
LAS PRUEBAS DE PREFERENCIA EN ALIMENTOS SON MÁS COMPLEJAS DE LO IMAGINADO

OFELIA ANGULO y MICHAEL O'MAHONY

RESUMEN

Las pruebas de preferencia son ampliamente utilizadas en el lanzamiento de nuevos productos y en estudios de mercado en la industria alimenticia. Tradicionalmente se aplican a productos diferentes. Sin embargo, la magnitud de la diferencia entre los productos no siempre se evalúa con anticipación a las pruebas de preferencia. Las pruebas de diferencia, a través del

modelo de Thurstone, permiten estimar el grado de diferencia entre dos productos. Así, resulta interesante conocer las similitudes entre las pruebas de preferencias y las pruebas de diferencia a fin de tomar ventaja de los avances que existen en estas últimas para mejorar la utilización y aplicación de las pruebas de preferencia.

Las empresas de alimentos frecuentemente requieren predecir si sus productos serán comprados por los consumidores. Esta predicción constituye una de las mediciones más relevantes para el éxito de la compañía. No tiene sentido establecer toda una infraestructura para la producción de alimentos que los consumidores no comprarían. Existen diversas metodologías para hacer estas predicciones, siendo la prueba de preferencia por pares considerada normalmente como la más sencilla; sin embargo, trabajos recientes han demostrado que no es tan sencilla. La mayoría de los consumidores reporta falsas preferencias. Al evaluar la preferencia entre dos productos, muchos consumidores indican preferir un producto por sobre otro, aún cuando los productos son idénticos (fueron tomados del mismo lote). Este comportamiento indica que los consumidores emiten su juicio de preferencia en base al tipo de pregunta y no en base a los productos degustados. Es así que los consumidores tien-

den a reportar preferencias hacia productos que prácticamente no prefieren. Las compañías de alimentos registran preferencias que no existen. En esta revisión se analiza la evidencia de la existencia de estas falsas preferencias y presenta el uso de pares 'placebo' para resolver el problema. Además, se analiza las ventajas y desventajas en el uso de la escala hedónica para estimar preferencias y, finalmente, se hace un análisis de la importancia de diseñar métodos que reporten preferencias reales que permitan a la industria de alimentos tomar decisiones exitosas.

Pruebas de Diferencia

El propósito de las pruebas de diferencia es determinar, con un panel o con grupos de paneles, la magnitud de la diferencia percibida entre dos estímulos confundibles. Las formas para estimar la magnitud de la diferencia percibida son diversas. Una de ellas es d' (d prima), derivada de la teoría de detección de señales y del

modelaje thurstoniano (Thurstone, 1927a, b; Green y Swets, 1966; MacMillan y Creelman, 2005). La aplicación de este parámetro en evaluación sensorial ha sido documentada (O'Mahony *et al.*, 1994; Lee y O'Mahony, 2004). Es una medición de la diferencia entre dos estímulos confundibles en unidades (desviaciones estandares) de la variabilidad percibida en un estímulo a la vez. Además, es una medición básica o fundamental, pues el resultado es el mismo independientemente del método utilizado (e.g. duo-trío, triangular, 2-AFC, 3-AFC, tetradas, etc). Finalmente, d' es lo que el ingeniero llamaría la relación señal-ruido.

Otra forma, más sencilla, de representar la magnitud de la diferencia entre estímulos confundibles es el índice R de John Brown (Brown, 1974; O'Mahony, 1992, 2007; Delwiche, 2007). El índice R expresa la magnitud de la diferencia en términos de probabilidad; es la probabilidad de que un juez o un grupo de jueces discriminen exitosamente dos estímulos en pruebas

PALABRAS CLAVE / Evaluación Sensorial / Placebo / Pruebas de Diferencia / Pruebas de Preferencia / Thurstone /

Recibido: 13/05/2008. Modificado: 20/02/2009. Aceptado: 25/02/2009.

Ofelia Angulo. Ingeniera Bioquímica en Alimentos, Instituto Tecnológico de Culiacán, México. MSc. en Alimentos y Nutrición, Universidad de Purdue, EEUU. Doctora en Ciencias de los Alimentos, Universidad de París, Francia. Profesora, Instituto Tecnológico de Veracruz, México. Dirección: Unidad de Investigación y Desarrollo en Alimentos, Instituto Tecnológico de Veracruz, Veracruz, México. e-mail: oangulo@itver.edu.mx

Michael O'Mahony. Psicólogo y Químico, Universidad de Bristol (UB), Reino Unido. Doctor en Psicología y Química, UB, Reino Unido. Profesor, Departamento de Ciencias de los Alimentos y Tecnología, Universidad de California en Davis, EEUU. e-mail: maomahony@ucdavis.edu

de comparación por pares (2-AFC). La probabilidad se estima a partir de datos de identificación de estímulos o de pruebas de ordenación. Aun cuando no se considera una medida básica o fundamental (el resultado varía ligeramente en función de la metodología utilizada) es ampliamente utilizado y puede ser empleado en lugar de las pruebas hedónicas y en alineación de conceptos. Además de la determinación de la magnitud de la diferencia percibida entre dos estímulos, la prueba sensorial también incluye la determinación de si el juez o el grupo de jueces realmente percibe la diferencia entre los dos estímulos (la diferencia es significativa) o si los jueces responden al azar. Para evaluar esto, las pruebas de diferencia se repiten y se evalúa el desempeño del juez. Si el desempeño del juez es muy bueno de tal manera que la probabilidad de respuesta al azar es muy pequeña se concluye que el juez puede discriminar entre los estímulos y que no responde al azar. Se dice entonces que el desempeño del juez fue por encima del aleatorio. Tradicionalmente, la estadística binomial se utiliza para evaluar estos datos y el desempeño del juez por encima del aleatorio se considera significativo.

Pruebas de Preferencia

La prueba de preferencia por pares (Stone y Sidel, 1992; Lawless y Heymann, 1998; Resurrección, 1998) se usa algunas veces sin la opción de 'No preferencia'. Esto se hace con el objeto de analizar los datos por medio de estadística binomial. Si los consumidores optan por la opción de No preferencia se han reportado formas de modificar esta opción, aunque no necesariamente se recomiendan (Odesky, 1967; Lawless y Heymann, 1998; Resurrección, 1998; Marchisano *et al.*, 2003; Alfaro-Rodríguez *et al.*, 2005; Angulo y O'Mahony, 2005). Estas formas incluyen: ignorar las respuestas de No preferencia, dividir las equitativamente entre las dos opciones de preferencia o dividir las proporcionalmente entre las opciones de preferencia en función de las frecuencias de preferencia correspondientes. Otra forma es asignar las respuestas de No Preferencia a una de las dos opciones de preferencia simplemente lanzando una moneda (Falk *et al.*, 1975). Cualquiera de las opciones sugeridas no alterarían las conclusiones del estudio siempre y cuando el número de consumidores que optan por la opción de No preferencia fuera pequeño. Sin embargo, no siempre es el caso (Alfaro-Rodríguez *et al.*, 2005; Angulo y O'Mahony, 2005).

Otra razón para utilizar la opción de No preferencia en las pruebas de preferencia es que genera información que facilita la interpretación de los resultados finales del estudio. Por ejemplo, en una prueba de preferencia sin la opción de No Prefe-

rencia en donde la mitad de los consumidores elige el producto A y la otra mitad elige el producto B no es fácil interpretar el resultado. Una interpretación puede ser que los consumidores no presentan preferencia por ninguno de los productos y aleatoriamente eligen uno y el otro producto. Pero también es posible que la mitad de los consumidores prefiera el producto A y la otra mitad prefiera el producto B. Es fácil comprender que la empresa actuaría de forma muy diferente en una y otra de las interpretaciones arriba expuestas. La inclusión de la opción de No preferencia permite resolver este problema.

Analogías entre Pruebas de Diferencia y de Preferencia

Al considerar las pruebas de diferencia es posible encontrar una analogía con las pruebas de preferencia desde el punto de vista estadístico. Para indicar la significación estadística en el desempeño de un juez; es decir, si su desempeño es estadísticamente mejor que el aleatorio, se aplica el análisis binomial. Para determinar la magnitud de la diferencia percibida entre el par de estímulos evaluados, entonces la teoría de detección de señales y el modelaje thurstoniano (Thurstone, 1927a, b; Green y Swets, 1966; Alfaro-Rodríguez *et al.*, 2005) permiten estimar la magnitud de la diferencia a través del parámetro d' . Así mismo, el parámetro d' puede utilizarse en las pruebas de preferencia para estimar la tendencia de la preferencia. El grado de preferencia también puede estimarse con métodos bayesianos (Draper *et al.*, 1969), pero d' aporta una medida básica y sencilla. A fin de establecer la significación estadística de los datos con respecto a la hipótesis de no preferencia puede aplicarse la prueba de chi-cuadrada, la cual es una prueba multinomial sencilla aplicada en otros estudios (Marchisano *et al.*, 2003; Alfaro-Rodríguez *et al.*, 2007).

Cabe mencionar que la prueba de chi-cuadrada compara el patrón de frecuencias esperadas (FE) bajo la suposición de no preferencia (hipótesis nula) contra el patrón de frecuencias observadas (FO) con los estímulos en estudio. Tradicionalmente, la hipótesis nula establece situaciones de igualdad de frecuencias en las tres opciones de respuesta, a saber: Prefiero A, No preferencia y Prefiero B. Esta igualdad de preferencias equivale a la situación en que un tercio de los consumidores reporta No preferencia y un tercio de los consumidores reporta preferencia por cada uno de los productos en estudio. La igualdad de frecuencias también podría interpretarse como si los consumidores aleatoriamente eligieran una de las opciones de respuestas sin considerar los productos. Ninguna de estas interpretaciones representa la situación en que los consumidores degustan los productos y re-

portan No preferencia. Es lógico pensar que las frecuencias esperadas (FE) cuando se asume una situación en la que no hay preferencia por ninguno de los productos, el 100% de los consumidores elijan la opción de No preferencia. Sin embargo, la misma prueba de preferencia induce a los consumidores a reportar preferencias por uno de los productos cuando en realidad no prefieren uno en particular.

Uso de la Condición Placebo

Ennis y Collins (1980) realizaron un estudio con cigarrillos (denominados A y B) evaluados por consumidores en sus hogares. La evaluación incluyó una variedad de atributos y pruebas de preferencia con la opción de No preferencia. El 40% de los consumidores reportó preferir uno de los productos y el 20% reportó No preferencia. Considerando que ambos cigarrillos fueron tomados uno del inicio y otro del final de la línea de producción, (cigarrillos idénticos), las preferencias expresadas corresponden a otros factores no relacionados con las notas sensoriales de los productos, como puede ser el sesgo en la respuesta misma. Es posible argumentar que este patrón de preferencias podría representar la frecuencia esperada bajo la condición de no preferencia (hipótesis nula), de la misma manera que los resultados de productos con características sensoriales diferentes representarían la frecuencia observada en la prueba de significación estadística con chi-cuadrada. Esta forma de obtención de las frecuencias esperadas es similar al uso de placebos en investigación médica. A partir de estos resultados, el patrón 40-20-40 parecería la frecuencia esperada (FE) apropiada o el 'placebo' para la situación de no preferencia de los consumidores. Sin embargo, otros autores (Marchisano *et al.*, 2003; Chapman y Lawless, 2005; Chapman *et al.*, 2006; Alfaro-Rodríguez *et al.*, 2007) han reportado frecuencias diferentes variando los productos, el tipo de opciones de respuesta, las condiciones experimentales y el tipo de consumidores. Por ello, a la fecha la mejor estrategia sería obtener las 'frecuencias esperadas placebo' correspondientes a cada producto investigado. Trabajos recientes reportan tanto las frecuencias esperadas como las observadas con el mismo grupo de consumidores (Alfaro-Rodríguez *et al.*, 2007), mientras que otros (Foley *et al.*, 2003; Marchisano *et al.*, 2003) las obtienen con diferentes grupos de consumidores. Esta última estrategia asume que ambos grupos de consumidores son equivalentes, suposición que no siempre se cumple, sobretodo si el número de consumidores es pequeño.

Así, el par placebo de muestras idénticas puede considerarse una medida de él o los efectos de las condicio-

nes experimentales *per se*. El par de muestras diferentes, objeto del estudio, genera resultados del efecto de la preferencia y de las condiciones experimentales. La comparación de ambas frecuencias con chi-cuadrada provee de una mejor estimación de las preferencias. Esta estrategia representa una preocupación metodológica, es decir, saber si el par placebo debe evaluarse primero o después del par diferente. El estudio con papas fritas de Alfaro-Rodríguez *et al.* (2007) parece indicar que es indiferente, aunque esto debe confirmarse con otros estudios antes de generalizar conclusiones.

Preferencias Falsas

El uso del par placebo (muestras idénticas) presenta una estrategia adicional para el análisis estadístico. Los consumidores que reportan No preferencia bajo estímulos idénticos, pueden ser considerados como no sesgados, al no verse afectados por la condición experimental. De tal manera solo los jueces no sesgados serían considerados en el análisis estadístico. La desventaja es que el tamaño de la muestra puede reducirse grandemente. Los trabajos reportados a la fecha (Ennis y Collins, 1980; Marchisano *et al.*, 2003; Alfaro-Rodríguez *et al.*, 2005, 2007; Angulo y O'Mahony, 2005; Chapman y Lawless, 2005; Chapman *et al.*, 2006) indican que cuando los estímulos comparados son prácticamente idénticos, la mayoría de los consumidores reportan preferencias por uno o el otro de los productos. Estas preferencias falsas indican un efecto muy marcado de las condiciones experimentales. Hay varias formas para explicar este comportamiento. La primera es que los consumidores realmente perciban ligeras diferencias entre los productos debidas a variaciones en el producto o a variaciones en las señales que los sentidos envían al cerebro durante la degustación. Este segundo argumento de la variabilidad momentánea de la percepción se utiliza en modelos derivados de la teoría de detección de señales con el modelaje thurstoniano (Green y Swets, 1966; Frijters, 1979; Ennis, 1990, 1993; O'Mahony *et al.*, 1994; O'Mahony y Rousseau, 2002; Lee y O'Mahony, 2004; Macmillan y Creelman, 2005).

Estas ideas han sido confirmadas por Chapman *et al.* (2006), quienes evaluaron si dos muestras de leche idénticas se percibían como iguales o diferentes antes de realizar pruebas de preferencia. Los consumidores que reportaban percibir las muestras como iguales se consideraban consumidores sin preferencia definida. Los consumidores que percibían las leches como diferentes, presentaban un porcentaje más alto de preferencias falsas. Los autores, asumen que el efecto thurstoniano no es suficiente para explicar completamente las altas frecuencias

de preferencias falsas. Es posible que los consumidores presenten una fuerte preconcepción de que las muestras deben ser diferentes para que se les haga una pregunta de preferencia, aún si las muestras parecieran idénticas. Chapman *et al.* (2006) reportaron preferencias falsas para muestras idénticas de relojes, cuando el código del color y el nombre del color, escritos de fábrica, fueron claramente visibles; los consumidores reportaron encontrar difícil creer que se les presentaran muestras con el mismo código y nombre de color en una prueba de preferencia. Finalmente, otra explicación de las preferencias falsas dada por Chapman *et al.* (2006) es que como la respuesta era tan obvia, el consumidor pudo haber tratado de hacer la prueba más compleja; esto sucede cuando los consumidores creen estar siendo evaluados por psicofísicos. Marchisano *et al.* (2003) notaron que algunos consumidores se esforzaban por demostrar cuan sensitivos eran, reportando preferencia hacia uno de los productos. Otros consumidores pensaron que la elección de la opción de No Preferencia podría no agrandar al evaluador, mientras que otros no querían pasar por débiles al no reportar preferencia hacia uno de los productos. Podría entenderse también como cuando un amigo presenta dos tazas idénticas de café para que la otra persona elija una para sí mismo, mientras que quien invita tomaría la taza remanente. Es equivalente a lo que Ullmann-Margalit y Morgenbesser (1977), llamarían 'colectar' (tomar entre dos muestras iguales) y 'elegir' (entre estímulos diferentes).

No todas las culturas presentan el mismo comportamiento. Los coreanos son más propensos a reportar preferencias falsas que los norteamericanos. En algunos casos los coreanos manifiestan pena por no reportar preferencias, interpretando esto como si la respuesta no fuese de utilidad para el experimentador. Waterman *et al.* (2001) han notado la presión experimentada por los consumidores al ser entrevistados por una figura de autoridad como puede ser un investigador. Es por ello, que Marchisano *et al.* (2003) evaluaron el efecto de la prueba de manera anónima, en la que el investigador no recibía la hoja de respuestas durante la prueba. Sin embargo, esta estrategia no produjo el efecto esperado y en algunos casos se reportan mayor frecuencia de preferencias falsas.

Independientemente del desconocimiento de las causas por las que los consumidores reportan preferencias falsas, el uso de las muestras placebo con estímulos idénticos resulta una buena estrategia. Cuando las frecuencias de preferencias de ambos productos evaluados son equivalentes pero difieren de las frecuencias reportadas con la condición placebo, es posible concluir que hay dos grupos demográficos de consumido-

res con preferencias diferentes y no el caso contrario, en donde los consumidores manifiestan no preferir uno de los productos en particular (Alfaro-Rodríguez *et al.*, 2007). Otra aplicación de la condición placebo es como una herramienta de selección de consumidores no sesgados por las condiciones de la prueba. Así, los consumidores que reportan No preferencia con los estímulos idénticos constituyen el grupo de estudio para los estímulos diferentes. Es posible que los consumidores no sean afectados por la prueba (no sesgados) en una ocasión y en la siguiente lo sean. Los resultados reportados con la muestra de consumidores completa y con la muestra de consumidores no sesgados ha sido la misma en dos reportes (Chapman y Lawless, 2005; Alfaro-Rodríguez *et al.*, 2007). Sin embargo, antes de dar recomendaciones generalizadas se requiere de más estudios a fin de evaluar el efecto del tipo de consumidor (cultura) y del tipo de producto. Al realizar la comparación estadística entre las frecuencias observadas contra las frecuencias esperadas de todos los consumidores se presenta la ventaja de muestras grandes pero incluye los consumidores con preferencias falsas. Por otro lado, al utilizar solo los consumidores no sesgados, se reduce el tamaño de la muestra y la potencia de la prueba. Es necesario estudiar metodologías que aumenten la muestra de la población, incrementando la frecuencia de no preferencia en la condición placebo.

Prueba Hedónica

Después de considerar la relevancia de las falsas preferencias como resultado de la condición experimental en las pruebas de preferencia, la prueba hedónica puede ser usada para hacer comparaciones de preferencia. Es posible que las preferencias expresadas en la escala hedónica impliquen preferencias menos fuertes que las indicadas en la prueba de preferencia *per se*. Los resultados de preferencia con las escalas indican menos falsas preferencias que las expresadas con la condición placebo de la prueba de preferencia por pares. Particularmente, la escala híbrida que combina nivel de agrado e intención de compra presenta el menor número de preferencias falsas en relación a otras estrategias. Aparentemente, la explicación deriva de la longitud menor de la escala (6 vs 9-puntos) y no de las etiquetas de la escala que inducen un cambio en la estrategia cognitiva. Es fácil argumentar que entre más corta la longitud de la escala, es más probable reportar calificaciones empatadas. Marchisano *et al.* (2003) encontraron que 42% de los consumidores reportan No preferencia con la escala de 9 puntos; 59% con la de 5 puntos y 70% con la de 3 puntos.

Simone y Pangborn (1957) reportaron una desventaja en el uso de la escala hedónica para evaluar preferencias. Durante las evaluaciones de duraznos, los consumidores manifestaron frustración al no poder expresar que aunque los duraznos recibían la misma calificación hedónica, no había opción para expresar su ligera preferencia hacia uno de los productos. Este problema puede resolverse fácilmente incluyendo particiones en cada categoría de la escala hedónica.

Lo anterior puede relacionarse con las escalas de intensidad. Durante la manipulación en la presentación de las muestras para ser evaluadas con escalas de intensidad, Jeon *et al.* (2004) postularon dos tipos de errores en el uso de las escalas. Estos errores recibieron nombres específicos más tarde (Park *et al.*, 2004). Cuando los estímulos con la misma intensidad de percepción reciben calificaciones diferentes se le denomina error del estímulo idéntico. Cuando los estímulos con diferente intensidad de percepción reciben la misma calificación se denomina error del estímulo diferente. Jeon *et al.* (2004) hipotetizaron que las escalas de mayor longitud (mayor número de categorías) pueden inducir mayor número de errores de estímulos idénticos y que las escalas más cortas inducen un mayor número de errores de estímulos diferentes. Estas hipótesis fueron evaluadas por Park *et al.* (2004, 2007) con resultados que tendían hacia la confirmación.

Estos conceptos de longitud de la escala y los errores con el uso de la escala pueden aplicarse en el contexto de la escala hedónica. El experimento de Marchisano *et al.* (2003), demostrando que la escala hedónica de 9 puntos exhibe mayor número de falsas preferencias para estímulos idénticos que la escala de 5 y 3 puntos, puede describirse en términos de longitud de la escala y error de estímulos idénticos.

Como se mencionó anteriormente, el uso del placebo en pruebas de preferencia debe considerar el efecto de presentar las muestras placebo antes y después del par de muestras en estudio. Los resultados de Alfaro-Rodríguez *et al.* (2007) indican que el efecto es mínimo pero concluyen que antes de hacer generalizaciones es necesaria mayor evidencia. El efecto esperado sería el de contraste. Los consumidores pueden ser más propensos a percibir las muestras placebo como iguales si se degustan después del par de muestras en estudio y *viceversa*. Si se perciben como iguales serán más propensos a utilizar la opción de No Preferencia (Chapman *et al.*, 2006). Sin embargo, el efecto contraste requiere que las muestras en estudio se perciban fácilmente diferentes en cada degustación. Al evaluar galletas de chocolate, percibidas en apariencia diferentes, papas fritas y jugo de naranja, percibidas no tan diferentes, los resultados

no mostraron efecto en el orden de presentación de las muestras placebo antes o después, con las muestras más similares (papas y jugo), pero sí lo hubo para las muestras diferentes (galletas). Aparentemente, el tipo de efecto está en función del tipo de alimento evaluado. Para las muestras más parecidas (papas, jugo), el efecto de contraste se confirma por utilizar con mayor frecuencia la opción de No Preferencia al presentar las muestras placebo después de las muestras en estudio. Para las muestras más diferentes (galletas), el efecto de contraste predice una mayor preferencia para el segundo par de muestras degustadas. Ésta fue la tendencia, aunque de manera no estadísticamente significativa. A la fecha, la recomendación es contrabalancear el orden de presentación de los dos pares de muestras, requeridas para el uso de la opción de No preferencia en las pruebas de preferencia con muestras placebo.

Otras Formas de Estimar Preferencias

Además del uso de la opción de No Preferencia, hay métodos alternos para decidir si los consumidores presentan o no preferencias. La estrategia más convincente es monitoreando las compras del consumidor; existe la tecnología para hacerlo. Rosas-Nexticapa *et al.* (2005) monitorearon las compras de yogurt durante 12 meses en los hogares de los consumidores, después de haber realizado pruebas de nivel de agrado y pruebas de intención de compra. Otra opción es utilizando la estrategia de las pruebas de diferencia: repetir las pruebas. Es lógico argumentar que si el consumidor cambia su respuesta durante la repetición de la prueba deberá ser considerado como si no presentara una preferencia operativa; es decir, no tendrá efecto en la compra del producto. Esta estrategia fue evaluada con y sin la opción de No Preferencia (Chapman y Lawless, 2005; Chapman *et al.*, 2006; Wilke *et al.*, 2006). Al aplicar la prueba de preferencia por pares cuatro veces con intervalos de siete minutos, 49,8% de los consumidores cambian sus preferencias al menos una vez en cereales con pasas y 71% de los consumidores lo hacen al evaluar refrescos de cola (Wilke *et al.*, 2006). La misma inconsistencia fue reportada para productos lácteos evaluados consecutivamente (Chapman y Lawless, 2005; Chapman *et al.*, 2006). Alfaro-Rodríguez *et al.* (2007) reportaron inconsistencia después de ocho pruebas repetidas con la opción de No Preferencia en el 100% de los consumidores de papas fritas fácilmente discriminables. Es necesario mencionar que en la repetición de las pruebas, la frecuencia de la utilización de la opción de No Preferencia varía para muestras idénticas (placebo; Alfaro-Rodríguez *et al.*, 2007; Chapman y Lawless, 2005). La tendencia para reportar preferencias con estímulos

idénticos varía con el tiempo conforme a lo esperado en el sesgo en la respuesta. El efecto de la repetición de las pruebas de preferencia merece mayor atención como línea de investigación.

Conclusiones

La suposición que muestras idénticas, tomadas del mismo empaque, no deben presentar diferencias en la experiencia sensorial suficientemente grandes como para que el consumidor las perciba es lógica. Si lo hiciera, entonces el programa de aseguramiento de la calidad del producto no está funcionando. Es posible que éste sea el caso, así que es muy importante la selección adecuada de los estímulos a evaluar en este tipo de investigaciones. Después de esta advertencia es razonable que, independientemente de las respuestas experimentales, los consumidores deben considerarse con preferencias no operativas (no afectan el comportamiento de compra). El uso de las muestras idénticas permite una condición placebo de investigación o de prueba de mercado. Sin embargo, aun para productos con diferencias, es posible que los consumidores reporten preferencias no operativas; es decir, pueden cambiar sus preferencias y decidir comprar uno u otro producto según la ocasión. El desafío es desarrollar una estrategia que permita medir esta falta de preferencia operativa, ya sea utilizando la pregunta correcta o por manipulación de las condiciones experimentales de la prueba. ¿Que tan diferentes deben ser los productos para que los consumidores reporten preferencias operacionales? Esta pregunta puede relacionarse con el criterio τ en las pruebas de diferencia (Rousseau, 2001; O'Mahony y Rousseau, 2002; Lee *et al.*, 2007). La zona de preferencia no operativa puede verse como la línea divisoria que define el criterio de nivel de agrado τ . De haber preferencias definidas, el siguiente desafío es desarrollar metodologías para determinar que tan fuerte debe ser la preferencia. Es decir ¿que tan grande debe ser la diferencia entre la preferencia y el criterio alrededor de la zona de falta de preferencia operacional?

Las pruebas de preferencia pueden encontrar ventajas en los avances documentados para las pruebas de diferencia. Las investigaciones que permitan precisar la forma correcta para establecer preferencias reales en los productos en desarrollo o en estudios de mercado son altamente recomendadas y requeridas.

REFERENCIAS

- Alfaro-Rodríguez H, O'Mahony M, Angulo O (2005) Paired preference tests: d' values from Mexican consumers with various response options *J. Sens. Stud.* 20: 275-281.
- Alfaro-Rodríguez H, Angulo O, O'Mahony M (2007) Be your own placebo: A double paired

- preference test approach for establishing expected frequencies. *Food Qual. Pref.* 18: 353-361.
- Angulo O, O'Mahony M (2005) The paired preference test and the 'No Preference' option; was Odesky correct? *Food Qual. Pref.* 16: 425-434.
- Brown J (1974) Recognition assessed by rating and ranking. *Br. J. Psychol.* 65: 13-22.
- Chapman KW, Lawless HT (2005) Sources of error and the no-preference option in dairy product testing. *J. Sens. Stud.* 20: 454-468.
- Chapman KW, Grace-Martin K, Lawless HT (2006) Expectations and stability of preference choice. *J. Sens. Stud.* 21: 441-455.
- Delwiche JF (2007) Thurstonian probabilistic approaches to new food product development. En MacFie H (Ed) *Consumer Led Food Product Development*. CRC. New York, NY, EEUU. pp.456-470.
- Draper NR, Hunter WG, Tierney DE (1969) Which product is better? *Technometrics* 1: 309-320.
- Ennis DM, Collins J (1980) *The Distinction between Discrimination and Splitting in Paired Testing*. Report #80-233. Philip Morris Research Center, Richmond, VA, EEUU. 50 pp.
- Ennis DM (1990) Relative power of difference testing methods in sensory evaluation. *Food Technol.* 44: 114-118.
- Ennis DM (1993) The power of sensory discrimination methods. *J. Sens. Stud.* 8: 353-370.
- Falk SN, Henrickson RL, Morrison RD (1975) Effect of boning beef carcasses prior to chilling on meat tenderness. *J. Food Sci.* 40: 1075-1079.
- Foley M, Williams A, Bade J, Lancaster B, Popper R, Carr BT (2003) Effect of preference-question format with and without sample differences Abstr. 5th Pangborn Sensory Science Symposium. July, 2003. Boston, MA, EEUU. Abstract #043.
- Frijters JER (1979) The paradox of discriminatory nondiscriminators resolved. *Chem. Senses Flavor* 4: 355-358.
- Green DM, Swets JA (1966) *Signal Detection Theory and Psychophysics*. Wiley. New York, NY, EEUU. pp. 53-116.
- Jeon SY, O'Mahony M, KIM KO (2004) A comparison of category and line scales under various experimental protocols. *J. Sens. Stud.* 19: 49-66.
- Lawless HT, Heymann H (1998) *Sensory Evaluation of Food: Principles and Practices*. Chapman and Hall. New York, NY, EEUU. pp. 430-475.
- Lee H-S, O'Mahony M (2004) Sensory difference testing: Thurstonian models. *Food Sci. Biotechnol.* 13: 841-847.
- Lee H-S, van Hout D, Hautus M, O'Mahony M (2007) Can the same-different test use a β -criterion as well as a τ -criterion? *Food Qual. Pref.* 18: 605-613.
- Macmillan NA, Creelman CD (2005) *Detection Theory: A User's Guide*. 2^a ed. Lawrence Erlbaum. Mahwah, NJ, EEUU. pp. 1-50
- Marchisano C, Lim J, Cho HS, Suh DS, Jeon SY, Kim KO, O'Mahony M (2003) Consumers report preference when they should not: A cross-cultural study. *J. Sens. Stud.* 18: 487-516.
- Odesky SH (1967) Handling the neutral vote in paired comparison product testing. *J. Market. Res.* 4: 199-201.
- 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. *J. Sens. Stud.* 7: 1-47.
- O'Mahony M (2007) Conducting difference testing and preference trials properly. En MacFie H (Ed) *Consumer-Led Food Product Development*. CRC. New York, NY, EEUU. pp. 434-455.
- O'Mahony M, Rousseau B (2002) Discrimination testing: a few ideas, old and new. *Food Qual. Pref.* 14: 157-164.
- O'Mahony M, Masuoka S, Ishii R (1994) A theoretical note on difference tests: models, paradoxes and cognitive strategies. *J. Sens. Stud.* 9: 247-272.
- Park JY, Jeon SY, O'Mahony M, Kim KO (2004) Induction of scaling errors. *J. Sens. Stud.* 19: 261-271.
- Park JY, O'Mahony M, Kim KO (2007) 'Different-stimulus' scaling errors; effects of scale length. *Food Qual. Pref.* 18: 362-368.
- Resurrección AVA (1998) *Consumer Sensory Testing for Product Development*. Aspen Publ. Gaithersburg, MD, EEUU. pp. 10-20.
- Rosas-Nexticapa M, Angulo O, O'Mahony M (2005) How well does the 9-point scale predict purchase frequency? *J. Sens. Stud.* 20: 313-331.
- Rousseau B (2001) The β -strategy: an alternative and powerful cognitive strategy when performing sensory discrimination tests. *J. Sens. Stud.* 16: 301-318.
- Simone M, Pangborn RM (1957) Consumer acceptance methodology; one vs two samples. *Food Technol.* 11: 25-29.
- Stone H, Sidel JL (1992) *Sensory Evaluation Practices*. Academic Press. San Francisco, CA, EEUU. pp. 243-270.
- Thurstone LL (1927a) A law of comparative judgment. *Psychol. Rev.* 34: 273-286.
- Thurstone LL (1927b) Psychophysical analysis. *Amer. J. Psychol.* 38: 368-389.
- Ullmann-Margalit E, Morgenbesser S (1977) Picking and choosing. *Social Res.* 44: 757-785.
- Waterman A, Blades MY, Spencer C (2001) Is a jumper angrier than a tree? *Psychologist* 14: 474-477.
- Wilke KD, Cochrane C-YC, Chambers IVE (2006) Multiple preference tests can provide more information on consumer preferences. *J. Sens. Stud.* 21: 612-625.

FOOD PREFERENCE TESTS ARE MORE COMPLEX THAN IT IS THOUGHT

Ofelia Angulo and Michael O'Mahony

SUMMARY

Preference tests are widely employed for launching new products and in market studies in the food industry. Traditionally, they are applied to different products. However, the magnitude of the difference among products is not always evaluated before the preference tests are carried out. Difference tests using Thurstone

model allow the estimation of the degree of difference between two products. Thus, it is of interest to know the similarities between preference tests and difference tests, so as to take advantage of the advances taking place in the latter and improve the use and application of preference tests.

AS PROVAS DE PREFERÊNCIA EM ALIMENTOS SÃO MAIS COMPLEXAS DO QUE SE IMAGINA

Ofelia Angulo e Michael O'Mahony

RESUMO

As provas de preferência são amplamente utilizadas no lançamento de novos produtos e em estudos de mercado na indústria alimentícia. Tradicionalmente se aplicam a produtos diferentes. No entanto, a magnitude da diferença entre os produtos não sempre se avalia com antecipação às provas de preferência. As provas de diferença, através do modelo de Thurstone, permitem

estimar o grau de diferença entre dois produtos. Assim, resulta interessante conhecer as similitudes entre as provas de preferências e as provas de diferença a fim de tomar vantagem dos avanços que existem nestas últimas para melhorar a utilização e aplicação das provas de preferência.