Experimental Study

The influence of time interval between products on the consumer evaluation in home use tests

WAGENINGEN
UNIVERSITY & RESEARCH

MSc Food Quality Management thesis
Code: FQD-80436
Date: 23-5-2018

Student:

Bart van Gorcum
Registration number: 931210270120

Supervisors:

Chair group Food Quality and Design
Bea Steenbekkers
Ruud Verkerk

Summary

A large number of the new products fail. Current literature underpins the importance of consumer sensory research (a.o. Kahn et al., 2006; Macfie., 2007). Literature about the influence of sensory methodology regarding in home use tests (HUT) is limited, and is focused on central location tests. The aim of this thesis is to investigate to what extent the time between samples is influencing the results in sequential monadic HUT. An experiment is executed to get insight into the effect of time between samples on the results, and to assess what time interval subjects chose between the samples (cookies) if it is up to the subject. The experiment shows that subjects generally choose to evaluate the samples directly after each other. It is found that the time between samples significantly influences the results. No time between samples shows significant differences for two out of the four rated attributes: appearance $F(1, 80) = 4.305$, $p= 0.04$ and crunchiness $F(1, 80) = 4.859$, $p= 0.03$. The 24 hour time interval does not show any significant results. Further analysis of the data shows that the order effect is present, but is the opposite direction in the 24 hour time interval for the vanilla cookie in comparison to the order effect which is described in literature (a.o. Kofes et al., 2009; Mantonakis et al., 2010; Welch & Swift, 1992). The adaptation level effect is generally not present in the experiment. The findings of the study show that the time interval between samples influences the results. The study underpins the importance of methodology in sensory consumer evaluation and stresses the need to develop valid methodologies.

Table of contents

Chapter 1: Introduction

1.1 Consumer involvement in new product development

Consumers are more demanding of products than ever before (Euromonitor, 2017), they want high quality products which satisfy their needs. Therefore, it is crucial for food producers to deliver high quality products to be able to sell them and to eventually make profit. In the food industry, quality is a term which is of major importance as well. But, what is actually meant with quality? Quality is a dynamic concept. Quality is defined as meeting or exceeding the consumers' expectations (Luning & Marcelis, 2009). So to know what quality of a certain food product is, it is crucial to get insight into the consumers' expectations and their perception. The common process of embedding quality into a product is called the new product development (hereinafter: NPD) process, or the quality design process (Luning & Marcelis, 2009). In figure 1, this process visualized.

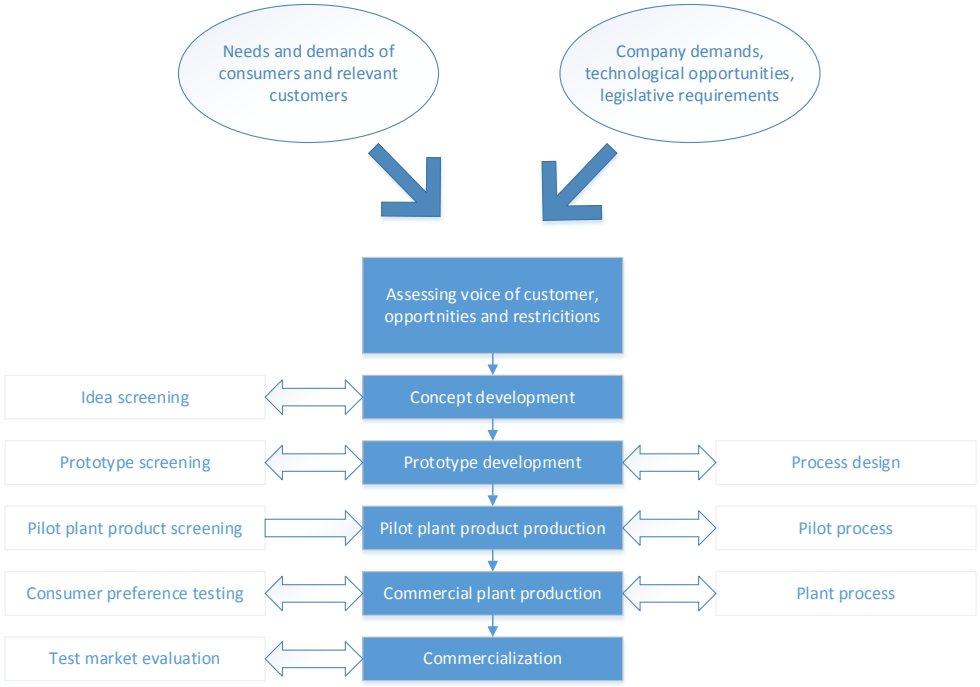


Figure 1. Quality design process, also known as the NPD process (Luning & Marcelis, 2009)

Between 40% and 90% of the new products fail because of not meeting their marketing objectives (Grunert & Van Trijp, 2014). So, it is not a matter of course that new products are successful. A large and growing body of literature laid down the success factors of products, the greater part of the literature underpins that consumer involvement plays a significant role in the successful development of new products (a.o. Kahn et al., 2006; Macfie., 2007). As can be seen in Figure 1, the consumer can be involved in the NPD process, by screening products and giving preferences. Current trends in sensory and consumer science are that consumers are increasingly involved in sensory tasks, and the change to more realistic assessment of the product perception (Giacalone, 2018). A clear indicator of this trend is the move from central location tests to home use tests in consumer tests. Home use tests are considered as a more realistic assessment of the product assessment since the products are consumed in a more natural environment.

1.2 Consumer involvement in sensory tasks: Central location tests versus home use tests

To measure affective responses, the consumer has to be consulted (Stone & Sidel, 2008). Sensory hedonic testing can be done in different environments. Due to the low costs and controlled environment, the most popular test environment is the central location tests (hereinafter: CLT). The central location test can be conducted at several places, for example a classroom in an university, a shopping mall, the canteen of a food company, in a mobile sensory test lab, or at the location of a field testing agency. CLT has its advantages. Outside influence can be minimized in tasting booths, the test can take place in a controlled environment and the samples can be prepared in an uniform way (Lawless & Heymann, 2010).

An often mentioned argument against CLT is that CLT tests are artificial and do not take factors into account like the preparation of the foods. Therefore, the home use test (HUT) receives increasing attention. In HUT, the consumer prepares the food in his/her own way and consumes the food in its own environment. Thereby, the consumer is able to evaluate the product at a moment of choice. A large number of studies has been done into the difference of results between CLT and HUT tests (a.o. Boutrolle et al.; 2007; Karin et al., 2015; Sosa et al., 2008).

Boutrolle et al. (2005) compared the results of several CLT and HUT studies. The majority of the studies shows that the liking scores between CLT and HUT differ. The scores in natural conditions (HUT) are often higher than the scores in standardized conditions. The review of studies between standardized situation tests (CLT and lab tests) and HUT tests is shown Appendix 1, and is completed with recent studies. The column “sample presentation” shows whether the products are tested simultaneous, monadic sequential or completely monadic. Monadic sequential is the research design in which every participant receives several products and evaluate them one by one. In a monadic research design, the respondent only evaluates one product (Alex et al., n.d.). In a simultaneous research design, the participant receivers two or more products at the same time.

In general, the reviewed literature in Appendix 1 shows higher liking scores for HUT than for CLT. Boutrolle et al. (2007) mentions five factors which may be potential explanations for the differences: (1) consumers choose the consumption moment, (2) consumers eat when they are hungry in HUT, (3) longer contact with the product in HUT, (4) different setting, in CLT the consumer may be more analytical, (5) the usual context of consumption in HUT.

When interpreting the table in Appendix 1, attention should be paid to the fact that the HUT tests are executed in different ways. Monadic and monadic sequential tests can be distinguished. Thereby, also the monadic sequential tests were executed differently (see column “Method of HUT”). The table shows a diversity of time spans between the testing of the different products. In a number of studies it was up to the participant to decide when to consume the product, and in other studies the participant had to leave a fixed time interval between the observations. Also, the majority of the studies did not mention the way the monadic sequential tests were executed. As Helson (1964) mentioned, the time between stimuli can influence the rating and thereby the way sequential monadic HUT is executed, can influence the results. It is not clear what time interval between the products were used in the studies. The current research will investigate the extent to which the time interval between the products in a sequential monadic design, influences the results of the sensory evaluation.

1.3 Biases in sensory evaluation

A weakness of the monadic sequential design is that the context of a sensory evaluation can influence the results. This are so called context effects. Context effects are defined as conditions in which the judgment about a product will shift, because of factors like other products tested (Lawless & Heymann, 2010). Two major context effects exist; the contrast effect and the order effect.

The contrast effect is that a product will be rated as more intense if a weaker product was tested before and that the product will be rated as less intense if a strong product was tested before (Lawless & Heymann, 2010; Rankin & Marks, 1991; Zellner et al., 2006). Contrast effects are common in hedonic ratings as well, and the effect does also occur with trained panels (Olabi & Lawless, 2008). The contrast effect is explained by the psychologist Helson (1964) in the adaptation level theory, he proposed that the reference is the average level of a stimulus which is shown before. He laid down that more recent items tend to have a stronger effect on the adaptation level.

Another often discussed effect in hedonic testing is that the first sample receives a higher score than subsequent samples (a.o. Kofes et al., 2009; Mantonakis et al., 2010; Welch & Swift, 1992). Mantonakis et al. (2010) found the order effect in sequential monadic wine tests. In a sequential monadic design with two to five different wines, it was found that the first tested product was more liked than subsequent tested wines. Welch & Swift (1992) investigated the order effect in carbonated drinks, they concluded that both the acceptance and intention to buy was higher for the first sample in the sequence.

However, the mentioned on the order and the adaptation level are based on CLT tests. Thereby, the effect of time between samples on the adaptation level and order effect is not laid down in the previously published studies.

1.4 Demarcation

Since a large percentage of the new product introductions fail, research has been done into the success and fail factors. Five different success factors were obtained by Kahn et al. (2006). The current research focuses on a key success factor in the NPD process, the consumer involvement in NPD. Consumer involvement is crucial to get insight into whether a product exceeds consumers expectations or not.

Preliminary desk research showed that a trend is occurring in which consumers are increasingly involved in sensory evaluation, and that a transition is occurring into more realistic product evaluation (Giacalone, 2018). In consumer evaluation, a distinction can be made between HUT and CLT. The transition described by Giacalone (2018) is visible in practice, HUT is receiving increasing attention.

The differences of HUT and CLT have been described extensively in literature, literature shows higher liking scores for HUT than for CLT. However, the execution of HUT tests received limited attention. The current research focuses on the execution of HUT tests, and the effect of time between products on the results. The demarcation is visualized in figure 2.

In figure 2, the consumer is shown which gives input into the different consumer evaluation methods. The results of these consumer evaluation methods (HUT or CLT) lead to go and no-go decisions within the quality design process (also known as: NPD process). The lower box distinguishes two ways of executing sensory consumer evaluation; CLT and HUT. Appendix 1 shows that the results between HUT and CLT differ. However, research within the sequential monadic HUT tests is limited. The focus of the current research is on the execution in the HUT method, namely on the time interval between samples.

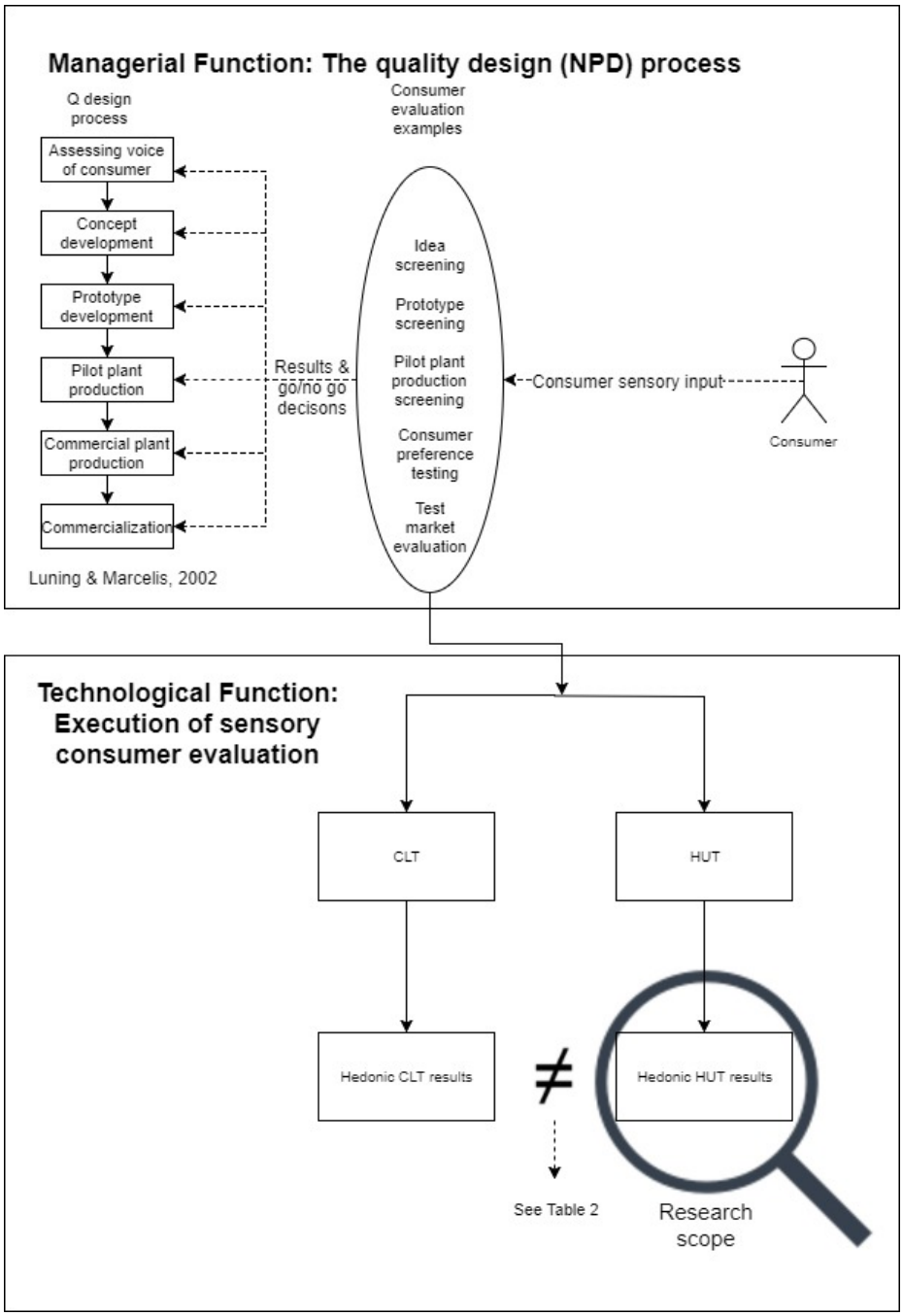


Figure 2. Demarcation of the research

In the present study, the effect of the time interval between products on the results in HUT will be investigated. Two effects are investigated: the order effect and the effect of the adaptation level. It is important to lay down the definition of those terms in the present study. The order effect is defined as that the effect that the first tested product is rated higher than the subsequent tested product(s) (Mantonakis et al., 2010). The adaptation level effect is defined as the effect of a stimulus which is shown before on the current stimulus (Helson, 1964). In the case of sensory evaluation, this is the effect of the previously tested product on subsequent products.

1.5 Research objective

In line with the background information stated above, there are currently two knowledge gaps concerning the time interval between (to be) tested products in sequential monadic HUT:

1. The actual behaviour of respondents in sequential monadic home use tests in terms of time between tasting the products, if there are no instructions on the time between the products.
2. The effect of time between samples in home use tests on the results, the order effect and the adaptation level effect.

The aim of this thesis is to investigate to what extent the time between samples is influencing the results in HUT, to get insight into the validity of the sequential monadic HUT.

1.6 Research questions

The following research questions will be answered in the research:

Main research question: To what extent does the amount of time between sample influence the results in monadic sequential HUT?
Research question 1: What time interval do participants choose between sample in monadic sequential HUT?
Research question 2: To what extent does the adaptation level effect occur during monadic sequential HUT?
Research question 3: To what extent does the order effect occur during monadic sequential HUT?

Chapter 2: Research method

To gain insights into the effect of time interval between monadic sequentially tested samples on the results, a consumer study is done. The sample is subdivided in six different groups, according to time between samples and order of samples. An overview of the design is shown in Table 1

Table 1: Overview of the study design

Group	Time between samples	Order of the samples V=Vanilla cookie C=Coffee cookie
1	Up to participant, between 0 and 24 hours (t=?)	VC
2		CV
3	No time, evaluation of second immediately after the first sample (t=0)	VC
4		CV
5	24 hours (t=24h)	VC
6		CV

2.1 Design

As mentioned by Helson (1964), more recent items tend to have a stronger effect on the adaptation level. However, the effect of the interval between products in HUT and the effect on the hedonic results has to my knowledge never been investigated so far. Research on the effect of time between products in discriminative tests has been done, an example is the research of Harker et al. (2002). They investigated the effect of time between samples on the discriminative power of consumers in apples, at different levels of firmness. They tested the time intervals of 1 minute and 1 day. The 1 minute delay is representing monadic sequential research as executed in practice, and the 1 day delay between products is representing the more realistic time between snacks.

A similar approach was done in the current research. From group 1 and 2 insights are gained about the subjects' tasting behavior if no time interval between the products is set. In group 3 and 4, the products are tested with no time interval between the products, which resembles the current practice in monadic sequential CLT and a part of the HUT studies laid down in Appendix 1. At group 5 and 6, the time interval between the two products is 1 day, which represents the expected actual eating behaviour of in-between snacks.

2.2 Subjects

Each of the 6 groups are targeted to consist of 40 subjects. So each hedonic evaluation of product A and B at the different time levels will consist of 80 subjects, which is in compliance with the guidelines by Stone and Sidel (2004). The subjects are recruited via social media and by contacting students at the university and student housing.

The time between samples is the main point of interest for the research. To minimize other noise in the data, it is important to have an uniform group. Thereby, it is important to get sufficient participants within the set time planning for the thesis. The cohort consists of students from Wageningen University. Students from Wageningen university have comparable backgrounds (age and education level) and this enables the researcher to get sufficient data within the set time planning, because of the researcher's network at the university.

2.3 Samples

The test will be done with two individually packed cookies. The products are packed individually and have the same packaging. The two cookies are sold under the Rioba brand. A picture of the coffee cookie (right), vanilla cookie (left) and its packaging are shown in the figure below. The samples were packed in white envelopes, and the two envelopes were put in a plastic bag. The instructions as shown in Appendix 2 were taped on the outside of the plastic bag.

Figure 3. Samples used in the research.

2.4 Questionnaire design

The data from the subjects is collected via the website www.cookietest.nl, which was linked to the questionnaire software Qualtrics (Qualtrics, Provo, USA). The questionnaire is shown in Appendix 3. To be able to check if a participant did what was instructed, each sample got a unique code. This code consists of the following sub codes:

XX | XX | XXX -> Condition | Participant number | Product number

An example of a code used is: 101742
This stands for condition 1, participant 01, product number 742.

After the first product was tested, the subject was asked some questions regarding demographics to control the confounding variables. After evaluating the first sample, instructions for the time interval until evaluating the next sample were shown. The survey flow is visualised in Figure 4.

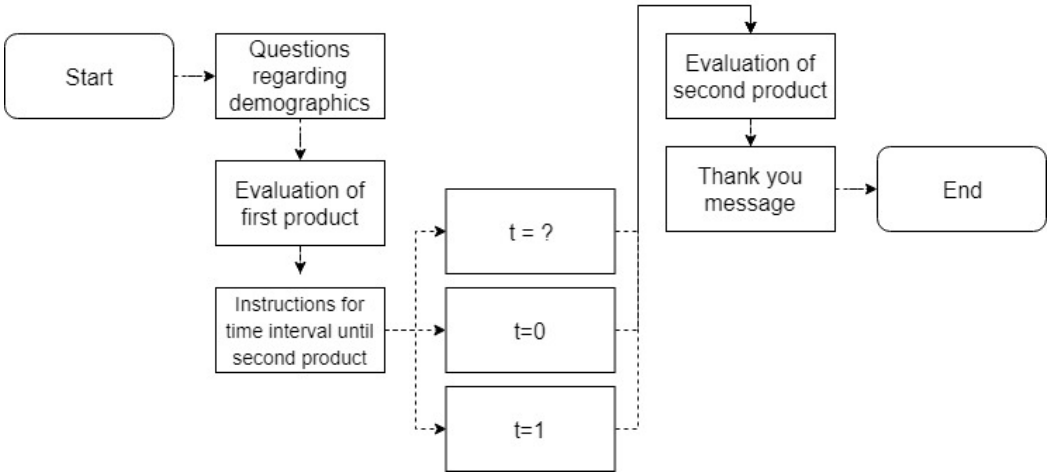


Figure 4. Survey flow

Each subject is asked to give details about his/her year of birth, country of birth and whether he/she is a student or not. The subjects which are in the t=? condition, are asked to give their motivation for choosing the consumption moment of the second tested cookie.

For rating each cookie, the subjects are asked to rate the appearance, the taste, the crunchiness and the overall liking of the cookie. The most relevant attributes for the consumer judgement of cookies are identified by consulting experts.

The attributes are rated on a 7-point scale since it is judged as easy to use and is most preferred by subjects in the research on scaling by Preston & Colman (2000). In the second part of the questionnaire, the overall liking of the cookie is rated on a scale from 1 to 10. This is placed on a separate screen, to minimize the effect of the previous attributes on the final judgement. Another remedy for the so called common method bias is to use a different scale for the overall rating (Podsakoff et al., 2003).

2.5 Data analysis

In this subchapter the plans of analyses are discussed.

2.5.1 Plan of analysis to get insight into the time interval between products.

In group 1 and 2 (see Table 1), data is obtained about the time interval between the two products. Subjects were asked to evaluate the product whenever they want, but within 24 hours. In analysing this time interval, a distinction can be made between no time between samples and longer times between the samples. No time between samples is defined as a time interval between the evaluation of 0 and 3 minutes, to give the subject time to remove the old cookie (as indicated by the instructions) and re-open the questionnaire.

2.5.2 Plan of analysis to get insight into the effect of time interval between products on the results

Linear mixed models are commonly used to compare different sensory methods (Boer, 2017). For each of the following three analyses, a linear mixed model is carried out to assess the existence of significant differences. Data from t=0 and t=24h are used to investigate the effect of time between products on the results. An overview of the data obtained is shown in Table 2.

Table 2. Schematic overview of the obtained data about the coffee and vanilla cookie at t=0 and t=24h.

	Sequence A Vanilla first Coffee second	Sequence B Coffee first Vanilla second
t=0	V_{a0} -> C_{a0}	C_{b0} -> V_{b0}
t=24h	V_{a24} -> C_{a24}	C_{b24} -> V_{b24}

Data about the variables gender, age and whether the subject is a student are collected. The age and student status are not taken into account into the mixed model. The majority of the target group (98%) is born between 1982 and 2002, so part of the millennial generation (Elam et al., 2007) and thereby considered as a homogeneous group. Also, the student status is not taken into account in

the mixed model because there is no proven causal relationship between student status and liking of cookie products. The confounding variable gender is taken into account in the mixed model, since gender influences the hedonic perception of foods (Ueland, 2007).
In the analyses, the cookie, order (test sequence) and gender are specified as fixed factors. The subjects are specified as random factors. The level of significance is set at α=0.05, and the liking of the attributes is analysed with the Least Significance Differences method. The analyses are done with SPSS software (IBM SPSS version 20, Chicago, IL, USA). The means as shown in the latter of the thesis are estimated means, indicating that the mean is adjusted for the fixed variables in the model (Grace-Martin, n.d.).

2.5.2.1 Ratings per time level

The differences in rating per attribute for the two cookies are calculated at the time level t=0 and t=24h. The data produced show how the attributes are liked, and whether the differences are significant between the cookies. This output resembles the output which is currently produced by sensory evaluation companies and lead to decisions within food companies. The results are shown per time interval (t=0 and t=24h).

2.5.2.2 Order effects per time level

The order effect is that the first sample receives a relatively higher score than subsequent samples (Mantonakis et al., 2010). The order effect for each time level and product are calculated. This is calculated per type of cookie. The estimated mean of the attributes from the first tested cookie is compared with the estimated mean of the attributes of the second tested cookie.

The following analyses are done (see Table 2).
V_{a0} is compared with V_{b0} to get insight in the order effect at t=0 for the vanilla cookie.
C_{a0} is compared with C_{b0} to bet insight in the order effect at t=0 for the coffee cookie.
V_{a24h} is compared with V_{b24h} to get insight in the order effect at t=24h for the vanilla cookie.
C_{a24h} is compared with C_{b24h} to get insight in the order effect at t=24h for the coffee cookie.

2.5.2.3 The effect of the previously tested product on the results

The second tested products from t=0 and t=24h are compared, per cookie type. By executing this test, insights are gained into the effect of testing the previous product in relation with time level.

The following analyses are done (see Table 2).
C_{a0} is compared with C_{a24} to get insight into the effect of the previously tested vanilla cookie on the second tested coffee cookie at t=0 and t=24h.
V_{b0} is compared with V_{b24} to get insight into the effect of the previously tested coffee cookie on the second tested vanilla cookie at t=0 and t=24h.

Chapter 3: Results and discussion

This chapter presents and discusses the results of the sensory evaluation.

3.1 Sample distribution

3.1.1 Overview of distributed samples

A pilot study was executed before the actual study is done. The results of the pilot study are used to optimize the consumer study. Since the pilot study showed that the drop-out was substantial, the final study was executed with a larger sample size to take this drop-out into account.

The samples were distributed in two time period. The first period was between 19 and 24 February 2018, and the second period between February 26th and March 2nd 2018. A second time period was necessary because the first period did not lead to a sufficient number of correctly evaluated samples for each condition. In total 412 subjects indicated to be willing to participate, in total 221 (54%) subjects finished the sensory evaluation correctly.

Reasons for the dropout are not starting the evaluation at all, evaluating only one product or using the wrong time interval between the two samples. The correct time intervals are defined in Table 3, in which also the corresponding labels are shown. The upper limit for the time interval at group 3 and 4 is set at 3 minutes, this is the time representing finishing the left-over of the first cookie.

Table 3. Overview of the groups and the corresponding time intervals

Group	Label	Time interval in instructions	Time intervals defined as correct
1 & 2	Time = ?	Up to participant, between 0 and 24 hours	Lower: directly
			Upper: 24 hours
3 & 4	Time = 0	No time, evaluation of second immediately after the first sample	Lower: directly
			Upper: 3 minutes
5 & 6	Time = 1	24 hours.	Lower: 16 hours
			Upper: 32 hours

Table 4 presents an overview of the correctly evaluated sample pairs. Correctly evaluated sample pairs are the number of subjects which used the instructed time interval between the cookies. The last column shows the number of sample pairs which are used for further analyses.

Table 4. Overview of the sample distribution and correctly evaluated samples.

Group	Time label	Samples distributed	Subjects who started with sensory evaluation	Correctly evaluated samples
			(As percentage of samples distributed)	*(As percentage of subjects who started with evaluation)*
1	Time = ?	64	43 (67%)	39 (91%)
2	Time = ?	68	37 (50%)	37 (100%)
3	Time = 0	58	41 (76%)	41 (100%)
4	Time = 0	59	43 (73%)	41 (95%)
5	Time = 24 h	69	58 (84%)	39 (67%)
6	Time = 24 h	94	57 (61%)	24 (42%)
Total		*412*	*279 (68%)*	*221 (54%)*

3.1.2 Discussion of sample distribution

It can be seen from table 4, that group 5 and 6 reported substantial higher dropout rates than the other groups. This indicates that subjects did not use the right time interval between the samples. The major cause for the higher dropout rate is that subjects in this group use the wrong time interval between the samples (less than 16 hours or over 32 hours) or that the second product is not evaluated. A likely cause for this is that subjects do not read the instructions well, or forget to evaluate the second sample.

In future research, methods can be designed to prevent a high dropout rate at the 24h time interval. A measurement which can be taken is sending the subjects a notification when they have to consume the second sample. Another possibility is giving the subjects' the samples in the 24h condition one by one, because then it is not able to consume both samples directly after each other. However, a drawback of the one by one distribution is that it requires more effort from the subject to hand over the second product which can lead to a lower willingness to participate in the study. Ideally, the subjects can receive a financial incentive when they successfully finish the study. This is likely to motivate the subjects to finish the study, but requires bigger budgets for the researcher.

3.2 Subject demographics

In total 221 participants successfully finished the study. The gender, year of birth and whether the subject is student are shown in Table 5.

Table 5. Demographics per group.

Group		**Gender**		**Age**		**Student**	
	n	**Male (%)**	**Female (%)**	**Mean ± SD**	**Lower and upper boundary of age**	**Yes (%)**	**No (%)**
1	39	23	67	24 ± 3.1	18-34	100	0
2	37	27	63	24 ± 2.8	18-32	97	3
3	41	32	68	23 ± 2.8	18-30	95	5
4	41	46	54	23 ± 4.2	16-34	98	2
5	39	33	67	24 ± 6.1	14-50	82	18
6	24	17	83	22 ± 2.6	18-28	100	0

As can be seen in the table above, the proportion of students/non-students in group 5 is not equal to the other groups. Closer inspection of the data shows that the majority (98%) of the non-students are born between 1982 and 2002 and are in the age interval of (PhD) students. This majority is part of the same generation, the millennial generation (Elam et al., 2007). Therefore, the groups are considered to be homogeneous and do not interfere with the goal of the research: comparing methods.

3.3 Time interval between samples

In group 1 and 2, the subject was free to choose the time between the samples. The majority of the subjects did the evaluation of the second sample directly after the first sample, as shown in Table 6.

Table 6. Time interval used between the samples, if the subject was free to choose the time interval.

Time interval between samples	Percentage	Cumulative Percentage
0 and 3 minutes (directly)	92%	92%
Between 3 minutes and 1 hour	4%	96%
Between 1 and 12 hours	3%	99%
Between 12 and 24 hours	1%	100%

The reasons mentioned are shown in Appendix 4. The reasons mentioned are classified by the following terms in Table 7.

Table 7. Classification terms and a description.

Term	Percentage	Description
Convenience	17 %	It was convenient to eat the second cookie at the chosen moment.
To make sure not to forget	7 %	It was evaluated at the moment because the subject wanted to make sure that it does not forget to evaluate the second cookie.
For methodological reasons	9 %	The subject choose to eat the second cookie at the chosen moment, because it believes that it will yield more valid results.
Curiosity	6 %	The subject was curious for the second cookie after tasting the first.
Hungriness	20 %	The subject was still hungry after evaluating the second product.
Eager for a cookie	10 %	The subject choose to evaluate the second product because it was eager for a cookie.
No reason mentioned	36 %	No motivation was mentioned.

As shown in Figure 5. The majority of the subjects did not mention any motivation (36%). Hungriness is mentioned as the most important reason for choosing the consumption moment of the second cookie (20%), followed by the reason that the evaluation moment for the second cookie is a convenient moment (17%), and because the subject was eager for a cookie (11%).

One subject argued that the cookies are not too big and that it was still hungry after eating the first cookie. Therefore, it might be possible that different products or bigger portions may lead to a different choice for time interval between two test products in a consumer evaluation.

3.4 Ratings per time level

This chapter aims to get insight into the extent to which the time between samples influences the results in sequential monadic presentation. In the analysis, the two expected mean liking on the attributes at t=0 and t=24h are shown. F-tests were used to analyse the difference in liking at both time levels. The statistical outputs are shown in Appendix 5. The unbalanced groups in the different time levels in terms of sequence and gender, was balanced by the use of a mixed model. Therefore, the numbers shown in the graphs are estimated means.

3.4.1 Evaluation with no time between cookies

If the sensory test is executed with the method of no time between the two products, It will lead to the results presented in Figure 6. There was a significant effect of the main effect cookie on the rating of appearance $F(1, 80) = 4.305$, $p= 0.04$, indicating that the appearance of the vanilla cookie is more liked than the appearance of the coffee cookie. There was also a significant effect of the main effect cookie on the rating of crunchiness $F(1, 80) = 4.859$, $p= 0.03$. This effect indicates that the crunchiness of the coffee cookie is more liked than the crunchiness of the vanilla cookie.

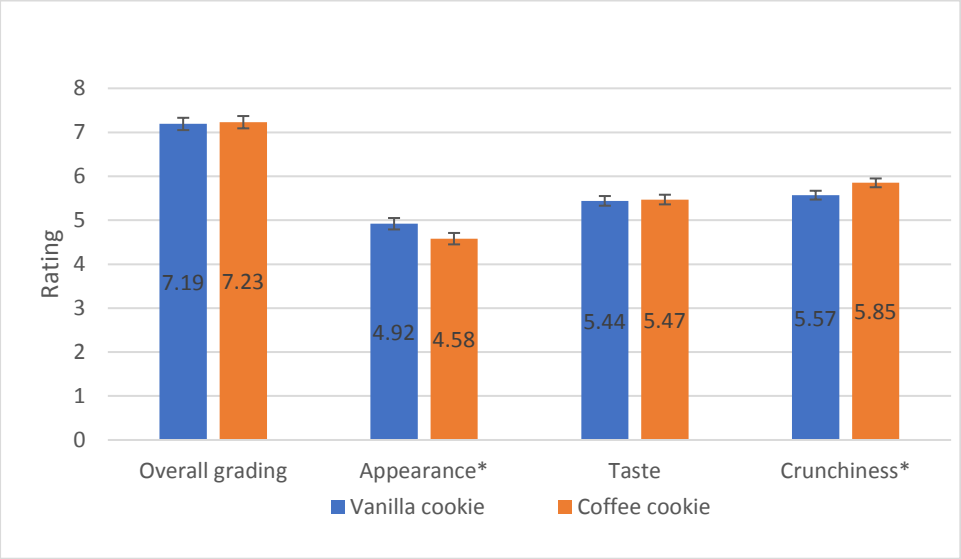


*Figure 5. Results of sensory evaluation, with no time between the two products. *: indicates a significant difference.*

3.4.2 Evaluation with 24 hours between cookies

Figure 7 gives the expected means and standard errors of the both cookies with the 24 hour time interval between the cookies. No significant differences were found between the two cookies, indicating that both cookies are equally liked on the evaluated attributes.

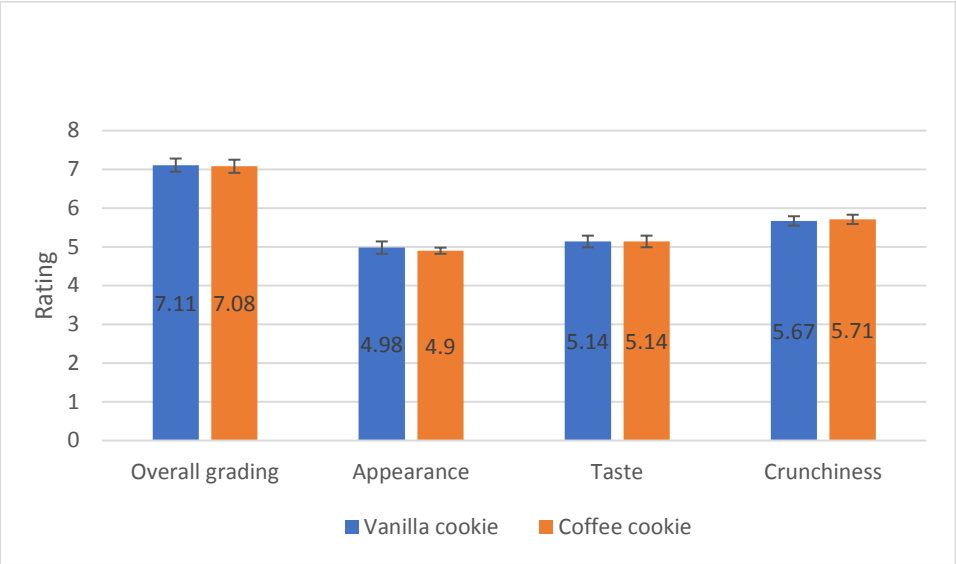


*Figure 6. Results of sensory evaluation, with 24 hours between the two products. *: indicates a significant difference.*

3.4.3 Discussion of the outcome at both time levels

The results presented above show that the sequential monadic method with no times between products, gives significant differences in appearance and crunchiness. Whereas the 24h time interval does not show any significant differences between the products. This implies that a research design with no time between samples is more suitable to find perceived differences between products.

An explanation for significant differences in the t=0 interval may be from a combination of two effects: an adaptation effect and a memory effect. A memory effect is that the subject remembers the previously tested product better in the no-time condition and therefore takes this into account with evaluating the second product. Almeida et al. (1999) investigated the discriminative power of judges with a soft drink product, they compared the discriminative power at time levels between direct and 24 hours. They found that an increase in delay between the two products, leads to a decrease in discriminative power. The research of Almeida et al. (1999) indicates that the memory on sensory attributes will fade over time. Taking this to account, the adaptation level for rating the second product is better memorized in the direct condition than in the 24h condition. With the adaptation level in mind, the subject is able to compare the rated attributes better in the t=0 interval than in the t=24h time interval. The ability of comparing the two products, may lead to differences in attribute liking.

The differences in outcome between the t=0 and t=24h may have implications for the external validity of the results obtained in the test. It proves that the time interval between samples influences the results of the sensory evaluation. As stated in Chapter 1, using consumer input in the NPD process is important to have a successful NPD process. A form of consumer input can be sensory evaluation. This evaluation can guide decisions in the NPD process, like go/no-go decisions with new products. Therefore, attention should be paid to the time interval which is used to produce the sensory data.

3.5 The adaptation level effect

To explore the effect of the previously tested product more in detail, the second evaluated samples were compared for each cookie in t=0 and in t=24h. The data used in the analyses are shown in Appendix 6.

3.5.1 Influence of previously evaluated coffee cookie on the second evaluated vanilla cookie

Figure 8 shows the rating of the second tested vanilla cookie on t=0 and t=24h. The figure shows that the second tested vanilla cookie at t=24h has a significant higher rating for the crunchiness than the second tested cookie at t=0, $F(1, 61) = 4.637$, $p= 0.04$.

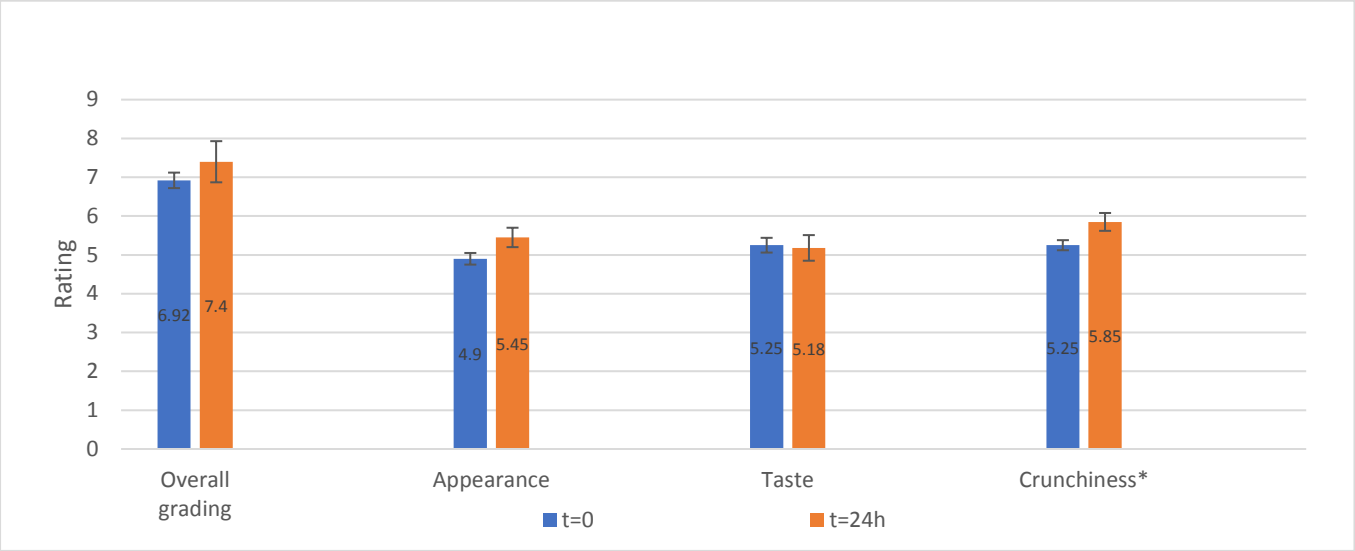


*Figure 7. Expected means of the second evaluated vanilla cookies at t=0 and t=24h. *: Indicates a significant difference.*

3.5.2 Influence of previously evaluated vanilla cookie on the second evaluated coffee cookie

As can be seen in Figure 9, no significant differences are shown in the coffee cookie.

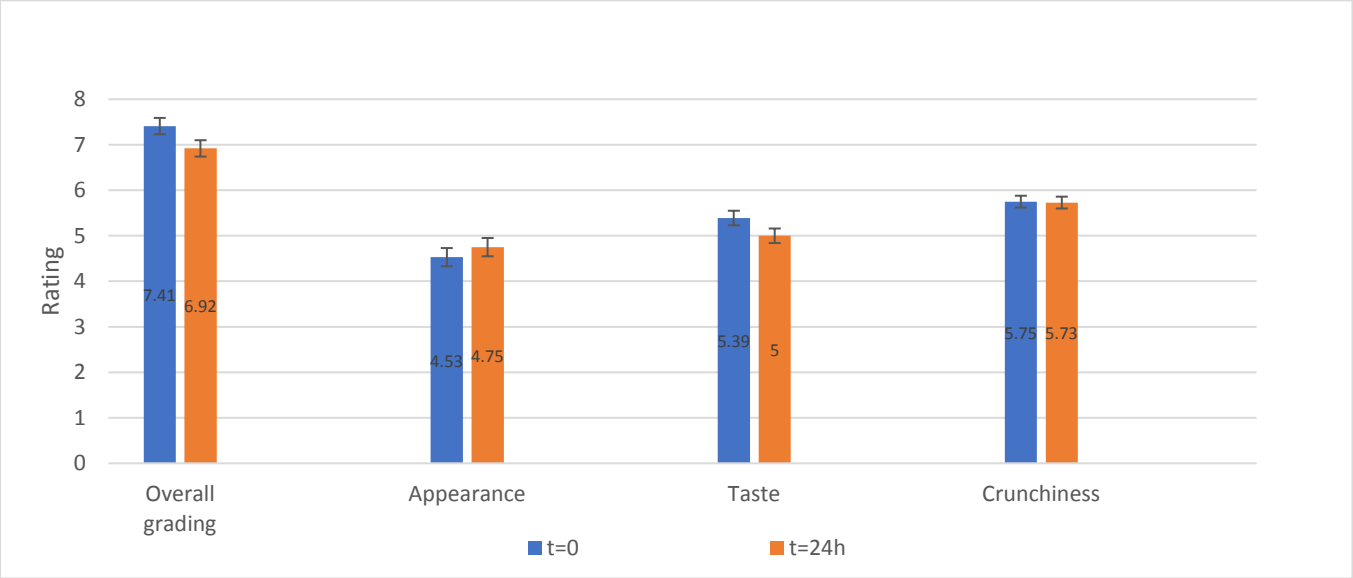


Figure 8. Expected means of the second evaluated coffee cookies at t=0 and t=24h.

3.5.2 Discussion of the adaptation level effect

As shown in the results, the second tested product does only show a significant difference in the crunchiness for the coffee cookie. The attributes in the second tested vanilla cookie do not show significant differences between when it is tested directly after the coffee cookie, or 24 hours after the coffee cookie. The fact that there is only one significant difference, indicates that the time between products generally does not influence the adaptation level effect.

3.6 Order effect

The set of analyses below assesses the extent to which the order effect occurred. The expected mean values of the first and second tested products are compared, for both the vanilla cookie and coffee cookie separately.

3.6.1 Order effect in vanilla cookies

The results obtained from the analysis on the order effect in the vanilla cookies can be seen in Figure 10 and 11. The data output is added in Appendix 7. An order effect is significant in crunchiness at t=0, F(1, 78) = 10,151, p= 0.002. In t=24h a significant difference is shown in appearance, F(1, 59) = 5,518, p= 0.022.

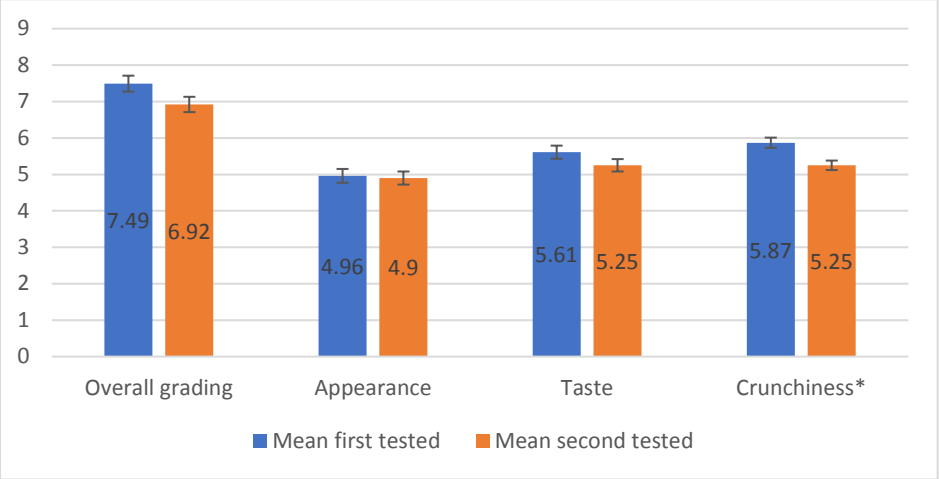


*Figure 9. Order effect of the vanilla cookie, with no time between the cookies. *: indicates a significant difference.*

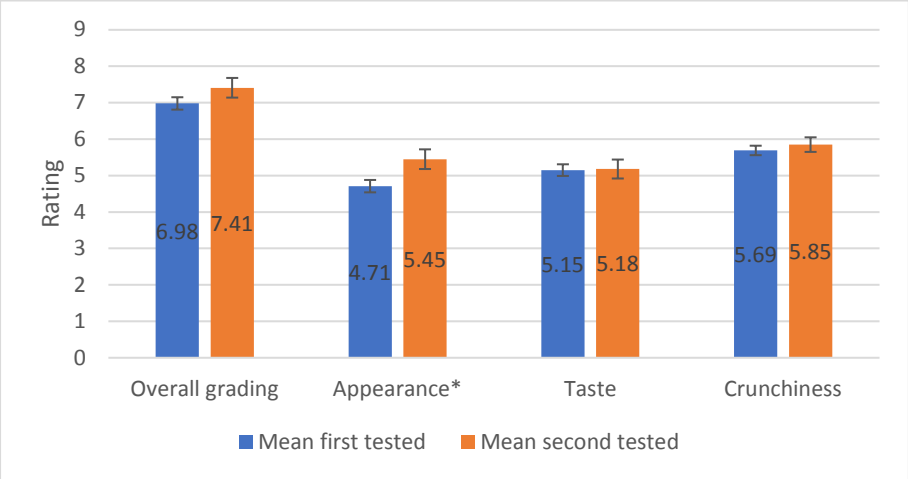


*Figure 10. Order effect of the vanilla cookie, with 24h between the cookies.. *: indicates a significant difference.*

3.6.2 Order effect in coffee cookies

For the coffee cookies, no significant order effects were found. Contrary to the results about the vanilla cookie, the trend is that the first product is rated higher for both the no time and 24 hours condition.

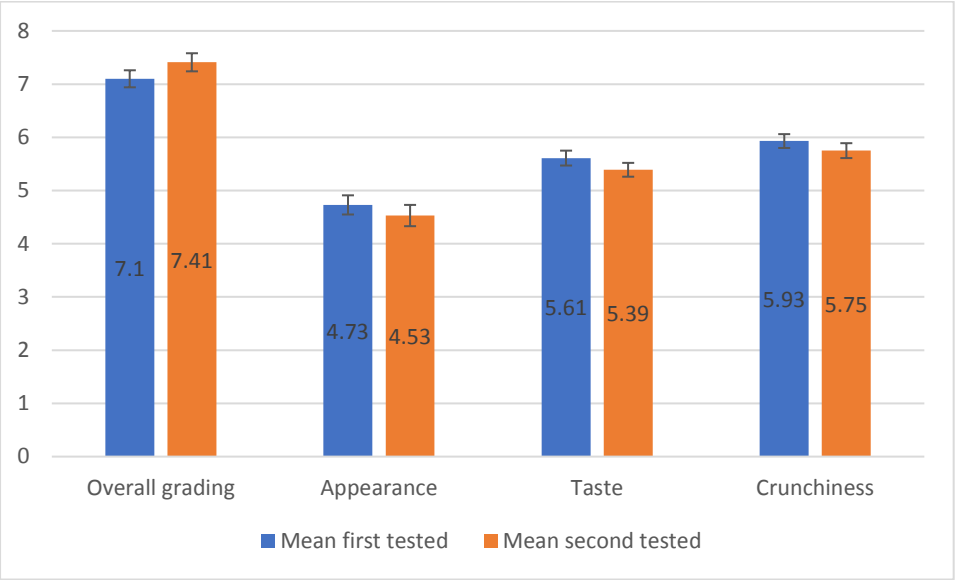


*Figure 11. Order effect of the coffee cookie, with no time between the cookies.. *: indicates a significant difference.*

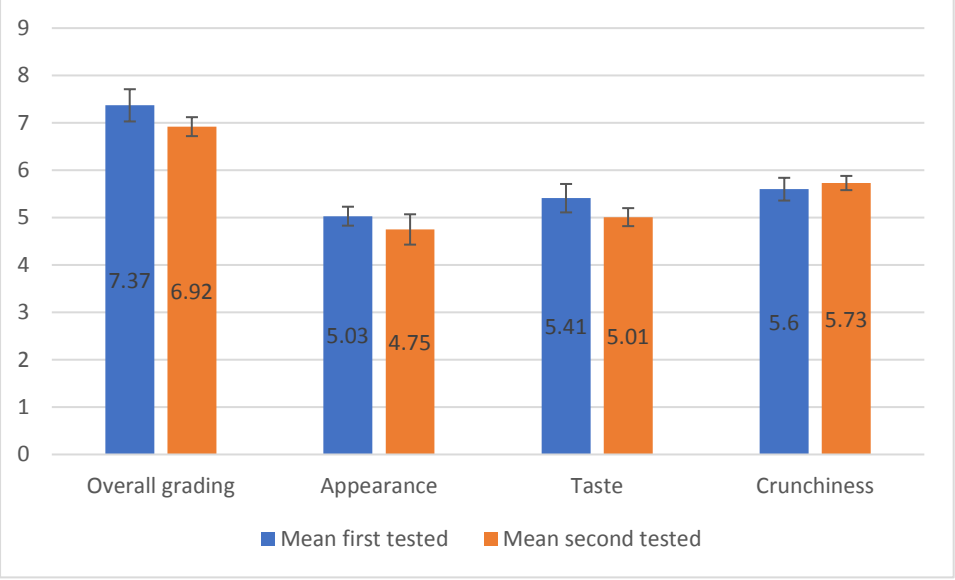


*Figure 12. Order effect of the coffee cookie, with 24 hours between the cookies.. *: indicates a significant difference.*

3.6.3 Discussion of the order effect

The most interesting finding is that there is a trend in the direction of the order effect in the vanilla cookie between the t=0 and t=24h condition. The order effect is known as that the first sample receives higher scores than subsequent samples (Mantonakis et al., 2010). This is the case if there is no time between samples, with a significant difference for crunchiness at the vanilla cookie with no time between the samples. However, when there is 24 hours between the samples, the opposite trend is visible with a significant difference for appearance in the vanilla cookie.

A possible explanation for the different direction of the order effect might be the creation of expectations. As mentioned by Deliza & Macfie (1996), the liking of a product is dependent on the expectations before consumption. In the 24 hours between before the second cookie is evaluated, there is more time available to create expectations based on the previously evaluated cookie. It might also be possible that the subjects forget the previously evaluated and thereby have lower expectations which meet or exceed the expectations.

The different direction in the order is not visible in the coffee cookie. For both t=0 and t=24h, it is the trend that the first product is rated higher. This is in line with previously published studies about the order effect (Mantonakis et al., 2010; Kofes et al., 2009).

Chapter 4: Conclusions

4.1 Conclusions

This study was to set out the effect of the time interval between samples on consumer evaluation. This was investigated by an experiment. It was found that in the case with small cookies, the consumer chooses to evaluate the cookies directly after each other. The current research showed that generally a monadic sequential design with no time between two samples, shows more significant differences than a monadic sequential design with 24 hours between the samples. So, a monadic sequential design with no time between the samples is more suitable to find differences between samples. However, the 24 hour between the samples may be a better way to resemble the actual consumption behaviour, because it is assumed not to be common to eat two different cookies directly after each other.

The most surprising finding in this study is the trend that the order effect has opposite directions in the t=0 and t=24h condition for the vanilla cookie. In a monadic sequential design with no time between the samples, the first evaluated vanilla cookie tended to be rated higher than the second evaluated vanilla cookie. In a monadic sequential design with 24h between the products, the first evaluated vanilla cookie tended to be evaluated lower than the second evaluated vanilla cookie. This is the opposite direction, than the order effect which is described in literature (a.o. Kofes et al., 2009; Mantonakis et al., 2010; Welch & Swift, 1992). However, the coffee cookie did not show an opposite direction in the order effect.

The effect of the previously evaluated sample on the subsequent sample was evaluated. In general, no significant differences in liking were observed between the second evaluated which was rated directly after the first sample and the second evaluated sample which was evaluated 24 hours after the first evaluated sample. This indicates that the time between the samples has a limited effect on the evaluation of the second sample.

In general, the findings underpin the importance of the methodology in measuring consumer hedonic response. Also the importance of a detailed methodology chapter in consumer sensory evaluation is stressed, since the methodology used influences the results.

4.2 Recommendations for further research

There is a strong need for more research on the effect of research methodology in sensory consumer research. The present study adds to the growing body of literature that the methodology significantly influences the results. As the results of sensory evaluation are used in NPD related decisions, it is important to know what the researcher is measuring in a sensory evaluation.

Currently, literature mentions consumer evaluation as an important success factor (Kahn et al., 2006). However, how this consumer evaluation should be executed is unclear because the link between consumer evaluation in a research setting and the actual consumer evaluation after purchase has received very limited attention. Therefore, it should be investigated to what extent consumer evaluation predicts consumer evaluation after actual purchase. HUT is a first step towards a more realistic evaluation of products, but still consumers are aware that they are being evaluated. New techniques and methodologies should be developed in which data can be obtained about the consumer liking of a product, and the consumer is not aware of being researched. This can lead to less biased results, and decreases the influence of the method on the results.

The last recommendation for further research is to investigate the underlying mechanisms in the order effect. The opposite order effect was found in the vanilla cookie at the 24 hours interval, but

was not found in the coffee cookie 24 hours interval. Further research can investigate why this only occurs at the vanilla cookie and what underlying mechanism leads to this difference.

LITERATURE

Alex, T., Gautreau, O., Kern, M., & Manfredi, P. (n.d.). Monadic vs. Monadic Sequential. Best practice guidance derivate from studies conducted in the period 2005 to 2012. Retrieved January 9, 2018, from http://www.samresearch.com/fileadmin/_migrated/content_uploads/2012_SAM_Monadic_Monadic_Sequential.pdf

Almeida, T. C., Cubero, E., & O'mahony, M. (1999). Same-different discrimination tests with interstimulus delays up to one day. *Journal of Sensory Studies*, *14*(1), 1-18.

Atradius. (2017, December 12). Market Monitor Food Netherlands 2017. Retrieved January 16, 2018, from Market Monitor Food Netherlands 2017

Barczak, G., & Kahn, K. B. (2012). Identifying new product development best practice. Business horizons, 55(3), 293-305.

Boer, E. (2017, September). Statistics in sensory science. Lecture presented at Sensometrics, HNE-30506 in Wageningen University, Wageningen.

Boutrolle, I., Arranz, D., Rogeaux, M., & Delarue, J. (2005). Comparing central location test and home use test results: Application of a new criterion. Food Quality and Preference, 16(8), 704-713.

Boutrolle, I., Delarue, J., Arranz, D., Rogeaux, M., & Köster, E. P. (2007). Central location test vs. home use test: Contrasting results depending on product type. Food Quality and Preference, 18(3), 490-499.

Calvin, L. D., & Sather, L. A. (1959). A comparison of student preference panels with a household consumer panel. Food Technology, 13, 469–472.

Daillant-Spinnler, B., & Issanchou, S. (1995). Influence of label and location of testing on acceptability of cream cheese varying in fat content. Appetite, 24, 101–106.

Deliza, R., & MacFie, H. J. (1996). The generation of sensory expectation by external cues and its effect on sensory perception and hedonic ratings: a review. Journal of Sensory Studies, 11(2), 103-128.

Elam, C., Stratton, T., & Gibson, D. D. (2007). Welcoming a new generation to college: The millennial students. Journal of College Admission, 195, 20-25.

Euromonitor. (2017). Top 10 Global Consumer Trends for 2017. Retrieved February 30, 2018, from http://go.euromonitor.com/white-paper-2017-top-10-global-consumer-trends-EN.html

Giacalone, D. (2018). Sensory and Consumer Approaches for Targeted Product Development in the Agro-Food Sector. In Case Studies in the Traditional Food Sector (pp. 91-128). ISO 690

Grunert, K. G., & van Trijp, H. C. M. (2014). Consumer-oriented new product development. Encyclopedia of agriculture and food systems, 2, 375-386.

Harker, F. R., Gunson, F. A., Brookfield, P. L., & White, A. (2002). An apple a day: the influence of memory on consumer judgment of quality. Food quality and preference, 13(3), 173-179.

Hellemann, U., Mela, D. J., Aaron, J. I., & Evans, R. E. (1992). Role of fat in meal acceptance. In: Proceedings of the 3rd R.M. Pangborn memorial symposium on advances in sensory food science, Helsinki, Finland, 2–6 August.

Helson, H. H. 1964. Adaptation-Level Theory. Harper & Rowe,New York.

Kahn, K. B., Barczak, G., & Moss, R. (2006). Perspective: establishing an NPD best practices framework. Journal of Product Innovation Management, 23(2), 106-116.

Grace-Martin, K. (n.d.). Why report estimated marginal means in SPSS GLM? Retrieved March 21, 2018, from Why report estimated marginal means in SPSS GLM?

Karin, W., Annika, Å., & Anna, S. (2015). Exploring differences between central located test and home use test in a living lab context. International Journal of Consumer Studies, 39(3), 230-238.

Kofes, J., Naqvi, S., Cece, A. and Yeh, M. 2009. Understanding Presentation Order Effects and Ways to Control Them in Consumer Testing. Paper presented at the 8th Pangborn Sensory Science Symposium, Florence, Italy.

Kozlowska, K., Jeruszka, M., Matuszewska, I., Roszkowski, W., Barylko-Pikielna, N., & Brzozowska, A. (2003). Hedonic tests in different locations as predictors of apple juice consumption at home in elderly and young subjects. Food Quality and Preference, 14, 653–661.

Lawless, H. T., & Heymann, H. (2010). Sensory evaluation of food: principles and practices. Springer Science & Business Media.

Luning & P. A., Marcelis (2009). Food quality management: a techno-managerial approach. Wageningen Pers.

MacFie, H. (Ed.). (2007). Consumer-led food product development. Elsevier.

Mantonakis, A., Rodero, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. Psychological Science, 20(11), 1309-1312.

McDaniel, M. R., & Sawyer, F. M. (1981). Preference testing of whiskey sour formulations: magnitude estimation versus the 9- point hedonic. Journal of Food Science, 46, 182–185.

McEwan, J. A. (1997). A comparative study of three product acceptability trials. Food Quality and Preference, 8, 183–190.

Miller, P. G., Nair, J. H., & Harriman, A. J. (1955). A household and a laboratory type of panel for testing consumer preference. Food Technology, 9, 445–449.

Mörlein, D., Schiermann, C., Meier-Dinkel, L., Trautmann, J., Wigger, R., Buttinger, G., & Wicke, M. (2015). Effects of context and repeated exposure on food liking: The case of boar taint. Food Research International, 67, 390-399.

Murphy, E. F., Clark, B. S., & Berglund, R. M. (1958). A consumer survey versus panel testing for acceptance evaluation of Maine sardines. Food Technology, 12, 222–226.

Olabi, A., & Lawless, H. T. (2008). Persistence of context effects after training and with intensity references. Journal of food science, 73(4).

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of applied psychology, 88(5), 879.

Posri, W., MacFie, H., & Henson, S. (2001). Improving the predictability of consumer liking from central location test in tea. In Proceedings of the 4th R.M. Pangborn memorial symposium: A sense odyssey, Dijon, France, 22–26 July.

Pound, C., Duizer, C., & McDowell, K. (2000). Improved consumer product development. Part one. Is a laboratory necessary to assess consumer opinion? British Food Journal, 102, 810–820.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta psychologica, 104(1), 1-15.

Rankin, K. M., & Marks, L. E. (1991). Differential context effects in taste perception. Chemical Senses, 17, 617–629

Sosa, M., Martinez, C., Marquez, F., & Hough, G. (2008). Location and scale influence on sensory acceptability measurements among low-income consumers. Journal of sensory studies, 23(5), 707-719.

Shepherd, R., & Griffiths, N. M. (1987). Preferences for eggs produced under different systems assessed by consumer and laboratory panels. Lebensmittelwissenshaft und Technologie, 20, 128–132.

Stone, H., Sidel, J., Oliver, S., Woolsey, A., & Singleton, R. C. (2008). Sensory evaluation by quantitative descriptive analysis. Descriptive Sensory Analysis in Practice, 23-34.

Sveinsdottir, K., Martinsdottir, E., Thorsdottir, F., Schelvis, R., Kole, A., & Thorsdottir, I. (2010). Evaluation of farmed cod products by a trained sensory panel and consumers in different test settings. Journal of sensory studies, 25(2), 280-293.

Ueland, Ø. (2007). Gender differences in food choice. In Understanding consumers of food products (pp. 316-328).

Van Kleef, E., Van Trijp, H. C., & Luning, P. (2005). Consumer research in the early stages of new product development: a critical review of methods and techniques. Food quality and preference, 16(3), 181-201.

Welch, J. L., & Swift, C. O. (1992). Question order effects in taste testing of beverages. Journal of the Academy of Marketing Science, 20(3), 265-268.

Zellner, D. A., Allen, D., Henley, M., & Parker, S. (2006). Hedonic contrast and condensation: good stimuli make mediocre stimuli less good and less different. Psychonomic Bulletin and Review, 13(2), 235–239

Appendix 1 – Overview of studies on HUT vs CLT tests

Study	**Tested products**		**Sample presentation**		**Method of HUT**	**Outcome**
		Lab	***CLT***	***HUT***		
Miller et al. (1955)	18 Paired samples of noodle soup	Simultaneous	x	Monadic sequential	Unknown	For different products there was no different outcome. For similar products there was a different outcome.
Murphy et al. (1958)	6 sardines	Simultaneous	x	Simultaneous	Unknown	Yes: home scores > lab scores
Calvin and Sather (1959)	Paired samples of 15 different food products.	Simultaneous	x	Monadic sequential	Unknown	Yes, for 4 pairs of products. Home use more discriminant.
McDaniel and Sawyer (1981)	6 pairs of whisky	Simultaneous	x	Simultaneous	Unknown	Yes, HUT more discriminant
Shephard and Griffiths (1987)	3 eggs	monadic sequential	x	Monadic sequential	Unknown	No differences.
Helleman n et al. (1992)	Real meal	N/A	x	N/A	Unknown	Yes, lab scores higher than home and cafetaria scores
Daillant-Spinnler and Issancho u (1995)	3 fat versions of cream cheese	monadic sequential	x	Monadic sequential	Unknown	Generally no, yes only for high fat version in which lab>home

Study	Tested products		Sample presentation		Method of HUT	Outcome
		Lab	***CLT***	***HUT***		
McEwan (1997)	10 crackers	monadic sequential	monadic sequential	Monadic sequential	Several visits when product were brought, time between is unknown	Yes, ranking order changes
Pound et al. (2000)	3 milk chocolate bars	monadic sequential	monadic sequential	Monadic sequential/simultaneous	A bag with the 3 samples brought home, up to consumers when to eat but within two weeks.	No
Posri et al. (2001)	4 teas	x	Unknown	Unknown	Unknown	Yes, ranking order changes
Kozlowksa et al. (2003)	5 apple juices	monadic sequential	monadic sequential	Monadic sequential	Drink three juices a day, rate day after	Yes, samples with no sugars. Home scores > lab scores and CLT
Boutrolle et al. (2007)	a pair of salted crackers, sparkling water and a milk beverage	x	Monadic sequential	Monadic sequential	Products were brought home several times, up to respondent when to eat.	Yes, higher liking for HUT.
Sosa et al. (2008)	2 powders soups and 2 chocolate milks	x	simultaneous	Simultaneous	Received paired samples	Yes, higher scores for HUT. CLT more discriminative.
Sveinsdottir et al. (2010)	Two samples of COD	x	Monadic sequential	monadic sequential	One sample per week, eat day when received	Yes, CLT rating in Liking lower than in HUT.
Morlein et al. (2014)	2 types of boar meat	monadic	monadic	x	Receive 1 piece of meat	Yes, higher dislike in sensory lab
Karin et al. (2015)	5 meal replacers	x	unknown	Unknown	Up to respondents, they receive them at once	Yes, higher scores for HUT.

Appendix 2 – Instructions to subjects

Instructions to the respondent - back

Hello!

Thank you for participating in the cookie study!
In this study we ask you to eat and evaluate two cookies.
You can start with this experiment at a moment of choice. But first read these instructions.

When you start the evaluation: first read the instructions with attention. Only open the envelope with the product if you are instructed to.
In the plastic bag are two cookies, which are individually packed.

If you have any questions, you can contact me at: bart.vangorcum@wur.nl.

Best regards,
Bart van Gorcum

When you are ready to start: **Please follow the instructions below.**
Keep in mind that both cookies should be evaluated in the same conditions. For example: if you dip the first cookie in your coffee during the evaluation, the second cookie should be evaluated in the same way.

Evaluation step 1

1. Open the plastic bag.
2. You will see two closed envelopes, **do not open them yet!**
3. Go to: www.cookietest.nl (do not forget to add ***www.***)
4. Open the packaging with code **570653** Open only this packaging.
5. Evaluate the product using the online form.
6. After the evaluation is finished, the left-over of the cookie has to be thrown away. The first tested cookie must not be in sight during the evaluation of the second cookie.

Evaluate the second cookie 24 hours after evaluating the first cookie.

Evaluation step 2

7. Go to: www.cookietest.nl (do not forget to add ***www.***)
8. Open the envelope with code **570742**
9. Evaluate the product using the online form.
10. The sensory evaluation is finished, thanks a lot for participating!

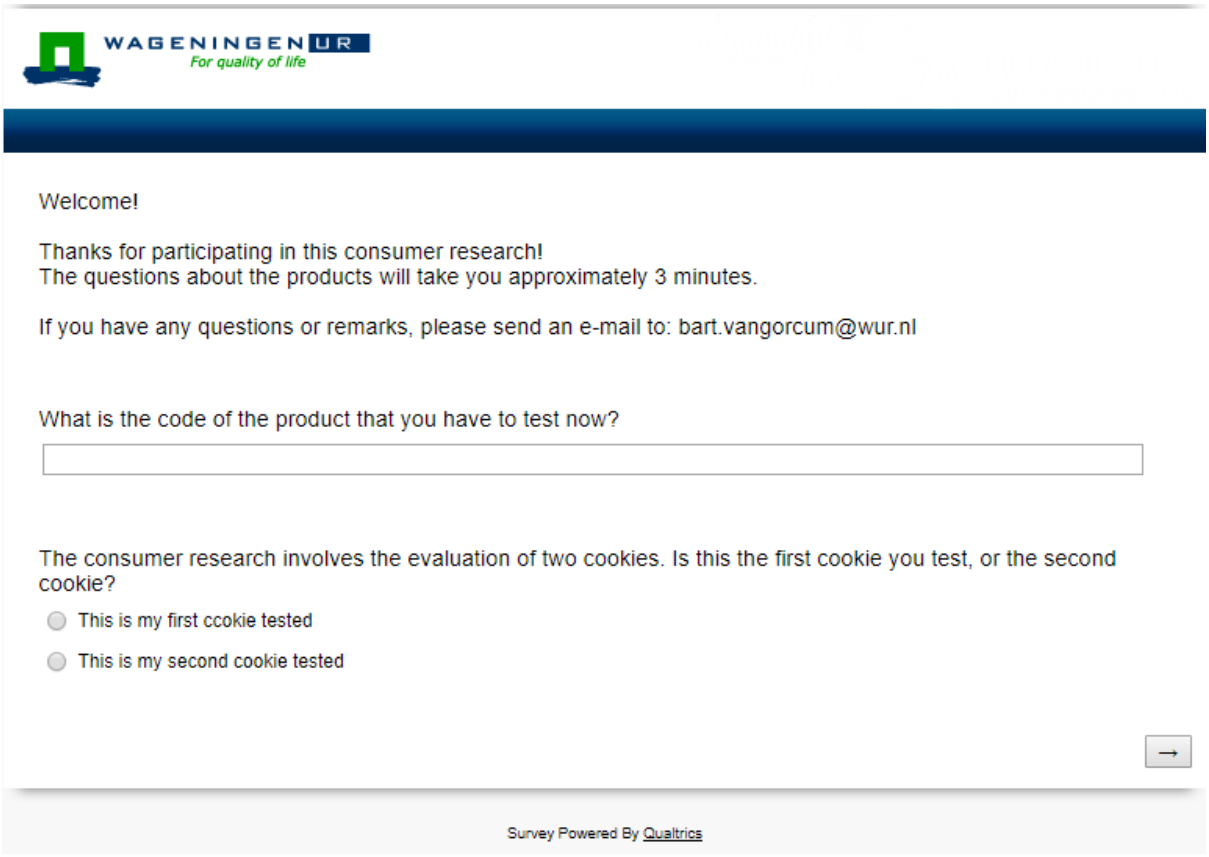
Instructions to the respondent - front

PLEASE READ THE OTHER SIDE OF THIS PAPER BEFORE OPENING THE PACKAGING

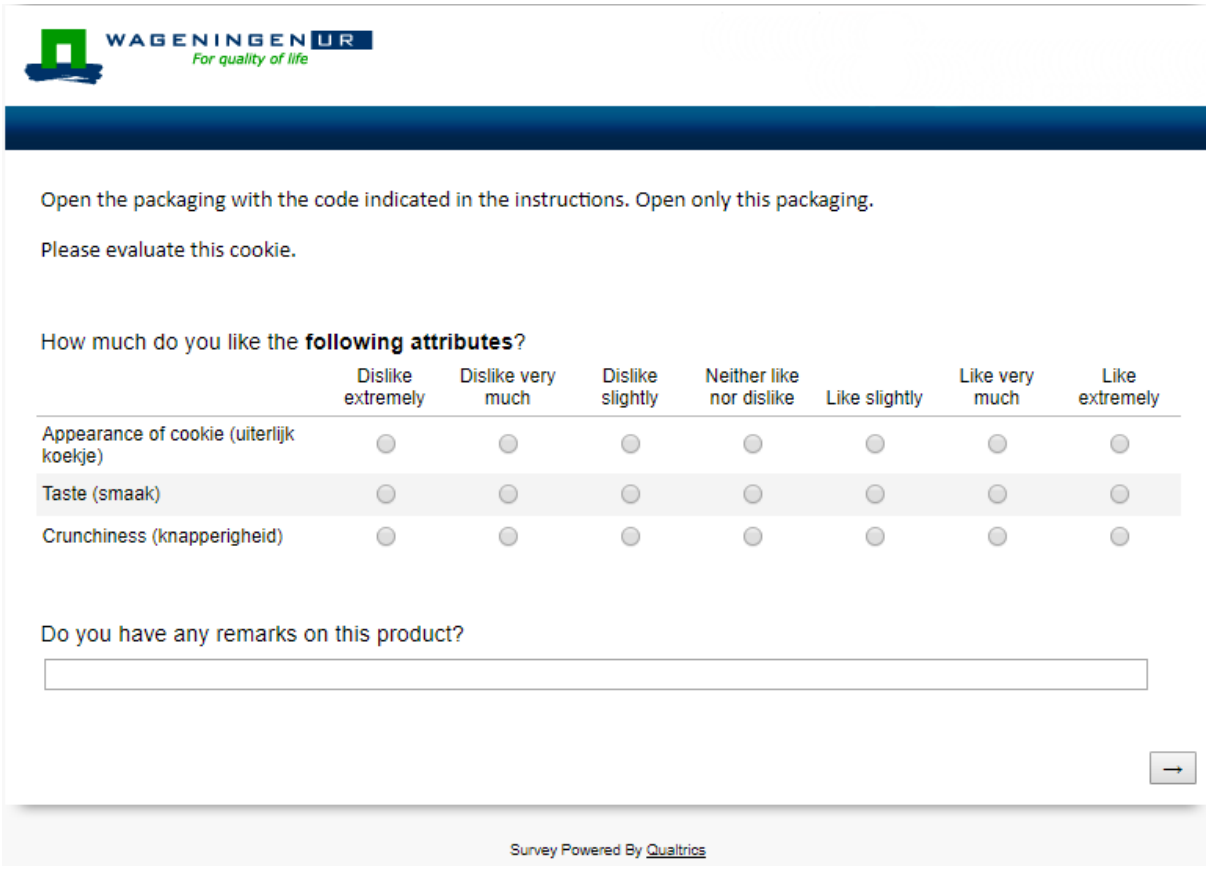
DO THE TEST INDIVIDUALLY AND START AT LATEST AT WEDNESDAY 28th OF FEBRUARY

Appendix 3 – Online evaluation form

Page 1

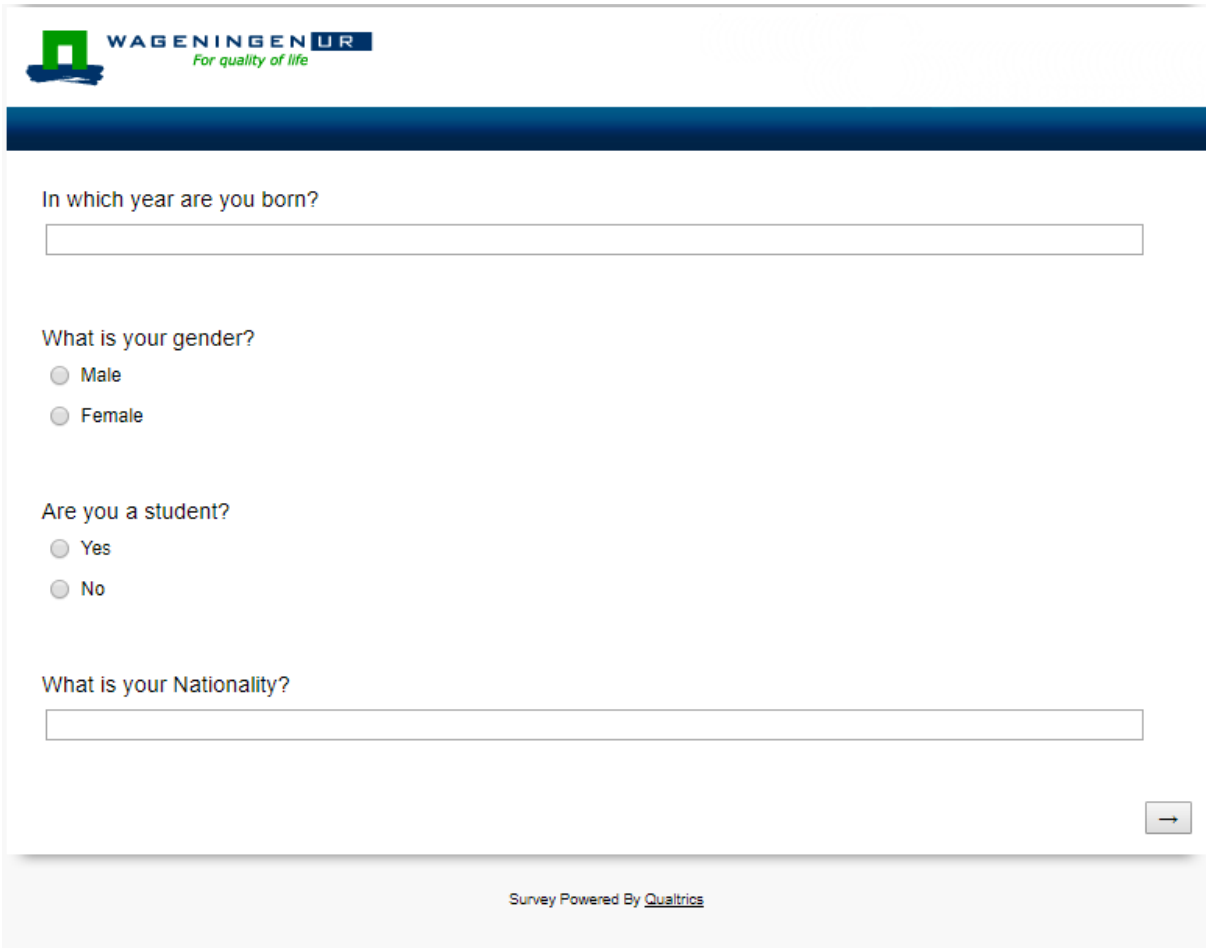


Page 2

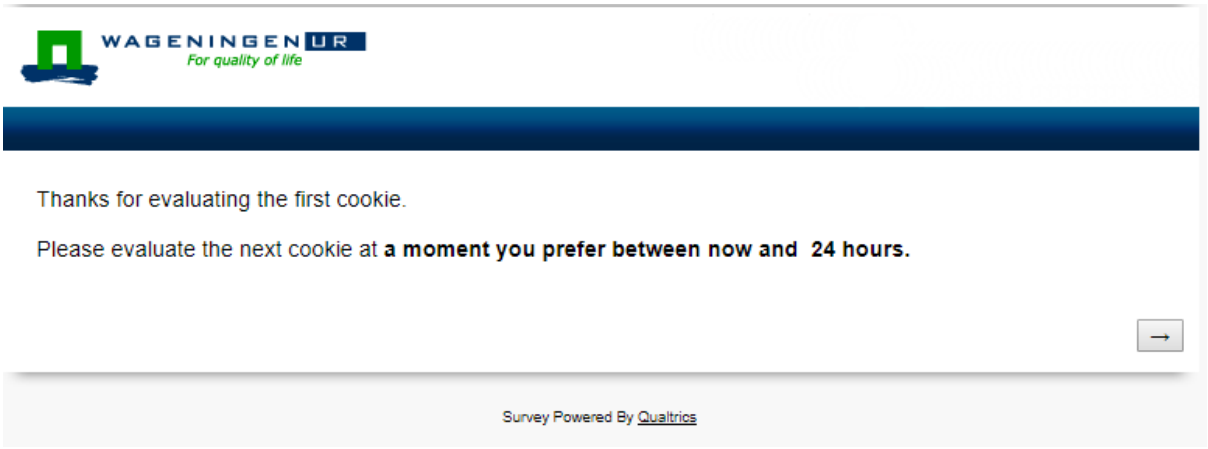


Page 3

Page 4



Page 5 – Instructions for the consumption of the second cookie (condition dependent)



Appendix 4 - Reason for time choice

Term	Code
Convenience	1
To make sure not to forget	2
For methodological reasons	3
Curiosity	5
Hungriness	6
Eager for a cookie	7
No reason mentioned	4

Label	**Motivation by subject**
1	right after the first cookie, since I was already busy and on my laptop
1	Right after, because it was most convenient for me. I was hungry and them I ate it
1	Break while writing my thesis I evaluated the second cookie right after, because of laziness. Perhaps partly hungry as well.
1	Right after;easiest/fastest
1	when I am free
1	just convenience
1, 6	I won't have much time within 24hours, so it was just the most practical to do it straight away. Besides, I was hungry.
1	After the 1st one, more practical
1	I forgot and when I tought about it, I tested the second cookie
1,7	When I had the time and was in the mood for a cookie, I thought it was a good moment to try the second one.
1	Based on time, it was easier to do it immediately
2	I waited a few minutes after evaluating the first cookie, then I evaluated the second one. I did this on purpose because I thought that if I waited too long I might forget to do the second evaluation.
3	Directly, since if I postpone to taste the second cookie, i might be forgot how the different taste from the first
1, 2, 7	I had time now, and otherwise I would probably forget tomorrow. And I felt like eating cookies :)
2	directly, to get it over with and not to forget.
2	I chose to do it right after the first cookie because I was afraid I would forget the second.
2	Immediately after the first cookie, because I was scared I would forget to evaluate the other cookie if I would do it later.
3	I chose to do the second test when the circumstances were as equal as when i did the first test
3	10 minutes after the first one and drank some water first
3	Took a bit of water and then right after the first
3	Right after the first one, after I drank some water
3	to look at the same vectors one trying the first one, like did i eat before tasting it etc
1, 2, 3	first cookie befor working and second after one or two hours,when i finished so that i dont forget to do it, and i drank water before so i dont have a different taste of something that i ate in advance.
4	straight after

Label	**Motivation by subject**
4	Directly after the first
4	Direct after first cookie
4	I choose to evaluate it right after the first cookie
4	Right after the first cookie
4	i just had a small break inbetween the two cookies
4	right after
4	Just after the first one, was taking some time to evaluate both after each other
4	the minute after
4	Directly after it
7	Liked the first cookie
4	I just did it right after
4	I decided to taste the second cookie just after the first one.
4	Just after tasting the first
4	I chose to do this directly after the first cookie.
4	just immediately
4	Just after the first evaluate.
4	After getting the cookies
4	Time
4	I didn't realy choose a time to evaluate the second, I did it just right after the first
4	Right after the first
4	Directly after eating the first cookie
4	After I finished the other cookie
4	I did it right after evaluating the first cookie
4	i did it right after the first one
4	Right after the other
4	2 minutes after
4	I choose to test the second cookie approximately 1 hour later than the first cookie
5	I just wanted to try both so I tried them within 5 minutes
5	Straight after, I was curious what the second cookie would be
5	directly, since I was curious for the second one
7	I really liked the first biscuit and I tried immediatly the otherone
5	Curiosity
6	I was bored one hour after the first cookie and got a little bit hungry again.
6	I was hungry
6	When I was in need for another cookie
6	I was hungry
6	Hungry
6	Ik had trek zo vlak voor het avondeten dus dacht ik test beide koekjes achter elkaar, want ze waren vrij klein en ik dacht zal ik die tweede vanavond nemen, nee ik neem m nu
6	I was hungry

Label	**Motivation by subject**
6	I did that few minutes afterwards because I got hungry
6	I tried the second cookie directly after the first one, because I was quite hungry and the cookies were not too big
6	right after the first one, because I was hungry :)

APPENDIX 5 – SPSS OUTPUT RATINGS PER LEVEL

5.1.1 Overall grading

5.1 Output at timelevel = 0.

Attribute	Scale	Vanilla cookie	St dev	Coffee cookie	St dev	Mean difference	F-Value	Significance
Overall liking	Grade from 1 to 10	7.19	0.14	7.23	0.14	-0.04	0.097	0.756
Appearance	7 point Likert	4.92	0.13	4.58	0.13	0.34	4.483	0.037
Taste	7 point Likert	5.44	0.11	5.47	0.11	-0.03	0.044	0.834
Crunchiness	7 point Likert	5.57	0.1	5.85	0.10	-0.28	4.914	0.030

5.2.1 Overall grading

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	79.000	4036.954	.000
Order	1	79.000	2.865	.094
Gender	1	79.000	.151	.699
Order * Product	1	79.000	1.007	.319
Product	1	79.000	.097	.756
Gender * Product	1	79.000	2.436	.123

a. Timelevel = Zero

b. Dependent Variable: Overall_grading.

5.1.2 Appearance

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	79	2111.907	.000
Order	1	79	.503	.480
Gender	1	79	.548	.461
Order * Product	1	79.000	1.368	.246
Product	1	79.000	4.483	.037
Gender * Product	1	79.000	.224	.637

a. Timelevel = Zero

b. Dependent Variable: Appearance.

5.1.3 Taste

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	79	4719.357	.000
Order	1	79.000	.026	.871
Gender	1	79.000	.002	.965
Order * Product	1	79.000	5.435	.022
Product	1	79.000	.044	.834
Gender * Product	1	79.000	1.569	.214

a. Timelevel = Zero

b. Dependent Variable: Tasty.

5.1.4 Crunchiness

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	79	6524.347	.000
Order	1	79	3.081	.083
Gender	1	79	.001	.972
Order * Product	1	79	9.770	.002
Product	1	79	4.914	.030
Gender * Product	1	79	1.688	.198

a. Timelevel = Zero

b. Dependent Variable: Crunchiness.

5.2 Output at time level = 24 hours

Attribute	Scale	Vanilla cookie	St dev	Coffee cookie	St dev	Mean difference	F-Value	Significance
Overall liking	Grade from 1 to 9	7.11	0.17	7.08	0.17	0.03	0.026	0.872
Appearance	7 point Likert	4.98	0.16	4.9	0.08	0.08	0.158	0.692
Taste	7 point Likert	5.14	0.15	5.14	0.15	0	0	0.989
Crunchiness	7 point Likert	5.67	0.12	5.71	0.12	-0.04	0.077	0.782

5.2.1 Overall grading

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	60	3242.677	.000
Order	1	60	1.033	.314
Gender	1	60	3.167	.080
Order * Product	1	60.000	.011	.916
Product	1	60.000	.026	.872
Gender * Product	1	60.000	.408	.526

a. Timelevel = 24 hours

b. Dependent Variable: Overall_grading.

5.2.2 Appearance

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	60	1579.374	.000
Order	1	60	2.764	.102
Gender	1	60	1.825	.182
Order * Product	1	60.000	.134	.716
Product	1	60.000	.158	.692
Gender * Product	1	60.000	2.914	.093

a. Timelevel = 24 hours

b. Dependent Variable: Appearance.

5.2.3 Taste

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	120	2244.700	.000
Order	1	120	.162	.688
Gender	1	120	3.776	.054
Order * Product	1	120	.473	.493
Product	1	120	.000	.989
Gender * Product	1	120	.392	.533

a. Timelevel = 24 hours

b. Dependent Variable: Tasty.

5.2.4 Crunchiness

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	60	3793.125	.000
Order	1	60	.196	.660
Gender	1	60	5.380	.024
Order * Product	1	60	.299	.587
Product	1	60	.077	.782
Gender * Product	1	60	.132	.718

a. Timelevel = 24 hours

b. Dependent Variable: Crunchiness.

APPENDIX 6 – Effect of previously evaluated product

6.1 The second rated vanilla Cookie

Attribute	Scale	Mean second tested at t=0	SE	Mean second tested at t=24h	SE	Mean	F-Value	Significance
Overall liking	Grade from 1 to 9	6.92	0.2	7.4	0.53	-0.48	1.19	0.280
Appearance	7 point Likert	4.9	0.15	5.45	0.25	-0.55	3.002	0.088
Taste	7 point Likert	5.25	0.19	5.18	0.33	0.07	0.04	0.842
Crunchiness	7 point Likert	5.25	0.13	5.85	0.23	-0.6	4.637	0.035

6.1.1 Overall grading

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	61	1037.444	.000
Condition	1	61	1.190	.280
Gender	1	61	.007	.933
Condition * Gender	1	61	.009	.924

a. Product = vanilla, Vanilla first or second? = Vanilla second

b. Dependent Variable: Overall_grading.

Estimates[a,b]

Condition	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
4.00	6.918	.221	61	6.476	7.359
6.00	7.403	.386	61	6.631	8.174

a. Product = vanilla, Vanilla first or second? = Vanilla second

b. Dependent Variable: Overall_grading.

6.1.2 Appearance

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	61.000	1041.841	.000
Condition	1	61.000	3.002	.088
Gender	1	61.000	.030	.864
Condition * Gender	1	61.000	.235	.630

a. Product = vanilla, Vanilla first or second? = Vanilla second

b. Dependent Variable: Appearance.

Estimates[a,b]

Condition	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
4.00	4.895	.159	61.000	4.577	5.213
6.00	5.450	.278	61.000	4.894	6.006

a. Product = vanilla, Vanilla first or second? = Vanilla second

b. Dependent Variable: Appearance.

6.1.3 Taste

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	61	727.862	.000
Condition	1	61	.040	.842
Gender	1	61	.023	.879
Condition * Gender	1	61	.567	.454

a. Product = vanilla, Vanilla first or second? = Vanilla second

b. Dependent Variable: Tasty.

Estimates[a,b]

Condition	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
4.00	5.252	.192	61	4.869	5.636
6.00	5.175	.336	61	4.504	5.846

a. Product = vanilla, Vanilla first or second? = Vanilla second

b. Dependent Variable: Tasty.

6.1.4 Crunchiness

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	61.000	1600.272	.000
Condition	1	61.000	4.637	.035
Gender	1	61.000	.919	.342
Condition * Gender	1	61.000	.015	.903

a. Product = vanilla, Vanilla first or second? = Vanilla second

b. Dependent Variable: Crunchiness.

Estimates[a,b]

Condition	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
4.00	5.252	.138	61.000	4.977	5.528
6.00	5.850	.241	61.000	5.368	6.332

a. Product = vanilla, Vanilla first or second? = Vanilla second

b. Dependent Variable: Crunchiness.

6.2 The second rated Coffee Cookie

Attribute	Scale	Mean second sample at t=0	SE	Mean second sample at t= 24h	SE	Mean difference	F-value	Significance
Overall liking	Grade from 1 to 9	7.41	0.18	6.92	0.18	0.49	3.342	0.071
Appearance	7 point Likert	4.53	0.2	4.75	0.201	-0.22	0.561	0.456
Taste	7 point Likert	5.39	0.16	5.00	0.16	0.39	2.484	0.119
Crunchiness	7 point Likert	5.75	0.13	5.73	0.13	0.02	0.017	0.896

6.2.1 Overall grading

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	76.000	2907.731	.000
Condition	1	76.000	3.342	.071
Gender	1	76.000	.340	.561
Condition * Gender	1	76.000	1.610	.208

a. Product = Cinnamon, Vanilla first or second? = Vanilla first

b. Dependent Variable: Overall_grading.

Estimates[a,b]

Condition	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
3.00	7.409	.187	76.000	7.037	7.781
5.00	6.923	.189	76.000	6.547	7.300

a. Product = Cinnamon, Vanilla first or second? = Vanilla first

b. Dependent Variable: Overall_grading.

6.2.2 Appearance

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	76	950.772	.000
Condition	1	76	.561	.456
Gender	1	76	.386	.536
Condition * Gender	1	76	.243	.623

a. Product = Cinnamon, Vanilla first or second? = Vanilla first

b. Dependent Variable: Appearance.

Estimates[a,b]

Condition	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
3.00	4.525	.211	76	4.104	4.946
5.00	4.750	.214	76	4.324	5.176

a. Product = Cinnamon, Vanilla first or second? = Vanilla first

b. Dependent Variable: Appearance.

6.2.3 Taste

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	76.000	1727.013	.000
Condition	1	76.000	2.484	.119
Gender	1	76.000	.120	.730
Condition * Gender	1	76.000	2.248	.138

a. Product = Cinnamon, Vanilla first or second? = Vanilla first

b. Dependent Variable: Tasty.

Estimates[a,b]

Condition	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
3.00	5.394	.176	76.000	5.044	5.744
5.00	5.000	.178	76.000	4.646	5.354

a. Product = Cinnamon, Vanilla first or second? = Vanilla first

b. Dependent Variable: Tasty.

6.2.4 Crunchiness

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	76	4165.459	.000
Condition	1	76	.017	.896
Gender	1	76	2.039	.157
Condition * Gender	1	76	.017	.896

a. Product = Cinnamon, Vanilla first or second? = Vanilla first

b. Dependent Variable: Crunchiness.

Estimates[a,b]

Condition	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
3.00	5.754	.125	76	5.505	6.003
5.00	5.731	.127	76	5.479	5.983

a. Product = Cinnamon, Vanilla first or second? = Vanilla first

b. Dependent Variable: Crunchiness.

APPENDIX 7 – Order effect

Vanilla cookies								
Time interval	**Attribute**	**Mean first tested**	**SE**	**Mean second tested**	**SE**	**Mean difference**	**F-value**	**Significance**
T=0	Overall grading	7.49	0.22	6.92	0.21	0.57	3.574	0.062
	Appearance	4.96	0.19	4.90	0.18	0.06	0.074	0.787
	Taste	5.61	0.18	5.25	0.17	0.36	2.182	0.144
	Crunchiness	5.87	0.14	5.25	0.13	0.62	10.151	0.002
Time interval	**Attribute**	**Mean first tested**	**SE**	**Mean second tested**	**SE**	**Mean difference**	**F-value**	**Significance**
T=24h	Overall grading	6.98	0.17	7.41	0.27	-0.43	1.832	0.181
	Appearance	4.71	0.17	5.45	0.27	-0.74	5.518	0.022
	Taste	5.15	0.16	5.18	0.26	-0.03	0.005	0.945
	Crunchiness	5.69	0.13	5.85	0.2	-0.16	0.431	0.514

7.1 No time between samples

7.1.1 Overall grading

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	78	3873.275	.000
Order	1	78	1.708	.195
Gender	1	78	.258	.613
Order * Gender	1	78	1.661	.201

a. Product = Cinnamon, Timelevel = Zero

b. Dependent Variable: Overall_grading.

7.1.2 Appearance

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	78.000	1202.663	.000
Order	1	78.000	.590	.445
Gender	1	78.000	.527	.470
Order * Gender	1	78.000	3.927	.051

a. Product = Cinnamon, Timelevel = Zero

b. Dependent Variable: Appearance.

7.1.3 Taste

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	78	3297.016	.000
Order	1	78	1.296	.258
Gender	1	78	.632	.429
Order * Gender	1	78	5.290	.024

a. Product = Cinnamon, Timelevel = Zero

b. Dependent Variable: Tasty.

7.1.4 Crunchiness

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	78.000	3705.464	.000
Order	1	78.000	.790	.377
Gender	1	78.000	.773	.382
Order * Gender	1	78.000	.322	.572

a. Product = Cinnamon, Timelevel = Zero

b. Dependent Variable: Crunchiness.

7.2 24 hours between samples

7.2.1 Overall grading

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	59.000	1308.378	.000
Order	1	59.000	1.279	.263
Gender	1	59.000	.128	.722
Order * Gender	1	59.000	.790	.378

a. Product = Cinnamon, Timelevel = 24 hours

b. Dependent Variable: Overall_grading.

7.2.2 Appearance

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	59.000	659.037	.000
Order	1	59.000	.522	.473
Gender	1	59.000	.000	.988
Order * Gender	1	59.000	.013	.908

a. Product = Cinnamon, Timelevel = 24 hours

b. Dependent Variable: Appearance.

7.2.3 Taste

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	59	829.039	.000
Order	1	59	1.226	.273
Gender	1	59	.131	.719
Order * Gender	1	59	.839	.364

a. Product = Cinnamon, Timelevel = 24 hours

b. Dependent Variable: Tasty.

7.2.4 Crunchiness

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	59.000	1651.719	.000
Order	1	59.000	.220	.641
Gender	1	59.000	2.786	.100
Order * Gender	1	59.000	.708	.403

a. Product = Cinnamon, Timelevel = 24 hours

b. Dependent Variable: Crunchiness.