REMark – Revista Brasileira de Marketing
e-ISSN: 2177-5184
DOI: 10.5585/remark.v13i2.2692
Data de recebimento: 07/01/2014
Data de Aceite: 10/03/2014
Editor Científico: Otávio Bandeira De Lamônica Freire
Avaliação: Double Blind Review pelo SEER/OJS
Revisão: Gramatical, normativa e de formatação

EXPERIMENTAL RESEARCH IN MARKETING

ABSTRACT

Considering the growing number of scientific studies published in the marketing field and the development of unique theories of the area (Hunt, 2010), using experimental designs seems increasingly appropriate to investigate marketing phenomena. This article aims to discuss the main elements in conducting experimental studies and also to stimulate researchers to adopt this research method. Several international journals (e.g., JCR, JCP, JMR, JR, JBR) have been publishing articles based on experiments that not only demonstrate a relationship between two events, but also elucidate how they occur by means of mediation and moderation analyses. This article intents to be a roadmap for novice researchers on how to conduct experiments and to offer new perspectives in experimental research for experienced researchers.

Keywords: Experimental Design; Marketing Research; Consumer Behavior Research; Causal Relationships.

PESQUISA EXPERIMENTAL EM MARKETING

RESUMO

Considerando o crescimento do número de estudos científicos publicados na área de marketing e o consequente desenvolvimento de teorias próprias (Hunt, 2010), o uso de experimentos parece ser cada vez mais pertinente para elucidar o funcionamento dos fenômenos mercadológicos. Este artigo pretende discutir os principais elementos para a realização de um estudo experimental, além de estimular os pesquisadores a adotarem este método de pesquisa. Vários periódicos internacionais (p. ex. JCR, JCP, JMR, JR, JBR) têm publicado artigos com base em estudos experimentais que evidenciam não somente a relação entre dois eventos, mas também como tais eventos ocorrem, utilizando análises de mediação e moderação. Este artigo pretende servir como um guia na condução de experimentos para pesquisadores iniciantes e oferecer novas perspectivas da pesquisa experimental para pesquisadores mais experientes.

Palavras-chave: Método Experimental; Pesquisa em Marketing; Pesquisa em Comportamento do Consumidor; Relações Causais.

José Mauro da Costa Hernandez[1]
Kenny Basso[2]
Marcelo Moll Brandão[3]

[1] Doutor em Administração de Empresas pela Fundação Getúlio Vargas – FGV. Professor da Universidade de São Paulo - USP e do Centro Universitário da FEI, Brasil. E-mail. **jmhernandez@uol.com.br**

[2] Doutor em Administração pela Universidade Federal do Rio Grande do Sul – UFRGS. Professor da Faculdade Meridional – IMED, Passo Fundo – RS, Brasil. E-mail. **bassokenny@gmail.com**

[3] Doutor em Administração pela Fundação Getúlio Vargas – FGV. Professor de Marketing do PPGA - Programa de Pós-Graduação em Administração, Universidade Nove de Julho – UNINOVE, São Paulo, Brasil. E-mail: **marcelo.brandão@uninove.br**

1 INTRODUCTION

Considering the remarkable growth in the number of scientific studies published in the marketing field and the development of unique theories of the area (Hunt, 2010), the use of experiments seems increasingly more appropriate to explain marketing phenomena. When the theory regarding a certain subject matter is already well developed, qualitative methods in general may be ineffective in generating newknowledge, as they are not able to generate subsidies to falsify existing propositions and hypotheses (Bonoma, 1985).

Once marketing theories are already well developed, the focal point for future research should be identifying and measuring causal relationships using such methods as experiments. Contrary to other research methods, experiments are distinguished by two main factors: (1) the manipulation of one or more independent variables; and (2) the control of extraneous variables through strategies such as the random assignment of subjects to experimental conditions. By manipulating one or more independent variables, and considering the necessary controls, the researcher can make inferences about the observable consequences on one or more dependent variables. This process enables them to understand the cause and effect relationships in marketing phenomena.

When we consider the volume of marketing publications based on experimental studies, the relevance of this type of research becomes quite clear. The preference for experimental studies of several periodicals, but mainly those focusing on consumer behavior, can be explained by some aspects of the marketing knowledge evolution. Periodicals like JCR, JCP, and JMR have been publishing experiment-based articles that not only exploit the relationship between two events (a necessary condition for cause and effect inferences), but also how such events occur using complex mediation and moderation analyses. Through these sort of analyses, marketing phenomena mechanisms can be more deeply explained in terms of how they occur and generate the observed effects.

Even periodicals with a more balanced stand between the academic and managerial views and without a clear preference for consumer behavior studies (e.g., JM, JR and JBR), have been publishing more and more studies aiming to test causal relationships through experimental research. Therefore, the experiment seems to be the most adequate causal research design to advance the knowledge of phenomena already investigated in correlational studies.

Despite the extremely favorable international scenario to experimental designs, Mazzon and Hernandez (2013) observed that experimental studies represented less than 5 percent of marketing articles published in Brazil between 2000 and 2009. Although the central objective of this article is to discuss the main elements of an experiment, we hope it will also stimulate more scholars to adopt this research method. To fulfill these objectives, initially we introduce the concept of causality, which is followed by the description of the elements of an experiment. Next we present a classification of experimental studies. Mediation and moderation analyses are then discussed, and so are the factors that may pose a threat for conducting a good experiment and analyzing its results. Finally we present trends and challenges for marketing experiments along with our concluding remarks.

1.1 The Concept of Causality

A causal research design is prescribed when the objective is to verify whether a cause and effect relationship exists among two or more variables. Nevertheless, the notion of cause and effect that people use on a daily basis is not the same as scholars hold. For example, when someone says that the sales increase of a candy bar X was due to an increase in advertising expenditures, the obvious implication seems that the variation in the advertising budget (the cause) was responsible for the sales volume variation (the effect). Even though the average person would feel comfortable with this notion of cause and effect, the academic researcher would probably say that we cannot be sure that the sales volume increase was due to higher advertising investments, if other possible explanations can not be ruled out. For example, sales of X candy bar could have soared due to decreases in competitors' advertising, or due to a better product distribution, or even because of changes in consumer preferences. In fact, we cannot even be sure that the alleged cause (the increase in advertising budget) occurred before the effect (the increase in sales volume). For example, if advertising expenses were set as a sales percentage, it is plausible that the sales volume increase was responsible for the rise in advertising, and not the opposite.

The meaning of causality in science has never ceased to bring about heated debates. In the previous example, observing that sales and advertising investiments varied at the same time do not imply causality but merely the existence of a concomitant variation. So, what differentiates the statement "X causes Y" from the statement "X and Y vary concomitantly"? According to Hunt (2010), four necessary and sufficient conditions are required to infer a causality relationship: time order of occurrence, concomitant variation, absence of spurious associations and theoretical support. Therefore, relationships among variables that do not fulfill all four criteria cannot be considered causal.

Time order of occurrence means that the "cause" variable, also called independent or predictor variable, must precede the "effect" variable, also called dependent or criterion variable. But in many situations it is not possible to clearly distinguish between what happened first from what happened afterwards. As in the candy bar example, a less careful observer, believing that sales increases are always the consequence of higher advertising expenditures, might conclude that this is what really happened. Yet it is possible that the opposite had taken place and, therefore, the causality relationship in this event cannot be definitely established. Although the premise held for many years was that attitude change precedes behavior change, research has shown that behavioral changes generally precede attitude change (Fishbein & Ajzen 1972; Ray 1973). In complex environments, the time order of events is practically impossible to determine. For example, is the higher unemployment rate the cause or the consequence of a rise in the interest rate? Is the currency devaluation the cause or consequence of the rising inflation rate? Probably, there are arguments in favor of both sides.

The second condition to infer causality – concomitant variation – implies that changes in or the presence of the cause-variable must be systematically associated to changes in or the presence of the effect-variable (Hunt, 2010). When there is a correlation (a statistical measure of association) between two variables, there is evidence in favor of causality; the absence of correlation, on the other hand, is normally sufficient reason to refute the hypotheses of causality.

To ensure that the cause and effect relationship is not spurious – the third condition to infer causality – there must be no other variable which, when introduced as a predictor variable, eliminates the systematic association between the cause and effect variables. Suppose that a researcher is analyzing the relationship between two variables: X (the cause variable) and Y (the effect variable). To verify the existence of an association between X and Y, a linear regression, or some equivalent method, is run, being X the predictor variable and Y the criterion variable. Initially, let us assume that the predictor variable coefficient turns out to be significant. Next, a new variable, Z, is added to the regression as an alternative explanation to X for the effect on Y. The relationship between X and Y can be considered spurious if the coefficient associated with variable Z is significant, and the one associated with X is no longer significant. If, on the other hand, the coefficient associated with variable X remains significant, it can be stated that the relationship between X and Y is not spurious; thus, the causality relationship cannot be discarded based on this argument.

The need to demonstrate that a relationship is not spurious has led many science philosophers to question whether it is actually possible to infer causality. The argument is that it is not feasible to rule out all the possible alternative explanations and, as a consequence, it would never be possible to eliminate completely the argument of a spurious relationship. Essentially, no hypothesis could be refuted, which led some marketing scholars to state that "science is relative" (Peter & Olson 1983, p. 120-21).

But, for those who champion scientific pragmatism (Hunt, 2010), theoretical support, the fourth condition of causality, should solve many of the problems raised by scientific relativism defenders. In the previous example of the X candy bar, the existing marketing theory is able to explain why increases in advertising investments may provoke increases in product sales and could be used to infer the proposed causality relationship.

Taken together, the four criteria reinforce the scientific value of experimental research design, especially when compared to other non-experimental methods such as surveys and exploratory research. In time, the concepts of experiment and causal research have become practically synonymous, to the point of causal research being referred to as experimental research, since this is the only research method recognized as capable of allowing inferences about causality. In the next section we analyze the concept and the elements of experimental research design.

2 ELEMENTS OF EXPERIMENTAL DESIGNS

In an experiment, the researcher manipulates levels of the independent variables and observes their results on the dependent variable while controlling for the effect of other variables that may offer alternative explanations. When we examine closely each element of experimental designs (Fig. 1), it is easier to understand why this is the only method that guarantees the necessary and sufficient conditions to infer a causality relationship.

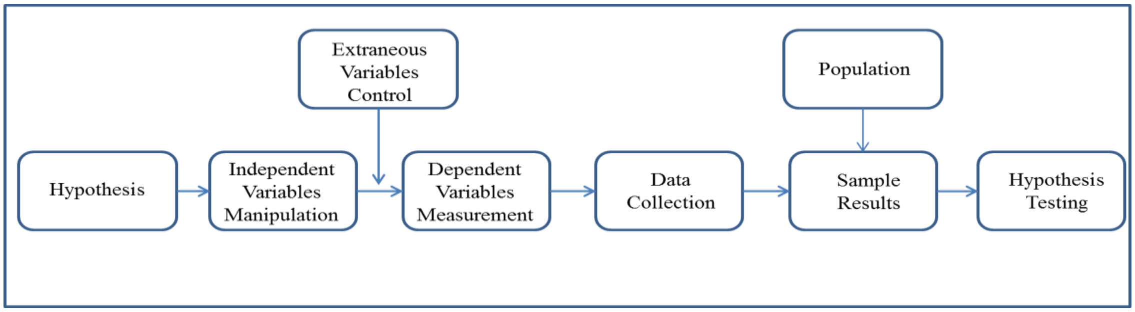

Figure 1 -Elements of experimental designs

One of the conditions to infer causality is theoretical support for the supposed relationship under analysis. When there is support in theory, hypotheses can be formulated and tested after data collection. Although exploratory and descriptive research designs do not have the formulation of a hypothesis as a necessary prerequisite, in scientific experimental studies the formulation of a theoretical hypothesis is mandatory.

The central element of experimental designs, the one which distinguishes it from other types of research, is the manipulation of independent variables. By doing so, the researcher guarantees the time sequence in which variables occur (the first condition to infer causality), as he or she controls when the stimulus corresponding to the independent variable will be introduced to research subjects, and at which point the dependent variable will be measured. Thus the researcher can ensure that the dependent variable is always measured only after the stimulus related to the independent variable has been presented.

By manipulating the independent variable, the researcher can also precisely control the nature of each treatment level, expecting to demonstrate that certain levels of the independent variable evoke different responses in the dependent variable. Generally the independent variable is manipulated along two levels, although as many levels may be used, as considered necessary. The aim is to ensure that the chosen number of levels is sufficient to produce different effects on the dependent variable.

For example, to check whether advertising skepticism moderates the effect of perceived brand extension similarity on brand extension attitude, Hernandez and Marinelli (2013) ran an experiment in which perceived brand extension similarity was manipulated. One of their main concerns was that brand extensions should be varied enough to evoke different responses, but not to the point of being considered absurd, which would diminish the experiment's external validity (the degree to which a study represents reality). One of the studies used as main stimulus the Dell computer brand; the brand extensions chosen were a multifunctional printer (very similar), sunglasses (moderately similar) and a coffee maker (very dissimilar). The appropriate choice of perceived similarity levels aided in verifying the expected effect on attitude towards the proposed brand extensions.

The same experimental design may have both manipulated and measured independent variables. Manipulated independent variables are always discrete, but measured variables may be discrete or interval. Suppose an experiment to verify the effects of advertising skepticism on the persuasive power of an ad featuring a celebrity. One of the choices for running this experiment is to use a scale to measure skepticism and then manipulate the message (with and without the celebrity). The greater problem with just measuring the independent variable in an experiment is that it can be strongly correlated to another variable; thus the alternative explanation - that the effect was caused by some other, unmeasured variable - cannot be ruled out. When the variable is directly manipulated, the alternative explanation becomes less plausible. For example, to verify if the need for cognitive closure influences how an individual searches for information (attribute-based search vs. choice-based search), Choi, Koo, Choic e Auh (2008) conducted two experiments. In the first experiment, the need for cognitive closure was measured, whereas in the second it was manipulated. This measuring-manipulating strategy is used mainly when the independent variable is an individual trait such as skepticism, self-esteem, involvement, mood, self-confidence, or information processing style.

So it is important to keep in mind that an experimental design requires at least one manipulated independent variable. In certain situations, independent variable levels are determined *post-hoc*, based on specific population characteristics (e.g., monthly income levels). When a study includes only measured variables, it is considered an observational or non-experimental study.

In the social sciences, making sure that the independent variable manipulation has been done successfully is not always a simple task. As this is a necessary condition to characterize an experiment, the usual procedure is to use manipulation check variables, whose sole objective is to ensure that the manipulation has worked as planned. For example, in studies utilizing various brand extensions as stimuli, subjects must be asked to indicate how much the extensions are similar to the parent brand; at the end of the experiment, the manipulation must be checked to see if itwas done as predicted. When the manipulation fails, it is advisable to start the experiment all over again to ensure that the independent variable is correctly manipulated.

Independent variables, when observable, may be directly manipulated; when they are non-observable, they can be manipulated indirectly. Generally, context variables such as price, promotion type, colors used in a retail environment or print ad, the number of salespeople in the store, the ad´s message type, or human density inside a retail outlet, may be directly manipulated. Their manipulation is relatively easy to be tested for reliability, since it is enough to ask subjects if they understood or noticed the manipulation. On the other hand, individual-related variables, such as the need for cognitive closure, regulatory focus, skepticism, ego depletion or self-esteem, are only manipulated by approximation and are harder to verify. In those cases, scales are normally used to check the manipulation, or individuals are indirectly inquired as to their state of mind. For example, there are many ego-depleting tasks, but no scale to measure this state. Therefore, what is usually done to check ego depletion manipulation is to ask individuals how difficult they found doing the task assigned.

Keeping in mind the cost and time involved in developing experiments, it is strongly recommended that manipulations are extensively pre-tested until they yield acceptable results. Although pretesting does not guarantee manipulations to succeed during the actual experiment, it decreases the odds of failure.

Extraneous variables represent something that may influence experimental results, as they offer alternative explanations to the independent variable. Thus, controlling for extraneous variables ensures that the only thing changing throughout the experiment are the treatment levels (Tabachnick & Fidell, 2006). In the natural sciences it is relatively easier to control for effects due to extraneous variables, because experiments can be done in the laboratory – totally isolated environments. But in the social sciences, it is much more difficult to control for that.

There are three ways of controlling extraneous variables effects. The first one is to keep constant, throughout the experiment, all the variables that may influence results. Lab experiments are better to keep environmental variables (e.g., light, temperature and noise) and testing conditions (e.g., stimulus application and data collection) relatively controlled. Even though the effect of individual variables (for example, self-esteem, self-confidence and experience) cannot be controlled through the experimental design, they are usually "controlled for" by the random assignment of subjects to experimental conditions.

Another way to control for extraneous variables effects is by randomly assigning participants to experimental conditions. Let us take an experiment in which subjects are to be exposed to three different commercials. They should not be exposed to them in the same order because doing that would create an order sequence effect. As there are six different ways to present the three commercials, each sixth of the sample will see them in a different order. At the end, it is necessary to check if the presentation order has had any effect on the dependent variable. If not, presentation order should not be considered in the analysis.

The last technique to control for extraneous variables is by means of statistical control. In most cases, statistical control is done by including the extraneous variables into the analytical model, as if they were another independent variable so their effect on the dependent variable can be isolated. Extraneous variables are also referred to as covariates in multivariate data analysis books and manuals.

When conducting experiments, great care must be taken to avoid respondents from becoming aware of the nature of the stimuli, and of the effects expected by researchers, as this consciousness may become an extraneous variable. For example, one of the best known effects in medical studies is the placebo effect, that is, the effect brought about by patients' expectation and motivation to get better after receiving treatment. To avoid its influence on results, part of the sample gets a medicament with the active principle and the other gets a placebo (a non-active agent). Although patients are warned of that, it is fundamental to make sure that they do not know which one they are taking. As a general rule, it is most advisable that the experiments are done in double bind condition, i.e., neither subjects nor researchers are aware of the experimental conditions to which each respondent has been exposed.

To avoid that participants become aware of the experiment objectives, in general it is told a fictitious story (also called cover story). For example, in study involving the placebo effect of the price of a product, Wright et al. (2013, experiment 1) told participants that a new beverage would launched in Brazil and the objective of the study was to test its effects. The new beverage allegedly increased intelectual performance and, in order to test for these benefits, participants should complete a cognitive performance task. To increase the story veracity, it was told that the beverage could take up to 10 minutes before one could feel its benefits and they should wait watching a TV documentary. In fact, the beverage was a mixture of guarana soda, orange juice, and lemmon soda and had no active principle.

From the ethical point of view, it is mandatory that participants are debriefed about the actual objectives of the experiment when it is over. This practice is universally accepted and constitues a requisite for experimental studies to be approved by the Ethical Committe of research organizations.

Measuring the dependent variable is the element most similar to other research designs. In marketing studies, a criterion variable may be of diverse nature (attitudes, emotions, behaviors, judgments, and choice, among others). There are practically no boundaries to creativity in measuring dependent variables. Waber, Shiv, Carmon e Ariley (2008) wanted to verify whether a pain killer price has influence upon its therapeutic value. In order to do so, they recruited volunteers to test a medicament and split them into two groups. All participants were told that the study aimed to check the therapeutic effect of a pain killer. One of the two groups was told that each pill cost US$ 10 cents, while the other one was told that the price of each pill was US$ 2.50. Each participant underwent two sessions of electric shocks administered to their wrists, one before taking the medicine and another afterwards; in both sessions the level of perceived pain was measured. Results showed that reduction on perceived pain (the dependent variable) in both sessions was greater for the group who had taken the pricier pills; besides, the more intense the shock, the more pronounced the reduction.

Once the experiment has been planned and data-collection tools are ready, the next step is actual data collection. Regardless of data being collected in the laboratory or in the field , in this phase it is important to ensure that all cases are randomly assigned to the experimental conditions. In other words, all individuals must have an equal probability of being selected for each of the experimental conditions. Randomization of subjects to experimental conditions has two goals (Tabatchnik & Fidell, 2006). First, to prevent the researcher to assign (even if unconsciously) the best, most intelligent, or most able subjects to the experimental condition he or she believes is the most effective. The second goal is to eliminate potential individual differences (e.g., gender, age, experience, motivation, self-esteem) among respondents. The random assignment of subjects to experimental conditions is expected to distribute individual differences in such a way that each condition can be considered equivalent before subjects are exposed to the stimulus. To guarantee randomization, testing for demographic, or any other differentiating characteristics, can be done among experimental groups. If no association between the groups and those characteristics is found, the sample can be considered randomly distributed.

Once data is collected from a previously selected sample, the last step in the experiment is hypothesis testing. One feature that differentiates experiments from other research methods is its goal of always testing a hypothesis that allows some inference about the population represented in the sample.

In conclusion, if there is enough theory to support an association between an independent and a dependent variable; if subjects are randomly assigned to experimental conditions, so that there is no association between individual characteristics and the independent variable; if the researcher has control of all environmental variables; if the researcher can control in what moment the stimulus is presented, and the dependent variable is measured; then differences in the dependent variable can be naturally attributed to the differences in treatment levels, since there is no other explanation for the systematic variation in the dependent variable. Therefore, it is said that experiments are the only way to demonstrate a cause-and-effect relationship.

3 A CLASSIFICATION OF EXPERIMENTAL DESIGNS

Experiments may be classified as true experimental, quasi-experimental, or pre-experimental types. Depending on where they are conducted, they may be laboratory (controlled environment) or field (real-life environment) studies. In terms of experimental designs, they may be classified as between-subjects, within-subjects, or both. In this section, all these classifications are discussed and finally the factorial design, largely used in marketing studies, is presented.

3.1 Types of Experimental Design: True Experimental vs. Quasi-experimental vs. Pre-experimental

True experimental studies have as a premise the random assignment of subjects to treatments. This type of study reduces the possibility of finding alternative explanations to its findings, since randomization eliminates biases due to subject assignment to experimental conditions – something that cannot be said about pre- or quasi-experimental designs.

Differently from true experiments, quasi-experimental studies do not involve the random assignment of subjects to treatments. They are usually utilized when the researcher has no control over the randomization process, or when studies are conducted in the field. Besides, in quasi-experimental designs it is assumed that the researcher cannot fully control subject exposure to manipulations of the independent variable. For example, Chae, Li and Zhu (2013, study 3), utilized a quasi-experimental study to verify the effects of proximity of product image (close to or distant from potential effects described in an ad) on judgments of product effectiveness. As independent variables, the authors manipulated one factor (spatial proximity) and measured another (knowledge level). Through knowledge measurement, authors were able to classify subjects into two groups (more vs. less knowledgeable). The method for creating the second factor does not allow random assignment of subjects, since the measurement already defines to which group each individual should be assigned; thus, this study design has to be classified as quasi-experimental.

Pre-experiemntal designs, on the other hand, are characterized by the absence of a control group. In other words, in pre-experimental studies a single group is analyzed; it is not compared to an equivalent group, which has not been exposed to the manipulation (control group). Let us suppose that we want to analyze consumer perceived unfairness after a price change. In order to do so, first a price change is made (X) and then perceived unfairness is measured (Y); thus, consumer perceived price unfairness levels are obtained, but there is no basis for comparison with other consumers (the control group), who were exposed to a normal price condition.

In another example, Estrela and Botelho (2006) used a pre-experimental design to verify the effect of price discounts on retail sales. They reduced product prices in one week and returned to previous price levels in the following week. During the study, they did not control when the subjects were exposed to the treatment, nor did they know if subjects noticed the price variations. Neither there was a control group to be compared to the group exposed to the treatment (the price variations). Such characteristics define a pre-experimental design.

3.2 Environment: Field vs. Laboratory

Experiments may be conducted in a real environment (field experiment) or in an artificial one (laboratory experiment). To decide in which environment the study should be done, Bonoma (1985) suggests analyzing if the phenomenon may be satisfactorily studied outside its natural environment, because, once withdrawn from where it naturally occurs, it may become distorted. Calder, Phillips and Tybout (1981) posit that, if the study's aim is to apply and test theoretical propositions, laboratory experiments are the best choice, since they are able to offer higher degree of control over variables, besides allowing the randomized assignment of participants to experimental groups.

In order to investigate the effects of product size and product format distortion in recycling behavior, Trudel and Argo (2013) used laboratory experiments. Despite that, it is worth mentioning that their experimental procedures stimulated subject involvement and brought close proximity to real life situations. Participants were given several sheets of paper and a pair of scissors and were told that the purpose was testing the scissors. During study procedures, participants were instructed to cut the paper in different sizes – the manner which the scholars devised to manipulate product format and size. In the end, participants were asked to evaluate the scissors along a scale. After finishing a series of other, unrelated tasks, participants left the laboratory and came upon two garbage bins, one with and one without a recycling sign on it. The amount discarded in each bin was used as the measurement for recycling behavior.

According to Sawyer, Worthing and Sendak (1979), laboratory experiments have several advantages, such as higher flexibility to verify non-behavioral measures like attitudes and beliefs; lower costs; shorter time spans; and higher secrecy regarding data and the experiment itself. On the other hand, field experiments offer more realism.

A field experiment was developed by Baca-Motes, Brown, Gneezy, Keenan and Nelson (2013) to evaluate if a hotel's commitment to environmental protection generates behavioral change in guests. With that in mind, the authors trained the receptionists to present to guests, during check-in, a letter of commitment to environmental protection issued by the hotel. With the letter, the researchers manipulated the independent variable (environment protection commitment) along two levels: in general terms (feeling responsible for the environment) or in specific terms (reusing towels during stay). They also used a control group who was not exposed to the manipulation. The study lasted for 31 days and 4,325 individuals were exposed to the experimental conditions.

3.3 Between-Subjects, Within-Subjects and Mixed Designs

When an experiment has more than one experimental treatment, the researcher faces three options: (1) exposing each subject to a single experimental treatment and comparing the measurements between subjects exposed to different treatments (between-subjects design); (2) exposing each subject to all experimental treatments and comparing the measurements related to the same subject (within-subjects design); and (3) mixing both options, i.e., exposing subjects to different treatments of one or more factors and to all treatments of other factor(s).

Between-subjects designs are the most frequent in the marketing literature and the most recommended when demand artifacts are likely to occur. This design is used when the experimental context for a treatment may influence another treatment, when the experimental context involves a single decision or when there are several variables to manipulate and/or control for (Greenwald, 1976; Charness, Gneezy & Kuhn, 2012). For example, to analyze the influence of food presentation and type on flavor perception, Poor, Duhachek and Krishnan (2013) resorted to a between-subject experimental design, in which subjects were exposed to images of healthy food (almonds) or unhealthy food (fries), either alone or in consumption situations. In this study, each subject was exposed to only one of the four possible situations.

Within-subjects designs, though used to a lesser extent than between-subjects designs, are generally used: in studies where subject response bias is to be eliminated, since different treatments are not confused with different subjects; when the experimental context involves a series of decisions; when the researcher is trying to increase external validity of results; and when internal validity is sought, no matter what the process of subject random assignment to treatments (Greenwald, 1976; Charness et al., 2012). For example, Ge, Häubl and Elrod (2012), tested whether the choice of a particular alterative is greater when favorable information to the product is delayed (vs. no delay) during the pre-purchase. To this end, the authors manipulated the effects of delayed presentation of information, in the pre-purchase situation, exposing individuals to both conditions: information delay (decision A) and information presented all at once (decision B); thus, they used a within-subjects design.

Even though the within-subjects design offers a series of possible applications, the psychological effects generated by successive treatments may bias results. Among these psychological effects, of greater concern are demand artifacts (Charness et al., 2012), a phenomenon in which the inclusion of one treatment influences how subjects will behave and respond to subsequent treatments. According to Charness et al. (2012) and Greenwald (1976), individuals tend to respond to subsequent stimuli based on how they perceived previous ones, that is, participants manipulate their responses according to what they guessed about the hypotheses of the experiment. Unless researchers are seeking to study learning and practice effects, demand artifacts may be prejudicial to study results.

Trying to align the benefits from each experimental design, some scholars have been using both types, either in the same study (e.g., McQuarrie & Mick, 2003; Thomas, Dsai & Seenivsan, 2011; Thomas & Tsai, 2012) or in different ones (e.g., Chan & Sengupta, 2010).

3.4 Factorial Design

A factorial design is used to test the effect of two or more independent variables on the dependent variable. A complete factorial study involves the combination of all factor conditions. So, a study including two factors, one with 3 conditions and the other with 2, results in six possible combinations, that is, six experimental groups. Factorial studies enable researchers to explore all the possible combinations between factor levels, and so the main effects of each factor, as well as all of the possible interactions between them, can be analyzed. This experiment design is recommended when an interaction analysis is needed to test the proposed hypotheses.

Let us take the following hypothesis: high-power consumers (a social influence measure) perceive stronger price unfairness when they pay more than other consumers do (other- comparison), whereas low-power consumers perceive stronger price unfairness when paying more than they did in previous transactions (self-comparison). Jin, He and Zhang (2014) used a 2 (power state: low vs. high) x 2 (reference: self-comparison vs. other-comparison) factorial study to test this hypothesis. The combination of the two factor levels yielded four experimental conditions. Each participant was exposed to a single condition in a between-subjects design. The authors identified a significant effect of the interaction between the factors on price unfairness perception, providing support for the proposed hypothesis.

Since the complexity level in an experimental study is determined by the number of factor and factor levels, the researcher must be aware that some studies may become unfeasible in terms of application, costs and execution complexity. Therefore, the suggestion is to start with simpler studies, whose aim is to analyze the main effect under investigation; next, other variables may be added to help explain the phenomenon.

In this sense, Chan and Sengupta (2010) used two studies (1A and 1B), less complex in terms of factor manipulation, to check the effect of flattery on explicit and implicit attitudes. Then the authors ran other factorial studies in which factors interacted to further explain the phenomenon. With such a strategy, researchers may be able to grasp and empirically evidence how a phenomenon occurs, extending the knowledge.

4 MEDIATION AND MODERATION

Besides learning if there is a cause-and-effect relationship between X and Y, often the investigator also wants to find out how and when X has an influence upon Y. The first question (how?) refers to the psychological, cognitive and biological processes that relate X to Y; the type of analysis that addresses this question is called mediation. The second question (when?) relates to the phenomenon's boundary conditions, i.e., under which conditions or to what kind of people does X influence Y – or not (Hayes, 2013). The analysis used to answer this question is called moderation. Next, we will discuss both.

4.1 Mediation

When a researcher wants to describe a phenomenon and its mechanism, or, in other words, how the independent variable, X, affects the dependent variable, Y, usually they propose a model in which one or more intervening variables, M_1, M_2,...,M_k , hold a cause-and-effect relationship with X and Y (see Figure 2). In Figure 2, a, b and c represent the effect corresponding to each path. This effect can be measured, for example, using linear regressions.

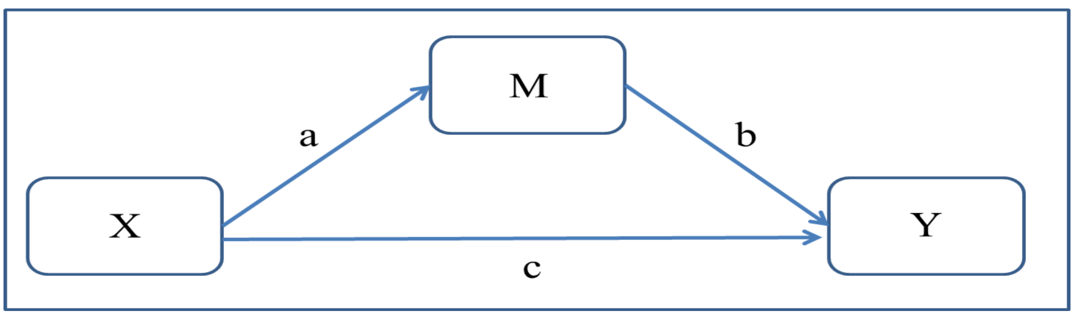

Figure 2 - Mediation conceptual framework

Lee, Keller and Sternthal (2010) noticed that consumers with a prevention regulatory focus form more favorable brand attitudes when the language of the advertising is more concrete (low-level construal), whereas consumers with a promotion regulatory focus form more favorable brand attitudes when the language of the advertising message is more abstract (high-level construal). To explain the mechanism behind the brand attitude formation, Lee, Keller and Sternthal (2010, experiment 4) tested the effect of two mediator variables: processing fluency (the degree of difficulty in understanding a message) and engagement (motivation to understand the message). The results indicate that, when regulatory focus and message language match (prevention regulatory focus with low-level construal, or promotion regulatory focus with high-level construal), both processing fluency and engagement increase, leading to more favorable brand attitudes. When regulatory focus and language do not fit (promotion regulatory focus with low-level construal, or prevention regulatory focus with high-level construal), processing fluency and engagement diminish, leading to less favorable brand attitudes.

Baron and Kenny (1986), who proposed the method for testing mediation effects most commonly used in marketing research, suggest that "mediators explain how external physical events take on internal psychological significance" (p. 1176). In Baron and Kenny´s method, a variable, M, is a mediator between X and Y when; (a) variations in the independent variable (X) levels cause variations in the mediator variable M (path a, Figure 2); (b) variations in the mediator variable, M, are responsible for variations in the dependent variable Y (path b, figure 2); and (c) when variable M effects are controlled for, the previously significant relationship between X and Y (path c, Figure 2) is no longer significant. There is the strongest evidence in favor of mediation when path c (Figure 2) is zero or non-significant, which allows to infer total mediation. When path c remains significant, a partial mediation is to be inferred.

In the last 10 years or so, though, this mediation testing procedure has been gradually substituted by an alternative method (see, for ex., Hayes, 2009; Zhao, Lynch & Chen, 2010). The main problem with Baron and Kenny’s (1986) method is its assumption of a necessary causal relation between X and Y, that is, path c in Figure 2 must be significant. In other words, if there is no causal relationship between X and Y, there cannot be mediation between X and Y. Yet, according to Bollen (1989, p. 52), “the absence of correlation (between two variables) does not invalidate causality” and “correlation is neither a necessary nor a sufficient condition to demonstrate causality”. Though totally opposed to what was then believed (refer to the second condition to demonstrate causality – concomitant variation – in the first section of this article), this new vision has been quickly adopted. Its main argument is that, if it is possible to demonstrate the existence of one or more mediator variables, M_1, M_2,...,M_k , then a relationship of causality between X and Y can be inferred (see also the third condition for causality, the absence of spurious correlation, as discussed in the first section).

So, the only necessary condition for variable M to mediate the relationship between X and Y is that path a*b in Figure 2 (the indirect effect of X on Y) be different from zero. Sobel´s (1982) test is used to calculate the confidence interval for a*b. This test has long been used to make inferences about the corresponding value in the population, often as a complement to Baron and Kenny’s method. Sobel´s test limitation is assuming the a*b distribution is normal, when it has been demonstrated to be irregular for small samples – generally the ones used in experiments (Bollen & Stine, 1990). Since it is not possible to know for sure if the a*b distribution can be approximated by the normal distribution in a given situation, nowadays the bootstrapping method is used to calculate the confidence interval for the value of a*b.

Bootstrapping is especially recommended to make inferences about a value whose distribution is unknown. Through this method, the original *n*-size sample is treated as a small representation of the original population. The observations in the original sample are then submitted to random resampling with replacement a very large number of times (generally between five and ten thousand resamples) and the statistics of interest (in the present case, the value of a*b) is calculated for each resample. This way, a hypothetical distribution is built for the statistic, which permits the desired confidence interval to be obtained. Strictly speaking, this procedure is not the same as hypothesis testing. But in practice, interpreting the bootstrapping confidence interval is the same as in hypothesis testing: if it contains zero, then the null hypothesis (the statistic is equal to zero) cannot be rejected; if it does not, the null hypothesis is rejected and the conclusion is that the statistic is different from zero.

Ferraro, Kirmani and Matherly (2013) used the method described above to test whether attitude toward a brand user mediates the relationship between conspicuous brand usage and brand attitude. They hypothesised that, if the brand is used conspicuously by someone, observers may develop a negative attitude toward this person and, as a consequence, toward the brand. Their findings showed that conspicuous brand usage negatively impacts the attitude toward the user; this, in turn, hurts the attitude toward the brand, but only for people with a low self-brand connection. For individuals with a high self-brand connection, conspicuous brand usage has neither a direct nor an indirect effect on brand attitude. Thus, Ferraro and colleagues (2013) were able to conclude that conspicuous brand usage may hurt the brand only when consumers have a low self-brand connection, and that this effect is mediated by attitude toward the brand user. One point that must be stressed in this study is that its results not only explain “how”, but also “when” the effect of conspicuous usage on brand attitude is mediated by attitude toward the brand user – when the self-brand connection is low, but not when it is high.

If the point of investigation is “when” a phenomenon happens, the involved analysis is the moderation, topic discussed in the following section.

4.2 Moderation

Moderation analysis is used to find out the phenomenon´s boundary conditions. The relationship between two variables, X and Y, is said to be moderated when its nature (magnitude and signal) depends on a third variable, W. The usual representation for moderation is shown in Figure 3.

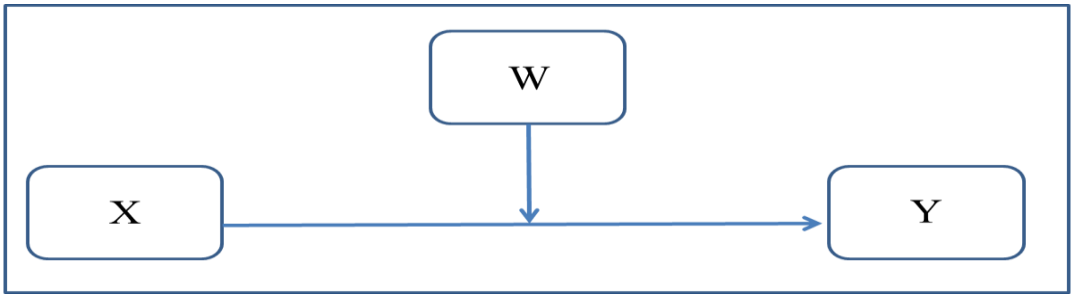

Figure 3 - Moderation conceptual framework

The moderation effect is also known as interaction and can be tested by means of a regression analysis, when X or Y are interval scaled, or by means of an ANOVA, when both X and Y are nominal or ordinal scaled. Once the existence of an interaction effect is verified, the next step is testing how moderation occurs by checking the effect of X on Y for each value of W (this process is called probing).

Hernandez, Han and Kardes (2014) tested the hypothesis that the effect of brand familiarity on product evaluation, when only a few attributes are shown, is moderated by the consumer´s objective knowledge. The point is that individuals with low objective knowledge evaluate products heuristically, emphasizing the brand and disregarding other attributes. As a consequence, products with a familiar brand are more favorably evaluated than products with unfamiliar brands. On the other hand, individuals with high objective knowledge make systematic evaluations considering both the brand and all the other attributes. Therefore, if attributes are insufficient for an adequate evaluation, individuals with high objective knowledge will equally evaluate products with a familiar or an unknown brand. In order to test this hypothesis, Hernandez and colleagues (2014) asked subjects to evaluate a digital camera, whose description had four attributes. Half the sample saw the description of a Sony camera and the other half saw an unfamiliar brand. Next, respondents were asked to answer ten questions that measured their objective knowledge about digital cameras. The results showed that less knowledgeable subjects evaluated the Sony camera more favorably than the other one, whereas high knowledgeable subjects evaluated both cameras similarly. Another analysis showed that the effect of brand familiarity on the evaluation was mediated by information sufficiency in the descriptions.

When the model´s independent variables (X and W) are discrete, moderation analysis is relatively simple. Using Y as the dependent variable, an analysis of variance is run to verify whether the interaction between X and W is significant. If so, the next step is testing the differences in Y for different values of X, holding constant each value of W.

But things are not so easy when the independent or moderator variables are continuous. In this case, the problem cannot be solved through analysis of variance, a method most researchers are familiar with. Until recently, the usual procedure was to dichotomize a continuous variable using its mean or median. The major problem with this is loss of information: when a variable measured along five, seven or nine points is transformed into a two-level variable, the power of the analysis is reduced. Besides, the results gotten through dichotomization have been shown to differ from the analysis of complete data (for a more thorough discussion, refer to Irwin & McClelland, 2001, 2003; MacCallum, Zhang, Preacher & Rucker, 2002; Maxwell & Delaney, 1993; Fitzsimons, 2008).

Conversely, an analysis without dichotomization, though not exactly trivial, is not so complicated. Let us take a simple example. In an experiment, the independent variable, X, has been manipulated along two levels; the moderator, W, and the dependent variable, Y, have been measured by supposedly interval scales (e.g., a Likert-type or semantic differential scale). First, a regression analysis is run with Y as the dependent variable, and X, W and X*W (the interaction between X and W) as independent variables. If the interaction coefficient is significant, then there is an interaction effect between X and W, that is, W moderates the effect of X on Y.

Once the existence of a moderation effect has been demonstrated, the next step is probing its nature. This is done by running two new regressions, similar to the previous one, but transforming the moderator variable into convenient values, generally one standard deviation above and below the mean of the moderator variable. Changes in significance and magnitude of the coefficient associated to the independent variable, X, point to the nature of the moderation. This type of analysis has been called in the marketing literature spotlight analysis and further details can be found in

Aiken and West (1991), Irwin and McClelland (2001), Spiller, Fitzsimons, Lynch Jr. and McClelland (2013). Hayes (2013) developed PROCESS, a syntax for both SPSS or SAS that can be used for both mediation and moderation analysis (see http://afhayes.com/).

5 THREATS TO EXPERIMENTS

Researchers have long been discussing and criticizing some aspects of experiments. Besides, several factors may reduce the impact of experimental studies to science and, consequently, their ability to make theoretical contributions. Among those factors are the validity of experiments, the power of the analysis and sampling methodology. Scholars who wish their experimental findings accepted and disseminated should pay close attention to such aspects, both in designing and conducting their studies.

5.1 Validity in Experimentation

Validity is a term to describe how close to the truth an inference really is. It is a property of inferences; different methods have distinct impacts on inference validity (Shadish, Cook & Campbell, 2001).

There are four types of validity (Calder, Phillips & Tybout, 1982; Shadish et al., 2001): statistical validity, associated to the relationship between treatments and results; construct validity, how well the manipulation or the measurement items represent the theoretical concept; internal validity, how solid inferences on the causal relation are; and external validity, or how far the causal relation can be generalized to other contexts, persons or situations.

Statistical Validity. Two main types of error may interfere in the statistical validity : Type I and Type II errors. Montgomery (2005) cites that Type I error happens when the researcher wrongly rejecting the null hypothesis, inferring a causal relation when it actually does not exist. To avoid making a Type I error, in psychology and related fields (e.g., consumer behavior), Kelley and Preacher (2012) recommend obeying the established convention: for *p* values equal or below 0.05 (level of significance), the null hypothesis should be rejected.

Type II errors happen, says Montgomery (2005), when the researcher does not reject a false null hypothesis and, consequently, infers that there is no causal relation when, in fact, there is one. Therefore, the lower the significance level utilized (to reduce Type I error), the higher the probability of a Type II error. In the same sense, Fern and Monroe (1996), and Ye, Marionova and Singh (2011) posit that small samples may increase the likelihood of Type II error, since larger samples are clearly related to statistical significance.

Construct Validity. Several marketing studies use constructs (for a deep discussion on the construct concept, see Peter, 1981). In experiments, construct validity means how the construct is understood and evaluated, be it used as the independent or dependent variable. When the construct is a categorical independent variable, that is, a factor being manipulated in the experiment, construct validity can be evaluated though peer-reviewed construct validity analysis. Another tool to ensure that construct is really being manipulated is the manipulation check.

When the construct is a dependent variable, validity may be assessed through content analysis of the scale items used to measure it, but also through statistical testing. For that purpose, we suggest assessing construct validity through the analysis of: internal consistency (Cronbach, 1951; Garver & Mentzer, 1999; Iacobucci & Duhachek, 2003); convergent validity (Bagozzi, Yi & Phillips, 1991; Garver & Mentzer, 1999); and discriminant validity (Fornell & Larcker, 1981; Bagozzi et al., 1991; Garver & Mentzer, 1999). These analyses must be used as many as possible, but most marketing experiments report only internal consistency measure, especially Cronbach´s Alpha (e.g., Mattila & Wirtz, 2001; McFerran, Dahl, Fitzsimons & Morales, 2010).

Internal Validity. In Winer´s (1999) opinion, internal validity should be the main focus of experimental studies. Internal validity is mainly related to causality inferences (McQuarrie, 2004), when the researcher is able to infer that treatment X caused the observed effect on Y. Thus, according to Anderson and Bushman (1997), if the design and structure of a study are such that one can confidently conclude that the independent variable caused systematic changes in the dependent variable, then the study is said to have high internal validity. On the other hand, if a study leaves plausible alternative interpretations of the observed relation between the independent and dependent variables, then it is said to have lower internal validity. Anderson and Bushman (1997) also mention that a way to reduce alternative explanations is using mediational processes to explain the mechanisms by which the independent variable causes the observed effects on the dependent variable.

To Zhang and Shrum (2009), field studies have lower internal validity than laboratory studies, executed with greater control and random assignment of subjects to experimental conditions. In this sense, Zourrig, Chebat and Toffoli (2009) state that studies using scenarios and laboratory controls help to apply homogeneous stimulus to all subjects and to control their effects on participants, increasing the researcher´s ability to make inferences on the causal relation. Yet, this type of study may pose a threat to external validity.

Schram (2005), and Roe and Just (2009) believe there is a trade-off between internal and external validity, so that an increase in internal validity leads to a decrease in the study realism and generalizability (lower external validity), whereas more closeness to reality and higher generalizability (higher external validity) may lead to a loss of control and consequently to a larger number of plausible alternative interpretations of the findings (lower internal validity).

External Validity. Ecological or external validity (Shadish et al., 2001) is the most discussed in marketing studies (Winer, 1999; McQuarrie, 2004). According to Winer (1999) and Lynch Jr. (1999), three main perspectives affect external validity in experiments. First, problems with the statistical generalization of findings, derived from using a limited sample which does not allow considering the findings generalizable to the population of interest. The second perspective concerns the robustness of findings, or the possibility to project the effects to other contexts, subjects or time periods. Finally, realism is also considered a possible threat to external validity in experiments, as tasks, stimuli and treatments may be distant from reality and thus make it difficult to project the findings into the real world.

Regarding the statistical generalizability of experimental findings to the population, Lynch Jr. (1999; 1982) implies that it can be unfeasible due to the usual sample sizes. Nevertheless, during sample selection, the researcher may decide to choose subjects, whose traits are more likely to influence the dependent variables. Contemplating such characteristics by using quota sampling (Lynch Jr., 1999) could be an effective way to increase external validity.

Most experimental designs keep constant the research context, but, to Lynch Jr. (1999; 1982), the context may bias results. Therefore, contexts can be alternated between studies, or even within the same study, so that results can be replicated in different contexts of interest. It is important that the researcher controls the context and tests its interactions with the other independent variables in the study.

In terms of realism, Aronson, Ellsworth, Carlsmith and Gonzales (1990) state that experimental treatments, besides their role as conceptual variables, must relate to reality as perceived by subjects. So, both in marketing and in social psychology, experiments must replicate real-world situations and possibly avoid participant guessing and interpreting the research objective. Yet, asking participants to think about a real situation (like their relationship with a brand) may cause different previous experiences to interfere with the relationships observed in the present study. Thus, Lynch Jr (1999) warning that realism should not be mistaken for carrying field or laboratory studies, but with the realism of the study context.

Pan and Siemens (2011, experiment 2) manipulated retail density using photographs depicting different numbers of people to assess its effects on consumers. In order to do so, first the authors developed a pretest with three photos, each one with a different density level – low, medium and high. Next, participants filled in a density perception scale used in previous studies. Using pictures to manipulate density has realism, or ecological validity (for a discussion on ecological validity in manipulating retail environments, see Bateson & Hui, 1992).

Summarizing, to increase external validity Winer (1999) suggests: (1) running multiple studies, with different manipulation schemes, subjects and procedures, in an attempt to replicate findings; (2) identifying conditions that limit or reverse the effects found; (3) measuring subject motivation and involvement with experiment tasks to eliminate subjects who do not show engagement; (4) bringing experimental tasks close to reality; and (5) using covariates to control for effects of individual traits and situational characteristics on dependent variables.

5.2 Power of a Test

Analysis of the power of a test can be found in books and articles on social, behavioral and health sciences. According to Faul, Erdfelder, Lang and Buchner (2007), the power of a statistical test is the probability of rejecting the null hypothesis when it is actually false, reducing Type II error. If β represents the probability of Type II error, then power can be calculated as 1 - β (Sawyer & Ball, 1981). According to Faul et al. (2007), power depends on the chosen level of probability of Type I error, on the sample size, and on the effect size defined for the verifiable hypothesis.

Sawyer and Ball (1981) suggest that marketing studies try to answer two often neglected questions: is the study sensitive enough to identify covariation? And, if there is evidence of it, how strong is the covariation between cause and effect?

The effect strength can be measured using Cohen's (1988) *d*, and/or partial squared *eta;* both assess the size of the effect of the independent variable on the dependent one. To Kelley and Preacher (2012, p. 140), the effect size is "a quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest". Cohen's *d* is calculated as the difference between two means divided by the standard deviations of two groups

(Thalheimer & Cook, 2002). As defined by Cohen (1992), effect size should follow certain conventional rules:, values under 0.2 are considered small, being 0.5 the desirable reference values between 0.2 and 0.8 are considered medium, whereas values above 0.8 are large.

The partial squared eta measures how intense the independent-dependent variable association is (Cohen, 1988). It is calculated as the sum of the squares of the effect, divided by the sum of the squares of the effect plus the sum of the squares of errors. Cohen's (1992) suggestion is that partial squared eta values around 0.01 should be considered small effect sizes; values around 0.06 are medium, and values around 0.13 are considered large effects.

Faul et al. (2007) introduced software G*Power 3 to analyze the power of statistical tests using the normal distributions t, F, $\chi 2$ (parametric tests). The software can be downloaded from the following address: http://www.gpower.hhu.de/.

5.3 Sample Selection

When a study´s aim is theory application, Calder et al. (1981) recommend using a sample with homogenous respondents, except when individual characteristics are necessary and may not be manipulated (e.g., personality traits). Two reasons justify using a homogenous sample to run experiments. First, as they reduce the likelihood of factors that may interfere with results, a homogenous sample allows for more accuracy in theoretical hypothesis testing than heterogeneous samples (Calder et al., 1981; Lynch Jr., 1982). Second, when respondents possess heterogeneous characteristics that may affect their responses, the variance error may increase and the sensitivity of statistical tests in identifying relationship significance may decline (Calder et al., 1981).

Falk and Heckman (2009) cite that the use of student samples to experimental studies in the social sciences are fairly discussed. Thus Calder et al. (1981) recommend the uses of homogeneous sample to test theoretical propositions, the study may replicated the findings using heterogeneous samples with closer similarity to the population. Replications using a heterogeneous sample may warrant greater generalizability to the findings (Calder et al., 1981).

Referring to sample sizes, even if modern statistical tools (such as non-parametric analyses) offer solutions for small samples, it is recommended that studies employ larger samples, with more cases per group, respecting the sample size recommendations mentioned in this article (e.g., effects of sample size on power). Hair, Anderson, Tatham and Black (2005) argue that samples with more than 30 cases already tend to show characteristics of a normal distribution; so, we recommend research designs with at least 30 cases assigned to each experimental condition.

6 PERSPECTIVES ON USING EXPERIMENTS IN MARKETING RESEARCH

In this section we present some trends and challenges concerning the use of experiments in the marketing field, based on what we have seen published in the main marketing journals, and on comments by scholars and critics. We do not expect that our observations are unanimously accepted, since they are not empirically-based in scientific terms. But we argue that evidence is sufficient to say that our remarks make sense in the present state of marketing research.

6.1 Trends in marketing experiments

It is no news for academics that getting published in recognized periodicals is an increasingly competitive endeavor. Observing works published in the main marketing and consumer psychology periodicals during the last 30 years (JM, JCR, JMR, JAMS, JCP, IJMR), we were able to detect some trends, such as: the growing number of experimental studies reported in each article; usage of simpler experimental designs, but with much more rigorous procedures; and the increased number of field studies. Each trend is briefly discussed and then the challenges in using experiments in marketing studies.

6.1.1 The growing number of experimental studies reported in each article

Articles reporting one single experiment used to be the rule, even for the most prestigious journals. Today, articles describe at least three experiments, not only to demonstrate the main phenomenon, but also to rule out possible alternative explanations, by means of increasingly complex analyses (Pham, 2013).

As an illustration to our argument, we compared the December, 1985 issue (volume 12) of the *Journal of Consumer Research* to the December, 2013 issue (volume 40). Among the 8 articles published in the 1985 issue, only 2 reported experimental findings, and each one had only one experiment. In the last issue of 2013, among the 12 articles published, 9 reported findings from 41 experiments (an average of 4.5 experiments per article). And, JCR is not an exception, for the same trend was observed in consumer behavior periodicals like JM, JMR, JCP and IJMR.

As Tabachnick and Fidell (2006, p.4) warn: "even if a study is properly experimental in all respects, inference of a causal relationship between a depedent variable and an independent variable is hazardous until the study is successfully replicated one or more times. Researchers only begin to believe that there is a causal relationship between two variables when a study is repeated and similar results are found. After that, the argument for a causal relationship is strengthened by asking the research question using a variety of research strategies and achieving similar results each time."

Contrary to what happens in several disciplines, replications of studies are not accepted for publication in the most prestigious marketing journals (the exception being IJMR, which recently opened up a new section exclusively for replications). Thus, the only solution for researchers is to replicate their own studies several times. A typical article reports at least three experiments: the first shows the causal effect; the second, using a different strategy (a different manipulation of the independent variable, the dependent variable measured in a different way, or distinct data collection procedures), demonstrates the effect of a moderator variable; the third study, using still another strategy, shows the effect of a mediator variable. A fourth and a fifth one could eliminate alternative explanations. It is only natural that the number of reported experiments has grown from 1 - 1.5 to 4.5 - 5 per article in the last 30 years.

6.1.2 Simpler Experimental Designs

Although the number of experiments per article has increased, it also seems that they have become a little simpler. But that does not mean less rigorous. To illustrate our point, we took two highly-cited articles on memory published in the 1980's. To study the influence of attribute recall on product choice, Lynch Jr., Marmorstein and Weigold (1988) ran an experiment manipulating only 2 independent variables. Yet, in order to test the four hypotheses of the study, 108 subjects were exposed to eight, quite complexly-arranged, experimental conditions. Data were collected in two different moments, with a two-day interval, and four dependent variables were analyzed. In another study, published around the same period, Keller (1987) wanted to verify the effect of advertising retrieval cues on brand evaluations. To do so, the author ran an experiment, in which 3 independent variables were manipulated using a mixed (two between-subjects and one within-subjects) design, with 2 replications. Sixteen product ads were devised to serve as stimuli for the 4 experimental conditions and a control group. The 200 participants, adults recruited among a group of student parents, were exposed to 12 ads for different brands from 4 product categories. Six dependent variables were measured and analyzed.

On the other hand, it is much more likely that a recently-published article will describe between 3 and 6 experiments with relatively simpler designs, with one or two hypotheses tested per experiment. Nevertheless, the rigor in reporting experiments has not diminished. Because the number of experiments per article has increased and the limit number of pages has not changed substantially, the solution found by more and more periodicals is to publish a portion of the article on their websites. As an example, JCP has recently issued new publishing guidelines, expressly demanding that authors submit "a detailed description of methods, analyses and study findings in the text or in an appendix, to permit study replication; a methodological appendix describing all the stimuli and measures used in the study; and a statement on the role each author played in the study" (Pechmann, 2013, p.1). We conclude that, to be accepted by a recognized periodical, several studies with varied complexity levels are necessary, carefully reported in order to be replicated.

6.1.3 Increased number of field studies

We were not sure whether to classify this topic as a trend or a challenge, because, despite the growing number of field experiments published in the last couple of years, it is also true that the number of laboratory experiments is still much higher. Although field or laboratory studies may be used separately, Falk and Heckman (2009) argue that mixing both is recommended to investigate most social sciences phenomena, since laboratory experiments complement the deficiencies in field experiments, and vice-versa.

Several authors have written about the need to amplify realism and external validity in marketing studies (for example, Pham, 2013). Tough still tentatively, more and more authors have tried to face this challenge through field experiments, which, despite their higher external validity, are more difficult to execute. For example, to verify if the fit between framing and construal level in a message promoting recycling would influence the volume of recycled garbage, White, MacDonnell and Dahl (2011) conducted a field experiment in partnership with the city of Calgary, Canada. They created three measures for the recycled garbage volume and observed it among the 390 homes included in the sample, in 3 different moments: during the first week, the third week, and after six months. The stimulus, a print folder promoting recycling, was distributed to the homes during the second week of the experiment. It is important to stress how complex this type of experiment was, as it involved behavioral measures quite difficult to obtain (how much and what kind of garbage was destined to recycling), observed in 3 different moments, using a relatively spread-out probabilistic sample, during a six-month period.

Despite being more difficult to execute, quite probably the number of field experiments published in the main marketing journals will rise in the future. Even in laboratory studies, the level of realism of the stimuli employed will be increased. Not only will stimuli have to help subjects to imagine a situation or to perceive it as real, but they will also lead participants to get involved in the study. How much participants relate to the tasks and how much they get involved will help researchers to get more reliable and more realistic results.

6.2 Challenges to using experiments in marketing

In this section we discuss two of the main challenges to marketing experiments: more representative samples and the increase in using behaviors as dependent variables.

6.2.1 More representative samples

According to Pham (2013), one of the seven sins in consumer psychology studies is the use of convenient samples. Pham criticizes not only the excessive confidence in student samples, but also resorting to consumer panels like Mturk, or relying exclusively on samples composed with North-American citizens. In his opinion, consumer psychology (and, as we say, the whole marketing field) should encourage studies with real consumers, with a wider variety of social economic backgrounds and living in different places.

Actually, published studies with real consumer samples are quite rare. As a matter of fact, student samples are relatively easy to reach and much less costly. Besides, college students have all the necessary cognitive abilities to understand and respond to the ever more complex scenarios that marketing researchers develop. On the other hand, criticism on studies carried out exclusively with student samples have grown to such an extent, that several periodicals no longer accept them (the *Journal of Advertising Research* is one example). Thus, more and more studies with actual consumer samples are to be expected, and fewer with student samples. But this a huge challenge, considering the costs and difficulties in getting actual consumer samples.

6.2.2 Studies with observation of behavior

Quite recently, Baumeister, Vohs and Funder (2007) argued that studies investigating natural behaviors were rare to find. When this was the case, human behavior was almost always observed “in a seated position, usually in front of a computer … Finger movements, as in keystrokes and pencil marks, constitute the vast majority of human action” (p. 397). Although Baumeister and colleagues (2007) were criticizing social and personality psychology studies, their criticism would apply to marketing studies as well. The authors (2007) continue: “though psychology is the science of behavior, behavior is most difficult to be found in publishing in the field... Social psychology has turned in recent times to the study of reaction times and questionnaire responses. Sometimes these questionnaires ask people to report what they have done, will do, or would do. More often, they ask people to report what they think, how they feel, or why they do what they do.”

Just like in psychology, today most marketing experiments ask people to report their thoughts, feeling, memories and attitudes. Nevertheless, thanks to various studies and personal observations, it is known for a fact that people do not always do what they say they will, or have not done what they say they have. One of the most important issues in marketing is learning about consumer behavior: whether or not they bought, used, discarded a product, where and how they stored something, whether or not they saw a commercial, if they bought something as a gift or for themselves, how well or how badly they spoke of a brand. Yet, as in psychology, very often marketing studies report measurements of intentions, attitudes, emotions and thoughts. This is also a huge challenge, because experimental studies involving behavioral measurements are more prone to errors and more difficult to carry out. Nevertheless, it is quite likely that in the future journals demand more studies involving behavioral measurements.

7 CONCLUSIONS

This article aimed to discuss the main aspects of experimental designs, not to put an end to discussions on how to improve marketing experiments. Our expectation, as is the case of marketing academia, is that experimental studies become ever more rigorous, without losing their relevance. Thus, articles on research methods are crucial as a foundation for new researchers, and also as a matter of discussion among more experienced scholars. We strongly recommend new articles on experimental designs, bringing different viewpoints and addressing the gaps left in our work.

It is our expectation that further studies deepen the knowledge on moderation and mediation testing (e.g.., using variables measured along category or interval scales, manipulated or stated, with more adequate tests, etc.), as well as on experimental design building (the need for several pilot-tests to find the ideal design for hypothesis testing, the need for planning and measuring samples due to experimental conditions, among others,).

There is a need not only to discuss experimental methods, but to use them with methodological rigor, in order to understand marketing phenomena more deeply. It is crucial that researchers are aware of the aspects mentioned here, as well as of challenges and trends in experimental studies in marketing.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage Publications.

Anderson, C. A. & Bushman, B. J. (1997). External validity "trivial" experiments: The case of laboratory aggression. *Review of General Psychology*, 1(1), 19-41.

Aronson, E., Ellsworth, P. C., Carlsmith, J. M., & Gonzalez, M. H. (1990). *Methods of Research in Social Psychology*. 2 ed. Reading, MA: Addison-Wesley.

Baca-Motes, K., Brown, A., Gneezy, A., Keenan, E. A., & Nelson, L. D. (2013). Commitment and behavior change: Evidence from the field. *Journal of Consumer Research*, 39(5), 1070-1084.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991) Assessing construct validity in organizational research. *Administrative Science Quarterly*, 36(3), 421-458.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182.

Bateson, J. E. G., & Hui, M. K. (1992). The ecological validity of photographic slides and videotapes in simulating the service setting. *Journal of Consumer Research*, 19(2), 271 281.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, 2(4), 396-403.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley and Sons.

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology*, 20, 115-140.

Bonoma, T. V. (1985). Case research in marketing: Opportunities, problems, and a process. *Journal of Marketing Research*, 22(2), 199-208.

Calder, B. J., Phillips, L. W., & Tybout, A. M. (1981). Designing research for application. *Journal of Consumer Research*, 8(2), 197-207.

Calder, B. J., Phillips, L. W., & Tybout, A. M. (1982). The Concept of External Validity. *Journal of Consumer Research*, 9(3), 240-244.

Chae, B., Li, X., & Zhu R. (2013). Judging Product Effectiveness from Perceived Spatial Proximity. *Journal of Consumer Research*, 40(2), 317-335.

Chan, E., & Sengupta, J. (2010). Insincere Flattery Actually Works: A Dual Attitudes Perspective. *Journal of Marketing Research*, 47(1), 122-133.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1), 1-8.

Choi, J. A., Koo, M., Choic, I., & Auh, S. (2008). Need for cognitive closure and information search. *Psychology & Marketing*, 25(11), 1027-1042.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2. Ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power prime. *Psychological Bulletin*, 112(1), 155-159.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.

Estrela, V. B., & Botelho, D. (2006). Efeito de atividades promocionais no varejo. *Revista de Economia e Administração*, 5(3), 297-311.

Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326, 535-538.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. L. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods,* 39(2), 175-191.

Fern, E. F., & Monroe, K. B. (1996). Effects-Size estimates: Issues and problems in interpretation. *Journal of Consumer Research*, 23(2), 89-105.

Ferraro, R., Kirmani, A., & Matherly, T. (2013). Look at Me! Look at Me! Conspicuous brand usage, self-brand connection, and dilution. *Journal of Marketing Research*, 50(4), 477-488.

Fishbein, M., & Ajzen, I. (1972). Attitudes and opinions. *Annual Review of Psychology*, 23. 487-543.

Fitzsimons, G. J. (2008). Death to dichotomizing. *Journal of Consumer Research*, 35(1), 5-8.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, 18(3), 382-388.

Garver, M. S., & Mentzer, J. T. (1999). Logistics research methods: Employing structural equation modeling to test for construct validity. *Journal of Business Logistics*, 20(1), 33-57.

Ge, X., Häubl, G., & Elrod, T. (2012). What to say when: Influencing consumer choice by delaying the presentation of favorable information. *Journal of Consumer Research*, 38(6), 1004-1021.

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin*, 83(2), 314-320.

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (2005). *Análise multivariada de dados*. 5. ed. Porto Alegre: Bookman.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* New York, NY: The Guilford Press.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, 76(4), 408-420.

Hernandez, J. M. C., Han, X., & Kardes, F. R. (2014). Effects of the perceived diagnosticity of presented attribute and brand name information on sensitivity to missing information. *Journal of Business Research*, 67(5), 874-881.

Hernandez, J. M. C., & Marinelli, F. M. (2013). Seeing is not believing: The influence of advertising skepticism on brand extension evaluations. In XXXVII Encontro Nacional da Anpad, Anais, Rio de Janeiro.

Hunt, S. D. (2010). *Marketing theory: Foundations, controversy, strategy, resource-advantage theory*. Armonk, NY: M.E. Sharpe.

Iacobucci, D., & Duhachek, A. (2003). Advancing alpha: Measuring reliability with confidence. *Journal of Consumer Psychology,* 13(4), 478-487.

Irwin, J. R., & McClelland, G. H. (2001). Misleading heuristics and moderated multiple regression models. *Journal of Marketing Research*, 38(1), 100-109.

Irwin, J. R., & McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, 40(3), 366-371.

Jin, L., He, Y., & Zhang, Y. (2014). How power states influence consumers' perceptions of price. *Journal of Consumer Research*, 40(5), 818-833.

Keller, K. L. (1987). Memory factors in advertising: The effect of advertising retrieval cues on brand evaluations. *Journal of Consumer Research*, 14(3), 316-333.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, 17(2), 137-152.

Lee, A. Y., Keller, P. A., & Sternthal, B. (2010). Value from regulatory construal fit: The persuasive impact of fit between consumer goals and message concreteness. *Journal of Consumer Research*, 36(5), 735-747.

Lynch Jr., J.G., Marmorstein, H., & Weigold, M.F. (1988). Choices from sets including remembered brands: Use of recalled attributes and prior overall evaluations. *Journal of Consumer Research*, 15(2), 169-184.

Lynch Jr., J. G. (1982). On the external validity of experiments in consumer research. *Journal of Consumer Research*, 9(3), 225-239.

Lynch Jr., J. G. (1999). Theory and external validity. *Journal of the Academy of Marketing Science*, 27(3), 367-376.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1), 19-40.

Mattila, A., & Wirtz, J. (2001). Congruency of scent and music as a driver of in-store evaluations and behavior. *Journal of Retailing*, 77(2), 273-289.

Maxwell, S. E. & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, 113(1), 181-190.

Mazzon, J. A., & Hernandez, J. M. C. (2013). Produção científica brasileira em marketing no período 2000-2009. *Revista de Administração de Empresas*, 53(1), 67-80.

McFerran, B., Dahl, D. W., Fitzsimons, G. J., & Morales, A. C. (2010). I'll have what she's having: Effects of social influence and body type on the food choices of others. *Journal of Consumer Research*, 36(6), 915-929.

McQuarrie, E. F. (2004). Integration of construct and external validity by means of proximal similarity: Implications for laboratory experiments in marketing. *Journal of Business Research*, 57(2), 142-153.

McQuarrie, E. F., & Mick, D. G. (2003). Visual and verbal rhetorical figures under directed processing versus incidental exposure to advertising. *Journal of Consumer Research*, 29(4), 579-587.

Montgomery, D. C. (2005). *Design and analysis of experiments.* 7. ed. New York: John Wiley & Sons, Inc.

Pan, Y., & Siemens, J. C. (2011). The differential effects of retail density: An investigation of goods versus service settings. *Journal of Business Research*, 64(2), 105-112.

Pechmann, C. (2013). Editorial regarding the new submission guidelines at the Journal of Consumer Psychology. *Journal of Consumer Psychology* 24(1), 1-3.

Peter, J. P. (1981). Construct validity: A review of basic issues and marketing practices. *Journal of Marketing Research*, 18(2), 133-145.

Peter, J. P., & Olson, J. C. (1983). Is Science Marketing? *Journal of Marketing*, 47(4). 111-125.

Pham, M. T. (2013). The seven sins of consumer psychology. *Journal of Consumer Psychology*, 23(4), 411-423.

Poor, M., Duhachek, A., & Krishnan, H. S. (2013). How images of other consumers influence subsequent taste perceptions. *Journal of Marketing*, 77(6), 124-139.

Ray, M. (1973). Marketing communication and the hierarchy of effects. In Clark, P., *New Models for Mass Communications Research*, v. 2, Beverly Hills: Sage, 147-176.

Roe, B. E., & Just, D. R. (2009). Internal and external validity in economics research: Tradeoffs between experiments, fields experiments, natural experiments, field data. *American Journal of Agricultural Economics*, 91(5), 1266-1271.

Sawyer, A. G., & Ball, D. (1981). Statistical power and effect size in marketing research. *Journal of Marketing Research*, 18(3), 275-290.

Sawyer, A. G., Worthing, P. M., & Sendak, P. E. (1979). The role of laboratory experiments to test marketing strategies. *Journal of Marketing*, 43(3), 60-67.

Schram, A. (2005). Artificiality: the tension between internal and external validity in economic experiments. *Journal of Economic Methodology*, 12(2), 225-237.

Shadish, W., Cook, T., & Campbell, D. (2001). *Experimental and quasi-experimental designs for* *generalized causal inference*. Boston: Houghton Mifflin.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In Leinhart, S. (ed), *Sociological methodology*, San Francisco, CA: Jossey-Bass, 290-312.

Spiller, S. A., Fitzsimons, G. J., Lynch Jr., J. G., & McClelland, G. H. (2013) Spotlights, fllodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research*, 50(2), 277-288.

Tabachnick, B. G. e Fidell, L. S. (2006). *Experimental design using ANOVA*. California: Cengage.

Thalheimer, W., Cook, S. (2002). *How to Calculate Effect Sizes from Published Research articles: A Simplified Methodology*. Documento digital, 2002. Disponível em: <http://www.bwgriffin.com/gsu/courses/edur9131/content/Effect_Sizes_pdf5.pdf>. Acesso em: 01/11/2012.

Thomas, M., & Tsai, C. I. (2012). Psychological distance and subjective experience: how distancing reduces the feeling of difficulty. *Journal of Consumer Research*, 39(2), 324-340.

Thomas, M., Desai, K. K., & Seenivasan, S. (2011). How credit card payments increase unhealthy food purchases: Visceral regulation of vices. *Journal of Consumer Research*, 38(1), 126-139.

Trudel, R., & Argo, J. J. (2013). The effect of product size and form distortion on consumer recycling behavior. *Journal of Consumer Research*, 40(4), 632-643.

Waber, R. L., Shiv, B., Carmon, Z., & Ariely, D. (2008). Commercial features of placebo and therapeutic efficacy. *Journal of the American Medicine Association*, 299(9), 1016-1017.

White, K., MacDonnell, R., & Dahl, D. W. (2011). It's the mind-set that matters: The role of construal level and message framing in influencing consumer efficacy and conservation behaviors. *Journal of Marketing Research*, 48(3), 472-485.

Winer, R. S. (1999). Experimentation in the 21st century: The importance of external validity. *Journal of the Academy of Marketing Science*, 27(3), 349-358.

Wright, S., Hernandez, J.M.C., Sundar, A., Disnmore, J., & Kardes, F. (2013). If it tastes bad it must be good: Consumer naïve theories and the marketing placebo effect. *International Journal of Research in Marketing*, 30(2), 197-198.

Ye, J., Marinova, D., & Singh, J. (2012). Bottom-up learning in marketing frontlines: conceptualization, processes, and consequences. *Journal of the Academy of Marketing Science*, 40(6), 821-844.

Zhang, Y., & Shrum, L. J. (2009). The influence of self-construal on impulsive consumption. *Journal of Consumer Research*, 35(5), 838-850.

Zhao, X., Lynch Jr., J.G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197-206.

Zourrig, H., Chebat, J. C., & Toffoli, R. (2009). Consumer revenge behavior: a cross-cultural perspective. *Journal of Business Research*, 62(10), 995-1001.