Ipsos InnoQuest
The Innovation and Forecasting Specialists

Understanding the Order Effects in Sequential Monadic Product Tests

A White Paper from Ipsos InnoQuest
May 2013

Dr. Nikolai Reynolds
Director, Ipsos InnoQuest Germany

Understanding the Order Effects in Sequential Monadic Product Tests

Abstract

This document details an analysis of 22 product tests in which two products have been evaluated by consumers in each study.

The objective of the study was to understand if the adoption of sequential designs improves data sensitivity and discrimination in comparison to a monadic design. It also examines how order effects can be exploited to improve recommendation in product development.

In addition, a "second position effect" is revealed which can be further utilized to improve product testing results.

Introduction

In the market research industry, it is a wide-spread assumption that a sequential monadic design can improve data sensitivity and discrimination in product testing (e.g. Gacula 1987; Komanska, 1989) in comparison to a pure monadic design.

The reasoning behind the assumption is that such a design, for instance, enables consumers to make direct comparisons testing the products after each other, allowing a more critical comparison than having separate consumers testing single products, i.e., monadic.

However, critique to sequential monadic designs surface often with respect to order effects (Welch and Swift, 1992; Friedman and Schillewaert, 2012). Order effects are caused by context or the environment in which the product is evaluated. When a product is seen and tested in the first position, a respondent's context may be the last usage experience with other products in the same category. When seen and evaluated in the second position, the context may well be the product seen and tested first. To take account of the order effect, the order of presentation is commonly rotated such that each product is presented equally often in each ordered position (i.e., seen first or second) (see, e.g., McBurney and White, 2009).

1. Data and methodology

To challenge the assumption that a sequential monadic design will improve data sensitivity and discrimination between 2 products tested in comparison to monadic designs, 22 studies conducted in 2010 and 2011 (see Table 1) were analyzed.

In each study 2 products were tested blind in a sequential monadic order. To account for country effects, a broad geographical coverage, encompassing ten different countries, was considered (Brazil, France, Germany, India, Mexico, Philippines, Russia, Turkey, UK, and USA). In addition, five different industries within fast moving consumer goods were included (beverages, butter/margarine, desserts, personal care, and soups/seasoning/sauces) to consider industry specific effects.

Each study comprised a sample size of n=400, i.e. n=200 monadic evaluations from consumers.

Table 1: Overview of studies analyzed

Study	*Category*	*Country*
Study 1	Butter/Margarine	US
Study 2	Butter/Margarine	Philippines
Study 3	Beverages	India
Study 4	Beverages	Russia
Study 5	Beverages	India
Study 6	Beverages	India
Study 7	Personal Care	France
Study 8	Personal Care	US
Study 9	Personal Care	UK
Study 10	Personal Care	US
Study 11	Personal Care	Germany
Study 12	Personal Care	UK
Study 13	Soups/Seasoning/Sauces	Mexico
Study 14	Soups/Seasoning/Sauces	Brazil
Study 15	Desserts	Mexico
Study 16	Desserts	Turkey
Study 17	Desserts	UK
Study 18	Desserts	USA
Study 19	Desserts	USA
Study 20	Desserts	Mexico
Study 21	Desserts	Turkey
Study 22	Desserts	USA

Each study consists of around n=400 (total data), i.e. n=200 of monadic evaluations.

Source: own compilation

1.1. Data

The data was collected in central locations by conducting computer or paper assisted personal interviews. For comparability reasons, a common variable ("Overall Liking") across all studies was chosen for this analysis. Consumers were asked for their overall liking of the test product after having tested it. In all studies the same rating scale was used (seven point overall acceptance scale ranging from "very poor" to "excellent").

Most of the obtained samples were street recruitment, i.e., consumers on the street were screened and invited to test blinded products at a central location.

The representativeness of the obtained samples for the respective target groups of each country and industry was mainly assessed using four criteria for which complete target population information was available across the different countries and industries: first, non-rejecters of the category, second, gender specific share of the category, third, household and, fourth, age distribution. Despite some differences across the countries, overall the collected responses generally reflected existing target group variations with regard to these four indicators.

1.2 Methods

The common approach in assessing the performance of products in sequential monadic studies encompasses a significance testing on the total aggregated (first and second position) data. The primary assumption is that through rotation and aggregation order effects are distributed more evenly and not, as often mistakenly stated as being, "reduced", "cancelled out" or "avoided" (e.g. Stevens, 2006). To investigate the order effects, it is required to split the data in the order the products have been evaluated, i.e. in first position evaluations and second position evaluations. Four data points are derived out of this split each consisting around 200 consumer responses.

First position evaluations:

- Product A evaluated first
- Product B evaluated first,

Second position evaluations:

- Product B evaluated second, i.e., after the consumers had tested product A.
- Product A evaluated second, i.e., after the consumers had tested product B.

Statistical differences can then be tested among first position evaluations (dependent sample, A vs. B) second position evaluations (dependent sample, A vs. B), and the difference between the same products in the different positions (independent sample A vs. A, B vs. B).

2. Empirical results

The survey data on the performance of the individual products was examined by splitting up the total sequential aggregated data into first and second position evaluations. Table 2 exhibits the arithmetic means of the variable "overall liking". The lowest score (4.36) is provided by product B in study number three, whilst the highest (5.83) by product A in study number 13.

Table 2: Overview of arithmetic means – Overall Liking – 7 point scale

	Product A Total Data	Product B Total Data	Product A 1st Position	Product A 2nd Position	Product B 1st Position	Product B 2nd Position
Study 1	5.48	5.43	5.52	5.44	5.69	5.17
Study 2	5.10	5.24	5.55	4.65	5.62	4.86
Study 3	5.56	4.71	5.47	5.64	5.05	4.36
Study 4	5.13	4.91	5.04	5.22	4.94	4.87
Study 5	5.56	5.35	5.61	5.51	5.44	5.26
Study 6	5.70	5.31	5.67	5.73	5.23	5.38
Study 7	4.37	4.47	4.47	4.27	4.43	4.51
Study 8	5.78	5.60	5.95	5.60	5.55	5.65
Study 9	5.50	5.58	5.48	5.51	5.80	5.35
Study 10	5.72	5.95	5.79	5.65	6.13	5.78
Study 11	5.17	5.09	5.37	4.98	5.30	4.87
Study 12	5.54	5.06	5.64	5.43	5.63	4.56
Study 13	5.80	5.04	5.83	5.77	5.31	4.78
Study 14	5.49	5.15	5.59	5.39	5.16	5.14

Arithmetic means of consumer responses on a 7 point "Overall Liking" scale
Source: own data analyses

Table 3: Aggregated vs. first and second position ratings

	Total sequential aggregated data Scale Point Difference	1st position monadic data Scale Point Difference	2nd position monadic data Scale Point Difference
Study 1	0.05	-0.17	0.27*
Study 2	0.13	-0.07	-0.21*
Study 3	0.85*	0.43*	1.28*
Study 4	0.22*	0.09	0.35*
Study 5	0.21*	0.17*	0.26*
Study 6	0.39*	0.44*	0.35*
Study 7	0.10	0.04	-0.24
Study 8	0.18*	0.40*	-0.05
Study 9	0.08	-0.32*	0.16
Study 10	0.23*	-0.34*	-0.13
Study 11	0.09	0.07	0.11
Study 12	0.48*	0.01	0.88*
Study 13	0.76*	0.53*	0.99*
Study 14	0.34*	0.43*	0.25*

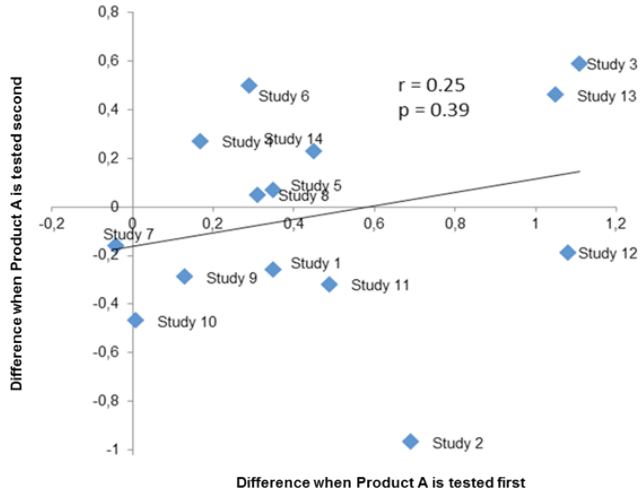
**Significant at the 95% confidence level*
Source: own data analysis.

In the second step of the analysis, arithmetic means were compared between the aggregated evaluation, i.e., sequential data, of products A and B. Then the monadic first position evaluations were compared. Table 3 exhibits the results of this comparison for the first 14 studies. Significances exist at a scale point difference larger than 0.17 on the aggregated level. A comparison of monadic (first position) with total sample ratings suggests some consistency: differences between average overall opinion ratings of roughly .2 or more tend to be found (are consistent) for both data examinations.

In general, there does not appear to be any added discrimination by virtue of the sequential monadic design. Further, use of the sequential monadic design does not add to statistical sensitivity, as often posited in the statistical literature. The second position ratings provide more discrimination between product A and B. In most cases, scale point differences are larger than 0.1.

While the above table (Table 3) discloses the level of discrimination between total aggregated and monadic evaluations it does not deliver a comparison of the difference for those consumers having evaluated first product A then product B and those consumers who have tested the other order, i.e., first product B, then product A. Figure 1 compares scale point differences of both rotations. The figure exhibits no significant correlation between how products have been evaluated in both rotations. The comparison of the scale point differences of the two rotations, at first deliver more discriminating results. However, there seems no universal pattern underlying the results

Figure 1: Comparison of the same product tested first vs. second position



Source: own data analyses

The above figure (Figure 1) raises the question: to what extent do second position ratings differ from first position ratings for the respective product. The analyses, shown in the next table (Table 4), revealed that the arithmetic means of second position ratings are generally lower, for some few exceptions. From this, a "second position effect" is derived and it is hypothesized that a true good product will decrease less than the worse product when examined in the second position.

Table 4: Differences between first and second positions of the same products

Scale Points-Difference between first and second positions of the respective product

	Scale point difference between 1st and 2nd position evaluations of product A	Scale point difference between 1st and 2nd position evaluations of product B
Study 1	-0.08	-0.52
Study 2	-0.90	-0.76
Study 3	0.17	-0.69
Study 4	0.18	-0.07
Study 5	-0.10	-0.18
Study 6	0.06	0.15
Study 7	-0.20	0.08
Study 8	-0.35	0.10
Study 9	0.03	-0.45
Study 10	-0.14	-0.35
Study 11	-0.39	-0.43
Study 12	-0.21	-1.07
Study 13	-0.06	-0.53
Study 14	-0.20	-0.02

Source: own data analyses

To investigate further the stability of the "second position" effect, eight studies were examined of the same food category but across countries. The next table (Table 5) supports the existence of a "second position" that occurs independently of country. Whilst the arithmetic mean shows lower averages in second position, the standard deviation increases, reflecting that consumers tend to consider more scale points and discriminate more in product performance.

Table 5: Detailed comparison of first vs. second positions

	Product 1			Product 2		
	1st position A	2nd position C	Scale difference	1st position B	2nd position D	Scale difference
Mexico						
Mean	5.41	5.03	-0.39	5.54	5.15	-0.39
Sign. different at 95% to	C			D		
Standard Deviation	0.93	1.14	0.21	1.02	1.12	0.10
Mexico						
Mean	5.52	4.80	-0.72	5.65	4.82	-0.82
Sign. different at 95% to	C	-		D	-	
Standard Deviation	0.85	1.17	0.32	0.82	1.04	0.22
Turkey						
Mean	4.92	4.89	-0.03	4.93	4.92	-0.01
Sign. different at 95% to	-	-		-	-	
Standard Deviation	0.91	1.10	0.19	0.89	0.89	0.00
Turkey						
Mean	5.37	5.31	-0.06	5.27	4.95	-0.32
Sign. different at 95% to	-	D		D	-	
Standard Deviation	0.93	0.94	0.01	0.85	1.09	0.24
UK						
Mean	5.29	4.97	-0.32	5.24	4.87	-0.37
Sign. different at 95% to	C	-		D	-	
Standard Deviation	1.06	1.31	0.25	0.94	1.04	0.10
US						
Mean	5.81	5.60	-0.21	5.98	5.37	-0.61
Sign. different at 95% to	C	d		AD	-	
Standard Deviation	0.92	1.20	0.28	0.82	1.20	0.38
US						
Mean	5.92	5.83	-0.09	5.88	5.50	-0.38
Sign. different at 95% to	-	D		D	-	
Standard Deviation	0.90	1.02	0.12	0.95	1.22	0.27
US						
Mean	6.00	5.27	-0.73	6.26	6.45	0.2
Sign. different at 95% to	C	-		A	CB	
Standard Deviation	1.03	1.46	0.43	0.83	0.75	-0.08

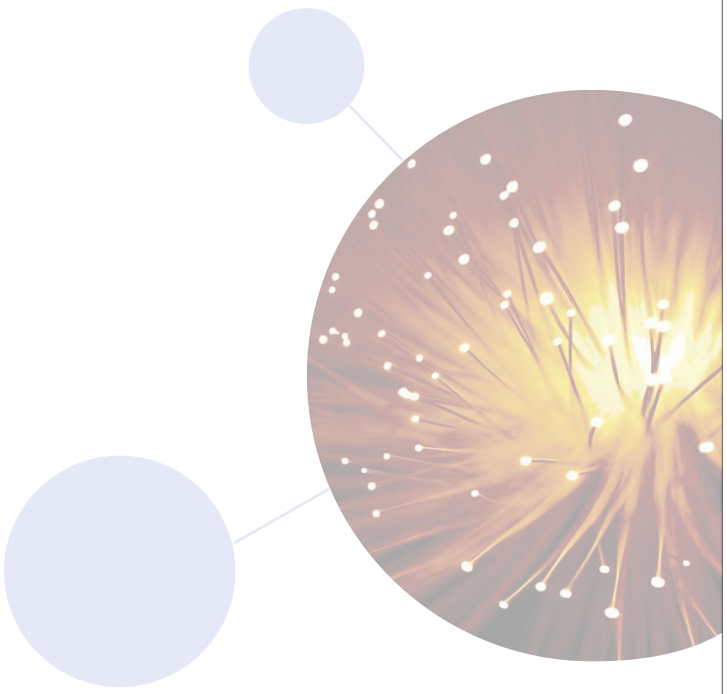
Source: own data analyses

3. Conclusions and discussion

The main conclusion of the paper is that sequential design does not offer demonstrable improvement in discrimination or statistical sensitivity when examining data in total or by first position. However, its virtue is in supplying detail of context or order effects.

There is no apparent increase or improvement on statistical test sensitivity due to use of sequential monadic design. These conclusions immediately suggest little advantage to use of sequential monadic design. However, dissecting the sequential

data in first and second position evaluations reveals more discriminating findings. Taking order effects into account, product perceptions change or are affected by the position in which they are evaluated. The comparison of the second position ratings provides greater product differentiation then either first position or total sample comparisons. Comparison between products dependent on order provides useful perspectives on consumer perceptions and the basis for defining characteristics of "best" product. However, there is need for further research as the research findings are limited to product tests in which 2 products have been tested. Further research is necessary to understand how applicable these findings are when testing more than two products in sequential designs. Also it is necessary to investigate how applicable these research findings are with respect to the use of concepts or other stimulus material.



References

Friedman, M./Schillewaert, N. (2012): Order and Quality Effects in Sequential Monadic Concept Testing: Methodological Details Matter in Concept-Testing Practice. In: The Journal of Marketing Theory and Practice Volume 20, Number 4, Fall 2012.

Gacula, JR., M. C. 1987. Some issues in the design and analysis of sensory data: Revisited. J. Sens. Stud. 2, 169-185.

Komanska, H. (1989): Sequential Monadic Designs: Some Theoretical Considerations. In: Journal of Sensory Studies Volume 4, Issue 3, pp. 201-211.

McBurney, D./White, T. (2009): Research Methods. Cengage Learning.

Stevens, R. (2006): The Marketing Research Guide. Routledge.

Welch, J. L./Swift, C. O. (1992): Question order effects in taste testing of beverages. In: Journal of the Academy of Marketing Science, Summer 1992, Volume 20, Issue 3, pp. 265-268.

Dr. Nikolai Reynolds
Director, Ipsos InnoQuest Germany
Fon +49 69 2474702745
Nikolai.Reynolds@ipsos.com

Ipsos InnoQuest
The Innovation and Forecasting Specialists

Ipsos InnoQuest is the Ipsos brand dedicated to meeting the innovation and forecasting needs of our clients.
Ipsos InnoQuest helps clients to maximize the ROI of their innovation processes through a unique, global end-to-end offer and deep expertise across a wide array of sectors. Central to our offer is our ability to provide financial metrics – even at very early stages – and clear direction for maximization of initiative potential.
We offer simple, intuitive tools built on a consistent philosophy of what drives innovation success, a comprehensive suite of tools for product research and development offering solutions throughout the product's lifecycle, and powerful simulation capabilities to help clients reach their financial targets. From the fuzzy front end to mix optimization, launch and beyond, we help our clients to maximize their innovation I.Q.
Ipsos InnoQuest is a specialized practice of Ipsos, a global market research company which delivers insightful expertise across six research specializations. With offices in 84 countries, Ipsos has the resources to conduct innovation research wherever in the world our clients do business.

Ipsos GmbH
Sachsenstraße 6
20097 Hamburg

www.ipsos.de