UNIVERSIDAD SAN FRANCISCO DE QUITO

COLEGIO DE CIENCIAS E INGENIERÍA

A CROSS-CULTURAL STUDY COMPARING THE USE OF THREE 9-POINT HEDONIC SCALES BETWEEN ECUADORIANS AND AMERICANS USING A RANK-RATING PROTOCOL

Javier Preslav Amores Burbano

Michael Koziol, DPhil., Director de Tesis
Mario Caviedes, Ph.D., Codirector de Tesis

Tesis de grado presentada como requisito
para la obtención del título de Ingeniero en Alimentos

Quito, febrero de 2014

UNIVERSIDAD SAN FRANCISCO DE QUITO

COLEGIO DE CIENCIAS E INGENIERÍA

HOJA DE APROBACION DE TESIS

A CROSS-CULTURAL STUDY COMPARING THE USE OF THREE 9-POINT HEDONIC SCALES BETWEEN ECUADORIANS AND AMERICANS USING A RANK-RATING PROTOCOL

Javier Preslav Amores Burbano

Michael Koziol, DPhil.
Director de Tesis _____

Mario Caviedes, Ph.D.
Codirector de Tesis _____

Javier Garrido, MSc.
Coordinador de Ingeniería de Alimentos y
Miembro del Comité de Tesis _____

Lucía de los Ángeles Ramírez, Ph.D.
Miembro del Comité de Tesis _____

Stalin Santacruz, Ph.D.
Miembro del Comité de Tesis _____

Ximena M. Córdova, Ph.D.
Decana de la Escuela de Ingeniería _____

Quito, febrero de 2014

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído la Política de Propiedad Intelectual de la Universidad San Francisco de Quito y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo de investigación quedan sujetos a lo dispuesto en la Política.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo de investigación en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma: _____

Nombre: Javier Preslav Amores Burbano

C. I.: 171506510-6

Fecha: Quito, febrero de 2014

HOJA DE APROBACIÓN DE PRESENTACIÓN DE TESIS

En calidad de Decana de la Escuela de Ingeniería, autorizo la presentación de la tesis en inglés y el texto presentado con uno y medio espacios interlineales.

_____

Ximena M. Córdova, Ph.D.
Decana de la Escuela de Ingeniería

DEDICATION

I dedicate this project to
my father, my mother,
my siblings and my family
in return for their efforts
and constant support.

If anyone deserves
some dedicatory lines,
they come to my mind
for walking next to me
throughout this intricate
yet rewarding path.

AKNOWLEDGMENTS

The author would like to thank Michael Koziol and Mario Caviedes, codirectors of this thesis, for overseeing this project as well as for their advices and reviews of the manuscript, Yamila Álvarez for the initial push that has resulted in this project, the future food engineers Carlos Barreiro, Natalia Castro, Sofía Maldonado, Gino Merino, Leandro Monar, Javier Morán, Daniela Naranjo, Mireya Padilla, Cristian Parra, Valeria Puga, Vanessa Reyes, Luís Fernando Silva, Emilia Torres, Andrés Torrico and María Paz Villalba for their invaluable help during data collection and Claudia Gutierrez for her help in the recruitment of American students. Finally, the author also thanks Universidad San Francisco de Quito for allowing the use of its Sensory Evaluation laboratory throughout the whole experiment.

RESUMEN

En un estudio intercultural, noventa y seis consumidores, 48 ecuatorianos y 48 estadounidenses, reclutados en las instalaciones de la Universidad San Francisco de Quito, probaron cuatro bebidas hidratantes disponibles en el mercado local. Las cuatro muestras fueron categorizadas utilizando escalas hedónicas de 9 puntos de 'Categorías y números', 'Solo categorías' y 'Solo números' siguiendo un protocolo Rank-Rating. Los datos obtenidos a partir del ordenamiento y categorización de las muestras fueron analizados mediante ANOVA/pruebas post-hoc, prueba t de Student, δ de Thurstone e Índice R. No se encontró diferencia estadística entre consumidores ecuatorianos y estadounidenses ($p \geq 0,05$), pero este resultado solo es aplicable a este experimento en particular, pues este puede variar dependiendo de las muestras utilizadas. Se determinó el poder discriminativo de los cinco análisis estadísticos empleados, mostrando el Índice R una ligera ventaja sobre las otras cinco pruebas. Finalmente, se encontraron diferencias en el uso de las tres escalas hedónicas, especialmente al usar la escala 'Solo números', pero estas diferencias no resultaron estadísticamente significativas ($p \geq 0,05$).

Palabras clave: Análisis de varianza; Estudio intercultural; Rank-Rating; Índice R; δ de Thurstone; escala hedónica de 9 puntos

ABSTRACT

In a cross-cultural study, ninety-six consumers, 48 Ecuadorian and 48 American, recruited from the Universidad San Francisco de Quito campus, tasted four sport drinks available in the local market. The four samples were rated using 'Words and numbers', 'Words only' and 'Numbers only' 9-point hedonic scales and following a Rank-Rating protocol. Data obtained from the ranking and rating of the samples were analyzed using ANOVA/post-hoc tests, Student's t-test, Thurstonian δ and R-Index. No statistical difference was found between Ecuadorian and American consumers ($p \geq 0.05$). However, these results are only applicable to this particular experiment, since these may vary depending on the samples used. The discriminating power of the five statistical analyses was determined, showing a slight advantage of the R-Index over the other tests. Lastly, differences in the use of the three hedonic scales were found, especially when using the 'Numbers only' scale, but such differences were not statistically significant ($p \geq 0.05$).

Keywords: Analysis of variance; Cross-cultural; Rank-Rating; R-Index; Thurstonian δ; 9-point hedonic scale

TABLE OF CONTENTS

INTRODUCTION

Consumers modify their behavior according to their experiences and the sensations provided by their surroundings. Such modifications may seem subjective and unquantifiable, but can actually be evaluated in accordance to the intensity of the stimuli that generated them. With the purpose of quantifying how a stimulus is perceived, scales are used. These scales allow the generation of a psychophysical response that reflects the intensity in which each person perceives a particular stimulus. Psychophysical models indicate that when the stimulus grows in intensity, the sensation of it grows in a similar way (Lawless & Heymann, 2010).

In order to measure the degree of liking, the most commonly used scale is the 9-point hedonic scale. First published by Peryam and Pilgrim (1957), the 9-point hedonic scale was originally introduced in 1949 as a menu planner for American soldiers (O'Mahony *et al.*, 2011) and has subsequently been used in both academic and industrial consumer research throughout the world (Yeh *et al.*, 1998). The 9-point hedonic scale comprises nine categories ranging from 'Like extremely' to 'Dislike extremely'. The subsequent statistical analysis is performed by giving each category a numerical value, transforming 'Like extremely' to 9 and 'Dislike extremely' to 1 (Stone *et al.*, 2012). Regarding data analysis, besides the usual ANOVA/post-hoc tests used to determine differences between mean hedonic scores, all data derived from the present study will also be analyzed using Student's t-test, Thurstonian δ and R-Index. Being parametric analyzes, ANOVA/post-hoc and Student's t-tests make the assumption that data obtained follows a normal distribution, unlike hedonic data which is most likely skewed or binomial, hence they are only useful to provide approximate differences between products (O'Mahony, 1986).

Signal Detection Theory, also called Thurstonian modeling, is based on the variability in the perception of food. Different environmental changes, such as fluctuations of the nervous system affect the way the intensity of a food is perceived after repeated tasting. Traditional parametric statistics do not account these variations, but Signal Detection Theory does (Lee & van Hout, 2009). Bi *et al.* (1997) define Thurstonian δ (the population parameter is referred as δ while d' is its estimate) as the difference between the means of intensity distributions for two products measured in perceptual standard deviations. A smaller δ means closer distributions and more similar products, a larger δ

means distributions are further apart resulting in seemingly more different products (ASTM International, 2003). However, it should be noted that for the estimation of d' it is assumed that both distributions are normal and have equal variances (Bi *et al.*, 1997; Kim *et al.*, 1998).

The R-Index, developed by Brown (1974), has become an important tool in sensory analysis and consumer research. The R-Index is an estimator of the area under the Receiver Operating Characteristic (ROC) curve used in Signal Detection Theory (Bi, 2006). When comparing two products, for example, in terms of preference, purchase intent or sensory intensity, the area under the ROC curve shows the probability of one of the products to be chosen over the other. Also, its relationship with signal detection and Thurstonian δ allows the R-Index to be used as a powerful nonparametric test statistic and a good measure of sensory difference (Bi & O'Mahony, 2007), but, unlike Thurstonian δ, the R-Index analysis makes no distribution assumption (H. Park *et al.*, 2007).

Occasionally, consumers using the 9-point hedonic scale and following a serial monadic protocol like two samples equally but prefer one sample over the other. In the frustration caused by their inability of showing which sample they prefer, consumers may rate the samples incorrectly, leading to the determination of false preferences (Villegas-Ruiz *et al.*, 2008). Subsequently, this derives in an incapability of the researcher to draw valid conclusions from the resulting data. Another problem that might occur during the administration of the test is memory loss; this means that consumers may give a sample they like better a lesser score because they forgot the intensities of the stimuli presented before it (Kim & O'Mahony, 1998). As an alternative, in order to minimize memory errors, the judge should merely retaste the samples, thereby reminding himself of how intense the stimuli are and then rank accordingly (Cordonnier & Delwiche, 2008). On the other hand, to address the inability of showing preferences, the judge is also presented a printed scale in which he can place the samples and move them up and down until he is satisfied with his answer (Koo *et al.*, 2002), and in case of a tie among two or more samples, the consumer is asked which samples he prefers over the other. This alternative to the serial monadic protocol is known as Rank-Rating or Positional Relative Rating protocol.

Because this is a cross-cultural study, the first goal is to determine if the cultural differences between Ecuadorian and American consumers affect the responses given when

using different hedonic scales. These cultural effects have been evaluated by Yao *et al.* (2003) and Yeh *et al.* (1998) and have shown that consumers from the Pacific Rim, specifically Japanese and Koreans, avoid the use of extreme and negative categories, which was explained to be caused by translation effects (Curia *et al.*, 2001) and a possible cultural politeness. Taking these studies into account, the hypothesis of the present study is that the cultural differences between Ecuadorian and Americans would result in a different distribution of scores.

The second goal of the study is to compare the discriminating power between five statistical analyses. It was first hypothesized by O'Mahony *et al.* (2004) that an R-Index analysis would elicit fewer significant differences than the usual ANOVA with multiple comparisons. A posterior study by H. Park *et al.* (2007) determined that both analyses were comparable, but with a slight and nonsignificant advantage for ANOVA. Similar to both studies, the present study compared both statistical analyses, but as it was previously mentioned it also introduced a comparison with Tukey test, Student's t-test and Thurstonian δ. It should be mentioned that, while a greater number of significant differences seems desirable when comparing two products using hedonic scales, a statistical test that yields fewer significant differences is by no means a less acceptable protocol. This occurs because hedonic rating is not a direct measure of preference, which means that a consumer who rates a product with a higher hedonic score may not necessarily prefer that product over another. However, unlike studies like those of Rosas-Nexticapa *et al.* (2005) product preference is the matter of different testing and it is not addressed in the present study.

The 9-point hedonic scale was selected for this experiment over other simpler scales, like the 5-point or 7-point hedonic scales for two main reasons. It has been hypothesized by Jeon *et al.* (2004) and J.-Y. Park *et al.* (2007) that, as the number of categories decreases and more samples are used, a tendency of giving stimuli of different intensities a score that represents the same intensities or a reversal of intensities appears ('different-stimulus' scaling errors). Also, in order to compare the present study to the previously mentioned cross-cultural studies, three 9-point hedonic scales were needed: the first scale, labeled as 'Words and numbers', showed all verbal categories as well as their corresponding numerical value; the 'Words only' scale showed only the verbal categories; and the 'Numbers only' scale showed only the numerical values of each category. The

studies of Nicolas *et al.* (2010) and O'Mahony *et al.* (2011) have shown that consumers respond differently when using a 'Words and numbers' or 'Words only' scales than when using a 'Numbers only' scale, so the numerical data obtained are not interchangeable. Therefore, the final goal of this study is to obtain data that supports the evidence that a 'Numbers only' hedonic scale elicits different results than a verbal hedonic scale due to the different cognitive strategies being applied by consumers using them.

MATERIALS AND METHODS

Participants

Ninety-six consumers of sport drinks were tested. Forty-eight consumers were Ecuadorian (23 male, 25 female, age range 17-26 years) while forty-eight consumers were American (15 male, 33 female, age range 19-23 years). All consumers were students recruited from the campus of the Universidad San Francisco de Quito. American consumers were exchange program students from different universities in the United States.

Stimuli

Four stimuli, consisting of four sport drinks available in the local market, were presented to each consumer during each session. Thus each consumer evaluated a total of twelve stimuli throughout the whole experiment. In order to avoid a stimulus error, caused by the consumer inferring about what the samples taste like based on their appearance, only clear-colored drinks were used. Also, since one of the objectives of the study is to evaluate the use of hedonic scales regardless of the stimuli presented, brand names are not shown and are hence identified in this study as "A", "B", "C" and "D". Stimuli were presented as 30 mL aliquots in 120 mL polystyrene cups at constant room temperature (20 ± 3 °C).

Distractor stimuli

In order to further minimize a possible memory effect affecting consumer response throughout each session, distractor stimuli were also presented. These distractors consisted of toothbrushes used for visual inspection only. Because these were mere distractors, their data was not analyzed. Distractors were assessed using the same three hedonic scales used with the target stimuli and followed the same procedure.

Scales

Since a Rank-Rating protocol was being used, all consumers were presented three 9-point hedonic scales printed on a cardboard strip; depending on whether the consumer was Ecuadorian or American, the scales and instructions were given in Spanish or English respectively. Verbal English scales were based on the original 9-point hedonic scale (Lawless & Heymann, 2010; Stone *et al.*, 2012) while the Spanish verbal scales were liberal translations in accordance to everyday language in Ecuador. All scales were 100 cm long and 20 cm wide and were divided into nine equal categories. For the 'Words and numbers' and 'Words only' scales, categories were labeled in capital letters (letters and numbers were 2 cm high). For the 'Numbers only' scale, categories were labeled from 1 to 9 (numbers were 8 cm high). Two additional cardboard strips with arrows printed on them were used alongside all scales to indicate the direction of the categories that correspond to the least and most liking. All writing was in black on a white background.

Procedure

Consumers were seated at an individual table inside a sensory evaluation testing room with the experimenter behind to provide help, if needed, and to give necessary instructions. The experimenter took care to be inconspicuous and not to 'crowd' the consumers. After consumers filled in their personal data, instructions were given and the experiment begun. Consumers were presented with four polystyrene cups coded using a three-number code and filled with one of the four beverages, which they could sample *ad lib*. They tasted the beverages and then were asked to rank them according to their degree of liking. They responded by placing the four cups in a row, with the beverages they liked the most on their right and the least to their left. When they had completed this task, the experimenter checked to make sure that the cups were ranked in the appropriate direction. After ranking, consumers were asked to rate their samples by placing them on the cardboard strips in the category that fitted their degree of liking. In case two or more samples were tied within a category, consumers were asked to indicate which sample they preferred over the other.

Special instructions were given when consumers used the 'Numbers only' scale. In this case, after ranking, consumers were asked if they liked, neither liked nor disliked or disliked each sample. Then they were shown how to indicate the degree of difference between the rankings they had just performed by visually varying the distance between the

first and second ranked stimuli. After this, consumers responded by placing the cups in appropriate positions along the scale. After all beverages were ranked and rated, all distractor stimuli were presented. These were ranked and rated following the same procedure used for the beverages.

Each consumer participated in three sessions: no fewer than five days elapsed between sessions to minimize a possible memory effect in the responses given. During every session each judge used only one of the three scales. The presentation order of the scales was counterbalanced to minimize any order effect. Every session lasted from 5 to 10 minutes.

RESULTS AND DISCUSSION

The numerical scores for the 'Numbers only' and those derived from the 'Words and numbers' and 'Words only' scales were noted for each consumer. For the 'Words only' scale, the numerical values 1-9 were assigned in the usual manner ('Dislike extremely', '1'; 'Like extremely', '9'). When computing ANOVA for multiple factors to analyze the results obtained by rating of the samples (MINITAB® Statistical Software., version 14.20, release 2005; O'Mahony, 1986), no significant differences were found between Ecuadorians and Americans ($P \geq 0.05$). However, significant differences were shown between samples ($P \leq 0.05$), therefore all possible pairwise comparisons between them were made (Tables 1 and 2) and significant differences between mean hedonic scores were determined by using post-hoc tests, Student's t-test and Thurstonian δ (ASTM International, 2003; Bi *et al.*, 1997). For each set of stimuli the rank order of preference was determined and all possible R-Index values (R_{MAT}) were computed (Lee & van Hout, 2009; O'Mahony, 1992); significance was then determined using the tables of Bi and O'Mahony (2007).

Table 3 shows the number of nonsignificant differences between pairs of samples (total possible 18); the greater the number of nonsignificant differences, the less discriminating the analysis. As in the experiments of O'Mahony *et al.* (2004) and H. Park *et al.* (2007), where comparisons between Least Significant Difference (LSD) tests and R-Index values were made, it was suggested that the difference between both analyses was small and not significant, which is also applicable in the present study. In this study, when conducting a binomial test to compare between statistical analyses, it can be seen that significant differences between products indicated by LSD were also indicated by the

R-index, except for one case shown in the Ecuadorian 'Numbers only' scale (Tables 1 and 3) and one case in the American 'Words only' scale (Tables 2 and 3), where R-Index showed significance while LSD did not, implying that both tests behave in a similar way (18 versus 20 from 38; binomial P=0.87).

The same is not applicable to Tukey and R-Index. When comparing the total number of significant differences in both American and Ecuadorian responses, Tukey showed nonsignificant differences in seven cases that the R-Index did not (Table 3), indicating that the former is less discriminating (25 versus 18 from 43; binomial P=0.36), yet the difference between both analyzes is still not significant. Student's t-tests behaved similar to LSD tests (19 versus 20 from 39; binomial P=1.00) resulting in every significant difference shown by LSD to be also shown by t-tests, except one case in the American 'Words only' scale (Tables 2 and 3), where t-test showed significance while LSD did not. Finally, a comparison between both signal detection related tests, R-Index and d', was made. Here, while d' showed that it is less discriminating than R-Index (22 versus 18 from 40; binomial P=0.64), statistically both tests behaved in a similar way. This however, should be approached with caution, because although a study by Lee & van Hout (2009) mentions the R-Index as being interchangeable with d', studies made by Rousseau (2006; 2011), indicate that, while both indices are useful to estimate sensory differences between samples, the R-Index does not account decision rules used by subjects, making both indices not interchangeable.

When comparing the number of times each scale was used differently depending on the sample analyzed, 73/96 to 80/96 consumers gave a different set of ratings on the 'Words and numbers' and 'Numbers only' scales while 75/96 to 81/96 gave a different set of ratings on the 'Words only' and 'Numbers only' scales (Table 4). However, it is also worth noting that 69/96 to 78/96 consumers gave a different set of ratings on the 'Words and numbers' and 'Words only' which may imply that a possible mixed cognitive strategy is involved when comparing this particular set of scales.

As in the experiment of Nicolas *et al.* (2010), these results suggest that consumers were using different cognitive strategies when dealing with numbers or words; the same experiment also suggests that words in the 'Words and numbers' scale might be more notorious than numbers.

Table 1: Comparison of mean hedonic scores and determination of t, d' and R-Index values in data obtained through the rating and ranking of four beverages by Ecuadorian consumers.

'Words and numbers' (n=48)								
Beverages	D		B		A		C	
Mean hedonic scores	$6.52^{a}{}_{a}$		$6.25^{ab}{}_{a}$		$5.71^{b}{}_{ab}$		$4.92^{c}{}_{b}$	
		D-B		B-A		A-C		
t-test		0.636		1.496		2.193*		
d'		0.351		0.356		0.587*		
R-Index		54.58		58.31		64.28*		
			D-A		B-C			
t-test			2.173*		3.652*			
d'			0.637*		0.881*			
R-Index			64.30*		69.18*			
				D-C				
t-test				4.525*				
d'				0.939*				
R-Index				74.87*				
'Words only' (n=48)								
Beverages	B		D		A		C	
Mean hedonic scores	$6.75^{a}{}_{a}$		$6.13^{ab}{}_{a}$		$6.00^{b}{}_{a}$		$4.56^{c}{}_{b}$	
		B-D		D-A		A-C		
t-test		1.661		0.372		3.570*		
d'		0.435		0.050		0.507		
R-Index		56.81		57.96		75.17*		
			B-A		D-C			
t-test			2.249*		3.839*			
d'			0.569*		0.801*			
R-Index			65.65*		77.60*			
				B-C				
t-test				5.966*				
d'				1.183*				
R-Index				81.42*				
'Numbers only' (n=48)								
Beverages	D		B		A		C	
Mean hedonic scores	$6.19^{a}{}_{a}$		$6.00^{a}{}_{a}$		$5.69^{a}{}_{a}$		$4.23^{b}{}_{b}$	
		D-B		B-A		A-C		
t-test		0.418		0.738		3.098*		
d'		0.302		0.000		0.749*		
R-Index		55.01		58.01		67.19*		
			D-A		B-C			
t-test			0.988		3.945*			
d'			0.302		1.131*			
R-Index			61.26*		76.17*			
				D-C				
t-test				3.952*				
d'				1.131*				
R-Index				77.47*				

a: Means with different superscripts (LSD) are significantly different (P≤0.05). Means with different subscripts (Tukey) are significantly different (P≤0.05).

*: Indicates that t (P<0.05), d' and R-Index values show significant difference for the pair of beverages.

Table 2: Comparison of mean hedonic scores and determination of t, d' and R-Index values in data obtained through the rating and ranking of four beverages by American consumers.

'Words and numbers' (n=48)								
Beverages	B		A		D		C	
Mean hedonic scores	$6.42^{a}{}_{a}$		$5.79^{ab}{}_{ab}$		$5.48^{bc}{}_{bc}$		$4.81^{c}{}_{c}$	
		B-A		A-D		D-C		
t-test		1.789		0.837		1.721		
d'		0.440		0.187		0.353		
R-Index		60.42		52.21		59.55		
			B-D		A-C			
t-test			2.325*		2.726*			
d'			0.253		0.353			
R-Index			61.50*		62.37*			
				B-C				
t-test				3.975*				
d'				0.755*				
R-Index				69.75*				
'Words only' (n=48)								
Beverages	A		B		D		C	
Mean hedonic scores	$5.92^{a}{}_{a}$		$5.63^{a}{}_{a}$		$5.48^{a}{}_{a}$		$5.21^{a}{}_{a}$	
		A-B		B-D		D-C		
t-test		0.856		0.342		0.608		
d'		0.051		0.132		0.455		
R-Index		56.81		50.95		54.99		
			A-D		B-C			
t-test			1.055		0.964			
d'			0.183		0.639*			
R-Index			58.20		55.86			
				A-C				
t-test				2.095*				
d'				0.585*				
R-Index				64.15*				
'Numbers only' (n=48)								
Beverages	A		B		D		C	
Mean hedonic scores	$5.75^{a}{}_{a}$		$5.71^{a}{}_{a}$		$5.38^{ab}{}_{a}$		$4.77^{b}{}_{a}$	
		A-B		B-D		D-C		
t-test		0.112		0.776		1.336		
d'		0.351		0.490		0.335		
R-Index		53.91		53.30		56.08		
			A-D		B-C			
t-test			0.819		2.140*			
d'			0.100		0.825*			
R-Index			56.94		61.02*			
				A-C				
t-test				2.225*				
d'				0.435				
R-Index				64.15*				

a: Means with different superscripts (LSD) are significantly different ($P \leq 0.05$). Means with different subscripts (Tukey) are significantly different ($P \leq 0.05$).

*: Indicates that t ($P < 0.05$), d' and R-Index values show significant difference for the pair of beverages.

Table 3: Number of nonsignificant differences between pairs of beverages using five statistical analyses.

Consumers	Statistical computation	'Words and numbers'	'Words only'	'Numbers only'	Total
Ecuadorians	LSD	2	2	3	7
	Tukey	4	3	3	10
	t-test	2	2	3	7
	d'	2	3	3	8
	R-Index	2	2	2	6
Americans	LSD	3	6	4	13
	Tukey	3	6	6	15
	t-test	3	5	4	12
	d'	5	4	5	14
	R-Index	3	5	4	12

Table 4: Number of times each pair of scales was used differently when rating four beverages.

Beverages	'Words and numbers' vs. 'Words only'	'Words and numbers' vs. 'Numbers only'	'Words only' vs. 'Numbers only'
A	70	79	81
B	69	77	75
C	78	80	81
D	74	73	77

Table 5: Number of times each scale was used differently when rating four beverages.

Beverages	All scales used differently	'Words and numbers' used differently	'Words only' used differently	'Numbers only' used differently	All scales used identically
A	44	12	14	23	3
B	41	15	13	21	6
C	53	12	13	15	3
D	42	14	18	17	5

When contrasting these results with those obtained by Nicolas *et al.* (2010) and O'Mahony *et al.* (2011), where 79 to 100% and 97.3% of consumers respectively, gave different ratings on the 'Words only' and 'Numbers only' scales, a similar pattern can be also observed here, but the complications of using different samples and a different experimental design in the present study makes these results not comparable. The results shown in Table 5 show the number of times consumers rated samples differently depending on the scale used. It can be seen that the majority of consumers (41/96 to 53/96) gave different scores to each sample on all three scales and very few consumers (3/96 to 6/96) gave at least one sample three identical scores.

Table 6: Mean hedonic ratings for the 'Numbers only', 'Words only' and 'Words and numbers' scales.

Consumers		'Words and numbers'	'Words only'	'Numbers only'	Total mean hedonic scores
Ecuadorians	Mean	5.849	5.859	5.526	5.745
(n=48)	St.Dev	1.834	1.927	2.260	2.018
	Min-max	1-9	1-9	1-9	1-9
Americans	Mean	5.625	5.557	5.401	5.528
(n=48)	St.Dev	1.863	1.943	2.070	1.959
	Min-max	1-9	1-9	1-9	1-9
Total (n=96)		5.737	5.708	5.464	5.636

However, not a single consumer gave all four samples identical scores throughout the whole experiment. This shows that, besides the different cognitive strategies used by consumers during rating, a memory effect is also taking place and consumers might be simply remembering or forgetting their responses between sessions (Nicolas *et al.*, 2010). Finally, further comparisons between scales are shown in Table 6, where it can be seen that the mean hedonic ratings for the 'Numbers only' scale are lower than the mean hedonic ratings for the other two scales when used by both Ecuadorians and Americans, but these differences are not statistically significant ($P \geq 0.05$).

CONCLUSIONS

When using ANOVA for multiple factors, no statistical differences were found between Ecuadorians and Americans ($P \geq 0.05$). However, this should be addressed carefully because in this study only sport drinks were evaluated and results may vary depending on the sample used. Tukey was the least discriminating test followed by d', while the most discriminating test was the R-Index, and unlike the analysis made by H. Park *et al.* (2007), which showed a small advantage for ANOVA/LSD test when compared to the R-Index, in this study the slight advantage went for the R-Index, yet the differences were still not significant. As expected, a greater number of consumers rated samples differently when using the 'Numbers only' scale than when using the 'Words and numbers' or 'Words only' scales and although in past studies the 'Numbers only' scale has proven to yield statistically different results than the 'Words and numbers' and 'Words only' scales, data obtained in this study shows no statistical difference between scales, making these results inconclusive. This can be corrected in future studies by increasing the number of consumers or using a wider array of samples. Also, the majority of consumers used the three scales differently throughout the experiment indicating that, despite using a

Rank-Rating protocol, memory effects are a strong component of the variability in the perception of stimuli. To summarize, the results obtained in this study do not have the intention to render a protocol or analysis inadequate, but rather to show how the mind of consumers is complex and adapts according to its surroundings and that the tools used to analyze these changes are better or worse suited depending on the circumstances.

REFERENCES

ASTM International. (2003). *ASTM Standard E 2262-03, "Standard Practice for Estimating Thurstonian Discriminal Distances"* (Vols. 10.1520/E2262-03R09). West Conshohocken, PA: www.astm.org. Retrieved from www.astm.org

Bi, J. (2006). Statistical Analyses for R-Index. *Journal of Sensory Studies, 21*, 584-600.

Bi, J., & O'Mahony, M. (2007). Updated and extended table for testing the significance of the R-Index. *Journal of Sensory Studies, 22*, 713-720.

Bi, J., Ennis, D. M., & O'Mahony, M. (1997). How to estimate and use the variance of d' from difference tests. *Journal of Sensory Studies, 12*, 87-104.

Brown, J. (1974). Recognition assessed by rating and ranking. *British Journal of Psychology, 65*, 13-22.

Cordonnier, S. M., & Delwiche, J. F. (2008). An alternative method for assesing liking: Positional Relative Rating versus the 9-point hedonic scale. *Journal of Sensory Studies, 23*, 284-292.

Curia, A. V., Hough, G., Martínez, M. C., & Margalef, M. I. (2001). How Argentine consumers understand the Spanish translation of the 9-point hedonic scale. *Food Quality and Preference, 12*, 217-221.

Jeon, S.-Y., O'Mahony, M., & Kim, K.-O. (2004). A comparison of category and line scales under various experimental protocols. *Journal of Sensory Studies, 19*, 49-66.

Kim, K.-O., & O'Mahony, M. (1998). A new aproach to category scales of intensity I: Traditional versus Rank-Rating. *Journal of Sensory Studies, 13*, 241-249.

Kim, K.-O., Ennis, D. M., & O'Mahony, M. (1998). A new approach to category scales of intensity II: use of d' values. *Journal f Sensory Studies, 13*, 251-267.

Koo, T.-Y., Kim, K.-O., & O'Mahony, M. (2002). Effects on forgetting on performance on various intensity scaling protocols: Magnitude Estimation and Labeled Magnitude Scale (Green Scale). *Journal of Sensory Studies, 17*, 177-192.

Lawless, H. T., & Heymann, H. (2010). *Sensory Evaluation of Food: Principles and Practices.* New York: Springer.

Lee, H.-S., & van Hout, D. (2009). Quantification of Sensory and Food Quality: The R-Index Analysis. *Journal of Food Science, 74*, 57-64.

Nicolas, L., Marquilly, C., & O'Mahony, M. (2010). The 9-point hedonic scale: Are words and numbers compatible? *Food Quality and Preference, 21*, 1008-1015.

O'Mahony, M. (1986). *Sensory Evaluation of Food. Statistical Methods and Procedures.* New York: Marcel Dekker Inc.

O'Mahony, M. (1992). Understanding discrimination tests: a user-friendly treatment of response bias, rating and ranking R-index tests and their relationship to signal detection. *Journal of Sensory Studies, 7*, 1-47.

O'Mahony, M., De Mingo, N., Mendez Martin, J., Perret, M., & Bodeau, J. (2011). 9-point hedonic scales: Data derived from versions using words or numbers give different results, while serial monadic presentation can reduce power. *9th Pangborn Sensory Science Symposium.* Toronto: Elsevier.

O'Mahony, M., Park, H., Park, J.-Y., & Kim, K.-O. (2004). Comparison of the statistical analysis of hedonic data using analysis of variance and multiple comparisons versus an R-index analysis of the ranked data. *Journal of Sensory Studies, 19*, 519-529.

Park, H., O'Mahony, M., & Kim, K.-O. (2007). A comparison of the discriminating power of ANOVA and R-index analyses of hedonic data for various products and experimental protocols. *Journal of Sensory Studies, 22*, 281-292.

Park, J.-Y., O'Mahony, M., & Kim, K.-O. (2007). 'Different-stimulus' scaling errors; effects of scale length. *Food Quality and Preference, 18*, 362-368.

Peryam, D. R., & Pilgrim, F. J. (1957). Hedonic scale method of measuring food preferences. *Food Technology, 11*, 9-14.

Rosas-Nexticapa, M., Angulo, O., & O'Mahony, M. (2005). How well does the 9-point hedonic scale predict purchase frequency? *Journal of Sensory Studies, 20*, 313-331.

Rousseau, B. (2006). Indices of Sensory Difference: R-Index and d'. *IFPress, 9*(3), 2-3.

Rousseau, B. (2011). Measuring product similarities: Are two indices, R-Index and d′, interchangeable? *2011 IFT Meeting.* New Orleans: IFPress.

Stone, H., Bleibaum, R. N., & Thomas, H. A. (2012). *Sensory Evaluation Practices.* San Francisco: Academic Press.

Villegas-Ruiz, X., Angulo, O., & O'Mahony, M. (2008). Hidden and false "preferences" on the structured 9-point hedonic scale. *Journal of Sensory Studies, 23*, 780-790.

Yao, E., Lim, J., Tamaki, K., Ishii, R., Kim, K.-O., & O'Mahony, M. (2003). Structured and unstructured 9-point hedonic scales: a cross cultural study with American, Japanese and Korean consumers. *Journal of Sensory Studies, 18*, 115-139.

Yeh, L. L., Kim, K. O., Chompreeda, P., Rimkeeree, H., Yau, N. J., & Lundahl, D. S. (1998). Comparison in use of the 9-point hedonic scale between Americans, Chinese, Koreans, and Thai. *Food Quality and Preference, 9*(6), 413-419.