Concept testing: the role of concept formulation for new product success

Master's Thesis by Giorgia Cianfrone

Student Number: 140594 - 3510

Supervisor:

Prof. Claus Varnes

Master of Science in Strategic Market Creation

Department of Economics and Business Administration

Copenhagen Business School

DATE, January 2019

Pages: 88

Characters: 152.367

ABSTRACT

The role of pre-market analyses has become crucial before the commercialization of a new product. In this phase, even more crucial is the role of concept testing, which is used by firms to get an initial reaction from customers before committing substantial funds to the product. However, companies are not satisfied with the results conducted with concept tests because there is a low accuracy of market predictions, hence it is not possible to clearly state whether the product would be a success or not. A way to overcome such issues, is to not underestimate the way in which the concept is presented to the respondents. This work is focused on stimuli design, namely the role of concept formulation when doing concept tests. It has been demonstrated that depending on the way in which the concept is presented, the overall evaluation varies. This has been proved conducting an online survey and statistically analysing the results. Through this research, an electronic device developed by Amazon and not available in Europe has been chosen and proposed to a selected audience. It has been presented in four different ways: statement, statement combined with visual and online reviews, visual and virtual prototype. The subsequent analyses confirm the different outcomes yield by each stimuli design.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND OF THE STUDY

In 1985, Marty McFly and Doc Emmett Brown were imagining the year 2015 as a new era with an increased use of robot technology, dog-walker drones, self-drying jacket, flying cars, self-tying shoelaces and so on. Even though we are three years behind the "Back to the Future" movie's expectations, we cannot deny that technology is advancing rapidly anyway.

Despite we have not had the chance to see flying cars yet, we are still living in a world that is continually evolving and that is bringing us new ways of living. We do not have time to familiarize with our current devices that we already need to implement new products in our everyday lives. The current scenario shows that goods are becoming obsolete faster, new markets are emerging, competition is spreading, and customers are continually asking for more.

Bearing this in mind, the importance of innovation is growing swiftly. Companies that want to gain a competitive advantage not only need to adapt to the existing environment and to their competitors' moves, but also they need to manage the constantly changing customers' needs. To survive, firms need to keep up and if they refuse the change, the risk is to gradually become extinct. Firms that fail to manage their product development activities start with a disadvantage and they are also risking the future (Fitzsimmons et al. 1991).

A way to be one step ahead is to develop and launch new products. The first step for a company is to evaluate the market response and the likability that the product will be a success: this can be done through a concept testing, an essential tool in the new product development. But before proceeding with a test that will state whether the market is ready or not to accept an innovation, some considerations about how to design the concept test should be made.

1.2 PROBLEM STATEMENT

A large amount of concept tests is made by companies every year, but the results of those tests or the ways in which they are conducted remain unknown. This information tends to be hide due to confidentiality reasons: not all the data coming from a company can be of public domain. Moreover, there is a lack of data about how concept tests should be designed (Peng and Finn, 2008). Therefore, it results difficult to provide a general blueprint on how to develop a concept test that every firm can use as a guide.

Concept tests are usually administered through a focus group or in-depth face-to-face interviews: in this way, it can be gathered a useful feedback for the product. Besides this great advantage, qualitative research is risky because sample size is small and participants often do not represent the entire market (Cooper, 2011). Furthermore, despite the apparently good responses, the majority of firms is not satisfied with the predictive performance of their concept tests (Peng and Finn, 2008).

An essential part of concept test is its design, the way in which the concept is shown. This is called "stimuli design", and it can be a statement, an image, a prototype. Typically, companies use one type of stimuli design at a time because using more than one simultaneously is time-consuming and expensive. Rarely it has been tested the same concepts on the same respondents, also because collecting data with traditional methods (e.g. face-to-face interviews) is quite difficult, especially from a logistic point of view (Peng and Finn, 2010).

Another issue is related to contradictory findings of a same topic. The term "contradictions" it is referred to the fact that there is no common agreement between the main authors who have focused on stimuli design. For instance, some of them come up with the same opinion on the fact that virtual testing environments and traditional testing environments yield the same responses, while others declare that the two different settings provide diverse results. Or, different findings have been gathered when the topic is changing the tone of stimuli design, some authors insist on the importance of the manner of communication while others do not agree. Due to this lack of agreement, it is not possible to generalize current findings. Maybe, most of this "dissonance" is caused by the fact that many findings on stimuli design have not been revised: some authors are stuck with their research done during the 80's while the most recent document on this topic has not been updated since 2011.

To sum up, here is the current scenario on concept testing:

- Lack of information on how concept tests are held and designed
- Small sample used
- Abounding in using traditional interview methods
- Low accuracy of market predictions
- Different stimuli design of a same product concept are not tested simultaneously
- Contradictions between current findings

1.3 RESEARCH QUESTION

With all this considered, the main purpose of this research is to further develop the scope of stimuli design. It would be interesting to administer a concept test made by four different stimuli designs: those will show the same product in different ways and they will be tested to the same audience in order to see if this could lead to diverse outcomes. As a result, the research question would be:

"Do different stimuli design of a same concept provide different results? If yes, how?"

To deepen the theme of the stimuli design, it has been chosen to hold an online survey that includes a virtual environment. The reason for this choice comes from a study made by Peng and Finn (2008) who suggest improving concept testing using

a wider sample, an online test and virtual prototypes. However, it is not the main intention of this project to advance a new theory on concept testing, the aim is to test a current one trying to overcome the issues stated in paragraph 1.2.

1.4 RESEARCH OBJECTIVES

The main purpose of this research is to understand whether the role of stimuli design might be relevant for new product success. To achieve this objective, it should be first analysed the exiting literature to understand the current findings on concept testing and stimuli design. The literature review should function as a starting point. The second step is to gather data on which further conduct the analysis: those data are the responses collected through an online survey. As a third step, the data gathered should be analysed to test the hypotheses previously mentioned and to give further considerations about the research question. Lastly, the final step is to draw conclusions, provide managerial implications based on the findings and then try to give suggestions for future research.

1.5 DELIMITATIONS

It may be useful to list the main delimitations of this research that limits the scope and define the boundaries of this study.

First, the aim of the thesis is not to define which one is the preferred stimuli design that must be chosen when doing a concept test. The purpose of the research is just to test whether different stimuli designs can lead to different reactions when evaluating a concept.

Second, it will be analysed just one product and not various. Therefore, the product area is delimited. Furthermore, it is not the intention of this research to see if the product tested can be introduced in the market. Of course, some considerations about its possible launch can be done after having analysed the data gathered, but it is not the main purpose of the research.

Third, it has been chosen the European market, Italy in particular, due to the author's acquaintances. Moreover, considering the research tools, most of the answers come from respondents belonging to the group age of 18-24 and coming from Italy.

Fourth, there are some limits on tools and methodology. This means that a quantitative research has not been combined with a qualitative study and a physical prototype has not been provided to the respondents, so they had not been able to see and try the actual product.

Finally, it won't be possible to generalize the findings due to the participants selection and the specific industrial sector chosen. Therefore, it is not possible to develop a general theory that can be applied to every circumstance.

1.6 THESIS STRUCTURE

The structure of the thesis consists in five parts:

- **Chapter 1**: introduction. After giving a first insight of the background and objective of the study, part one of the thesis follows with the definition of the problem statement and the subsequent research question. Next, a small brief concerning hypotheses and the theoretical framework is provided. Finally, research objectives and delimitations are explained.
- **Chapter 2**: literature review. First, it will be given a brief introduction to the NPD process, explaining at which stage the concept testing takes place. Then, it is introduced the definition of concept testing with a special emphasis on stimuli design. Finally, there will be a focus on the five hypotheses.
- **Chapter 3**: methodology and research design. Here, information on the empirical setting, which includes the market and the product chosen, will be given. Subsequently, it will be explained sample selection and survey design. This part will be concluded with an introduction to data analysis and the limitations of the research.
- **Chapter 4**: here the data gathered are analysed and the hypotheses previously mentioned are tested to understand whether they are confirmed or not. Moreover, it will be done a general overview on the sample with related analyses and some final remarks on the likely market success of the product would be done.
- **Chapter 5**: the ending part of the thesis will be focused on the conclusions drawn from the results coming from the survey. Managerial implications, limitations of the study and some suggestions for further research will be provided too.

CHAPTER 2 – LITERATURE REVIEW

2.1 INTRODUCTION

Through this chapter, the aim is to understand previous findings about concept testing and stimuli design. First, it will be given a quick overview on the role of innovation nowadays, and then two common New Product Development (NPD) models will be explained. Since concept testing plays a critical role in the NPD process (Ozer, 1999), it might be interesting to realize where does the concept testing phase is placed in such procedure. Additionally, many companies have indicated that the NPD stage that needs most improvements is the pre-launch market analysis, where concept testing takes place (Reid and De Brentani, 2004). In fact, the high failure rate of new products happens in the concept testing phase: here, poor ideas might pass the concept test and proceed in the development step, while potentially valuable ideas are discarded. However, the focus of this work is not explaining all the features of an NPD process, therefore only limited information about it will be presented.

Consequently, the literature review about concept testing will be exposed, defining notions, objectives and limitations about this topic. This will function as an introduction of the stimuli design part, which will describe its definition and its configuration, giving the actual findings of current literature on the effectiveness of the different types of stimuli design. Finally, the hypotheses will be deepened.

2.2 THE NEED FOR INNOVATION

In a fast-moving economy, a continuous innovation becomes something necessary to face the change. Firms with a positive attitude towards innovation are the ones with better chances to survive in this evolving world. Not only the word *innovation* is referred to the introduction of new products or new services, but also to a new way of doing things. In this case, the dissertation will solely focus on developing new products.

There are many reasons behind the decision of a company to innovate: markets can be entered faster, it is a great way to stand out from competitors, variable customer needs can be accessed easily, solutions might be provided before end-users realize to have a problem, and so on. Therefore, pursuing the "innovation path" can bring many advantages for a firm.

Developing new products or improving existing ones has turned out to be a common procedure to enhance sales. However, commercializing a new product does not always lead to a certain success. For example, the lack of information on end-users can bring a lot of uncertainties when proceeding with NPD activities. Despite the low rate of success, companies should pursue with their innovative vein to keep up with evolution and change.

2.3 NEW PRODUCT DEVELOPMENT PROCESS

A new product is something that provides new benefits, features or functionalities clearly visible for the customer (Cooper, 2011). It can be new to the market, new to the company or new for both of them. When developing a new product, the risks associated are high: the product might not satisfy customers' needs, it has been done a scarce market research, the product has nothing as unique or valuable compared to its competitors, the marketing plan has not been properly developed or executed, etc. There can be many reasons for a product for not becoming a success, but maybe the existing high failure rate can be attributed to the lack of precise pre-launch analysis (Dahl and Moreau, 2002; Zien and Buckler, 1997). This is a risk that each firm may encounter, therefore they should not be aware of developing innovations, otherwise they will face a slow death (Cooper, 2011). Their strategy should be on managing the risks and do early diagnosis rather than run away from problems (Keizer et al., 2002).

Due to those uncertainties, there is the need to follow a process divided into phases: the New Product Development process. This can be defined as a series of activities which a company employs to conceive, design and commercialize a product (Ayag, 2014). The NPD process is widely recognized as a core product strategy with high risk and one of the most challenging components of enterprise competitiveness (Thai, 2017). Next paragraph will show two common NPD model in which the Concept Testing phase takes place.

2.3.1 CRAWFORD AND DI BENEDETTO'S MODEL

Crawford and Di Benedetto (2010) propose a simple New Product Process: according to the authors, if this combination of activities is performed well, it will churn out the new products the company needs (fig. 2).

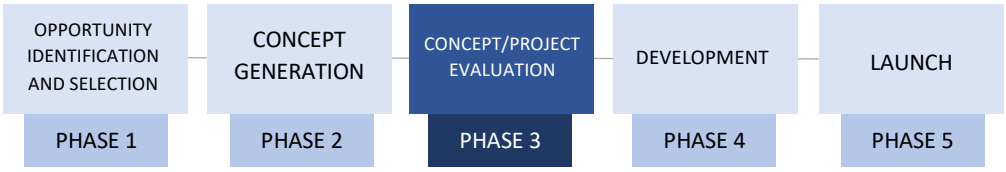

Figure 1 – New Product Process by Crawford and Di Benedetto

The model is characterized by periodic evaluations (go/no go decisions) all the way through the process. In Phase 1 the company needs to find and select the opportunities from which a new product can be developed. Those opportunities could come from company's internal and external resources or from completely new sources (Vicari et al., 2013). The next step is to transform the ideas previously gathered into a product concept. The product itself is not yet fully developed, but a descriptive draft is done to better conceive features and benefits associated. As Phase 3 occurs, the most promising product concepts need to be selected while the others need to be rejected. It is a sort of a screening phase done through the Concept

Testing. In the Development phase, the physical product is finally created, but before its launch product and market tests must be completed. Finally, in Phase 5 the commercialization of the new product begins.

As previously portrayed, the lack of precise pre-launch analyses might be one of the causes of new-products-failures. When considering this model, the pre-launch analyses are detected in Phase 2 (Concept Generation) and Phase 3 (Concept/Project Evaluation). However, Phase 2 is just the definition of the concept, no tests are done. Therefore, Phase 3 is what should be considered as the heart of Concept Testing since in this step the company can evaluate with consumers the product concept (fig. 2) and decide whether to pursue with the development phase.

2.3.2 THE STAGE GATE MODEL

Cooper (2011) suggests another system, a different blueprint for managing NPD process and improve its effectiveness and efficiency:

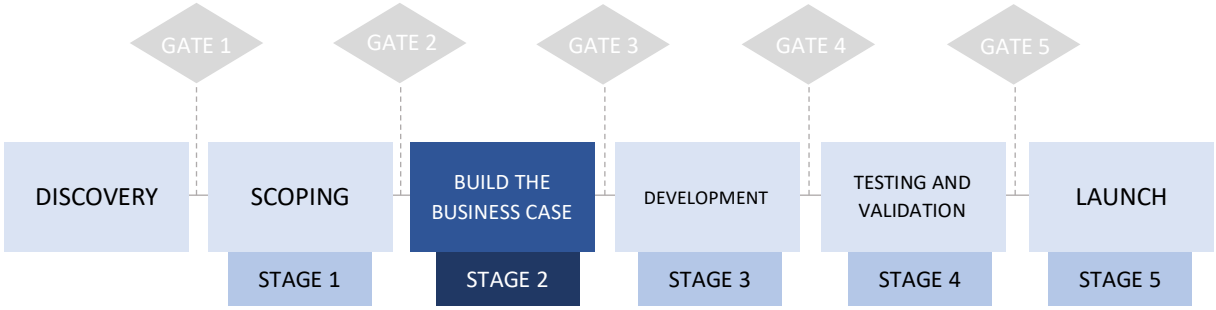

Figure 2 – Concept Testing in Stage-Gate Model by Cooper

This model is called *Stage-Gate* and it is defined as a conceptual and operational map that helps products to be developed from idea to launch (Cooper, 2011). The model is built upon stages and between them lies a check-point (*gate*) where it is decided whether to pursue with the next phase or not. Through the stages it can be seen how the work is done, while the control of the process is ensured by the gates, which are a chance for the company to review the project carefully and to bring together the experience and the expertise of the organization (Goffin and Mitchell, 2017). This model is a bit more articulated than the one previously portrayed: what differs is the Scoping part (preliminary market and technical assessments plus a delimitation of the project), Build the Business Case (a more detailed investigation phase in which it is clarified the role of the product in terms of positioning, attributes, benefits, etc. and a market research is held to detect customers' needs and wants) and Testing and Validation (simulated tests, lab tests, field tests, etc. to ensure the practicability of the project). This is just an example of the number of phases that can be held because the needs of the company define the amount of stages required. Usually, a Five-Stage-Gate model is necessary for complex product development processes (Cooper, 2011).

When considering the model by Cooper, Stage 1 and Stage 2 become essential before entering the Development phase. Actually, Cooper (2011) identifies six elements as part of an "integrated product and project definition" that should be defined before entering Stage 3: project scope, target market, product concept, product benefits, product features and positioning strategy. As we can see, the product concept belongs to the pre-launch analyses. After having defined the concept itself, companies should administer a concept test to determine the likely market acceptance. In this case, the Concept Testing phase is not clearly displayed as in the previous model. However, since Stage 1 (Scoping) is still a preliminary and not yet fully conceived investigation (it is a sort of a preview), concept testing is one of the actions included in Stage 2 (Build the Business Case) where the proposed product concept is tested with consumers (fig. 3).

2.3.3 FINAL REMARKS

After the launch stage, the process still goes on with a Post-Launch phase whose aim is to detect the efficacy and efficiency of the product after its commercialization in the market. Moreover, both models insist on the role of consumers, which should be addressed throughout the entire process. However, since the focus of this dissertation is not the NPD process, those characteristics have only been mentioned.

In this paragraph, it has been done a brief introduction to NPD in order to explain where the concept testing lies in such process. This has been done considering the model provided by Crawford and Di Benedetto (2010), which is a simple process with regular phases, and the Stage-Gate system by Cooper (2011), which is a common procedure used by firms representing an improvement of the regular practice. With all this considered, it can be now defined and clearly explained the notion of Concept Testing, an essential part of the pre-screening process.

2.4 CONCEPT TESTING

Launching a new product in a new or in an existing market is not an easy task. Before physically creating the product, it is necessary to understand whether the market is ready or not for this kind of innovation: in this case, concept testing plays an important role in the pre-launch phase of the new product development (see par. 2.3).

2.4.1 DEFINITION OF CONCEPT

Before starting to describe the concept testing itself, it should be first highlighted the definition of what is a concept. Kotler et al. (2012) distinguish between idea and concept. An *idea* consists in a generic understanding that a company can offer to the market. A *concept* is an elaborated version of the idea which is expressed in terms of specific benefits and features considered relevant for the customer and different from the competitors. Cooper (2011) simply defines the concept as what the product will be and do (Cooper, 2011), while for Crawford and Di Benedetto (2010) a concept is a stated relationship between product features and consumer benefits. Tauber (1972) propose a notion similar to Cooper, defining a concept as

"a mental image of a thing formed by generalization from particulars; also, an idea of what a thing in general should be". Developing a product concept is necessary but not sufficient: firms should test their concept to choose what to maintain and what to reject. Therefore, a concept evaluation through a test is required.

2.4.2 CONCEPT DEVELOPMENT PROCESS

Ulrich and Eppinger (2000) have identified a typical concept development process with the following activities (fig. 4):

Figure 3 – Concept development process

Everything starts with the identification of customer needs, which is the necessary basis for developing the product concept. During the second phase (Establish Target Specifications), the needs previously identified are translated into technical terms. Then, it follows the concept generation, which is based on the previous consideration and on the purposes of the company. This phase would produce a variety of concept (usually from 10 to 20) which need to be selected to choose the most promising ones. When the decision has been done, it starts the concept testing step where firms are ensuring that customer needs are met. With Phase 6 (Set Final Specifications), earlier target specifications are reviewed after the concept selection and testing. Finally, when the concept chosen seems to satisfy customer needs and wants, the company can proceed with the development plan of the product. Therefore, this process is later followed by the development phase of the NPD process. In each step, firm can use various forms of models and prototypes to let the concept be perceived as much real as possible. In the current master thesis, the focus will be on Phase 5 of this procedure (Testing Product Concepts).

2.4.3 CONCEPT TESTING: HOW IT WORKS

Concept testing might be thought of as a search for the "best" design, pricing, positioning and manufacturing of a new product (Dahan and Mendelson, 2001). It is a tool that helps to assess the likely market demand and the best customers to target with potential new products early in their development (Ozer, 1999; Page and Rosenbaum, 1992; Crawford and Di Benedetto, 2006). The most common definition is the one provided by Moore (1982, p. 279), who states that concept testing helps estimating customer reactions to a product idea before committing substantial funds to it. Page and Rosenbaum (1992) have identified a more specific definition of concept testing: "…a variety of marketing research-based approaches employed to assess the marketability of a product or a service idea prior to its actual development. Its purpose is to provide early feedback from the market about the perceived attractiveness of a proposed new product before its development has even begun" (p. 269).

To select feasible new product concepts, companies need to conduct concept testing by asking end-users to evaluate numerous new product ideas (Anschuetz, 1996; Moore, 1982). According to Tauber (1981, p. 169), the general procedure goes as follows: a stimulus is presented to consumers, who must evaluate it to let researchers try to predict a behavioural response (such as the purchase intention or the likely interest in the product). This stimulus is the concept, which can be a descriptive text or a visual (see par. 2.5.2). Through a concept testing, companies are testing a product concept with the user to ensure liking and purchase intent (Reidenbach and Grimes, 1984). One part of concept testing is the concept evaluation, through which consumers evaluate concept stimuli using a set of evaluation items (Moore, 1982). The evaluation is done through a set of questions (open- and/or closed-ended) which help measuring the following items: the importance of the product features, the perceived benefits and the predisposition towards the price (Krishnan and Ulrich, 2001). Companies might think to set an open question that asks consumers to give suggestions on the presented product concept. This could be a good way to easily understand whether the future product would meet consumers needs and wants.

In a typical concept testing study, firms try to assess the viabilities of their new products before making major financial and nonfinancial commitments into their development phase: this is done by asking the potential buyers the likelihood of their new product purchase and usage behaviour (Crawford and Di Benedetto, 2006; Moore, 1982; Page and Rosenbaum, 1992). Popular methods of administering a concept test are through a focus group or in-depth interviews because it can be collected a useful feedback for the product. However, doing this qualitative research is risky because of the small sample size and because participants often do not represent the entire market (Cooper, 2011). A way to overcome such issue is to use an online survey that can reach a wide range of consumers. Peng and Finn (2008) consider three main advantages when deciding to pursue and online questionnaire: cost-effectiveness, takes less time and ease of use.

Cooper (2011) considers concept testing as the final test prior the Development phase: through that, it is verified whether consumer needs are correctly understood, perceived and translated. Moreover, this phase is essential because after consumers have evaluated a concept, if features and benefits provided are not enough and/or they are not aligned with their needs and wants, there is still time to modify the original idea before further proceed in the NPD process. Basically, companies are gathering feedback on the product before advancing into the development stage.

Concept testing is typically administered through a *monadic test*, which means that the end-user will not test multiple products at once but only one. The monadic test is simple and easy to administer and to set up, but it is expensive and less reliable (Cui et al., 2015). However, there is a second type of test that companies might do, which is the *comparative test*. In this case, more than one product is tested at the same time, as if in a real decision-making situation the end-user compares one product with another. Trebbi and Flesch (1980) have administered a survey with 11 different product concepts of

various goods belonging to the same product category. Despite the effort to do a sort of a comparative test, the majority of academic literature enlightens the usage of monadic tests, probably for economic and logistic reasons (time consuming and difficult to propose different concept). In addition, it is not common for researchers and firms to test diverse concept formulations of the same product for the just mentioned motives.

2.4.4 CONCEPT TESTING KEY DIMENSIONS

Dolan (1993) has identified some key dimensions that are necessary when administering a concept test: this means, that they are typical questions that need to be included in the test. Those key dimensions are usually measured through a five-point scale which evaluates the overall positive or negative perception of the product concept. Those key dimensions are (and the corresponding questions which are highlighted in italics) are:

- Overall liking: it is done to assess if consumers like and want to have the product shown (*how much would you like to have this product?*)
- Importance: the perceived importance of features presented with the product (*how important do you think the functions and features of this product are?*)
- Uniqueness: whether the product can distinguish itself from the others, it is asked to assess if it is perceived as different from competitors' goods (*how would you rate this product in terms of being unique from the products currently sold?*)
- Solve problem: the product might be unique, but consumers should find it useful to make their life easier (*how sure are you that this product would solve a problem for you?*)
- Believability: whether the features stated by the product concept are perceived as realistic (*how believable to you were the things said about this product?*)
- Purchase intent: according to Dubas et al. (1999), it provides a direct estimate of the customer's belief when choosing a new product, it is a predictor of the purchase behaviour (*assuming this product was available in a store where you shop, how likely would you be to buy it?*).

In addition to this list, other necessary questions are related to the overall interest, which is a measure considered by Schwartz (1984) as essential to improve the results of concept testing, it is a diagnostic tool.

2.4.4.1 RED MEASURES

Concept test can be a tool to predict new product success, which may be the ability to remain on the shelves for more than two years in a row (Kotler et al., 2015). The way in which concept testing is done has been revised by IPSOS InnoQuest

research company[1]. IPSOS is one of the world's leading independent market research company which helps its clients (namely firms) to assess their market potential, to interpret new trends, to build long-term relationships with customers, etc. It is a company similar to GFK.

They have made a research on concept testing too (that's why they have been included in this project) which has highlighted that there is one way to predict new product success and the consequent purchase behaviour using measures that go beyond the simply purchase intention. Those measures (called RED measures) have been identified since they help choosing the high potential consumer goods concepts. Let's see them in detail:

- Relevance: the extent to which an innovation meets customer needs. Those needs are both emotional and functional. This is a good indicator of whether the problem solved by the concept is perceived as important for consumers;
- Expensiveness: the extent to which an innovation is perceived to be higher-priced than competitors. It is a pure measure of price, so there is nothing in common with the Value for Money (i.e. the relationship between the quality of the product and its related price);
- Differentiation: the extent to which the innovation provides unique benefits versus competitors. It can show whether a concept is a better solution than the currently available products.

Instead of solely relying on Purchase Intent, with those dimensions it is easier to understand which concept to accept and which one to reject. Actually, Purchase Intent can only provide an end-results, i.e. if consumers are interested in buying the product. On the contrary, RED measures can be analysed to determine the chances of new product success. According to IPSOS, RED measures are more able at "concept screening" and they avoid two typical drawbacks of the Purchase Intent indicator:

1. They are less sensitive to executional differences in concepts (e.g. elements related to packaging, price) than Purchase Intent. This means that they are more likely to provide objective responses to those elements in the concept that are not directly associated with the insights or benefit (www.ipsos.com);
2. RED measures are less likely to support familiar concepts such as line extensions like the Purchase Intent indicator.

RED measures are able to avoid those pitfalls and they can be used as a diagnostic tool to understand what stands behind the concept performance (www.ipsos.com). IPSOS has demonstrated that products high in Relevance were more successful than products high in Purchase Intent and vice versa. At the same time, if a product is perceived to be not expensive, it has great chances to be a success; otherwise, if it is perceived as expensive, chances weaken. As far as differentiation is

1 https://www.ipsos.com/en-us

concerned, the more the product is perceived as different and unique, the more the chances of its success enhance. Combining differentiation with expensiveness can determine the success too (i.e. low/medium expensiveness with high/medium differentiation).

2.4.5 AIMS AND CHARACTERISTICS OF CONCEPT TESTING

According to Crawford and Di Benedetto (2003), concept testing has three main objectives:

- It should help in identifying which one are the poor concepts that should be eliminated
- It should help in further developing the original product idea
- It should provide an initial trial and sales estimation

Moore and Pessemir (1993) specify other purposes of concept testing: choosing the most appropriate target segments, obtaining robust estimates of demand for the new product and determining if there is sufficient consumer appeal that guarantee the further product development.

As noticed, concept testing can have multiple aims. In a study conducted by Peng and Finn (2008), the authors have highlighted which ones are the most common aims chased by companies when testing a concept: further developing the original idea (81% of respondents), estimating concept's market potential (70%), eliminating poor concepts (66%), identifying the value of concept features (66%) and identifying the highest potential customer segment (53%).

Peng and Finn (2008) have identified some characteristics of a typical concept test:

1. There is not one main objective but various, even though the most important is to further develop the original idea;
2. It is more suitable for incremental innovations;
3. Is equally likely to be a monadic or a comparative test;
4. Pricing information are included;
5. Concepts are formulated with a stripped tone (see par. 2.5.2);
6. Data are collected through face-to-face interviews;
7. Structured and unstructured measures are used;
8. Each concept has been tested among at least 92 respondents;
9. It has been employed a scale with both numerical and verbal labels;
10. Respondents are selected by class usage, specific product usage or lead user criteria;
11. Outcomes are assessed with Top 2 Box Scores and/or rating scale mean.

This list highlights what companies were doing in 2008 when testing their concepts. But, even though some firms have decided to participate to the study made by the authors and shared their knowledge, they still do not have provided information on how is effectively developed a concept, how it is formulated and how its consequent testing phase is administered. Still, most of the data remains unrevealed: the reason behind this could be the fear of a security leak that may bring a great advantage to competitors. Therefore, it is not recommended to share all the secrets, but in this way it is not possible for scholars and for other firms, who have never done a concept test before, to follow a scheme or a guideline.

2.4.6 LIMITATIONS OF CONCEPT TESTING

When doing a concept test, the aim is to get a reaction from consumers to assess the marketability of the product. It is considered a low-cost and swift method to gather feedback before committing substantial funds to a specific product. However, according to a study made by Peng and Finn (2008) only 19% of companies that use concept test are satisfied with the results. The majority (62%) cited forecast inaccuracy as the main problem of concept testing. Actually, it is hard to predict the market success of a product through concept testing (Tauber, 1974; Urban et al., 1996; Gourville, 2005). The reason behind this could lie in the ability of the researcher to conceive and administer a test and its results, who should be educated in how to properly use testing results (Duke, 1994). Otherwise, the poor predictive capability might result from the fact that the original product concept is different from when is actually developed. Another problem faced is the one linked to radical innovations. In fact, concept testing might not be able to provide correct results when dealing with extremely new products that consumers cannot compare to anything already experienced. End-users do not know their purchase intention having little knowledge of an innovation (Duke, 1994). This would obviously discourage the development and launch of those kinds of products. However, some authors agree on the fact that consumers with product expertise should not be involved when developing innovations. It is worth mentioning that most of the times different authors have different considerations about the same topic, hence it is not possible to generalize.

2.4.7 FINAL REMARKS ON CONCEPT TESTING

Despite most of the time sales and trial predictions are not very accurate, concept testing can show whether to pursue in the next phase of the NPD process or not (Lees and Wright, 2004). It is still one of the most popular tests, in the pre-launch analyses, when a firm needs to assess and potentially modify a product before its development. Forecasts drawn by concept testing are essential in defining if the potential product should advance on the next stage of the NPD process (Klink and Athaide, 2006). It is helpful in providing early inputs on the usefulness, acceptability and design of a product concept. In addition, companies are under pressure when they need to boost the quality of their products and decrease the time to market without cost overruns (Crawford and Di Benedetto, 2010). Therefore, concept testing helps saving time and ensuring that the product is perceived by consumers to be a quality product.

To see whether a product would be a success or not, it is essential to not underestimate how the concept test is designed. For this reason, there are three main decisions concerning the design that should be made which, if altered, might have a negative impact on the entire process (Klink and Athaide, 2006):

1. Stimuli design
2. Respondent selection
3. Response measurement

Respondent selection is referred to who should do the concept testing. They usually represent the potential consumers to address, the ones who would potentially use the product and its benefits. On the contrary, *response measurement* represents the way in which consumers' answers are measured. This list is similar to the one provided by Crawford and Di Benedetto (2015) which highlights the importance of deciding what to test, who should participate and how and when carry out the test.

2.5 STIMULI DESIGN

Tauber (1972) identifies two principal elements that must be considered when doing a concept testing. First, the form of presentation. The product concept might be a visual, a simple text, a package or even a physical prototype, what matters is that the end-user should be able to perceive it. Second, the execution or design of what is presented. Therefore, in this case is essential the work of a copywriter, the overall creativity, the photography skills and so on. Crawford and Di Benedetto (2003) agree on the fact that, when doing concept testing, there are many decisions that needs to be done which belongs to non-concept factors: whether to offer competitive information, selecting the response situation, whether to express the price of the product, definition of the respondent group and the presentation format. It is worth noticing that both agree on the importance of the stimuli presentation when doing a concept test, which can be labelled as *stimuli design*. While concept is the idea that stays in people's mind, stimuli design is referred to how the product is shown to respondents, thus the way in which the concept is presented and formulated.

2.5.1 THE IMPORTANCE OF DESIGN

Improving the design of concept testing can improve the new product development process efficiency as well (Peng and Finn, 2009). It is almost common knowledge that the aesthetic profile of a product can guide the final purchase decision of the buyer. According to Creusen and Schoormans (2005), a product's appearance or its aesthetic design might influence inferences about the product's quality and/or functional performance. At the beginning, the aesthetics was predominant in women's and men's apparel, but now its importance has spread into other industries as well (Schmitt and Simonson, 1997), even though they were not associated with fashion, such as the automotive industry, tech industry, and so on. Nowadays,

even electronic devices are labelled as “fashionable accessories” (Peters, 1992). This leads to the following consideration: product aesthetics becomes another marketing tool that would help a firm to gather a competitive advantage. Therefore, it might happen that when consumers need to choose between two products, the visual characteristics can influence the final choice, regardless of the quality and the function of the other good. Similarly, the way in which a concept is presented to respondents might influence the predisposition towards the product and, consequently, the overall liking and the purchase intention.

2.5.2 STIMULI DESIGN TONE AND NARRATIVE TRANSPORTATION

There are two ways through which a concept can be formulated: stripped or embellished. On one hand, *stripped* descriptions are regular descriptions of the central concept idea in which a non-emotional tone is used. Usually, a stripped description can be a text made by few sentences. A stripped or factual tone could let many concepts to be tested and it is easy to use, however in this way few attributes of the product are presented. On the other hand, *embellished* definitions have a more emotional role, the tone is persuasive because they work as if they are commercial advertisement. Through this method, it becomes more realistic to evaluate the concept for end-users (Crawford and Di Benedetto, 2000). There is a third label highlighted by Lees and Wright (2004) which is *visual*: in this case, concept statements have an embellished tone and they are combined with line drawings, pictures, visual portrayals. This proves that according to some studies the written statement alone is not enough.

In a study conducted by Peng and Finn (2008) it is shown that companies usually present the concept as a stripped description (38%) or stripped with visual presentation (31%). Only the 9% of the respondents has declared that they use fully finished advertisement as a concept because it might be expensive and time-consuming. As previously mentioned, there are different opinions among authors on the same topic.

Another topic that should be addressed when talking about the tone used in stimuli design is the *narrative transportation*. Product concept presented in a narrative way means that a storyline, protagonists and a series of actions are employed to reach the respondents attention. By doing so, information about the concept are presented through this narrative, creating a more realistic situation in which consumers identify themselves. Van den Hende and Schoormans (2002) argue that testing done in a narrative format would yield more accurate evaluations. In a concept testing environment, this happens describing a person using the product and encountering attributes and benefits of it. The study conducted by Van den Hende et al. (2012), however, does only show the effect of narrative transportation when respondents are reading the concept, nothing is portrayed in a context of virtual environment.

2.5.3 DIFFERENT TYPES OF STIMULI DESIGN

With all this considered, it can be now described the different types of stimuli designs currently analysed by scholars. Concepts can be presented in various ways, from a simple statement that highlights some attributes of the product to a complete advertisement showing the main benefits. As it will be noticed, there are inconsistencies between findings from different authors on the same topic:

1. **Statement**

   According to Crawford and Di Benedetto (2003), a concept statement should be clear and realistic, without trying to oversell the product. Usually, firms use a simple statement to test the consumers' reaction to the pure concept (Moore, 1982). This means that the textual description should be about features and benefits of the product. Cooper (1993) recommends writing approximately 100 words, otherwise the text becomes too long, respondents may find it boring to read and they will consequently miss essential information. Choosing this way is cost effective and it takes considerably less time to be created. As previously mentioned, the concept description can be done with a stripped or with a embellished tone. According to Moore (1982), stripped statements should be the most used since they are more indicated for radically new concepts, tests are easy to do, and they are recommended for evaluating many concepts. On the other hand, an embellished concept is recommended when a concept is going into an existing product class (Moore, 1982). However, according to Schwartz (1984), the job of a concept is to display benefits and features in a persuasive way.

   Haley and Gatty (1971) showed that concept presented with different tones gave different results. Findings from Tauber (1972), however, have demonstrated that a concept testing on three products with two different tones have led to the same results, arguing that a product concept is evaluated regardless the manner of communication. Later on, Dolan (1993) added that a change from a factual to a persuasive tone would have not altered the purchase intention, while Marder (1997) and Crawford & Di Benedetto (2000) did not agree. In conclusion, it must be noted that the majority of researches have tested stripped descriptions rather than using an embellished tone.

2. **Statement with visual**

   This type of stimuli design belongs to the visual categorization proposed by Lees and Wright (2004) in which a statement is combined with a picture, a drawing, etc. Again, the statement can be presented with a factual or a persuasive tone.

   Theoretically, end-users might have different reactions when evaluating a verbal concept statement and a concept made with visual (Holbrooke and Moore, 1981; Vriens et al., 1998). Most of the studies agree on the fact that a visual concept test yields a more positive outcome than a written concept statement (Lees and Wright, 2004), probably because consumers might have a different reaction when they actually see the product stated in textual description, they do not have to imagine it anymore. Rossiter and Percy's (1980) argue that presenting a

persuasive statement might lead to similar results to those coming from the visual concept test. Seeing a picture of a product concept let the consumer be in a more realistic situation, but then again there is no common agreement between scholars when those results are compared to a statement with no visual. As far as the narrative transportation is concerned, Van den Hende and Schoormans (2002) have demonstrated that a narrative with draw images might provide similar results as a prototype demonstration.

3. **Statement with visual and customer reviews**

Peng et al. (2012) have tried an "unconventional" testing environment in which the classic visual with a statement is combined with online customer reviews. In this case, the authors have left the reviews coming from Amazon.com to provide a more realistic setting, in which consumers were evaluating Hi-Tech products already available to purchase. Klink and Athaide (2006) suggest the use of online reviews for later adopters: those users are more aware of choosing a new product because they need to receive positive information about it from non-marketing sources, such as friends and family. Trust issues are difficult to overcome, but a good help might come from using customer reviews when presenting the concept. Actually, this kind of stimuli design is not common in concept testing, therefore there are not enough findings specifically on adding a customer review part.

4. **Virtual prototypes**

Current trends on concept testing includes that companies should invest more on using virtual prototypes (Peng and Finn, 2008) to yield more accurate answers from respondents. In a virtual environment, the product is presented in all its features, consumers can visually evaluate it without being able to try it (unless it is an online service or a mobile application). Virtual concept testing on the web enables product development teams to test concepts without actually building the product (Dahan and Srinivasan, 2000). Those types of concept designs, along with visuals, help a company to inform customers on the additional benefits and features. The virtual prototype can be developed with the aid of 3D technologies or it can be a video: they can both give a better preview of what the product would be thanks to the help of virtual animation, sounds, visuals, etc. Therefore, a virtual setting can combine all the advantages of textual description and visuals (Peng et al., 2010). Findings on differences between a virtual and a traditional environment are not homogeneous. Haley and Gatty (1970) argue that using working or virtual prototypes should be avoided, however Crawford and Di Benedetto (2003) and Cooper (1993) state that showing the most accurate presentation of the product (such as a visual) could lead to more accurate results. Furthermore, some authors argue that a virtual testing environment provides positive results compared to a traditional setting (Dahan and Srinivasan, 2000; Von Hippel and Katz, 2002; Füller et al., 2006), while at the same time Peng et al. (2011) state that both environments yield identical results. Some authors have suggested to combine virtual products representation with a virtual shopping experience to better predict whether the new product would be a success or not. However, this method, called *information acceleration*, require a great

amount of information to build the virtual shopping environment and it is very expensive in terms of generating and testing each prototype.

5. **Physical prototypes**

   Not only images or virtual environments, but also actual products can be tested with consumers. For the food industry, for instance, it would be better for end-users to taste the product before letting an organization commercialize something with a bad taste. However, when the prime benefit is a personal sense, or when something related to art and entertainment industry should be evaluated, concept test usually fails (Crawford and Di Benedetto, 2010). A physical prototype might be useful for tech products (such as smartphones or notebook) so that consumers are able to actual see and try what are the main benefits and features of the tested goods.

   Again, there are different considerations from current literature on the effectiveness and efficacy of proposing a physical prototype. In their analysis, Dickinson and Wilby (1997) have found no significant variations in the responses when consumers where presented with concept statements and product trial for the same product. On the other hand, Lewis (1984) and Dahan and Srinivasan (2000) stress on the positive outcome provided by physical prototypes compared to a negative purchase intention of those end-users exposed to just the verbal statement. Actually, ready-prototypes provide customers additional and non-attribute-based information (Srinivasan et al., 1997).

2.5.4 REASONS FOR USING A VIRTUAL ENVIRONMENT

According to Dahan and Srinivasan (2000), there are numerous reasons for considering the virtual environment as suitable and opportune. First, there are no significant budget issues since creating a virtual prototype is cheaper than developing a physical prototype. Second, accessing respondents becomes simpler avoiding logistic issues related to the physical distance. Third, with the help of new technologies and the Internet, it can be generated a good experience even though consumers cannot physical touch the prototype and it can be generated a realistic environment. Finally, in a virtual environment is possible to test more than one concept at once, reducing costs and problems associated with the same situation done with physical prototypes. The advantages of using an online setting is shared by Wilke et al. (1999) too, who have shown that data gathered from the Internet were sensitive, reliable and projectable as the ones collected from in-depth interviews.

2.5.5 FINAL REMARK ON STIMULI DESIGN

A final remark should be done on stimuli design. What emerges is that companies do not usually hold a concept test of a same product concept. This procedure is time consuming, expensive and there are logistic issues because it is not easy to reach a wide range of consumers at once asking them to evaluate different concept of the same product answering the same questions. In a study conducted by Srinivisan et al. (1997), it has been demonstrated that it would be optimal to carry multiple product concepts into the prototyping and testing phase. Researches depict inconsistent findings; therefore, it is

not possible to generalize and to detect whether end-users might have different reactions based on the stimuli design proposed rather than based on the product itself. Moreover, the most common formulation used are simple statements and statements with visuals, which provide a fictional situation rather than the one yield by a virtual testing environment. Thus, it would be interesting to see if difference responses are gathered when the stimuli are presented in different ways and to insist on the use of virtual environments.

2.6 DEFINING HYPOTHESES

Five different hypotheses have been tested in relation to stimuli design and concept test evaluation. This procedure helps in trying to answer the research question. Let's now see them in detail.

- **Hypothesis 1**: "*Different stimuli design of a same concept would lead to different concept evaluations*"

  The aim of the first hypothesis is to demonstrate that stimuli design might be relevant in the final concept evaluation: once presented to different designs, consumers might respond in different ways when evaluating the concept. As seen in the literature review, statements and visuals are the most common stimuli designs used in concept testing and is not conventional for firms to test the same stimuli through different formulation. Therefore, with this hypothesis the research aims to test whether customers will increase their purchase intention, overall liking, interest etc. in the product concept considering the different presentation of it.

- **Hypothesis 2**: *"Product expertise might influence the overall concept evaluation"*

  Sometimes the concept evaluation might be influenced not only by its design, but also by consumers' characteristics. For instance, there might be a correlation between product expertise and the overall concept evaluation depending on the way in which it is presented. Product expertise is defined as the ability to perform product-related tasks successfully (Alba and Hutchinson, 1987), an important characteristic when evaluating a new product concept. Consumer with a high product expertise can understand product information faster and deduct attributes and benefits easily compared to people with lower product expertise (Schoormans et al., 1995). Jacoby et al. (1986) argue that product expertise involves a qualitatively higher level of knowledge or skills when compared to some external standards. The purpose of this hypothesis is to see whether there is a correlation between product expertise and the overall concept evaluation.

- **Hypothesis 3**: "*Innovators are more likely to provide more objective responses to new concepts*"

  Consumer innovativeness is the propensity of consumers to adopt new products (Hauser et al., 2006). Hirschman (1980) conceptualizes it as the desire or willingness to try new and different experiences. Olshshavsky and Spreng (1996) argue that when innovative consumers are presented with a new product, they are more able to identify evaluative criteria, to form expectations about product performance and to find reasons to reject innovation.

Therefore, it may be interesting to see if innovative consumers might provide more objective evaluation than non-innovative consumers.

- **Hypothesis 4**: "*Consumers are influenced by the word of mouth when evaluating the concept*"

  When deciding to buy a product, consumers might ask the help of their family, friends or they can check online reviews. All of this can be labelled as Word of Mouth (WOM), which can be addressed by consumers even when they are evaluating a concept. Typically, consumers are more influenced by information coming from nonmarketing sources such as word of mouth and product in use (Bass, 1969; Gatignon and Robertson, 1985; Mahajan and Muller, 1998). According to Dellarocas et al. (2007), survey remains one of the most popular methods to analyse WOM. Therefore, this hypothesis wants to verify whether the online reviews (namely WOM) would affect the final concept evaluation in a positive way. Moreover, it is interesting to see the role of WOM on the general predisposition towards the concept of non-expert users.

- **Hypothesis 5**: "*A move from a factual to a persuasive tone would increase the interest*"

  It is interesting to analyse whether the change of tone would affect the interest in the product concept. As stated in paragraph 2.5.3, there are different findings on this topic according to scholars: some argue that a stripped or an embellished tone yield the same results, while others argue the opposite. With this project, the four different stimuli design tone proposed in a selected order would be tested to see whether they would increase or decrease the interest in the device.

Considering those hypotheses, it can be developed a theoretical framework to see the links among each construct. This framework works as a guideline for the project to investigate and interpret findings:

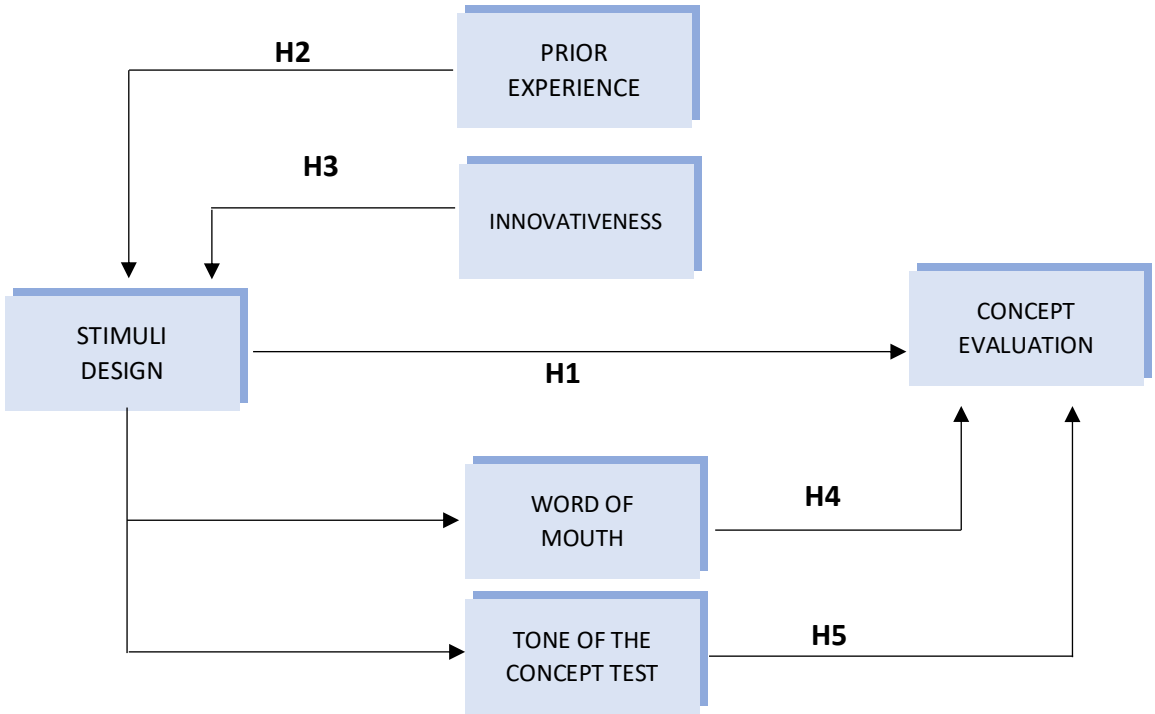

Figure 4 – Theorethical framework

CHAPTER 3 – METHODOLOGY

3.1 INTRODUCTION

The methodology part is focused on how the project has been conducted. Considering the research question, the aim of the thesis is not to develop a new theory but to test a current one. For doing so, it will be used a hypothetical-deductive reasoning. With this method, five hypotheses are generated and then tested through a quantitative method; if the predictions are correct, the hypothesis is confirmed, otherwise it is disconfirmed.

This chapter starts with an overview of the empirical setting, in which the market and the product chosen for the concept test will be enlightened. Second part will be about data collection: here, the focus is on the respondents' selection and on how the survey has been designed. In the final part, data analyses methods would be provided.

3.2 EMPIRICAL SETTING

As previously anticipated in paragraph 1.3, the research will be based on testing a product concept presented in four different ways with the same respondents. The audience tested will answer to the same questions for each concept test: by doing so, we will see if different outcomes will be generated when the concept is portrayed differently.

Before explaining how the research has been conducted, it is important to describe the market and the product chosen in the following two sections.

3.2.1 THE MARKET

Nowadays, time is speeding faster, and we are always asking for the easiest, cheapest and most comfortable aids to live our lives as consumers. Among these, e-commerce plays a pivotal role that is growing quicker since now we can buy literally every kind of product, at affordable prices, from anywhere that can be directly delivered to where we want. We can hardly get enough of online shopping, it is now part of our habits.

As stated, we can literally buy everything on Internet, even food and fresh groceries. Let's think, for example, of AmazonFresh, the grocery delivery service that let people be able to order fresh food to be carried in few hours at their home. This is aligned with nutrition trends that are currently propagating: for instance, all the food blogger websites and the cooking tv-shows are increasing the interest of people in food and cooking. Therefore, consumers are more aware of the way in which they are preparing food.

E-commerce would not be possible without the help of electronic devices, whose development is a spreading trend. Besides letting us be able to do online shopping, tech industry is providing us products that can basically do everything at our place, and it is increasingly consistent the integration of voice assistants in most them. Such tools are very important in terms of

saving time (Bengston, 2014). The aid of voice assistants and the continuous development of tech devices increment our need for immediacy and remote control (e.g. smart home).

For this reason, the focus of this dissertation will be on these three industries (e-commerce, food and tech), which are expanding their activities in our everyday lives. Let's see in the next paragraph how they can be integrated in just one product.

3.2.2 THE AMAZON DASH WAND WITH ALEXA

In 2017, Amazon decided to improve its consumer goods ordering service (called *dash*) introducing the Amazon Dash Wand with Alexa. Actually, it was already developed in 2014 but it has encountered major improvements with its third generation. This electronic device is a Wi-Fi connected barcode scanner through which people can let the process of buying groceries become faster and simpler. It must be pointed to a product barcode and the Wand will add the grocery to the Amazon cart. It is integrated with AmazonFresh, which means that fresh groceries can be bought as well, such as fruits and vegetables.

Maybe the most attractive feature is that the Dash Wand has Alexa, Amazon's voice assistant, built in it. Through its help, users can order groceries and home supplies directly speaking to the device, they can ask advice for recipes, for general questions, to create a shopping list, they can ask to convert ounces into cups and to control the smart home devices. It is a quite small tool that can be stuck to the fridge and brought everywhere (see fig. 5).

Figure 5 - Amazon Dash Wand with Alexa

Currently, the Amazon Dash Wand with Alexa is only available in the US for Prime members. Therefore, it is interesting to propose this "new-product" to the European market through different stimuli designs and see whether the overall concept evaluation would change or not. Moreover, considering that nobody in the sample has never seen or used this device, it would be a good way to see what happens when innovators and consumers with product expertise are placed in front of such device. Certainly, considerations about a possible launch of the product on the EU market can be inferred, however this is not the main purpose of the project.

3.3 DATA COLLECTION

The research conducted by Peng and Finn (2008) has shown that the companies who have taken part to the study suggest using an online environment rather than traditional qualitative research methods (see par. 2.4.3). They stress on the consistent use of online testing and virtual testing, adding that a virtual testing environment would provide more accurate data from consumers because of a more realistic setting. Therefore, the data have been collected through an online survey, which has gathered 246 responses.

3.3.1 THE RESPONDENTS

Mainly, respondents have been selected among the students staying at Bocconi University and at Copenhagen Business School University; in addition, it has been chosen to send the survey to other Facebook Groups, such as Amazon addicts, voice-assistants users, food lovers and similar. This has been done to see the different responses concerning product expertise and the degree of innovativeness. Since students and Facebook Groups have an international background, only answers obtained by European respondents were considered valid. This limitation has been taken since the original aim is to present the Amazon Dash Wand with Alexa only in those countries where the product is not available. Therefore, due to the limited resources, it has been done a convenience sampling belonging to the non-probability sampling methods.

3.3.2 DESIGNING THE SURVEY

As previously anticipated, according to the data available the majority of concept testing has been done through a qualitative research, leading to some drawbacks concerning sample size (see par. 2.4.3). Since many authors insist on administering an online survey to gather a wider sample, it has been decided to conduct this research with a quantitative method on the Internet. By doing so, it has been possible to experiment a "not so conventional" way of doing concept testing.

The survey takes about 15-20 minutes to be completed and the majority of questions have been set as multiple choice and with a five-point Likert scale (1=Strongly disagree, 5=Strongly agree). Therefore, almost the entire survey is designed with closed-ended questions because letting respondents be able to fill a blank space would have brought to incomplete or not relevant answers. Before seeing the results, it must be highlighted how the questionnaire has been designed. It is divided in two parts: the introductory part and the concept test part.

In the **introductory part**, respondents are asked to answer to some questions which defines their behavioural characteristics. Besides trying to infer their predisposition towards Amazon, some questions are intended to perceive their degree of innovativeness, their knowledge and usage of voice assistants, their interest in online shopping, etc. In the final part of the introduction, some questions related to food habits have been asked to see their purchase behaviour and if they are familiar with cooking and eating at home. This has been done considering that one of the features of the Amazon device is to let people be able to order fresh food. Therefore, it might be interesting to see if respondents enjoy cooking.

The **concept test part** is the main part of the survey. Here, the Amazon Dash Wand is proposed with four different designs and the same questions are asked for each stimulus presented.

1. **Concept test 1 – Statement**. The first concept is a verbal description of features of the Amazon Dash Wand. The tone used is factual:

   "*The Amazon Dash Wand is a tool for shopping groceries and home supplies on Amazon.com. It's equipped with Wi-Fi and a barcode-scanner. To use it, you simply point it at an item's barcode, and if the wand recognizes the item, it will make a noise, flash off the light and add the item to your Amazon cart. Now, it is provided with Alexa (Amazon's voice assistant) that can answer to your general questions, convert cups to ounces and help you look up recipes. Moreover, it can be integrated with your smart home. It is a small handheld device, water-resistant and magnetic, so it can be stuck to your fridge.*"

2. **Concept test 2 – Statement with visual and online reviews**. The previous statement with the same stripped tone is proposed again for this second type of test. What differs is that the image of the product is revealed, and it is combined with real online customer reviews found on Amazon.com. Therefore, in this case it has been created a real virtual-shopping experience: customers are looking for the device on Amazon.com and they found other end-users reviews, which may help their final purchase decision (fig. 6).

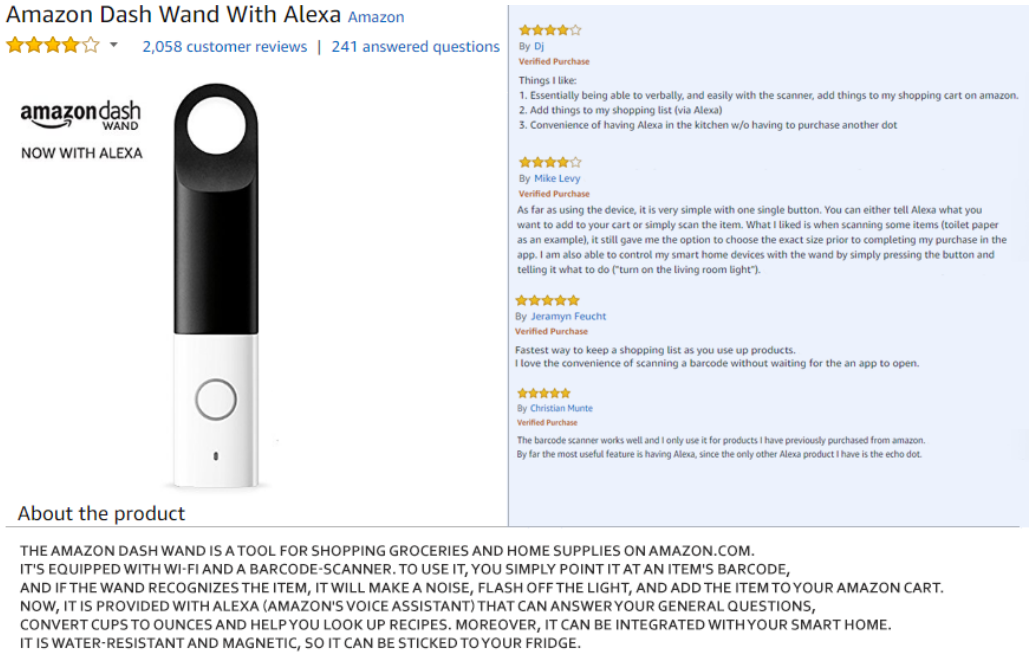

Figure 6 – Concept test 2 with statement, visual and online reviews

3. **Concept test 3 – Visual**. This third concept propose the image of the product with a description of its features and benefits in an embellished tone (fig. 7):

Figure 7 - Concept test 3

4. **Concept test 4 – Virtual prototype**.

Figure 8 - Screenshots of the Amazon Dash Wand with Alexa commercial

It has been used the Amazon Dash Wand with Alexa commercial, available on the Amazon YouTube Channel, that shows how the product works in a real-life situation[2]. The purpose of the video, which lasts 1.20 minutes, is to show how people can benefit from using the device. In the first scene, there is a woman in front of her fridge who asks Alexa a recipe with shrimps and the voice assistant gave her some recommendations. The woman speaks to the Wand stating the missing food that has to be ordered from AmazonFresh (i.e. "tomatoes", "mushrooms"),

[2] https://www.youtube.com/watch?v=s7IExS483wE

and afterwards, she scans the barcodes of the products that are running out of her fridge. In the following scene, the woman checks her Amazon cart before proceeding with the payment: after few hours, the groceries ordered have been delivered at her home. In the third scene, the protagonist is preparing the meal, but to continue she must know how many ounces are in a cup: hence, she asks Alexa, who immediately responds. In the meanwhile, her husband passes by the kitchen and asks Alexa to buy white wine from PrimeNow because they are expecting guests to dinner. So, Alexa assures that his favourite white wine will be at home by 4 and 6 pm. In the final scene, the male protagonist asks Alexa to diminish the lights of the kitchen, and after few minutes their friends come over. Those activities have all been done with the Dash Wand.

For each concept test, it has been provided the same set of questions, which included the six key dimensions provided by Dolan (1993) and the RED measures already mentioned in par 2.4.4.

The design stimuli have been proposed in this order to test $\mathbf{H_5}$, therefore the tone used goes from stripped (Concept 1 and 2) to persuasive **(**Concept 3 and 4). The survey ends with some demographic questions concerning gender, age, country, education and occupation.

3.4 USING AN ALREADY EXISTING PRODUCT

To conduct this research, it has been chosen to analyse the different customer reactions towards an already existing product. This might be in contrast with the original aim of conducting a concept test in a New Product Development process, which helps a company developing and then further launch a product that does not exist in the market (or, that already exists but it is an improvement of the current one, hence something changes anyway). Through this project, respondents were presented with the Amazon Dash Wand with Alexa. Even though this is an already existing product, it was perceived as new by European respondents: actually, the device is only available in the US and no launching campaign has been conducted in the European market. Therefore, this device has been presented to a sample who has never experienced such device.

A similar research was done by Peng et al. (2011). In their study, the authors have randomly selected four different heterogeneous products (i.e. Apple iPod Touch, Omron Pedometer, Sony E-Reader, Polaroid Pocket Instant Camera and Sony Ericsson Smartphone) which were high and less innovative. By that time, those devices were newly available in the market, so the sample was presented with something that already existed. The products were presented through two environments: traditional (a picture with a brief description paragraph) and virtual (a 3D picture with the same description). To further characterize the virtual environment, they have included the online reviews of customers who have already used the product shown with the help of Amazon.com. This was also useful to assess the impact of WOM. That's why in this project it has been chosen to do the same thing adding the online customers' reviews of the Amazon

Dash Wand with Alexa to the static picture of the device in order to assess whether the Word of Mouth might have had an influence on the final overall concept evaluation.

3.5 DATA ANALYSIS METHODS

To check the hypotheses, descriptive univariate statistical analyses have been used. Means, modes and standard deviation are the most common ones and they have been adopted even in this work. Moreover, top-box and top-two-box scores have been used to see the percentage of positive answers among the all sample. Those measures have been used especially to test $\mathbf{H_1}$, but they will be proposed again for other hypotheses too. Mean comparison has been done as well. Apart from those analyses, in this research it has been assessed the correlation between two variables when checking $\mathbf{H_2}$ doing descriptive bivariate statistical analyses. All the analyses have been conducted through Excel and SPSS.

CHAPTER 4 – ANALYSING THE DATA

4.1 OVERVIEW OF THE SAMPLE

Before starting to test the hypotheses, it would be interesting to have a general overview on the responses gathered. It has been chosen to test a product not available and not known in the European countries in order to see their reaction on something they have never experienced.

246 answers have been collected through the survey, in which 198 come from Italy while the others come from Denmark, UK, Germany and Czech Republic. Male respondents represent 56% of the sample, therefore the remaining 44% is composed by females. Concerning the age, 78% belongs to the range 18-24, 16% belongs to the interval 25-34 while only 6% is more than 35 years old. The majority of the sample is composed by students (80%), while 17% are employed, 1% are freelance professionals and 2% are retired.

Talking about Amazon, only 7% had never used this site. For the remaining 93%, respondents had a multiple choice question to indicate which were the products usually bought by them: accessories for electronic devices (70%), books and eBooks (58%), electronic devices (57%), clothing, shoes and jewellery (39%), beauty and personal care (30%) video games (14%), grocery and gourmet food (8%), musical instruments (5%) and furniture (5%). Among the sample, only 35% of respondents were Prime Members who use their membership to access Prime Day, getting exclusive discounts and for fasting shopping delivery.

Voice assistants have different results: 76% of people have already use them and the most known is Siri (89% of respondents). Alexa is famous for 60%, not bad considering that the majority of respondents come from Italy in which Amazon Echo is not fully spread. The ones who have tried voice assistants (76%) have used it to play some music (63%), get information (58%), send messages (56%), schedule an appointment (40%), book reservations (27%), create to do/shopping list (15%), interact with the smartphone (9%), order food (5%) and pay/buy something online (1%). Finally, when talking about food habits, 73% enjoys cooking and goes out to eat often (49% once a week, 40% 2-3 times per week), while buying groceries is done once a week for 57%.

Having done this overview of the sample, it might be interesting to do some related analyses.

4.1.1 COUNTRY DIFFERENCES

As previously mentioned in par. 2.4.4, we have used the six key dimensions provided by Dolan (1993) and considered to be essential to measure concept testing: overall liking, importance, uniqueness, solve problem, believability and purchase intention. Therefore, it has been done an analysis to see whether there are differences in stimuli designs reactions among different countries. Those key dimensions have been proposed with a five-point scale in which 1=extremely negative and

5=extremely positive (e.g. 1=dislike a great deal, 5=like a great deal). Comparing the means of Italy (80% of respondents), UK (13%), Denmark (4%), Germany (2%) and Czech Republic (1%), here are the results:

CONCEPT TEST 1 – STATEMENT						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Italy	2,47	2,48	2,80	1,67	3,30	2,05
UK	2,68	2,71	3,03	2,12	2,68	2,34
Denmark	3,22	2,22	3,00	2,67	3,44	3,00
Germany	3,50	2,00	3,50	2,00	2,00	3,00
Czech Republic	4,00	4,00	4,00	4,00	4,00	4,00
CONCEPT TEST 2 – VISUAL + STATEMENT + ONLINE REVIEWS						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Italy	2,76	2,36	3,18	2,10	3,33	2,51
UK	2,56	3,09	3,34	2,37	3,25	2,59
Denmark	3,22	2,78	3,11	3,00	3,44	2,78
Germany	2,00	2,00	3,00	2,00	2,50	2,50
Czech Republic	4,00	4,00	3,00	4,00	4,00	4,00
CONCEPT TEST 3 – VISUAL						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Italy	2,91	2,65	3,08	2,27	3,29	2,92
UK	3,00	2,56	3,31	2,31	3,12	2,50
Denmark	3,00	2,78	3,00	3,33	3,78	3,00
Germany	2,50	3,00	3,00	2,50	2,50	2,50
Czech Republic	4,00	3,00	4,00	5,00	5,00	5,00
CONCEPT TEST 4 – VIRTUAL PROTOTYPE						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Italy	3,92	3,90	4,57	3,70	4,69	3,76
UK	3,21	3,31	3,31	3,40	3,46	2,90
Denmark	3,33	3,00	3,00	3,33	3,11	3,33
Germany	3,00	2,50	3,00	2,00	3,00	2,00
Czech Republic	4,00	2,00	3,00	4,00	5,00	4,00

Table 1 - Country differences in the overall concept evaluation

Italy is the only one who has yield the most negative evaluations in Concept Test 1. However, it can be noticed an improvement from Test 1 to Test 4, in which answers gathered are mostly grouped in the interval 3-4, highlighting a positive feedback in the device. The major improvements have been encountered by solve problem which has experienced an increase of 2 points from Test 1 to Test 4. Overall, the other key measures have enhanced by 1,40 points and in some cases even more (e.g. purchase intention augmented of 1,71).

UK has shown a less negative evaluation in Test 1 than Italy, but still the general predisposition is not positive. Going on, the means improve: actually, results are better than Italy, reaching a neutral position. However, the situation is not like the Italian one because UK respondents continue to have a general neutral predisposition even in Concept Test 4, where means are included in the interval 2,90 – 3,46. This highlights that the product is experiencing positive feedbacks, however it does not reach an extreme propensity towards 4 or 5 as the Italian country.

Things are different for the remaining nations. Respondents coming from Czech Republic are the only ones who were overall convinced on the device throughout the entire survey, their responses were mostly positive. On the other hand, Denmark is the only one who has overall given neutral responses, most of them belonging to the interval of 2-3 without extremely propending to 1 or 5. Same considerations can be expressed for Germany, which has however shown a more negative predisposition than Denmark: apart from overall liking in Test 1, the other key measures' means in the four concept tests do not exceed the value of 3. Anyway, those three countries together represent only the 7% of the sample.

To sum up, there are differences among countries when evaluating a concept. Italy is the only one who has experienced an improvement from Concept Test 1 to Concept 4: this means that this country has a major propensity than the other countries in buying this device. The others, do not have a negative evaluation but their general propensity is not extremely positive as Italy.

4.1.2 DIFFERENCES BETWEEN PRIME MEMBERS AND REGULAR MEMBERS

Even though they only represent the 35%, it would be interesting to see if there are differences in the overall concept evaluation between Prime Members and Regular Members. This has been done considering that the Dash Wand is only available for Prime Member. Since this membership provides a lot of benefits such as AmazonFresh, maybe there are different results from respondents who have already experienced all the features of being a Prime member. Therefore, they might be more willing to buy this device compared to Regular members.

CONCEPT TEST 1 – STATEMENT						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Prime Members	2,54	2,47	2,95	1,80	3,12	2,23
Regular Members	2,58	2,54	2,81	1,80	3,27	2,12
CONCEPT TEST 2 – VISUAL + STATEMENT + ONLINE REVIEWS						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Prime Members	2,81	2,49	3,31	2,30	3,38	2,61
Regular Members	2,72	2,49	3,12	2,12	3,29	2,51
CONCEPT TEST 3 – VISUAL						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Prime Members	2,92	2,53	3,14	2,37	3,18	2,85
Regular Members	2,94	2,72	3,10	2,34	3,36	2.90
CONCEPT TEST 4 – VIRTUAL PROTOTYPE						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Prime Members	3,83	3,72	4,35	3,74	4,54	3,68
Regular Members	3,77	3,76	4,27	3,55	4,39	3,56

Table 2 – Differences between Prime Members and Regular Members

As seen in Table 2, there is a general improvement from Concept Test 1 to Concept test 4 as already experienced in par. 4.1.2. However, there are no big differences between Prime and Regular members: their general responses are similar in each concept test. Therefore, in this case having a particular membership has not influenced the general overall evaluation.

To further highlight this result, it has been done a bivariate analysis to state the correlation between the Amazon membership and the concept testing key dimensions. This has been done through the Pearson correlation coefficient adopting a p-value of 0,05 (see Table 3). As shown, there are no correlations between the two variables in each of the four concept tests. Hence, Amazon membership does not affect the overall concept test evaluation.

CONCEPT TEST 1 - STATEMENT						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Pearson correlation	0,015	0,037	-0,064	0,002	0,068	-0,058
Significance of correlation	Not significant	Not significant	Not significant	Not significant	Not significant	Not significant
CONCEPT TEST 2 – VISUAL + STATEMENT + ONLINE REVIEWS						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Pearson correlation	-0,052	-0,002	-0,125	-0,099	-0,052	-0,053
Significance of correlation	Not significant	Not significant	Not significant	Not significant	Not significant	Not significant
CONCEPT TEST 3 – VISUAL						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Pearson correlation	0,016	0,089	-0,023	-0,018	0,114	0,034
Significance of correlation	Not significant	Not significant	Not significant	Not significant	Not significant	Not significant
CONCEPT TEST 4 – VIRTUAL PROTOTYPE						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Pearson correlation	-0,034	0,024	-0,034	-0,113	-0,081	-0,066
Significance of correlation	Not significant	Not significant	Not significant	Not significant	Not significant	Not significant

Table 3 – Correlation between Prime Members and Regular Members

4.1.3 DIFFERENCES IN PRODUCT CATEGORIES BOUGHT ON AMAZON

We have mentioned that only 8% of respondents have already bought groceries and gourmet food on Amazon (see par. 4.1). Therefore, let's see if their feedback on the device is more positive than people who have never bought groceries online. Moreover, we will also see if there is a particular product category in which Amazon should insist to promote this device.

CONCEPT TEST 1 - STATEMENT						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Books	2,63	2,42	2,86	1,76	3,22	2,18
Clothing	2,67	2,52	2,85	1,91	3,20	2,30
Electronic devices	2,57	2,50	2,86	1,75	3,12	2,12
Accessories	2,54	2,53	2,84	1,71	3,23	2,09
Video Games	2,28	2,48	2,65	1,65	2,57	1,97
Groceries	2,10	2,21	2,42	1,68	2,42	1,78
Instruments	2,23	2,38	2,53	1,76	2,53	2,07
Personal care	2,35	2,57	2,67	1,79	3,02	2,01
Furniture	3,16	4,25	3,41	2,50	3,83	2,58
CONCEPT TEST 2 – VISUAL + STATEMENT + ONLINE REVIEWS						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Books	2,76	2,41	3,22	2,18	3,30	2,62
Clothing	2,83	2,52	3,21	2,25	3,39	2,64
Electronic devices	2,76	2,38	3,22	2,19	3,32	2,58
Accessories	2,69	2,45	3,18	2,13	3,34	2,49
Video Games	2,48	2,62	3,34	2,22	3,02	2,45
Groceries	2,73	2,21	2,94	2,47	3,26	2,63
Instruments	2,07	2,53	2,84	1,92	2,53	2,07
Personal care	2,76	2,57	3,23	2,20	3,30	2,52
Furniture	3,33	4,08	3,58	3,25	3,83	2,91
CONCEPT TEST 3 – VISUAL						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Books	2,93	2,48	3,01	2,32	3,23	2,88
Clothing	3,10	2,87	3,11	2,42	3,33	3,05
Electronic devices	3,01	2,56	3,11	2,33	3,24	2,89
Accessories	2,84	2,61	3,07	2,26	3,27	2,85
Video Games	2,80	2,40	3,11	2,25	2,91	2,40
Groceries	3,05	2,63	2,73	2,78	3,26	3,05
Instruments	2,07	2,38	3,00	1,92	2,86	2,07
Personal care	2,90	2,64	3,08	2,38	3,23	2,84
Furniture	2,91	3,75	3,58	3,08	3,67	2,91
CONCEPT TEST 4 – VIRTUAL PROTOTYPE						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Books	3,78	3,62	4,22	3,57	4,34	3,56
Clothing	3,97	3,88	4,40	3,65	4,50	3,73
Electronic devices	3,89	3,72	4,38	3,74	4,60	3,68
Accessories	3,89	3,85	4,44	3,64	4,56	3,65
Video Games	3,54	3,57	4,02	3,71	4,20	3,31
Groceries	3,78	3,78	4,21	3,78	4,73	3,78
Instruments	3,15	3,15	3,76	2,84	3,46	3,00
Personal care	3,97	3,97	4,50	3,82	4,64	3,67
Furniture	3,50	4,41	4,50	3,50	4,33	3,08

Table 4 – Concept tests evaluation considering product categories

Surprisingly, respondents who have already bought groceries on Amazon have provided the most negative feedback in Concept Test 1. For instance, purchase intention has the lowest value (1,78). While the other product categories are quite similar, the only one that has portrayed a more positive feedback is "furniture". However, they only represent 5% of the sample.

In Concept Test 2, "groceries" enhances a little even for purchase intention which becomes aligned with the other product categories. Still, the general evaluation is moderately negative, laying in the interval 2,21 – 3,26. This means that those respondents are still not fully convinced with the device features and benefits. In the other product categories, results are basically aligned apart from "furniture". A similar situation is encountered in Concept Test 3, in which there is a slight improvement for everyone, while in Concept Test 4 there is a major improvement for each product category.

To sum up, there is a general improvement from Test 1 to Test 4 for each product category. Despite the initial reluctance of giving a positive feedback, people who have already bought groceries on Amazon have changed their mind among the four tests. They have also shown the highest purchase intention (3,78). The only product category that has experienced a general positive feedback is "furniture", however it might not be recommendable to insist on them since they only represent the 5% of the sample, therefore a small portion. Apart from it, interesting results come from "clothing", "electronic devices", "accessories" and "personal care", which have overall provided the highest scores in the four tests. Therefore, those might be the leading categories for this device.

4.2 TESTING HYPOTHESES

After having done a preview of the sample with the related analyses, it is now time to test the hypotheses previously mentioned in par. 2.6. After each analysis, it will be stated whether the hypothesis is confirmed or not.

4.2.1 H1: DIFFERENT STIMULI DESIGN LEAD TO DIFFERENT OUTCOMES

To study **H1**, descriptive univariate analyses will be done. The key measure provided by Dolan (1993) have been used to test whether there are different outcomes in the overall concept evaluations. The key measures are: overall liking, importance, uniqueness, solve problem, believability and purchase intention. They have been measured with a scale from 1 to 5 in which 1=extremely negative and 5=extremely positive.

CONCEPT TEST 1 – STATEMENT						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Mean	2,56	2,52	2,86	1,80	3,21	2,16
Mode	2	2	3	1	3	2
St. Dev.	0,995	1,005	1,038	0,995	1,068	0,898
CONCEPT TEST 2 – VISUAL + STATEMENT + ONLINE REVIEWS						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Mean	2,76	2,49	3,19	2,19	3,32	2,55
Mode	3	2	3	2	3	3
St. Dev.	0,785	0,946	0,758	0,908	0,871	0,914
CONCEPT TEST 3 – VISUAL						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Mean	2,93	2,65	3,11	2,35	3,29	2,89
Mode	3	2	3	2	3	3
St. Dev.	0,756	1,041	0,851	0,872	0,738	0,740
CONCEPT TEST 4 – VIRTUAL PROTOTYPE						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Mean	3,79	3,75	4,30	3,62	4,45	3,60
Mode	4	4	5	4	5	4
St. Dev.	0,866	0,866	1,062	0,822	0,932	0,829

Table 5 - Univariate statistics for the four concept tests

4.2.1.1 MEANS

There is a general negative predisposition towards the product when it comes to Concept Test 1. The solve problem indicator is the most negative one with a mean of 1,80, showing that end-users do not think that this device would solve their problem. Purchase intention is quite low too (2,16) indicating that the intention to not buy the product is quite high. All the other indicators are similar in terms of being perceived as almost intermediate values (moderately important, moderately unique, etc.) apart from believability, which specifies that overall features and functionalities proposed by the device are perceived as somewhat believable (3,21) even though is quite close to the intermediate value of "neither believable or unbelievable".

In Concept Test 2, the values are similar to Concept Test 1, apart from uniqueness (3,19) and solve problem (2,19), which have gained more positive answers compared to the first concept. Therefore, the product is perceived as more unique and more problem-solver than the previous concept and the key measures of Concept Test 2 are better than Concept Test 1, but values are not so different. For instance, purchase intention has slightly increased (from 2,16 to 2,55). Same considerations on similarity can be stated for Concept Test 3, whose means are even more closer to Concept Test 2, even though purchase intention (2,89) has slightly increased. The three tests show a resemblance among their final evaluations, although Concept Test 1 is the worst in terms of overall willingness to buy the product.

Everything changes with Concept Test 4: each mean has varied of 1 point or more compared to the previous tests. The ones that have encountered the major change are uniqueness (4,30), solve problem (3,62) and purchase intention (3,60). This is quite interesting considering that Concept Test 4 is the only one that lets end-users to experiment a "real-life" context.

4.2.1.2 MODES

In the first three concept tests the indicators of importance, uniqueness and believability maintain the same mode of, respectively, 2, 3 and 3. This means that respondents do not perceive the device with important features and functionalities and what is stated is considered to be moderately unique and moderately believable. Therefore, there is no change in terms of the most frequent value in the three concept tests when it comes to importance, uniqueness and believability. This is highlighted in the mean, in which the overall believability indicator shows no considerable changes (see Table 5). However, despite the frequency of the same values among the three tests, the mean shows that there is a small improvement for importance and uniqueness. For instance, in Concept Test 1, uniqueness has a mean of 2,86 while in Concept Test 2 the mean is 3,19: in both cases, though, the mode continues to have the same value of 3.

Overall, modes of Concept Test 2 and 3 show values similar to the ones presented in Concept Test 1. However, comparing the first test with the other two, it can be seen a small improvement of 1 point-scale for overall liking, problem solving and purchase intention indicators. Therefore, the most frequent values show more positive evaluations towards the device compared to the first test. But, when we see the means, it can be noticed that the improvement is less than 1 point: for example, the enhancement for overall liking goes from 2,56 in Test 1 to 2,93 in Concept Test 3. Even though there are minor changes among the three tests, the more the respondents are seeing different designs, the more they are changing their evaluations from less to more favourable.

When comparing Concept Test 1 to Concept Test 4, it is easier to see the positive changes in the evaluations. Each indicator displays a transformation of 2 point-scale, except for problem solving which has encountered the major improvement from a mode of 1 in Concept Test 1 to a mode of 4 in Concept Test 4, therefore a change of 3 point-scale. This highlights that the general evaluation towards the product is positively increasing in the final concept test. Same considerations can be inferred for the remaining indicators, which have portrayed a great enhancement.

4.2.1.3 STANDARD DEVIATION

Overall, Concept Test 1 and Concept Test 4 have standard deviation values higher than Concept Test 2 and Concept Test 3. In Concept Test 1, the highest standard deviation lies in uniqueness and believability. For instance, the uniqueness indicator shows a standard deviation of 1,038, which indicates that respondents fall in the interval 2,3 and 4 (mean of 2,86). The lowest standard deviation lies in purchase intention (0,898), in which with a mean of 2,16 the answers are grouped in the interval 2-3.

In Concept Test 2, the standard deviation parameters are more homogenous, with the highest at 0,946 (importance) and the lowest at 0,758 (uniqueness): even though there still is a consistent deviation from the mean in most of the answers given in the entire test, here the values are more similar. Same considerations can be inferred from Concept Test 3: here, however, the highest and the lowest values (1,041 for importance and 0,738 for availability) exceed the ones highlighted in Concept Test 2. In both concept tests, their overall deviations are lower than the first one.

Concept Test 4 has high values of standard deviation compared to Concept Test 2 and 3. Here, the highest parameter lies in uniqueness (1,062), while the lowest is problem solving (0,822). In Concept Test 1 and 4 answers are less homogeneous and grouped towards the same values than Concept Test 2 and 3. This highlights that in the first and in the last concept there is a wider deviation from the mean compared to the other two concept tests. Overall, those univariate statistics show that there is an improvement in the final evaluations from Concept test 1 to Concept test 4, even though values for the first three concepts can be considered similar.

4.2.1.4 TOP-BOX AND TOP-TWO-BOX SCORES

Top-Box and Top-Two-Box scores have been used to give a further explanation. The results of both measures are expressed in percentage and they have been rounded up to the next digit (see Table 6).

CONCEPT TEST 1 - STATEMENT						
	Overall liking	**Importance**	**Uniqueness**	**Problem Solving**	**Believability**	**Purchase Intention**
Mean	2,56	2,52	2,86	1,80	3,21	2,16
Top-Box	2%	6%	6%	8%	9%	10%
Top 2 Box	20%	13%	27%	8%	38%	10%
CONCEPT TEST 2 – VISUAL + STATEMENT + ONLINE REVIEWS						
	Overall liking	**Importance**	**Uniqueness**	**Problem Solving**	**Believability**	**Purchase Intention**
Mean	2,76	2,49	3,19	2,19	3,32	2,55
Top-Box	13%	5%	6%	2%	9%	1%
Top 2 Box	13%	14%	26%	9%	34%	9%
CONCEPT TEST 3 – VISUAL						
	Overall liking	**Importance**	**Uniqueness**	**Problem Solving**	**Believability**	**Purchase Intention**
Mean	2,93	2,65	3,11	2,35	3,29	2,89
Top-Box	2%	9%	9%	3%	2%	2%
Top 2 Box	14%	18%	21%	11%	36%	10%
CONCEPT TEST 4 – VIRTUAL PROTOTYPE						
	Overall liking	**Importance**	**Uniqueness**	**Problem Solving**	**Believability**	**Purchase Intention**
Mean	3,79	3,75	4,30	3,62	4,45	3,60
Top-Box	11%	9%	60%	3%	65%	1%
Top 2 Box	79%	76%	76%	74%	85%	74%

Table 6 – Top-box and top-two-box scores for the four concept tests

In the first three Concept Tests, the Top-Box and the Top-Two-Box scores are similar for the following indicators: problem solving, credibility and purchase intention. This confirms what stated with the previous univariate statistics (see Table 5), in which the results of the first three concept tests were similar. Things are different for the other indicators which are slightly homogeneous between each other. For some indicators, percentages are identical because it was not possible to include the top two boxes since there were not sufficient answers (e.g. respondents only answered "Probably buy" and no one answered "Definitely buy").

In the first three concept tests there is a less favourable predisposition towards the device, even though some measures portray minor improvements. For instance, importance slightly enhances from Concept Test 1 to Concept Test 3 for both top-boxes scores. In Concept Test 4 we can see a great improvement in every indicator. Top-box scores for uniqueness and believability have exceeded 50%, indicating that the majority has assumed a positive predisposition towards the device for those two indicators. On the other hand, top-two-box scores have consistently changed from the first concept test to the last one. Every indicator has surpassed 50%: overall liking and believability exceed the 80%, while importance, uniqueness and purchase intention have almost reached this edge. The major enhancement has been encountered by the purchase intention indicator which has started with a 10% of "Probably buy" in Concept Test 1 but has achieved a 77% of both "Definitely buy" and "Probably buy" in the end.

Considering the results, it has been noted that depending on the way in which the concept is presented, different perceptions on the product are expressed. Hence, $\mathbf{H_1}$ is confirmed.

4.2.2 H2: THE ROLE OF PRODUCT EXPERTISE

Product expertise has been analysed with bivariate statistics using the Pearson correlation coefficient. In the survey, prior expertise is tested with question 1.8. To state the significance of the correlation between product expertise and key measures (see par. 1.4), it has been adopted the p-value of 0,05.

CONCEPT TEST 1 – STATEMENT						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Pearson correlation	0,120	0,178	0,134	0,192	0,154	0,210
Significance of correlation	Not significant	Significant	Significant	Significant	Significant	Significant
CONCEPT TEST 2 – VISUAL + STATEMENT + ONLINE REVIEWS						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Pearson correlation	0,164	0,163	0,009	0,234	0,074	0,194
Significance of correlation	Significant	Significant	Not significant	Significant	Not significant	Significant

CONCEPT TEST 3 – VISUAL						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Pearson correlation	0,089	0,039	0,169	0,261	0,71	0,115
Significance of correlation	Not significant	Not significant	Significant	Significant	Not significant	Not significant
CONCEPT TEST 4 – VIRTUAL PROTOTYPE						
	Overall liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Pearson correlation	- 0,043	- 0,067	- 0,108	- 0,021	- 0,019	- 0,009
Significance of correlation	Not significant	Not significant	Not significant	Not significant	Not significant	Not significant

Table 7 – Bivariate descriptive statistics

In Concept Test 1, the correlation between product expertise and the key measures is significant; the only indicator that has resulted as not significant is the overall liking, which exceeds the p-value of only 0,01 (see Table 7). Therefore, this indicator is almost significant. A similar situation is presented in Concept Test 2 which, however, shows uniqueness and believability as not significant. Moreover, according to the Pearson coefficient, the degree of correlation (when it is significant) is mostly low because each value lies in the interval 0<p-value< 0,3.

As the concept tests change, the correlation varies too. It is interesting to note that for the first two concept tests the degree of correlation is significant for almost every indicator, but in Concept Test 3 it is significant only for uniqueness and solve problem, while in Concept Test 4 there is no significance at all. This highlights that for concept tests with a factual tone, prior expertise influences the key measures, probably because people who are familiar with the device do not need further explanation to evaluate the concept. The partially inexistent correlation in Concept Test 3 and the completely absence of correlation in Concept Test 4 show that prior expertise is not necessary when the product is fully shown, especially with an explicative video. Therefore, $\mathbf{H_2}$ is confirmed for concept tests with a factual tone and a descriptive purpose because, as shown in Table 7, there is interdependence between variables.

4.2.3 H3: THE DEGREE OF INNOVATIVENESS IN THE CONCEPT EVALUATION

To measure the degree of innovativeness, it has been analysed question 1.9. For doing so, the sample of 246 respondents has been split in five groups according to Roger's Innovation Adoption curve: innovators (7%), early adopters (50%), early majority (36%), late majority (5%) and laggards (2%). After that, means have been compared to assess the differences among the groups in terms of key measures (see Table 8).

CONCEPT TEST 1 – STATEMENT						
	Overall Liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Innovators	3,33	3,11	3,44	2,56	3,89	2,78
Early adopters	2,39	2,40	2,67	1,74	3,21	2,22
Early majority	2,61	2,64	3,07	1,80	3,18	1,96
Late majority	2,33	1,67	2,51	1,16	2,50	1,67
Laggards	4,00	3,00	3,00	2,00	3,00	3,00
CONCEPT TEST 2 – VISUAL + STATEMENT + ONLINE REVIEWS						

	Overall Liking	Importance	Uniqueness	Solve Problem	Believability	Purchase Intention
Innovators	3,00	3,11	3,44	2,56	3,89	2,89
Early adopters	2,91	2,44	3,24	2,34	3,32	2,83
Early majority	2,61	2,54	3,09	1,92	3,26	2,13
Late majority	2,16	1,83	3,16	1,67	2,67	2,00
Laggards	2,00	2,00	3,00	2,00	4,00	3,00
CONCEPT TEST 3 – VISUAL						
	Overall Liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Innovators	3,33	3,11	3,44	3,22	3,67	3,33
Early adopters	3,04	2,60	3,26	2,41	3,25	3,01
Early majority	2,84	2,70	2,82	2,20	3,34	2,72
Late majority	2,33	2,00	3,33	1,67	3,00	2,17
Laggards	2,00	3,00	3,00	2,00	3,00	3,00
CONCEPT TEST 4 – VIRTUAL PROTOTYPE						
	Overall Liking	**Importance**	**Uniqueness**	**Solve Problem**	**Believability**	**Purchase Intention**
Innovators	3,56	3,56	3,67	3,56	3,89	3,22
Early adopters	3,79	3,78	4,42	3,72	4,61	3,78
Early majority	3,89	3,76	4,26	3,52	4,37	3,42
Late majority	3,83	3,50	4,50	3,33	4,33	3,67
Laggards	4,00	4,00	4,00	4,00	4,00	4,00

Table 8 – Innovation groups' means

Innovators' group portrays an overall accordance among the four concept tests: actually, their general predisposition towards the concept is quite similar in each case. Moreover, it seems like they have perceived the possible benefits and features of the product not being influenced by the stimuli design. In fact, their general evaluations in each test lie in the interval 2,56-3,89, which indicates that apart for some measures (namely solve problem and purchase intention) they have found the potentiality of the device without propending to an extremely positive evaluation of 5. When examining this group with the rest, it is interesting to note that the others tend to enhance their evaluation from Concept Test 1 to Concept Test 4, exactly what happened when testing H_1. This means that the other groups are influenced by the stimuli design, which might be necessary for people who are not so willing to adopt a new product as soon as it launched in the market.

Therefore, innovative consumers do not need to have further explanation on the device because they are already perceiving features and benefits as stated, their willingness to try new products let them be more critical when evaluating the concept. The same cannot be assured for the other groups of end-users: as shown in Table 8, they seem not able to imagine the device benefits and how it works until someone does not show that to them (in this case, with a virtual prototype). Therefore, since innovators have demonstrated to be able to objectively judge a concept, $\mathbf{H_3}$ is confirmed.

4.2.3.1 RESPONDENTS WHO WERE FAMILIAR WITH THE DASH SERVICES

At the end of each test, there was a question asking "*with what similar products have you compared the Amazon Dash Wand with Alexa?*". Respondents were free to answers because it was an open-ended question. Most of them has found it difficult to give a comparison with an existing product (65%), while others have compared the Dash Wand with voice assistants (12%), smartphones (2%). The remaining 21% has compared the device with Amazon Echo (13%) and Amazon

Dash Buttons (8%). Amazon Echo shares with the Dash Wand the implementation of Alexa, while Dash Buttons belong to Amazon Dash consumer goods ordering service like the Dash Wand: it allows consumers to reorder home appliances just pressing a physical button. Hence, it is interesting to see if those respondents have been provided more positive answers than the other groups of innovation adopters. For doing so, it has been done a mean comparison with the previous table. Table 9 shows the means of respondents who have compared the device with Amazon Echo and Dash Buttons.

CONCEPT TEST 1 - STATEMENT						
	Overall liking	**Importance**	**Uniqueness**	**Problem Solving**	**Believability**	**Purchase Intention**
Amazon Echo	2,05	2,08	2,22	1,25	2,91	2,09
Dash Buttons	2,42	2,26	2,36	1,15	2,78	2,00
CONCEPT TEST 2 – VISUAL + STATEMENT + ONLINE REVIEWS						
	Overall liking	**Importance**	**Uniqueness**	**Problem Solving**	**Believability**	**Purchase Intention**
Amazon Echo	2,94	2,05	3,05	2,06	3,00	2,94
Dash Buttons	2,89	2,10	3,10	2,00	3,10	2,78
CONCEPT TEST 3 – VISUAL						
	Overall liking	**Importance**	**Uniqueness**	**Problem Solving**	**Believability**	**Purchase Intention**
Amazon Echo	2,88	2,05	2,94	2,07	2,96	3,00
Dash Buttons	3,10	2,21	2,89	2,31	3,10	3,00
CONCEPT TEST 4 – VIRTUAL PROTOTYPE						
	Overall liking	**Importance**	**Uniqueness**	**Problem Solving**	**Believability**	**Purchase Intention**
Amazon Echo	4,00	4,00	4,88	4,00	4,94	4,00
Dash Buttons	4,00	3,89	4,68	3,89	4,89	3,89

Table 9 – Respondents who know similar products

Overall, from Test 1 to Test 4 results look very similar. Therefore, this 21% of the sample has experienced a major improvement when encountering each stimuli design. If we compare those answers with the one provided by innovators (see Table 9), we can see that there is a difference: innovators have yield similar results in each concept, there is no major improvement from Test 1 to test 4, they have provided an objective evaluation. This means that people who have associated the device with something they consider similar, have not provided the same kind of answers of innovators. Hence, knowing which product is the most comparable does not provide objective results because respondents are still influenced by the stimuli design.

4.2.4 H4: THE POWER OF WORD OF MOUTH

WOM has been expressed through Concept Test 2. As already noticed in Table 5, the overall means and modes show that there is a small increment from Concept Test 1 to Concept Test 2: the indicators that have expressed the major variation

are uniqueness, solve problem and purchase intention. Therefore, it seems that customers online reviews have influenced the key measures compared to the first concept test. But when analysing Concept Test 3, results are quite similar. Actually, for some aspects they are even better, although the increment is minimum.

What is interesting to note is that results are better in Concept Test 4 rather than Concept Test 2 (in terms of key dimensions of Concept Testing). This highlights that WOM has not positively influenced the final concept evaluation compared to the other tests; it seems that consumers have not been conditioned from the words of people who have already used the device and have had positive experience with it.

It might be interesting to check whether non-expert consumers are positively influenced or not by WOM. Considering question 1.8, respondents have been divided into four groups: extremely experts (4%), experts (41%), moderately experts (48%) and non-experts (see Table 10). After this subdivision, their means have been compared.

CONCEPT TEST 1 –STATEMENT						
	Overall liking	**Importance**	**Uniqueness**	**Problem Solving**	**Believability**	**Purchase Intention**
Extremely Experts	3,55	3,33	3,56	2,67	3,56	3,33
Experts	2,53	2,59	2,91	1,90	3,28	2,19
Moderately Experts	2,56	2,45	2,83	1,70	3,23	2,10
Non-Experts	2,27	2,11	2,50	1,44	2,56	1,78
CONCEPT TEST 2 – VISUAL + STATEMENT + ONLINE REVIEWS						
	Overall liking	**Importance**	**Uniqueness**	**Problem Solving**	**Believability**	**Purchase Intention**
Extremely Experts	3,78	3,33	3,22	2,89	4,00	3,56
Experts	2,76	2,53	3,16	2,34	3,25	2,60
Moderately Experts	2,70	2,43	3,24	2,06	3,38	2,47
Non-Experts	2,62	2,16	3,00	1,78	3,00	2,22
CONCEPT TEST 3 – VISUAL						
	Overall liking	**Importance**	**Uniqueness**	**Problem Solving**	**Believability**	**Purchase Intention**
Extremely Experts	3,78	2,78	3,56	3,44	4,11	3,89
Experts	2,91	2,68	3,18	2,50	3,25	2,87
Moderately Experts	2,88	2,64	3,10	2,17	3,25	2,81
Non-Experts	3,00	2,56	2,62	2,17	3,39	3,00
CONCEPT TEST 4 –VIRTUAL PROTOTYPE						
	Overall liking	**Importance**	**Uniqueness**	**Problem Solving**	**Believability**	**Purchase Intention**
Extremely Experts	4,22	3,11	3,67	4,22	4,78	4,00
Experts	3,73	3,78	4,29	3,52	4,40	3,56
Moderately Experts	3,75	3,76	4,30	3,63	4,43	3,58
Non-Experts	4,22	3,83	4,67	3,83	4,67	3,83

Table 10 – The influence of WOM on expert and non-experts

Despite the focus is different because Table 10 shows whether the WOM might influence the opinions of non-experts, results are similar when compared to Table 5. WOM has demonstrated to change the perception from Concept Test 1 to Concept Test 2, but the second test has not provided positive results compared to the overall concept evaluation. Actually, in Concept Test 2 the overall evaluation is still negative, even though it has improved from the first concept test. Those considerations are valid for each group of consumers, meaning that even non-experts are not influenced by WOM. Hence, $\mathbf{H_4}$ is confirmed when Concept Test 2 is compared with Concept Test 1, but it is disconfirmed for the other two concept tests.

4.2.5 H5: FROM A FACTUAL TO A PROMOTIONAL TONE

The four concept tests have been placed in a selected order that goes from a stripped description (Concept test 1 and 2) to an embellished presentation (Concept test 3 and 4). As seen in paragraph 4.2.1 (see Table 5), the purchase intention increases from Concept Test 1 to Concept Test 4. Hence, moving from a factual to a persuasive tone leads to a variation. However, the purchase intention of Concept test 2 and 3 are quite similar, the major changes happens when comparing the first and the last concept.

So far, we have analysed the key measure provided by Dolan (1993) as seen in par. 2.6. But what about the overall interest? To verify $\mathbf{H_5}$, questions included in part 3.1 of the survey have been taken into consideration to assess the overall interest. Those questions have been asked with a five-point Likert scale in which 1 = Strongly Disagree and 5 = Strongly Agree. Comparing the means among the four concept tests, it is possible to see whether a change of tone would increase the interest in the presented device.

"I am more willing to buy groceries online"				
	CONCEPT TEST 1 (factual)	**CONCEPT TEST 2 (factual)**	**CONCEPT TEST 3 (persuasive)**	**CONCEPT TEST 4 (persuasive)**
Mean	2,14	2,30	2,34	3,61
Mode	2	2	2	4
"I am looking forward to use the Amazon Dash Wand with Alexa"				
	CONCEPT TEST 1 (factual)	**CONCEPT TEST 2 (factual)**	**CONCEPT TEST 3 (persuasive)**	**CONCEPT TEST 4 (persuasive)**
Mean	2,21	2,29	2,46	3,56
Mode	2	2	2	4
"I am still not interested"				
	CONCEPT TEST 1 (factual)	**CONCEPT TEST 2 (factual)**	**CONCEPT TEST 3 (persuasive)**	**CONCEPT TEST 4 (persuasive)**
Mean	3,08	3,02	2,65	1,88
Mode	3	2	2	1
"I plan to use this device in the near future"				
	CONCEPT TEST 1 (factual)	**CONCEPT TEST 2 (factual)**	**CONCEPT TEST 3 (persuasive)**	**CONCEPT TEST 4 (persuasive)**
Mean	2,19	2,21	2,46	4,10
Mode	2	2	2	4

Table 11 – Changes in means and modes from factual to persuasive tone

In the first question, there is no big change among the first three concept tests, the majority of respondents rather disagree on the statement. However, it can be noticed an increase of the interest in buying online groceries from Concept test 1 to Concept test 3 and a great difference between Concept Test 1 and Concept Test 4. Same considerations can be inferred from the second question, which highlights a slight increase of interest in the first three concept tests and a major increase when comparing the first to the final test, which has augmented of 1,35 points. When asking people to directly state whether they were interested in the device or not, the first variation is visible between Concept test 2 and 3: here comes the first change of tone (from factual to persuasive), portraying that the interest in the product is enhancing. Those values continue to rise until they reach the amount of 1,88 (1=Strongly Agree) in the final concept test, confirming that respondents have increased their interest in the product. The final question is aligned with the previous considerations: the interest slightly changes among the first three concept tests, while it definitely improves when it comes to Concept Test 4, achieving the final mean of 4,10 (4=Rather agree).

Overall, the current situation is that the interest in the device augment from Concept Test 1 to Concept Test 4. Moreover, going from Concept Test 2 to Concept Test 3 shows a greater improvement compared to the change that happens from Concept Test 1 to 2 (apart from the first answer). Even though for some answers the variation is minor (especially among the first three concepts), one final remark can be done: a change of tone from factual to persuasive enhances the interest in the product. Hence, $\mathbf{H_5}$ is confirmed.

4.3 PREDICTING PRODUCT SUCCESS

Firms complain the low predictive market and trial success of concept testing. A way to overcome such issue is by using RED Measures (Relevance, Expensiveness and Differentiation), which help in estimating if the new product would be a success (see par. 2.6). When each of these measures score a high percentage (more than 50%), the product can be considered a potential success.

Relevance was measured through a five-point scale (1 = not effective at all, 5 = extremely effective) in which consumers were asked to assess whether the device was effective in fulfilling the following needs: accessing food recipes, controlling smart home devices, having a virtual assistant, creating a shopping list, saving time and buying groceries online. Surprisingly, apart from “saving time”, considerations on effectiveness decreases from Concept Test 1 to Concept Test 3 when seeing Top-Two-Box scores. This has happened because many respondents were probably not sure of the potentialities of the Dash Wand, so they have preferred to give neutral responses. However, in Concept Test 4 the product has scored more than 80% for each need, so here it can be considered a success.

In Concept Test 1, 72% of respondents have perceived the device as “Somewhat inexpensive” and “Very inexpensive” (see Table 13). This amount enhances until reaching the edge of 81% in Concept Test 4. Considering that the perceived expensiveness is low and that it exceeds half of the sample, so far the product can still be considered a success.

When considering differentiation, in the first concept test 27% of respondents perceives the device as “very different” and “somewhat different” from other similar products they could buy (see Table 13). This percentage decreases of just 2 percentage points in Concept test 2 and 3. However, in Concept Test 4, 76% of the sample agrees on the differentiation.

Those measures asked end-users to assess some features of a product compared to similar items, which have been already analysed in par. 4.2.3.1.

RED measures, according to IPSOS research company, are better predictors of the new product success. Actually, in this case, the product achieved a great percentage in relevance and differentiation for Concept Test 4, while for expensiveness scores were high in each test. Both cases suggest that the device can be considered a potential success. Again, this confirms the main purpose of this work, which is to demonstrate that the role of stimuli design in concept testing is relevant, even for predicting market success.

RELEVANCE						
CONCEPT TEST 1						
	Accessing food recipes	Controlling smart home devices	Having a virtual assistant	Buying groceries online	Creating a shopping list	Saving time
Top-Box	2%	2%	2%	9%	13%	14%
Top 2 Box	20%	20%	26%	35%	29%	58%
CONCEPT TEST 2						
	Accessing food recipes	Controlling smart home devices	Having a virtual assistant	Buying groceries online	Creating a shopping list	Saving time
Top-Box	2%	2%	7%	7%	5%	11%
Top 2 Box	20%	12%	24%	32%	28%	70%
CONCEPT TEST 3						
	Accessing food recipes	Controlling smart home devices	Having a virtual assistant	Buying groceries online	Creating a shopping list	Saving time
Top-Box	4%	5%	4%	13%	7%	10%
Top 2 Box	18%	11%	18%	27%	26%	78%
CONCEPT TEST 4						
	Accessing food recipes	Controlling smart home devices	Having a virtual assistant	Buying groceries online	Creating a shopping list	Saving time
Top-Box	12%	13%	10%	15%	23%	67%
Top 2 Box	85%	81%	85%	86%	86%	87%

Table 12 – Relevance measure

EXPENSIVENESS				
	CONCEPT TEST 1	CONCEPT TEST 2	CONCEPT TEST 3	CONCEPT TEST 4
Top-Box	11%	29%	13%	11%
Top 2 Box	72%	76%	79%	81%
DIFFERENTIATION				
	CONCEPT TEST 1	CONCEPT TEST 2	CONCEPT TEST 3	CONCEPT TEST 4
Top-Box	6%	7%	2%	61%
Top 2 Box	27%	24%	24%	76%

Table 13 – Expensiveness and Differentiation

IPSOS states that RED measures are more efficient than solely purchase intention. If we analyse Top-Box and Top-Two-Box scores of Table 14, purchase intention goes beyond 50% only in Concept Test 4. Hence, in the other tests respondents are not extremely willing to buy this device. This is similar to what happens with Relevance, however here Top-Box and Top-Two-Box Scores are greater than the ones identified by purchase intention. When comparing purchase intention with Expensiveness, the latter has better scores, while when comparing with Differentiation the situation is similar when analysing purchase intention and Relevance. This highlights what argued by IPSOS, namely that purchase intention is not a good predictor of a product market success and other measures (i.e. RED measures) should be used and combined to gather more accurate results. As demonstrated, RED measures have provided better results than purchase intention.

PURCHASE INTENTION				
	CONCEPT TEST 1	CONCEPT TEST 2	CONCEPT TEST 3	CONCEPT TEST 4
Top-Box	10%	1%	2%	1%
Top 2 Box	10%	9%	10%	74%

Table 14 – Purchase intention top-box and top-two-box scores

CHAPTER 5 – CONCLUSIONS

5.1 THE RESULTS

It has been proposed a survey in which the sample had to evaluate a concept of a product that was not available in European countries. This product was the Amazon Dash Wand with Alexa. Before testing the five hypotheses, we have done some general analyses on the sample which aimed to assess if there were differences in the overall concept evaluation among countries, Amazon memberships and product categories usually bought on Amazon.

The countries of origin of respondents were: Italy (80%), UK (13%), Denmark (4%), Germany (2%) and Czech Republic (1%). Comparing their results, it has been noticed that only Italy has changed its overall concept evaluation from Test 1 to Test 4 : starting with a negative predisposition, Italians have ended with a positive feedback. This demonstrates that stimuli design does matters when respondents are presented to a product concept. Same considerations cannot be stated for Denmark, Germany and Czech Republic, which have overall maintained a neutral feedback during the four tests. However, it is worth highlighting that they only represent the 7% of the sample. UK is a bit different because has provided similar reactions as the Italian country for Test 1 and 2, but then in Test 3 and 4 their general predisposition remains neutral. Despite those considerations, the study has demonstrated that there are different outcomes depending on the country and for some countries concept formulation plays a pivotal role.

Then we have analysed what happened when confronting the answers from Amazon Prime Members (34%) and Regular Members. No big differences have been detected in the two kinds of membership, they provide almost identical results that start with a negative outcome in Test 1 and end with a positive outcome in Test 4. Moreover, it has been demonstrated that there is no interdependency between the general concept evaluation and the Amazon memberships. Hence, the different results based on stimuli design from Test 1 to Test 4 are not related with this variable.

Finally, there is an 8% of the respondents who had already bought groceries online. Analysing their responses compared to the other product categories bought on Amazon (e.g. electronic devices, accessories, etc.), we have noticed that they have yield the most negative answers in Test 1, enhancing their final evaluation until reaching a positive feedback in Test 4. The same improvement happened for the other product categories, however their outcomes are more positive than the ones provided by online-groceries consumers. This highlights that even in this case, stimuli design has influenced the general concept evaluation, while having a "prior experience" with buying groceries online has not given a more positive outcome compared to the ones who have never bought food online.

Finally, we have tried to predict the potential product market success through some particular measures of concept testing (i.e. RED measures: Relevance, Expensiveness and Differentiation). Those measure have been compared with the overall score of purchase intention, demonstrating that this indicator is not enough in predicting the product market success.

Actually, RED indicators have highlighted that the product might be a potential triumph in the market, especially when consumers were presented with Concept Test 4. Hence, stimuli design is relevant also for assessing the predictive market success.

5.1.1 HYPOTHESIS 1

Analysing $\mathbf{H_1}$ has shown that there is an improvement of the overall concept evaluation when considering its key dimensions (overall liking, importance, uniqueness, solve problem, believability and purchase intention), even though the first three concept tests have yield similar results. Means, modes, standard deviation and top-box scores have all portrayed that the final outcome enhances from Concept Test 1 to Concept Test 4. In the first three tests, findings show that there is a small increment for each key dimension. This means that concept test provided with just a statement yield similar results of concept test done with visuals. Moreover, it has been shown that when consumers are presented with a descriptive statement and an image combined with online reviews their overall evaluations tend to be negative: hence, Test 1 and 2 are very similar. However, in the third concept test the situation improves, and when it comes to Concept Test 4 the results are even better. This highlights that using a virtual prototype which presents a real-life situation can give more favourable and possibly accurate results. Therefore, $\mathbf{H_1}$ is confirmed because the way in which the concept is presented has led to different outcomes.

5.1.2 HYPOTHESIS 2

With $\mathbf{H_2}$ we have checked whether there is a correlation between product expertise and the overall concept evaluation, always using the six key measures provided by Dolan (1993). It has been demonstrated that there is a correlation especially for Concept Test 1 and 2, in which product expertise appears to be essential for evaluating a device that nobody has ever used or seen. However, as the concept test changes its design, the correlation decreases until it completely disappears in Concept Test 4. This means that product expertise is necessary for those concept formulations that insist only on the features and on descriptive designs, otherwise it might be not possible to understand what the product is offering. If consumers have product expertise, they would be more able to evaluate a product not shown for its benefits and usage. In the last two concept tests, this correlation decreases because an image with a promotional tone and a virtual prototype give all the missing information to consumers, hence there is no need for product expertise. Therefore, in this case $\mathbf{H_2}$ is confirmed for concept tests with a descriptive purpose because there is interdependence between those variables. Moreover, it is confirmed that stimuli design plays an important role even in this situation.

5.1.3 HYPOTHESIS 3

$\mathbf{H_3}$ wanted to do something similar to $\mathbf{H_2}$, but here the focus is on the role of innovators. During the research, innovators were the only ones that have given similar judgements among the four stimuli designs: their evaluations have never

propended to value extremely negative or extremely positive, they were quite neutral for each concept formulation. All the other groups of innovation's adopters had provided different results: for them, there has been an enhancement from Test 1 to Test 4, as already detected in the previous analyses. With all this considered, it has been demonstrated that innovators are more able to give objective evaluations throughout the four different concept formulation. On the other hand, the remaining groups have been influenced by stimuli designs.

There were respondents who were already familiar with the Dash Wand, even though they have never used or seen it. When asking to state with what similar products they have compared the Amazon device, 21% of them have answered "Amazon Echo" and "Dash Buttons". A further analysis has been conducted to detect whether there were different responses between this 21% of the sample and the innovation adopters' groups. Their responses were not aligned with the innovators because there was an increase of the judgement from Test 1 to Test 4 (as happened for the other groups). Hence, having a sort of a "prior experience" with the device has not been considered to be relevant for the final concept evaluation because everything is guided by the stimuli design. Moreover, having compared the Dash Wand with Amazon Echo and Dash Buttons have not produced more positive responses in the first two concept tests.

5.1.4 HYPOTHESIS 4

WOM has been tested through $\mathbf{H_4}$, in particular it has been presented to respondents with Concept Test 2. Results have shown that WOM has led to better results only when compared Concept Test 1 is compared with Concept Test 2. But when it comes to Concept Tests 2 and 3, the visual combined with online reviews have depicted worse results. Moreover, means and modes of Concept Test 3 for some indicators look similar in the first three concept tests. This means that consumers might not be interested in the reviews stated by other people because what influence their final judgements is the chance to actually see the product and how it works. Furthermore, it has been done a second analysis to see whether WOM might have influenced non-expert users. Those kinds of users have overall given similar evaluations to the other groups of experts, confirming that WOM does not influence their general predisposition towards the concept. Therefore, Word of Mouth has not been considered relevant compared to the other stimuli designs presented.

5.1.5 HYPOTHESIS 5

The final hypothesis $\mathbf{H_5}$ was focused on the change of the overall interest depending on the tone of the stimuli design. It has been proved that Concept Test 1 and 2 (factual tone) have provided very similar results, but when consumers were exposed to Concept Test 3 (persuasive) their evaluation has enhanced. Even though the first three tests have shown almost identical results, it can be noticed a small gap between Test 2 and 3, which is the exact part of the survey where the tone changes from stripped to embellished. Therefore, the change is not substantial but highlights the importance of the tone used. This importance is portrayed through Concept Test 4 in which there is a major improvement in the overall interest,

more consistent than Concept Test 3. Hence, changing from a factual to a persuasive tone has increased the interest leading to more positive evaluations.

5.1.6 ANSWERING THE RESEARCH QUESTION

"Do different stimuli design of a same concept provide different results? If yes, how?"

This was the research question originally developed for this research which have been tested through five different hypotheses. With all this considered, the answer for the research question is affirmative. Each of the hypotheses have shown that depending on the way in which the product concept is presented, consumers will give different outcomes. Those different outcomes depend also on the degree of product expertise and innovativeness of consumers. Moreover, an essential role is played by the tone of concept testing, which affects the overall evaluation in a positive (if embellished) or negative (if stripped) way. The only factor that does not influence in a consistent way the final outcome is WOM. Hence, stimuli design plays an important role in concept testing, and it should not be underestimated when firms are conducting pre-launch analyses. In Table 15 it is presented an overview of the five hypotheses, while Table 16 shows a comparison between the literature review findings and the results of this study.

H_1	*Different stimuli design of a same concept would lead to different concept evaluations*	**Confirmed**
H_2	*Product expertise might influence the overall concept evaluation*	**Confirmed**
H_3	*Innovators are more likely to provide more objective responses to new concepts*	**Confirmed**
H_4	*Consumers are influenced by the word of mouth when evaluating the concept*	**Not confirmed**
H_5	*A move from a factual to a persuasive tone would increase the interest*	**Confirmed**

Table 15 – Results of the study

5.1.7 A COMPARISON WITH THE LITERATURE REVIEW

In this section it will be done a comparison with what has been previously detected during the literature review and what it was actually found out when conducting this project. This comparison is useful to see whether there is a match or a mismatch between current findings and literature reviews' findings.

The findings taken into consideration and compared with the current conclusions were related to the topic of stimuli design, which is the main theme of this thesis. Therefore, the subjects analysed for this comparison concern the narrative transportation, the tone of stimuli design, its form and its setting. Table 16 shows a summary of what has been discussed in this paragraph comparing the findings coming from the literature review done in Chapter 2 and the findings highlighted during this work.

5.1.7.1 A COMPARISON: NARRATIVE TRANSPORTATION

Let's start with the narrative transportation. Van den Hende and Schoormans (2002) have studied what happens when conducting a concept test using a narrative form. Through a narrative format, consumers are able to see what the benefits and attributes of a product concept are thanks to the help of a storyline: what happens is that protagonists are employed to use the product as if they were creating a real-life context situation. When conducting this research, the narrative transportation has been portrayed through Concept test 4. In this part of the survey, the sample was presented with a video commercial showing how the Amazon Dash Wand with Alexa would have provided a high-value to the groceries' delivery habits. According to Van den Hende and Schoormans (2002), providing such setting in a narrative format would yield to more accurate results. Their findings have been confirmed in this study: it has been showed that when presented with Concept Test 4, the sample was more willing to evaluate the product concept in a positive way. This confirms that results are more accurate since respondents were able to see how the product actually worked and what benefits could provide, that's why they have given more positive results. Through the narrative format of a couple who was implementing the electronic device in their everyday life was a strong tool to see the real features of the product, without forcing the respondents to imagine how the device works (something that happened during the remaining Concept Tests).

5.1.7.2 A COMPARISON: TONE OF STIMULI DESIGN

Considering the tone of stimuli design, it has been highlighted that the product concept can be presented in two ways mainly: stripped or embellished plus a third way provided by Lees and Wright (2004). For more complete information please refer to par. 2.5.2. There are advantages and disadvantages related to the different tones used, but the current findings on literature do not agree on the same considerations. For instance, Haley and Gatty (1971) propose that depending on the tone through which the product concept is presented the evaluations would be different. On the other hand, Tauber (1972) states that the tone of the stimuli design is not relevant for gathering different results. Results of this project confirms what stated by Haley and Gatty (1971), namely that the tone of stimuli design might affect the final overall concept evaluation. Considering the four concept tests conducted, it has been noticed that in Concept Test 1 and Concept Test 2 results were quite similar. Those two tests represented the device in a stripped manner. When changing the tone from stripped to embellished thanks to Concept Test 3, results had improved, even though the alteration was not extremely radical compared to the previous tests. Finally, Concept Test 4 has demonstrated the most favorable results. During this test the device was presented through a persuasive tone (i.e. the video commercial) which has led the sample to evaluate the product in the most positive way compared to the other tests. This highlights that the tone of stimuli design has a predominant role when the sample has to evaluate the product concept, their responses had changed depending on the tone used. Therefore, what has been argued by Tauber (1972) has not encountered a match during this research. Actually, a product concept is evaluated depending on the tone of communication used, hence the communication part has a pivotal role in stimuli design.

Dolan (1993) shares Tauber's opinion arguing that a change from a factual to a persuasive tone would have not altered the purchase intention. This has not been confirmed as well since it has been encountered an improvement from Concept Test 2 to Concept Test 3, namely the exact moment when it has been done a switch of tone. The purchase intention indicator has been affected too: during the first two tests provided with a factual tone, the degree of purchase intention was included in the range "Probably not buy" and "Might or might not buy", while in the remaining tests with a persuasive tone the purchase intention has improved placing itself in the range "Probably buy" and "Definitely buy". Therefore, changing the tone would affect not only the overall interest in the product, but also the purchase intention. However, despite the general positive results of using an embellished tone, companies usually proceed with a factual description because of the related costs and time advantages (Peng and Finn, 2008).

5.1.7.3 A COMPARISON: FORM OF STIMULI DESIGN

As previously stated in par. 2.5.3, there are basically five ways to present a product concept when it comes to stimuli design: statement, visual, visual plus statement combined with online consumer reviews, virtual prototype and physical prototype. During this research, it was not possible to adopt the physical prototype since the device chosen is not available in European countries, therefore the remaining stimuli designs were used. Even in this case, there are some inconsistencies between current findings when it comes to judge the efficacy of each form of stimuli design. During the literature review, authors had different opinions on which was the most effective method between a statement and a visual. Holbrook and Moore (1981) and Vriens et al. (1998) state that the sample might have different reactions when evaluating a verbal concept statement and a concept made with visual. This has been partially not confirmed during this research: although results from Concept Test 1 (statement) and Concept test 3 (visual) were similar, they differed in terms of improvement. Actually, Concept Test 3 had better results than Concept Test 1 when looking at means and modes, however their final evaluations were not extremely different between each other. Here's why this topic has been partially not confirmed in this project.

Even though most of the studies agree on the fact that a visual concept test yields more favorable outcome than a written concept statement (Lees and Wright, 2004), there is still a minority of authors that does not come up with the same opinion. For instance, Rossiter and Percy (1980) argue that using an embellished statement might lead to similar results to those coming from the visual concept test. Considering the result of this research, those findings are partially confirmed. As previously mentioned, evaluations for Concept Test 1 and Concept Test 3 were similar in terms of numerical value but different in terms of improvement (there was an upgrade from Test 1 to Test 3 in terms of positive evaluation towards the product). Therefore, despite the similar results in the first three tests, it has been encountered an improvement from Test 1 to Test 3, which has ended up in a positive outcome in Concept Test 4 (provided with an embellished tone as well).

5.1.7.4 A COMPARISON: WORD OF MOUTH

As already stated in par. 3.4, this research wanted to assess the power of word of mouth as well. For doing so, it has been decided to show to the sample real online customers' reviews of people who have already used the Amazon Dash Wand with Alexa. Those reviews were taken from Amazon.com. This was helpful to see whether consumers were influenced by other consumers in the choice of possibly buying the Amazon Dash Wand with Alexa.

During the literature review, it was found out that online reviews should be used for later adopters because they need to receive positive information from non-marketing sources (Klink and Athaide, 2006). This statement has been questioned in the current study. The power of WOM was highlighted with the help of Concept Test 2, which was a statement and an image combined with real online consumers' reviews. It has been showed that the overall product concept evaluation was negative: actually, Concept Test 2 had similar negative results of Concept Test 1 (the stripped statement). This means that probably consumers are not interested in what other end-users thought of that specific product. Maybe they do not consider they reviews helpful because they might want to trust their instinct or their first experience with the device.

Furthermore, it was analyzed the impact of word of mouth on groups of extremely experts, experts, moderately experts and non-expert consumers. The same considerations were highlighted: Concept Test 2 has still provided negative results (even though they were slightly better than the ones depicted in Concept Test 1). It might be think that non-expert consumers are more likely to be influenced by others' opinions since they have never used a similar device. Actually, responses coming from non-expert users were aligned with the ones provided by all the other groups of experts. The same has happened when it was analyzed the answers given by later adopters: as seen in par. 4.2.3, this part of the sample has provided a negative feedback in Concept Test 2 towards the device compared to the remaining concept tests. Therefore, as depicted in this project, the hypothesis suggested by Klink and Athaide (2006) were not confirmed since WOM has been proved to be not a relevant item to influence the final overall concept evaluation.

5.1.7.5 A COMPARISON: THE SETTING

When conducting a concept test, the product concept can be shown in two different kinds of setting. The sample can evaluate a product through a traditional environment or through a virtual one. In a traditional environment, respondents are presented with a statement, an image or, if possible, a physical prototype. On the other hand, in a virtual setting the sample is not able to actually perceive the product concept as it occurs with a virtual prototype: what happens here is that the product is simulated through 3D pictures, a video, virtual reality. Basically, the sample is presented with a virtual prototype. In each of those cases consumers are not able to feel the product, it can only be simulated a real-life context situation. According to a study made by Peng and Finn (2008), companies suggest switching to a virtual environment because of the great amount of advantages included. As mentioned in par. 2.5.4, a virtual environment is cheaper, access

to respondents does not encounter logistic issues, the developing technologies have let the virtual trial experience become more realistic and multiple concept can be tested all at once.

Despite those benefits, there is still someone who does not believe in the potentialities of the virtual environment. Haley and Gatty (1970), for instance, suggest avoiding the use of virtual settings because they do not provide a realistic situation, while other academics such as Cooper (1993) recommend using a virtual environment because if consumers are presented with a representation of the product as real as possible, they would be more able to provide accurate results. On the other hand, Peng et al. (2011) argue that a traditional and a virtual setting provide identical results.

With all this considered, let's now see what has happened in the current project. When consumers were presented with the statement and the images, their feedback was quite similar: they have showed an increasing interest in the product, even though it was small, but they have overall evaluated the device in a negative way. However, when the sample was presented with the virtual environment (Concept Test 4), the feedback towards the device was more positive. Actually, they have increased the interest and the purchase intention by at least 50% (see chap. 4). Therefore, first it has been detected that a traditional and a virtual environment give different results, hence the statement proposed by Peng et al. (2011) it is not aligned with what has been found in this research. Second, as Cooper and other authors have suggested, the virtual environment is more likely to provide more positive results compared to a traditional setting.

LITERATURE REVIEW FINDINGS	CURRENT STUDY FINDINGS
Testing done in a narrative format would yield more accurate evaluations (Van den Hende and Schoormans, 2002)	***Confirmed*** It has been demonstrated when comparing Concept Test 4 results with the other tests results
Concept presented with different tones gave different results (Haley and Gatty, 1971)	***Confirmed*** It has been proved through H_5
A concept is evaluated regardless the manner of communication (Tauber, 1972)	***Not confirmed*** As shown in H_5, respondents gave similar results with factual tone, but when it comes to persuasive tone the general concept evaluation encounters an improvement
A change from a factual to a persuasive tone would have not altered the purchase intention (Dolan, 1993)	***Not confirmed*** It has been proved during the entire survey, the purchase intention indicator improves when it comes to tests with a persuasive tone
End-users might have different reactions when evaluating a verbal concept statement and a concept made with visual (Holbrook and Moore, 1981; Vriens et al., 1998)	***Partially not confirmed*** According to the study, the overall evaluation changed from less favourable to more favourable from Concept Test 1 to Concept Test 3. However, means and modes have shown that results were improving but they were still similar
A visual concept test yields a more positive outcome than a written concept statement (Lees and Wright, 2004)	***Partially confirmed*** Despite the results from the first three concept tests were similar, there was a gradual improvement from test 1 to 3. However, results were not extremely positive compared to test 1, but it was possible to see an increment towards positive evaluations
Online reviews should be used for later adopters because they need to receive positive information from non-marketing sources (Klink and Athaide, 2006)	***Not confirmed*** Results from Table 4 show no positive evaluations for later adopters when they were presented with online reviews
A virtual and a traditional testing environment yield identical results (Peng et al. 2011)	***Not confirmed*** There were consistent differences when comparing Concept Test 4 with the other tests
A virtual testing environment provides positive results compared to the traditional setting (Dahan and Srinivasan, 2000; Von Hippel and Katz, 2002; Füller et al., 2006)	***Confirmed*** Concept Test 4 has obtained the most positive result in each hypotheses test

Table 16 – A comparison of the literature review and the current findings

5.2 MANAGERIAL IMPLICATIONS

Considering the results of this research, some managerial implications can be done.

First of all, it has been demonstrated that four different concept formulations do not yield the same results. Of course, it is not recommendable for companies to do exactly the same with just one product because it is expensive and time consuming. However, it is important for organizations to choose the right stimuli design when presenting a concept since, as proved, its formulation has a role. As far as an electronic device is concerned, it has been portrayed that the most favourable results came from showing the virtual prototype and when a persuasive tone is used. Actually, Concept Test 4 is the one who has performed better in each hypothesis tested. This highlights that a simple statement or an image is not enough because respondents need to be able to get a complete perception and comprehension of features and benefits of an electronical device.

Second, target selection assumes a great importance. Innovators are more likely to provide objective responses, while consumers with product expertise are relevant only when the product benefits and features are not completely shown or

clear. Therefore, if it is not possible to present a prototype because maybe the product is not ready yet, the sample choice becomes essential for evaluating a written statement and/or an image. Choosing the right target helps in having more accurate and relevant data, especially when the concept formulation is very simple.

Third, it is not necessary to rely on word of mouth. It has been showed that the opinions of end-users who have already tried the product might not been taken into consideration when respondents are evaluating the concept. Comparing the evaluation of Concept Test 1 and Concept Test 2, it can be seen an improvement, however it is not enough for stating that stimuli designs that include word of mouth provide more favourable results compared to other type of concept formulations. Evidently, it is not necessary to invoke other consumers' help because respondents are interested in their personal feelings perceived when seeing the product. Additionally, they might be able to evaluate the concept autonomously thanks to their background, without considering others' opinions. Therefore, it is not recommendable to use customer reviews when designing how to present the concept.

Fourth, the tone does matter. The research has shown that using a persuasive tone would provide more positive results compared to the outcomes obtained from tests with a factual tone. Hence, it is recommendable for companies to use an embellished tone similar to a fully completed advertisement. In this way, consumers are able to better perceive benefits and features and to understand whether to adopt or reject the concept shown.

Fifth, show real-life context situations. This has been done through Concept Test 4, which has gathered the most positive results. In the video shown, a couple demonstrated how the device was used in an ordinary day. Being presented to a real-life environment, respondents are more able to yield accurate information and see how the product actually works. This confirms what stated in the literature review when talking about narrative transportation, namely that a concept testing done in this format provides more accurate evaluations. Authors have made those considerations when presenting a statement, but this study has demonstrated that this is applicable also for virtual environments.

Finally, the most obvious consideration is to choose the right information to show. If the target is not a tech-savvy, it is recommendable to present features but also benefits in a way understandable for them, otherwise the respondents will not be able to infer whether the product would be useful for them or not. Moreover, it is essential to be clear and to present sincere information.

Therefore, when trying to commercialize a product, a company needs to give a lot of attention on the pre-launch analyses. Here, a pivotal role is assumed by concept testing, whose design/formulation is essential for gather consumers' feedback as closest to reality as possible. When the product is an electronical device, there are some suggestions that can be taken into consideration: using a virtual environment, be careful on the target selection, avoid gathering the help of word of mouth, use a persuasive tone, show real-life situations of product usage and be as clear as possible.

5.2.1 LAUNCHING THE DASH WAND IN EUROPE

Even though concept statement is the cheapest way to present a concept, it is not enough for gathering an objective response on the concept since features, benefits and performances are not clear: consumers might reject it only because they don't understand how the device is improving their lives. A static image could not be enough too, actually it gave the same negative outcomes of the statement. Hence, when possible it is recommendable to use a virtual prototype.

Considering the previous analyses, some implications can be argued when deciding to launch this Amazon device in Europe. First, Italy and Czech Republic are the only countries that have expressed a positive interest in the Amazon Dash Wand, hence the company might think to commercialize the product here before spreading in other countries. The product has been appreciated but is necessary to combine it with a clear and involving advertising strategy (such as translating the actual commercial, which was the most appreciated stimuli design). Second, for UK, Germany and Denmark the product might need something more. Their general neutral evaluations might be related to the fact that the product is not interesting enough or it has nothing special compared to what is presented in the market. In their cases, stimuli design was relevant only for UK and only for the first two concept tests: hence, for those countries it is not important how it is presented but what the features and benefits are. Finally, considering that the majority of respondents belongs to the group 18-24, the company might think to try to invest in this target (overall, they have provided a positive interest when presented with a virtual prototype). Those implications regarding the launch have been done based on this study, but of course the sample is too small to generalize.

5.3 LIMITATIONS

One of the main limitations of this research is that it was not possible to use a real Amazon Dash Wand with Alexa: it would have been interesting to show people the real product and let them use it. Using a physical prototype could yield to better results because consumers are put in front of the real product, they can feel it and they can see how it works just right in their hands. However, using a physical prototype is time-consuming and expensive. In this particular case, logistic issues have not allowed the respondents to use a physical device. Furthermore, since the electronical device is not available in European countries, it would have not been possible to use it and/or to integrate it with smart home devices. Finally, using a physical prototype would have been against the original purpose of the research, which was to administer an online survey.

Using mostly closed-ended questions might be another limitation because respondents were not free to state their thoughts about the product because they had to choose from a set of predetermined answers. However, this has been done in order to collect more accurate and relevant data.

Another limitation to this study is that it has not been included with qualitative research to the online survey. Focus groups or face-to-face interviews would have helped to see the non-verbal communication and to deepen the positive or negative overall concept evaluation with in-depth analyses.

A final possible limit is that since the project is focused on an electronic device, results gathered cannot be generalized for any kind of new product that do not belong to tech industry.

5.4 FUTURE RESEARCH

This study has not taken into consideration to use a physical prototype due to lack of resources. It would be interesting to do a similar study that deepen the role of stimuli design taking into consideration a physical prototype. Letting consumers be able to actually try the product, might lead to different evaluations, maybe even better than the one gathered from Concept test 4. Hence, future research should try to test a physical product especially to see the differences with the virtual environment.

Another suggestion for future research is to try to better develop the sample. This project has been done considering the majority of responses coming from Italian students, therefore it is not possible to generalize. It might be interesting to expand the area to see whether different cultural backgrounds might influence the final overall concept evaluation. An additionally consideration on the sample is to try to gather more responses that may provide more reliable findings and more similar to the actual market.

Finally, since this project is focused on an electronic device, findings cannot be generalized for each industry. Therefore, it might be interesting for a future research to try to develop a sort of a blueprint that every company can use for each kind of product regardless to which industry they belong.

REFERENCES

- Ahad, P. (2015). "The Evolution of Concept Testing". Survey Magazine. http://www.surveymagazine.org/evolution-concept-testing/
- Amazon Official Website. https://www.amazon.com/
- Amazon YouTube Channel (2017). *Amazon Dash Wand with Alexa – Official Commercial*. https://www.youtube.com/watch?v=s7IExS483wE&t=2s
- Ayag, Z. (2016). "An integrated approach to concept evaluation in a new product development". *Journal of Intelligent Manufacturing*, 27(5), pp. 991-1005
- Betters, E. (2017). "Amazon Dash Wand: what you can and can't do with Alexa device". Pocket-Lint. https://www.pocket-lint.com/smart-home/reviews/amazon/141501-amazon-dash-wand-what-you-can-and-can-t-do-with-this-alexa-device
- Brady, K. (1986). "Concept testing should measure more than the intent to purchase". *Marketing News*, pp. 63-64
- Consumer Reports (2018). "Pros and Cons of Amazon Prime". https://www.consumerreports.org/online-shopping/pros-cons-amazon-prime/
- Cooper R. G. (2011), "Winning at new products: creating value through innovation". Philadelphia: Basic Books
- Cox, D. and A. (2002). "Beyond first impressions: the effects of repeated exposure on consumer liking of visually complex and simple product designs". *Journal of the Academic Marketing Science*, 30(2), pp. 119-130
- Crawford, M. and Di Benedetto, A. (2010). "New Products Management". New York: McGraw-Hill.
- Cui, G., Peng, L. and Florès, L.P. (2015). "Selecting ideas for new product development: comparison of monadic test and adaptive concept screening under the G theory framework". *European Journal of Innovation Management*, 18(3), pp. 380-396
- Cui, G., Lui, H. and Guo, X. (2012). "The Effect of Online Consumer Reviews on New Product Sales". *International Journal of Electronic Commerce*, 17(1), pp. 39-57
- Dahan, E. and Mendelson, H. (2001). "An Extreme-Value Model of Concept Testing". *Management Science*, 47 (1), pp. 102-116
- Dahan, E. and Srinivasan, V. (2000). "The Predictive Power of Internet-Based Product Concept Testing Using Visual Depiction and Animation". *Journal of Product Innovation Management*, 17(2), pp. 99-109
- De Bont, C.J.P.M. and Schoormans, J.P.L. (1995). "The effects of product expertise on consumer evaluation of new-product concepts". *Journal of Economic Psychology*, 16 (4), pp. 599-615

- Dellarocas, C., Zhang, X.M. and Awad, N.F. (2007). "Exploring the value of online product reviews in forecasting sales: the case of motion pictures". *Journal of Interactive Marketing*, 21(4), p. 23-45
- Dubas, K.M., Dubas, S.M adn Atwong, C. (1999). "Some difficulties in predicting new product trial using concept scores". *Journal of Product and Brand Management*, 8(1), pp. 50-60
- Duke, C.R. (1994). "Understanding Costumer Abilities in Product Concept Tests". *Journal of Product and Brand Management*, 3(1), pp. 48-57
- Friedman, M. and Schillewaert, N. (2012). "Order and Quality Effects in Sequential Monadic Concept Testing: Methodological Details Matter in Concept-Testing Practice". *Journal of Marketing Theory and Practice*, 20(4), pp. 377-390
- Goffin, K. and Mitchell, R. (2017). "Innovation Management. Effective Strategy and Implementation". London: Palgrave.
- Haley, R.I. and Gatty, R. (1971). "The trouble with concept testing". *Journal of Marketing Research*, 8, pp. 230-232
- Hende, E.A, Dahl, W., Schoormans, J.P.L. and Snelders, D. (2012). "Narrative Transportation in Concept Tests for Really New Products: The Moderating Effect of Reader-Protagonist Similarity". *Journal of Product Innovation Management*, 29, pp. 157-170
- Henderson, T. (2017). "Why Innovation is Crucial to your Organization's Long-Term Success". Forbes. https://www.forbes.com/sites/forbescoachescouncil/2017/05/08/why-innovation-is-crucial-to-your-organizations-long-term-success/#656c9c803098
- IPSOS Consulting Agency. https://www.ipsos.com/
- Jagpal, S., Jedidi, K. and Jamil, M. (2007). "A Multibrand Concept-Testing Methodology for New Product Strategy". *Journal of Product Innovation Management*, 24(1), pp. 34-51
- Klink, R. R. and Athaide G. A. (2006). "An illustration of potential sources of Concept-Test Error". *Journal of Product Innovation Management* 23, pp. 359-370
- Kotler, P., Keller, K.L., Ancarani, F. and Costabile, M. (2015). "Marketing Management". Milano: Pearson Italia.
- Lees, G. and Wright, M. (2004). "The Effect of Concept Formulation on Concept Test Scores". *Journal of Product Innovation Management*, 21 (6), pp. 389-400
- Levy, S. (2018). "Inside Amazon's Artificial Intelligence Flywheel". Wired. https://www.wired.com/story/amazon-artificial-intelligence-flywheel/

- Li, C., Peng, L. and Cui, G. (2017). "Picking Winners: New Product Concept Testing with Item Response Theory". *International Journal of Market Research*, 59(3), pp. 335-353
- Luchs, M.G., Brower, J. and Chitturi, R. (2012). "Product Choice and the Importance of Aesthetic Design Given the Emotion-laden Trade-off between Sustainability and Functional Performance". *Journal of Product Innovation Management*, 29(6), pp. 903-916
- Mitchell, A.A. and Dacin, P.A. (1996). "The Assessment of Alternative Measures of Consumer Expertise". *Journal of Consumer Research*, 23(3), pp. 219-239
- Moore, W. L. (1982). "Concept Testing". *Journal of Business Research,* 10 (3), pp. 279-294
- Ozer, M. (2010). "The Moderating Role of Prior Experience and Behavioural Importance in the Predictive Validity of New Product Concept Testing". *Journal of Product Innovation Management*, 28, 109-122
- Page, A.L. and Rosenbaum, H.F. (1992). "Developing an Effective Concept Testing Program for Consumer Durables". *Journal of Product Innovation Management*, 9(4), pp. 267-277
- Patel, N. (2017). "Amazon's new Dash Wand is half magic, half boring". The Verge. https://www.theverge.com/circuitbreaker/2017/6/20/15840814/amazon-dash-wand-alexa-review
- Peng, L., Cui, G. and Li, C. (2012). "Individual differences in consumer responses to traditional versus virtual concept testing". *Journal of Product and Brand Management*, 21(3), pp. 167-175
- Peng L. and Finn A. (2008). "Concept Testing: the state of Contemporary Practice". *Marketing Intelligence & Planning*, 26 (6), pp. 649 – 674
- Peng, L. and Finn, A. (2010). "How Cloudy a Crystal Ball: A Psychometric Assessment of Concept Testing". *Journal of Product Innovation Management*, 27(2), pp. 238-252
- Peng, L. and Finn, A. (2010). "Whose Crystal Ball to Choose? Individual Difference in the Generalizability of Concept Testing". *Journal of Product Innovation Management*, 27(5), pp. 690-704
- Peng, L. and Finn, A. (2010). "How far can you rely on a concept test?" The generalisability of testing over occasions". *International journal of Market Research*, 52(3), pp. 363-382
- Peng, L., He, Y. and Li, C. (2011). "Consumer Responses to Traditional vs Virtual Concept Testing: The Influence of Innovativeness, Change Seeking and Cognitive Effort". *Asian Pacific Advances in Consumer Research*, 9, pp. 251-252
- Peng, L., He, Y. and Wan, X. (2010). "Devil or Angel? How the virtual testing environment can affect product evaluations". *International Journal of Market Research*, 53(3), pp. 415-436

- Popular Science (2017). “Amazon Dash Wand review: Alexa, help me spend even more money on Amazon”. https://www.popsci.com/amazon-dash-wand-review-alexa-help-me-spend-even-more-money-on-amazon
- Qualtrics. “How to Create a Product Concept Test”. https://www.qualtrics.com/experience-management/product/how-concept-test/
- Reisinger, D. (2017). “Amazon Dash Wand review roundup: what critics love and hate”. Tom’s Guide. https://www.tomsguide.com/us/amazon-dash-review-roundup,news-25397.html
- Schwartz, D.A. (1984). “Concept testing can be improved and here’s how to do it”. *Marketing News*, pp. 22-25
- Shoard, C. (2015). “Back to the Future Day: what Part II got right and wrong about 2015, an A-Z”. The Guardian. https://www.theguardian.com/film/filmblog/2015/jan/02/what-back-to-the-future-part-ii-got-right-and-wrong-about-2015-an-a-z
- Siemens, J.C., Smith, S., Fisher, D. and Jensen, T.D. (2008). “Product expertise versus professional expertise: congruency between an endorser’s chosen profession and the endorsed product”. *Journal of Targeting, Measurement and Analysis for Marketing*, 16(3), pp. 159-168
- Silbert, S. (2018). “All About the Amazon Dash Wand with Alexa”. LifeWire. https://www.lifewire.com/amazon-dash-wand-4143355
- Simon, M. (2017). “Amazon Dash Wand review: a home shopping device made for a not-too-distant future”. TechHive. https://www.techhive.com/article/3202772/connected-home/amazon-dash-wand-review.html
- Tauber, E.M. (1972). “What is measured by concept testing?”. *Journal of Advertising Research*, 12, pp. 35-37
- Tauber, E.M. (1975). “Why Concept and Product Test Fail to Predict New Product Results”. *Journal of Marketing*, 39(4), pp. 69-71
- The Health Compass (2012). “How to Test Creative Concepts”. https://www.thehealthcompass.org/how-to-guides/how-test-creative-concepts
- Tredgold, G. (2018). “Four Reasons Why You Need to Focus on Innovation. Innovation is a differentiator”. Inc. https://www.inc.com/gordon-tredgold/4-reasons-why-you-need-to-focus-on-innovation.html
- Trebbi, G. and Flesch, E. (1983). “Single versus multiple concept tests”. *Journal of Advertising Research*, 23(3), pp. 21-26
- Vicari, S., Cillo, P. and Raccagni, D. (2013). “Product Innovation. Dall’idea al lancio del nuovo prodotto”. Milano: Egea.
- White, C. (2017). “Review of the Amazon Dash Wand powered by Alexa”. Neowin. https://www.neowin.net/news/review-of-the-amazon-dash-wand-powered-by-alexa

- Yan, H. and Ma, T. (2014). “A fuzzy group decision making approach to new product concept screening at the fuzzy front end”. *International Journal of Production Research*, pp. 1-29
- Ying, C., Li, Y., Chin, K., Yang, H. and Xu, J. (2018). “A new product development concept selection approach based on cumulative prospect theory and hybrid-information MADM”. *Computer & Industrial Engineering*, 122, pp. 251-261
- Zemeckis, R. (1989). “Back to the Future: Part II”. Universal Pictures.

APPENDIX

Appendix 1. Amazon Dash Wand with Alexa Concept Test Questionnaire.

Section 1: Online Shopping & Amazon

1.1 Have you ever purchased something on Amazon?

- ☐ Yes
- ☐ No

1.2 What kind of products do you usually purchase on Amazon? (Please select one or more options) [Show only if "yes" is selected in question 1.1]

- ☐ Books and eBooks
- ☐ Clothing, shoes and jewellery
- ☐ Electronic devices (computer, TV, smartphones, e-book readers, etc.)
- ☐ Accessories for electronic devices (cases, earphones, keyboards, etc.)
- ☐ Video games
- ☐ Grocery and gourmet food
- ☐ Musical instruments
- ☐ Beauty and personal care
- ☐ Furniture

1.3 How often do you buy products on Amazon? [Show only if "yes" is selected in question 1.1]

- ☐ Less than once per month
- ☐ Once per month
- ☐ 2-3 times per month
- ☐ More than 3 times per month
- ☐ Daily

1.4 To what extent do you agree with the following sentences:

	Strongly disagree	Rather disagree	Neither agree or disagree	Rather agree	Strongly agree
I would hesitate in giving my personal information (e.g. credit card, street address, etc.) on Amazon	☐	☐	☐	☐	☐
I'm not interested in buying products online	☐	☐	☐	☐	☐
I prefer buying items into a physical shop rather than through an online website	☐	☐	☐	☐	☐
I have had a bad experience when buying products online	☐	☐	☐	☐	☐
I'm not very informed about online shopping conditions and benefits	☐	☐	☐	☐	☐
I believe Amazon has sufficient expertise to do business on the Internet	☐	☐	☐	☐	☐
I like buying products on Amazon	☐	☐	☐	☐	☐
I find Amazon trustworthy	☐	☐	☐	☐	☐
Overall, I am satisfied with Amazon	☐	☐	☐	☐	☐
The shopping experience on the Amazon's website is excellent	☐	☐	☐	☐	☐
I am likely to switch to another online shopping site in the near future	☐	☐	☐	☐	☐
I often recommend Amazon to others	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐

1.5 Are you an Amazon Prime member? [Show only if "yes" is selected in question 1.1]

- ☐ Yes
- ☐ No

1.6 For which purpose do you use the Amazon Prime membership? (Please select one or more options) [Show only if "yes" is selected in question 1.5]

- ☐ Prime Day
- ☐ Faster shopping delivery
- ☐ Unlimited video streaming
- ☐ Access to Amazon compatible devices (e.g. Dash Buttons)
- ☐ Fresh grocery delivery
- ☐ Unlimited e-Books access
- ☐ Access to Amazon music streaming service
- ☐ Exclusive discounts

1.7 Which of the following electronical device have you already used? (Please select one or more options)

- ☐ Smart TV
- ☐ e-Book readers
- ☐ Smart Watch
- ☐ Smartphone
- ☐ Computer
- ☐ Action camera
- ☐ Tablet
- ☐ Smart home devices
- ☐ None of them
- ☐ Other _____

1.8 How do you rate your knowledge of electronical devices compared to the rest of the population?

- ☐ Far below average
- ☐ Somewhat below average
- ☐ Average
- ☐ Somewhat above average
- ☐ Far above average

1.9 To what extent do you agree with the following sentences:

	Strongly disagree	Rather disagree	Neither agree or disagree	Rather agree	Strongly agree
I like to own a new product that distinguishes me from others who do not own this new product	☐	☐	☐	☐	☐
It gives me a good feeling to acquire new products	☐	☐	☐	☐	☐
If a new product gives me more comfort than my current product, I would not hesitate in buying it	☐	☐	☐	☐	☐
Overall, I often buy the latest electronical devices	☐	☐	☐	☐	☐
When I see a new electronical device in a store, I often buy it because it is new	☐	☐	☐	☐	☐
If a new product makes my work/life easier, than this new product is a “must” for me	☐	☐	☐	☐	☐

1.10 Which of the following voice assistants do you know?

- ☐ Siri
- ☐ Google Assistant
- ☐ Cortana
- ☐ Alexa
- ☐ Bixby
- ☐ Other _____

1.11 Have you ever used a voice assistant?

- ☐ Yes
- ☐ No

1.12 In which of the following occasions have you ever used a voice assistant? (Please select one or more options) [show only if "yes" is selected in question 1.11]

- ☐ To order food
- ☐ To book reservations
- ☐ To create to do list/shopping list
- ☐ To play some music
- ☐ To send messages
- ☐ To pay/buy something online
- ☐ To schedule an appointment
- ☐ To get information
- ☐ To interact with my smart home
- ☐ Other _____

Section 2: Food Habits

2.1 Do you enjoy cooking?

- ☐ Yes
- ☐ No

2.2 How often do you go out to eat?

- ☐ Once a week
- ☐ 2-3 times a week
- ☐ More than three days a week
- ☐ Everyday

2.3 How often do you go in a grocery store per week?

- ☐ Once
- ☐ Twice
- ☐ More than twice
- ☐ Everyday

2.4 How often do you buy groceries online?

- ☐ Never
- ☐ Once a week
- ☐ More than once a week
- ☐ Everyday

Section 3: Concept test n.1 – Statement

The Amazon Dash Wand is a tool for shopping groceries and home supplies on Amazon.com. It's equipped with Wi-Fi and a barcode-scanner. To use it, you simply point it at an item's barcode, and if the wand recognizes the item, it will make a noise, flash off the light and add the item to your Amazon cart. Now, it is provided with Alexa (Amazon's voice assistant) that can answer to your general questions, convert cups to ounces and help you look up recipes. Moreover, it can be integrated with your smart home. It is a small handheld device, water-resistant and magnetic, so it can be stuck to your fridge.

3.1 To what extent do you agree with the following sentences:

	Strongly disagree	Rather disagree	Neither agree or disagree	Rather agree	Strongly agree
I am more willing to buy groceries online	☐	☐	☐	☐	☐
I am looking forward to use the Dash Wand with Alexa	☐	☐	☐	☐	☐
I am still not interested	☐	☐	☐	☐	☐
It looks too complicated	☐	☐	☐	☐	☐
I will learn how to cook thanks to Alexa	☐	☐	☐	☐	☐
I will use it with my smart home	☐	☐	☐	☐	☐
I am afraid I will not control how much I will spend on groceries	☐	☐	☐	☐	☐
It is not the same as going to a physical grocery store	☐	☐	☐	☐	☐
I will be more willing to cook for other people	☐	☐	☐	☐	☐
To the extent possible, I plan to use this device in my future	☐	☐	☐	☐	☐

3.2 How often do you think you would use the Amazon Dash Wand with Alexa?

- ☐ Never
- ☐ Rarely (once per month)
- ☐ Sometimes (2-3 times per month)
- ☐ Often (once per week)
- ☐ Frequently (more than once per week)
- ☐ Always (more than once per day)

3.3 How much would you like to have this product?

- ☐ Dislike a great deal
- ☐ Dislike somewhat
- ☐ Neither like or dislike
- ☐ Like somewhat
- ☐ Like a great deal

3.4 How important do you think the functions and features of this product are?

- ☐ Not at all important
- ☐ Slightly important
- ☐ Moderately important
- ☐ Very important
- ☐ Extremely important

3.5 How would you rate this product in terms of being unique from the products currently sold?

- ☐ Not at all unique
- ☐ Slightly unique
- ☐ Moderately unique
- ☐ Very unique
- ☐ Extremely unique

3.6 How sure are you that this product would solve a problem for you?

- ☐ Absolutely not sure it will
- ☐ Slightly sure it will
- ☐ Moderately sure it will
- ☐ Somewhat sure it will
- ☐ Absolutely sure it will

3.7 How believable to you were the things said about this product?

- ☐ Extremely unbelievable
- ☐ Somewhat unbelievable
- ☐ Neither believable or unbelievable
- ☐ Somewhat believable
- ☐ Extremely believable

3.8 Assuming this product was available in a store where you shop, how likely would you be to buy it?

- ☐ Definitely not buy
- ☐ Probably not buy
- ☐ Might or might not buy
- ☐ Probably buy
- ☐ Definitely buy

3.9 Considering the price of the Amazon Dash Wand with Alexa (20$), how do you think is its price compared to similar products you could buy?

- ☐ Very expensive
- ☐ Somewhat expensive
- ☐ About average
- ☐ Somewhat inexpensive
- ☐ Very inexpensive

3.10 How do you think is the Amazon Dash Wand with Alexa different compared to similar products you could buy?

- ☐ Absolutely not different
- ☐ Somewhat not different
- ☐ About average
- ☐ Somewhat different

☐ Absolutely different

3.11 How effective do you think the Amazon Dash Wand with Alexa is in fulfilling the following needs?

	Not effective at all	Slightly effective	Moderately effective	Very effective	Extremely effective
Accessing food recipes	☐	☐	☐	☐	☐
Controlling smart home devices	☐	☐	☐	☐	☐
Having a virtual assistant	☐	☐	☐	☐	☐
Buying groceries online	☐	☐	☐	☐	☐
Creating a shopping list	☐	☐	☐	☐	☐
Saving time	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐

3.12 With what similar products have you compared the Amazon Dash Wand with Alexa?

_____

Section 4: Concept test n.2 – Visual and online reviews

4.1 To what extent do you agree with the following sentences:

	Strongly disagree	Rather disagree	Neither agree or disagree	Rather agree	Strongly agree
I am more willing to buy groceries online	☐	☐	☐	☐	☐
I am looking forward to use the Dash Wand with Alexa	☐	☐	☐	☐	☐
I am still not interested	☐	☐	☐	☐	☐
It looks too complicated	☐	☐	☐	☐	☐
I will learn how to cook thanks to Alexa	☐	☐	☐	☐	☐
I will use it with my smart home	☐	☐	☐	☐	☐
I am afraid I will not control how much I will spend on groceries	☐	☐	☐	☐	☐
It is not the same as going to a physical grocery store	☐	☐	☐	☐	☐
I will be more willing to cook for other people	☐	☐	☐	☐	☐
To the extent possible, I plan to use this device in my future	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐

4.2 How often do you think you would use the Amazon Dash Wand with Alexa?

- ☐ Never
- ☐ Rarely (once per month)
- ☐ Sometimes (2-3 times per month)
- ☐ Often (once per week)
- ☐ Frequently (more than once per week)
- ☐ Always (more than once per day)

4.3 How much would you like to have this product?

- ☐ Dislike a great deal
- ☐ Dislike somewhat
- ☐ Neither like or dislike
- ☐ Like somewhat
- ☐ Like a great deal

4.4 How important do you think the functions and features of this product are?

- ☐ Not at all important
- ☐ Slightly important
- ☐ Moderately important
- ☐ Very important
- ☐ Extremely important

4.5 How would you rate this product in terms of being unique from the products currently sold?

- ☐ Not at all unique
- ☐ Slightly unique
- ☐ Moderately unique
- ☐ Very unique
- ☐ Extremely unique

4.6 How sure are you that this product would solve a problem for you?

- ☐ Absolutely not sure it will
- ☐ Slightly sure it will
- ☐ Moderately sure it will
- ☐ Somewhat sure it will
- ☐ Absolutely sure it will

4.7 How believable to you were the things said about this product?

- ☐ Extremely unbelievable
- ☐ Somewhat unbelievable
- ☐ Neither believable or unbelievable
- ☐ Somewhat believable
- ☐ Extremely believable

4.8 Assuming this product was available in a store where you shop, how likely would you be to buy it?

- ☐ Definitely not buy
- ☐ Probably not buy
- ☐ Might or might not buy
- ☐ Probably buy

☐ Definitely buy

4.9 Considering the price of the Amazon Dash Wand with Alexa (20$), how do you think is its price compared to similar products you could buy?

☐ Very expensive

☐ Somewhat expensive

☐ About average

☐ Somewhat inexpensive

☐ Very inexpensive

4.10 How do you think is the Amazon Dash Wand with Alexa different compared to similar products you could buy?

☐ Absolutely not different

☐ Somewhat not different

☐ About average

☐ Somewhat different

☐ Absolutely different

4.11 How effective do you think the Amazon Dash Wand with Alexa is in fulfilling the following needs?

	Not effective at all	Slightly effective	Moderately effective	Very effective	Extremely effective
Accessing food recipes	☐	☐	☐	☐	☐
Controlling smart home devices	☐	☐	☐	☐	☐
Having a virtual assistant	☐	☐	☐	☐	☐
Buying groceries online	☐	☐	☐	☐	☐
Creating a shopping list	☐	☐	☐	☐	☐
Saving time	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐

4.12 With what similar products have you compared the Amazon Dash Wand with Alexa?

_____

Section 5: Concept test n.3 – Visual

5.1 To what extent do you agree with the following sentences:

	Strongly disagree	Rather disagree	Neither agree or disagree	Rather agree	Strongly agree
I am more willing to buy groceries online	☐	☐	☐	☐	☐
I am looking forward to use the Dash Wand with Alexa	☐	☐	☐	☐	☐
I am still not interested	☐	☐	☐	☐	☐
It looks too complicated	☐	☐	☐	☐	☐
I will learn how to cook thanks to Alexa	☐	☐	☐	☐	☐
I will use it with my smart home	☐	☐	☐	☐	☐
I am afraid I will not control how much I will spend on groceries	☐	☐	☐	☐	☐
It is not the same as going to a physical grocery store	☐	☐	☐	☐	☐
I will be more willing to cook for other people	☐	☐	☐	☐	☐
To the extent possible, I plan to use this device in my future	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐

5.2 How often do you think you would use the Amazon Dash Wand with Alexa?

- ☐ Never
- ☐ Rarely (once per month)
- ☐ Sometimes (2-3 times per month)
- ☐ Often (once per week)
- ☐ Frequently (more than once per week)
- ☐ Always (more than once per day)

5.3 How much would you like to have this product?

- ☐ Dislike a great deal
- ☐ Dislike somewhat
- ☐ Neither like or dislike
- ☐ Like somewhat
- ☐ Like a great deal

5.4 How important do you think the functions and features of this product are?

- ☐ Not at all important
- ☐ Slightly important
- ☐ Moderately important
- ☐ Very important
- ☐ Extremely important

5.5 How would you rate this product in terms of being unique from the products currently sold?

- ☐ Not at all unique
- ☐ Slightly unique
- ☐ Moderately unique
- ☐ Very unique
- ☐ Extremely unique

5.6 How sure are you that this product would solve a problem for you?

- ☐ Absolutely not sure it will
- ☐ Slightly sure it will
- ☐ Moderately sure it will
- ☐ Somewhat sure it will
- ☐ Absolutely sure it will

5.7 How believable to you were the things said about this product?

- ☐ Extremely unbelievable
- ☐ Somewhat unbelievable
- ☐ Neither believable or unbelievable
- ☐ Somewhat believable
- ☐ Extremely believable

5.8 Assuming this product was available in a store where you shop, how likely would you be to buy it?

- ☐ Definitely not buy
- ☐ Probably not buy
- ☐ Might or might not buy
- ☐ Probably buy

☐ Definitely buy

5.9 Considering the price of the Amazon Dash Wand with Alexa (20$), how do you think is its price compared to similar products you could buy?

☐ Very expensive

☐ Somewhat expensive

☐ About average

☐ Somewhat inexpensive

☐ Very inexpensive

5.10 How do you think is the Amazon Dash Wand with Alexa different compared to similar products you could buy?

☐ Absolutely not different

☐ Somewhat not different

☐ About average

☐ Somewhat different

☐ Absolutely different

5.11 How effective do you think the Amazon Dash Wand with Alexa is in fulfilling the following needs?

	Not effective at all	Slightly effective	Moderately effective	Very effective	Extremely effective
Accessing food recipes	☐	☐	☐	☐	☐
Controlling smart home devices	☐	☐	☐	☐	☐
Having a virtual assistant	☐	☐	☐	☐	☐
Buying groceries online	☐	☐	☐	☐	☐
Creating a shopping list	☐	☐	☐	☐	☐
Saving time	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐

5.12 With what similar products have you compared the Amazon Dash Wand with Alexa?

_____

Section 6: Concept test n.4 – Video

https://www.youtube.com/watch?v=s7IExS483wE&t=2s

6.1 To what extent do you agree with the following sentences:

	Strongly disagree	Rather disagree	Neither agree or disagree	Rather agree	Strongly agree
I am more willing to buy groceries online	☐	☐	☐	☐	☐
I am looking forward to use the Dash Wand with Alexa	☐	☐	☐	☐	☐
I am still not interested	☐	☐	☐	☐	☐
It looks too complicated	☐	☐	☐	☐	☐
I will learn how to cook thanks to Alexa	☐	☐	☐	☐	☐
I will use it with my smart home	☐	☐	☐	☐	☐
I am afraid I will not control how much I will spend on groceries	☐	☐	☐	☐	☐
It is not the same as going to a physical grocery store	☐	☐	☐	☐	☐
I will be more willing to cook for other people	☐	☐	☐	☐	☐
To the extent possible, I plan to use this device in my future	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐

6.2 How often do you think you would use the Amazon Dash Wand with Alexa?

- ☐ Never
- ☐ Rarely (once per month)
- ☐ Sometimes (2-3 times per month)
- ☐ Often (once per week)
- ☐ Frequently (more than once per week)
- ☐ Always (more than once per day)

6.3 How much would you like to have this product?

- ☐ Dislike a great deal
- ☐ Dislike somewhat
- ☐ Neither like or dislike
- ☐ Like somewhat
- ☐ Like a great deal

6.4 How important do you think the functions and features of this product are?

- ☐ Not at all important
- ☐ Slightly important
- ☐ Moderately important
- ☐ Very important
- ☐ Extremely important

6.5 How would you rate this product in terms of being unique from the products currently sold?

- ☐ Not at all unique
- ☐ Slightly unique
- ☐ Moderately unique
- ☐ Very unique
- ☐ Extremely unique

6.6 How sure are you that this product would solve a problem for you?

- ☐ Absolutely not sure it will
- ☐ Slightly sure it will
- ☐ Moderately sure it will
- ☐ Somewhat sure it will
- ☐ Absolutely sure it will

6.7 How believable to you were the things said about this product?

- ☐ Extremely unbelievable
- ☐ Somewhat unbelievable
- ☐ Neither believable or unbelievable
- ☐ Somewhat believable
- ☐ Extremely believable

6.8 Assuming this product was available in a store where you shop, how likely would you be to buy it?

- ☐ Definitely not buy
- ☐ Probably not buy
- ☐ Might or might not buy
- ☐ Probably buy
- ☐ Definitely buy

6.9 Considering the price of the Amazon Dash Wand with Alexa (20$), how do you think is its price compared to similar products you could buy?

- ☐ Very expensive
- ☐ Somewhat expensive
- ☐ About average
- ☐ Somewhat inexpensive
- ☐ Very inexpensive

6.10 How do you think is the Amazon Dash Wand with Alexa different compared to similar products you could buy?

- ☐ Absolutely not different
- ☐ Somewhat not different
- ☐ About average
- ☐ Somewhat different

☐ Absolutely different

6.11 How effective do you think the Amazon Dash Wand with Alexa is in fulfilling the following needs?

	Not effective at all	Slightly effective	Moderately effective	Very effective	Extremely effective
Accessing food recipes	☐	☐	☐	☐	☐
Controlling smart home devices	☐	☐	☐	☐	☐
Having a virtual assistant	☐	☐	☐	☐	☐
Buying groceries online	☐	☐	☐	☐	☐
Creating a shopping list	☐	☐	☐	☐	☐
Saving time	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐

6.12 With what similar products have you compared the Amazon Dash Wand with Alexa?

_____

Section 7: Personal Data

7.1 Gender:

- ☐ Male
- ☐ Female

7.2 Age:

- ☐ 18-24
- ☐ 25-34
- ☐ 35-44
- ☐ 45-54
- ☐ 55 or older

7.3 Country:

_____

7.4 Education:

- ☐ Less than high school diploma
- ☐ High school diploma
- ☐ Bachelor's degree
- ☐ Master of science
- ☐ PhD

7.5 Occupation:

- ☐ Student
- ☐ Employed
- ☐ Freelance professional
- ☐ Unemployed
- ☐ Retired