and Preference*. 17(1), 188–198.

XLSTAT (2016) Penalty Analysis in Excel Tutorials. Available At: https://help.xlstat.com/customer/en/portal/articles/2062259-penalty-analysis-in-excel-tutorial?b_id=9283 (Assess: 31/05/16)

Table 3: Penalty Analysis run on Sensory Attributes of Plantain Chips on 5 – Points JAR Scale and Overall Liking of the Product on 5 – Points Likert Scale

Penalty table:

Variable	Level	Frequencies	%	Sum(Q1a - Overall Likerbility)	Mean(Q1a - Overall Likerbility)	Mean drops	Standardized difference	p-value	Significant	Penalties	Standardized difference	p-value	Significant
Q2b - Saltiness	Not salty enough	27	21.95%	123.000	4.556	0.011	0.092	0.927	No	0.112	1.006	0.316	No
	JAR	90	73.17%	411.000	4.567								
	Too salty	6	4.88%	24.000	4.000	0.567							
Q3b - Sweetness	Not sweet enough	18	14.63%	78.000	4.333	0.260				0.260	2.218	0.028	Yes
	JAR	96	78.05%	441.000	4.594								
	Too sweet	9	7.32%	39.000	4.333	0.260							
Q4b - Crunchiness	Not crunchy enough	18	14.63%	78.000	4.333	0.260				0.260	2.218	0.028	Yes
	JAR	96	78.05%	441.000	4.594								
	Too crunchy	9	7.32%	39.000	4.333	0.260							

Table 4: Penalty Analysis run on Sensory Attributes of Plantain Chips on 5 – Points JAR Scale and Overall Liking of the Product on 7 – Points Likert Scale

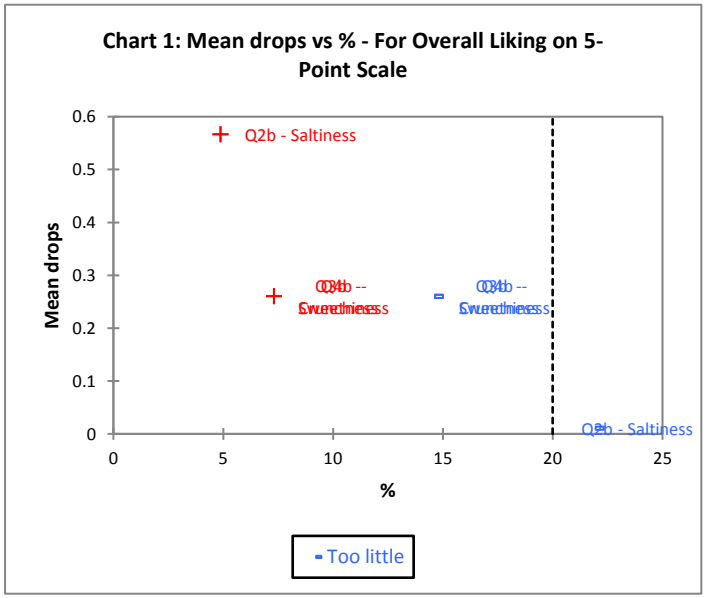
Penalty table: 7 Points Overall

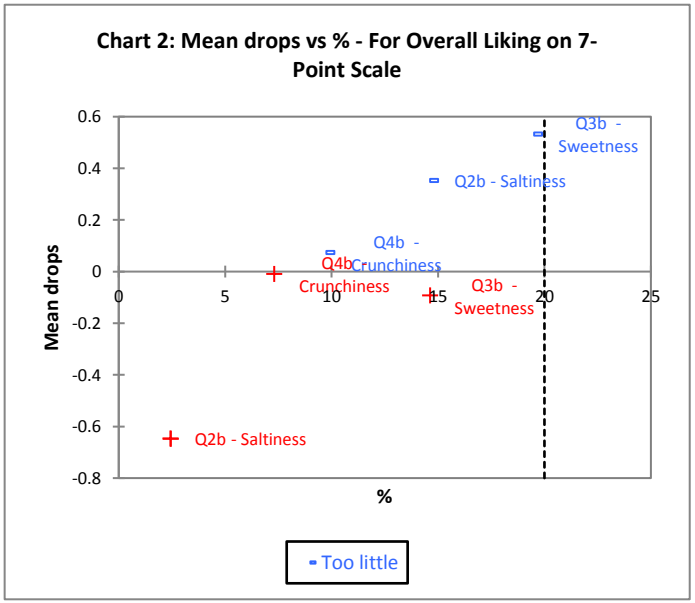
Variable	Level	Frequencies	%	Sum(Q1a - Overall Likerbility)	Mean(Q1a - Overall Likerbility)	Mean drops	Standardized difference	p-value	Significant	Penalties	Standardized difference	p-value	Significant
Q2b - Saltiness	Not salty enough	18	14.63%	108.000	6.000	0.353				0.210	0.740	0.461	No
	JAR	102	82.93%	648.000	6.353								
	Too salty	3	2.44%	21.000	7.000	-0.647							
Q3b - Sweetness	Not sweet enough	24	19.51%	141.000	5.875	0.532				0.265	1.178	0.241	No
	JAR	81	65.85%	519.000	6.407								
	Too sweet	18	14.63%	117.000	6.500	-0.093							
Q4b - Crunchiness	Not crunchy enough	12	9.76%	75.000	6.250	0.074				0.038	0.133	0.894	No
	JAR	102	82.93%	645.000	6.324								
	Too crunchy	9	7.32%	57.000	6.333	-0.010							

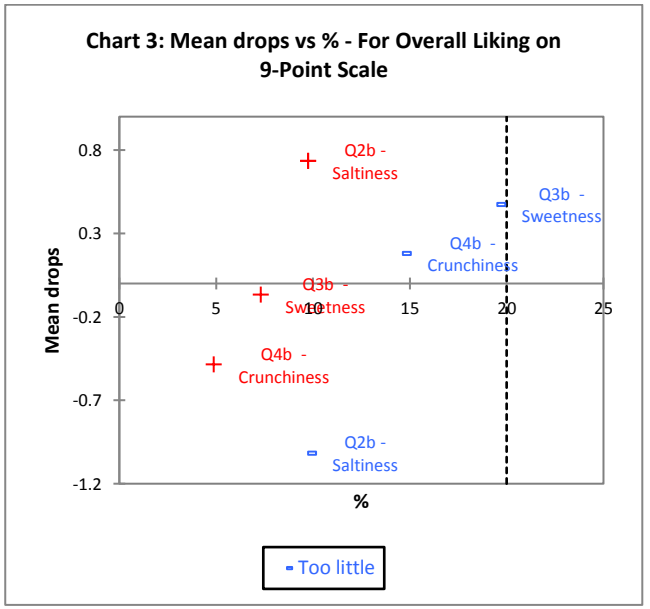
Table 5: Penalty Analysis run on Sensory Attributes of Plantain Chips on 5 – Points JAR Scale and Overall Liking of the Product on 9 – Points Likert Scale

Penalty table:

Variable	Level	Frequencies	%	Sum(Q1a - Overall Likerbility)	Mean(Q1a - Overall Likerbility)	Mean drops	Standardized difference	p-value	Significant	Penalties	Standardized difference	p-value	Significant
Q2b - Saltiness	Not salty enough	12	9.76%	102.000	8.500	-1.015							
	JAR	99	80.49%	741.000	7.485					-0.140	-0.618	0.538	No
	Too salty	12	9.76%	81.000	6.750	0.735							
Q3b - Sweetness	Not sweet enough	24	19.51%	171.000	7.125	0.475							
	JAR	90	73.17%	684.000	7.600					0.327	1.627	0.106	No
	Too sweet	9	7.32%	69.000	7.667	-0.067							
Q4b - Crunchiness	Not crunchy enough	18	14.63%	132.000	7.333	0.182							
	JAR	99	80.49%	744.000	7.515					0.015	0.067	0.947	No
	Too crunchy	6	4.88%	48.000	8.000	-0.485							







Mean drops vs %:

Variable	Level	%	Mean drops
Q2b - Saltiness	Not salty enough	21.951	0.011
	Too salty	4.878	0.567
Q3b - Sweetness	Not sweet enough	14.634	0.260
	Too sweet	7.317	0.260
Q4b - Crunchiness	Not crunchy enough	14.634	0.260
	Too crunchy	7.317	0.260

For Overall Liking on 5-Point Scale

Mean drops vs %:

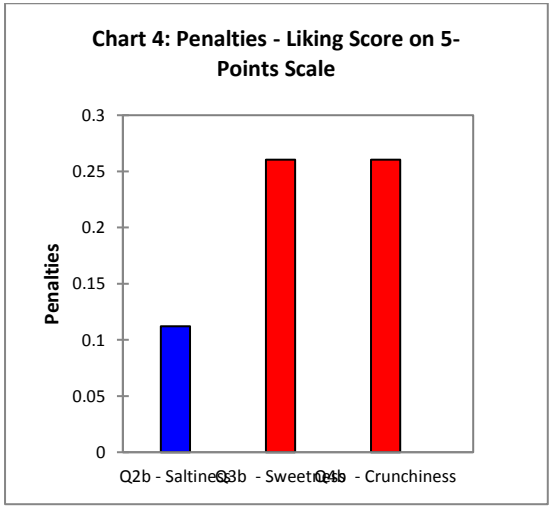
Variable	Level	%	Mean drops
Q2b - Saltiness	Not salty enough	14.634	0.353
	Too salty	2.439	-0.647
Q3b - Sweetness	Not sweet enough	19.512	0.532
	Too sweet	14.634	-0.093
Q4b - Crunchiness	Not crunchy enough	9.756	0.074
	Too crunchy	7.317	-0.010

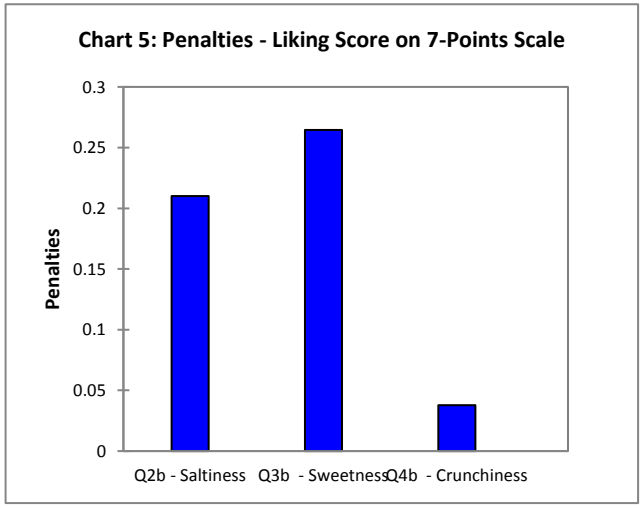
For Overall Liking on 7-Point Scale

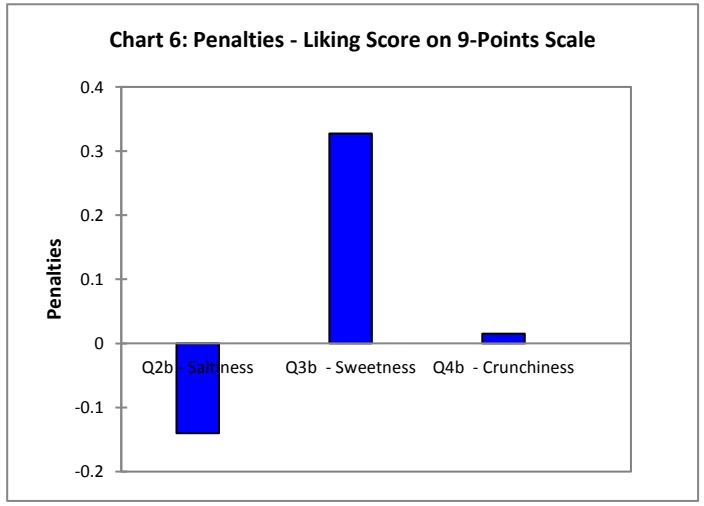
Mean drops vs %:

Variable	Level	%	Mean drops
Q2b - Saltiness	Not salty enough	9.756	-1.015
	Too salty	9.756	0.735
Q3b - Sweetness	Not sweet enough	19.512	0.475
	Too sweet	7.317	-0.067
Q4b - Crunchiness	Not crunchy enough	14.634	0.182
	Too crunchy	4.878	-0.485

For Overall Liking on 9-Point Scale







Note that Bars with Red Colour shows a Significant Penalty while those on Blue Colours do not have Significant Penalty at $P \leq 0.05$