Article:

Viglia, G, Zaefarian, G orcid.org/0000-0001-5824-8445 and Ulqinaku, A orcid.org/0000-0003-3456-6451 (2021) How to design good experiments in marketing: Types, examples, and methods. Industrial Marketing Management, 98. pp. 193-206. ISSN 0019-8501

https://doi.org/10.1016/j.indmarman.2021.08.007

Introduction

Experiments enable researchers to determine causal relationships between an independent variable and a dependent variable, by manipulating the independent variable with a high degree of control over the rest of the environment (Kirk, 2013). Assessing cause-effect relationships is a key motivation for experimental research compared to cross-sectional surveys. Although this latter approach is widely used in quantitative business marketing research, it can be problematic with regards to endogeneity (Ullah, Akhtar, & Zaefarian, 2018; Zaefarian, Kadile, Henneberg, & Leischnig, 2017). In general, failing to establish causality is a serious limitation in any research study (Zellmer-Bruhn, Caligiuri, & Thomas, 2016).

Experiments allow researchers to assess the effect of a predictor, *i.e.*, the independent variable, on a specific outcome, *i.e.*, the dependent variable, while controlling for other factors. As such, a key tenet of good experimental design is the accuracy of manipulation. Manipulation in an experiment refers to the procedure through which the researcher changes or alters the predicted cause (i.e., the independent variable) in a treatment group and a control group. Randomizing participants into these groups, it is possible to investigate how this change may affect the outcome (i.e., the dependent variable; Allen, 2017).

While experimental research remains one of the main methodological approaches in marketing and their related disciplines such as consumer behavior research (see Simester, 2017; Viglia & Dolnicar, 2020), its application in business marketing is scant. Our assessment of studies published in *Industrial Marketing Management* over the last two decades suggests that a growing number of studies published in recent years are benefitting from experimental research as their core methodology (e.g., Saab & Botelho, 2020; Zimmer, Salonen, & Wangenheim, 2020). However, most quantitative studies in business marketing research still often employ not experimental methods. Specifically, running a systematic search in Scopus of all the papers published in *Industrial Marketing Management* in the last decade (i.e., since 2010), we identified only 39 papers using this methodology, with a flat trend, except for a slightly increase in 2020 (see Figure S1 in Online Appendix).

Different justifications can be offered for the paucity of experimental studies in business marketing research. For example, business research is often multidisciplinary and looks at macro-level, long-term phenomena (Zellmer-Bruhn et al., 2016). Sometimes, random assignment is simply not a possibility because there is lack of sufficient samples to assign some firms to the treatment group i.e., to be exposed to the treatment, and other firms to the control group (Cuervo-Cazurra, Andersson, Brannen, Nielsen, & Reuber, 2016).

The difficulty in randomly assigning firms or senior or top-level managers to experimental groups may discourage researchers in industrial marketing from carrying experimental research. Relatedly, the nature of the sample used in business marketing (e.g., senior or top-level managers) collides with the use does not make the use of laboratory experiments with a convenience sample (. However, we suggest that most of the experiments in business research can be conducted in the field with real managers. Having high experimental realism and measuring actual behavior, this type of experiment provides direct industry recommendations. Specifically, compared to surveys or hypothetical laboratory experiments based on self-reported intentions, field experiments are particularly important because people lack perfect rationality, struggle to accurately predict their own preference and behavior, tend to misreport even unconsciously, and sometimes even construct ex-post explanations for behavior that makes sense to them (Kahneman & Egan, 2011). By measuring real behavior, field experiments close the so-called '*attitude-behavior gap'* (see Blake, 1999; Carrington, Neville, & Whitwell, 2014). A/B testing (see Anderson & Simester, 2011) is an actionable way to run field experiments with businesses. Practically, companies can run field experiments through A/B testing when evaluating if one technique (e.g., using one logo vs. another, a specific font in the contract vs. another, high vs. low levels of supplier monitoring or contract specificity, etc.) produces more favorable outcomes compared to another one.

Against this backdrop, the aim of this article is to provide an explanation of the value and importance of experiment methodology in business research and discuss different types and key tenets of good experimental design using multiple examples. More specifically, we explain the importance of the experimental method, how to decide between different types of experiments, how to design an experiment, the role of manipulation and attention checks, how to determine the sample size and where to possibly recruit respondents, and how to analyze and interpret the results of experimental data. Finally, we offer a checklist for authors and reviewers running or evaluating experimental studies in marketing.

Basics of Experiments

Experiments are defined as “a plan for assigning experimental units to treatment levels and the statistical analysis associated with the plan” (Kirk, 1995, p. 1). To accurately capture causality, it is important to i) manipulate the independent variable by having at least one manipulated group exposed to the treatment and one control group not exposed to the treatment, ii) have a randomized design where participants are assigned randomly to the conditions, iii) make sure that the independent variable is manipulated before the measurement of the dependent variable, and iv) test for differences in one (or more) dependent variables among conditions (Kirk, 2013). Randomization avoids respondents’ “selection” into treatment so that the only difference between groups is the intervention. The control tells what would have happened without intervention (counterfactual). Since respondents are randomly assigned to the two groups (or more in case of several treatment groups or several control groups) and respondents' characteristics are assumed to be normally distributed, there is no reason to expect that one group would be different than the other. Therefore, we can expect that the effect of the treatment is causal on the dependent variable. Moreover, individual characteristics that make one respondent different from the other are spread across the groups, not allowing for these aspects to affect the whole treatment or control group producing biased differences. A further beneficial aspect of experimental design is that of controlling for – and ruling out – alternative explanations that may cause the effect of one variable onto the other. If there are other factors which could be responsible for changes in the dependent variable, we cannot be confident that the presumed cause-effect relationship is correct.

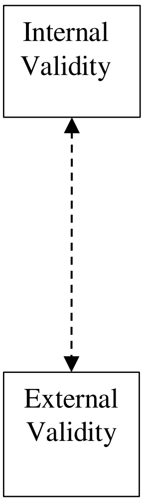
It is important to note here that there is a difference between random allocation of subjects in B2C and B2B research. Since it is fairly complicated and difficult to randomly allocate firms or buyers to the treatment and control conditions in B2B research, a solution is that of randomly allocating managers and employees to these conditions instead (Hada, 2021). This way, the researcher can study the differences in the outcomes across conditions (e.g., in decision-making).

There are different types of experiments, from the most conventional ones, such as field, online, and laboratory experiments, to quasi-experiments, and conjoint analysis. We discuss more in detail these types of experiments in the section below.

Different Types of Experiments

Internal and external validity are often understood to be opposing forces or competing with each other in experimental designs (Schram, 2005). Internal validity is the extent to which we can reliably conclude that the independent variable is the main responsible variable for the changes in the dependent variable(s) (Kirk, 2013). External validity refers to the extent to which the results can be generalized across populations (Kirk, 2013). Table 1 presents a taxonomy table of different types of experiments, with the advantages and disadvantages of application of each of the categories in Business to Business (B2B), but that apply to any other marketing areas too. The types of experiments are presented in a continuum that goes from the maximum level of the internal validity to the maximum level of external validity. Compared to conventional laboratory experiments and field experiments, experiments with increased behavioral realism are an intermediate category where there is a lower level of internal validity and a higher level of external validity. The last two categories presented in the table include types of studies that are sometimes referred to as experiments. The first type – quasi experiments/natural data – encompasses situations where data are organic, it is not possible to randomly allocate to treatment and control conditions, and there is no intervention by the experimenter (e.g., Garrett & Gopalakrishna, 2019; Laursen & Andersen, 2016; Ruiz & Kowalkowski, 2014). Longitudinal experiments are a type of quasi-experiments where the same participant (be this an organization or single respondent) is repeatedly examined over time, investigating possible changes in the dependent variable at any point in time or detecting trends (Zellmer-Bruhn et al., 2016). The second category – conjoint analysis – includes studies where participants have to express their evaluation and/or ranking order for a number of carefully designed attributes (e.g., Bendixen, Bukasa, & Abratt, 2004).

Table 1. Types of experiments



Type of experiment	Characteristics
Conventional laboratory experiment	High internal control for the experimenter; Generally, it presents an abstract framing; Imposed set of rules; Primarily homogenous subject pools
Experiment with increased behavioral realism	Experiment conducted in the lab or online, measuring some form of real behavior (e.g., simulating a real negotiation process in a lab, using game theory simulations online, or choosing real products online)
Field experiment	The experimenter wants to investigate the field context. Subjects may (vs. not) be aware of their participation in an experiment. Because of the field aspects, the researcher has less internal control.
Quasi experiments/Natural data	Same as natural field experiments except there is no intervention by an experimenter, but there is, however, some kind of external intervention that has occurred (e.g., a change in legislation, a natural disaster, etc.). Data is completely organic.
Conjoint analysis	Participants elicit their preferences on a series of manipulated factors (i.e., the attributes). This allows researchers to measure how much stakeholders value specific product features.

Why conventional laboratory experiments?

Laboratory experiments may be feasible when field experiments appear difficult to achieve when the researcher cannot plausibly acquire the necessary control. When focusing specifically on the mechanism behind the effect (i.e., theory application), convenience samples, such as students, are often used to investigate a vast array of matters (Calder, Phillips, & Tybout, 1981; Summers, 2019), also in business research (Bello, Leung, Radebaugh, Tung, & Van Witteloostuijn, 2009). While it is not wrong to use student samples in many cases (e.g., when testing general principles, consumer preferences and behavior, or personality traits), it is wrong to use them as surrogates of the general population for other matters that they are not representative of (e.g., market investments, delinquency, or decision-making in the B2B sphere; Flere & Lavrič, 2008). B2B studies frequently rely on executive MBA students as their sampling frame, arguing that these students have relevant background and experience (e.g., Jap, Robertson, Rindfleisch, & Hamilton, 2013), to overcome the student sample limitations, many studies with the characteristics of laboratory experiments are now conducted with a non-student sample (i.e., experiments with increased behavioral realism in Table 1).

Laboratory studies might be used to test the mechanisms behind an effect with more control, increasing the extent of internal validity of a study and enhancing the theoretical contribution (Calder et al., 1981). Laboratory experiments have been used in B2B research (e.g., An, Kreutzer, & Heidenreich, 2020; Liang, Kale, & Cherian, 2014; Oh, Peters, & Johnston, 2014). As mentioned in the taxonomy table, however, laboratory experiments present generalizability issues that should be carefully discussed by researchers. A possible way to address these concerns is to run a complementary field study to see whether there is consistency of the findings when measuring actual behavior with the target population (e.g., McCoy & Hargie, 2007).

Why field experiments?

Field experiments are experiments that study the actual population in the actual context, integrating into what is already taking place. Being conducted in natural settings, they are representative of the target population, and they allow for measuring actual behavior. Some of the findings coming from the use of field experiments in marketing question previous established relationships. For instance, McCoy and Hargie (2007) extend previous research by investigating the effects of personalization and envelope color on response rate, speed, and quality, by using real behavioral data with members of the Public Relations of the Institute's Membership Handbook in Northern Ireland.

Compared to laboratory experiments, field experiments are often weak when the goal is understanding mechanisms behind the effect, and they often involve a significant loss of control over the experimental procedure. For instance, Gneezy (2017) and Putnam-Farr and Riis (2016) underline risks to the perfect randomization of participants and slight changes in incentives as possible hiccups that may challenge the smooth ongoing of the field experiment.

Benefits of combining field and lab experiments

When the aim of the field experiment is to test previously established theories with greater behavioral realism, one way to approach it may be to first conduct the experiment in a controlled setting like the laboratory, before investing resources into the field experiment. In this case, researchers can start with a laboratory experiment to first support the theoretical evidence and then check the external validity by going into the field. Alternatively, the logical approach could be to first conduct a field experiment to investigate the main effects of the independent variable on the dependent variable, and then follow up with a laboratory experiment to investigate the potential mechanisms underpinning the effect or other factors that can make the relationship between variables stronger or weaker, or even reverse (e.g., from a positive effect of the independent variable on the dependent variable to a negative one or the opposite).

In some cases, the nature of one's research question directly dictates the experimental context. For example, when trying to seed information about their brands and products, marketers may need to rely on social media influential endorsers. Hence, marketers may be faced with the challenge of identifying the most appropriate set of influencers to collaborate with. When considering this category of collaborators (e.g., Valsesia, Proserpio, & Nunes, 2020), real data from social media in combination with a controlled experiment may provide results that are more valid and generalizable.

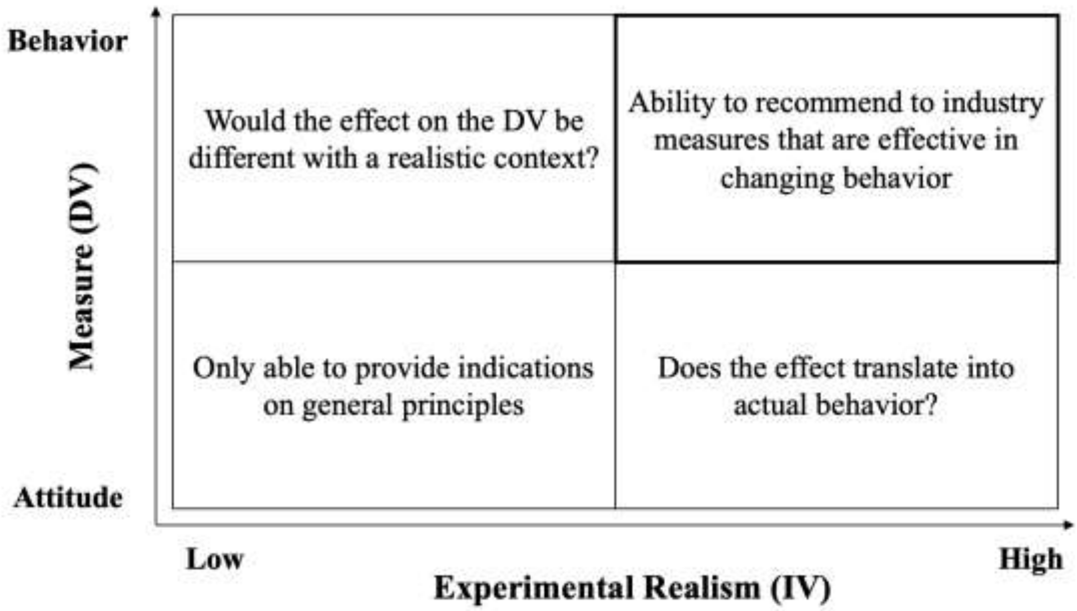
Improving Realism and Measuring Actual Respondents' Behavior

Experimental realism refers to the extent to which an experimental study reproduces a real behavioral situation. The experimental realism of a study ranges from very artificial to very realistic (Levitt & List, 2007), depending on the context used and on the choice of the independent variable. There is a growing emphasis on collecting data that are close to the real world and to the field of marketing (van Heerde, Moorman, Moreau, & Palmatier, 2021). Previous research suggests that the greater the inclusion of realism in the employed variables in an experiment, the greater the external validity and the generalizability of results (Morales, Amir, & Lee, 2017). This is consistent with Galizzi and Navarro-Martínez (2019) who show that, to have results that are externally valid, there is the need to test phenomena in natural settings. When the interaction between researcher and participant is mostly remote and anonymous, such as in online experiments, it is challenging to employ a realistic scenario, and it is usually easier to use a hypothetical one. However, this can be resolved by investing more efforts in the application of an experimental realism also online, having participants invest actual effort in making choices, searching, or acquiring extra information, for instance by using scenario-based manipulations and vignettes with managers and business decision-makers (e.g., Jap et al., 2013; Saab & Botelho, 2020; Tangpong, Li, & Hung, 2016). For instance, Saab and Botelho, (2020) manipulated purchase decision importance by creating vignettes that were randomly shown either portraying a high-importance purchase decision (i.e., the purpose of the good/service was administrative) or a low-importance purchase decision (i.e., the purpose of the good/service was customer use). Vignettes were used as manipulations of the independent variables and after being exposed to those, respondents could be assessed on the dependent variables (e.g., assessing functional risks of the decision).

Another important way to increase experimental realism is the use of technological tracking in field experiments. One example here is the study conducted by Ferguson and Mohan (2020). The authors employ eye-tracking monitoring to investigate the effects of influencers in B2B ads on managers' attention, ad recall, and ad attitudes, by featuring B2B advertisements to managers and tracking their eye movements. Actions could range from choosing to purchase an item to moving ones' eyes to a particular location (Morales et al., 2017). Behavior carries some consequences (e.g., social, financial, effort, time, self-efficacy) that extends beyond indicating one's thoughts about a given matter. It is easier and safer to draw more information on real behavior of respondents when researchers use dependent variables that are behavioral and consequential. Responding to a scale, or even declaring behavioral intentions in a hypothetical manner, does not fit this criterion. This is because reporting one's theories about behavior, emotions, or intended actions does not directly translate to actual behavior. This attitude-behavior gap phenomenon is defined as *"the differences between what people say and what people do"* (Blake, 1999, p.275). One possible explanation for this phenomenon is the tendency for survey respondents to answer questions in a manner that will be viewed favorably by others, the so-called social desirability bias (Grimm, 2010).

Figure 1 summarizes graphically the discussion on the two identified dimensions, experimental realism of the independent variable (IV) and behavioral measure of the dependent variable (DV), clarifying the concerns (or benefits) of each quadrant of the matrix.

Figure 1. Experimental realism of independent and behavioral measure of the dependent variable



For instance, one way to increase experimental realism online could be by increasing the interaction of participants with the stimuli (e.g., product, brand, company, different investment opportunities, strategic decisions they would make for the company, etc.), using a natural context, and incentivizing greater information seek and engagement. Examples of such techniques have been used in previous research (e.g., Barasch, Levine, Berman, & Small, 2014; Bruine de Bruin & Ulqinaku, 2020; Liang et al., 2014; Sarial-Abi, Vohs, Hamilton, & Ulqinaku, 2017; Ulqinaku, Sarial-Abi, & Kinsella, 2020).

There are some papers in B2B research where one aspect is optimal, and the other aspect has some room for improvement. For instance, on the one hand, Taylor, Hajmohammad, and Vachon (2020), respectively, have high experimental realism (IV is a vignette that presents a situation where respondents are "asked to assume a specific role and react to the information presented as if they themselves were in that situation;" Taylor et al., 2020, p. 4), but low on behavioral measures (DV just measuring intentions and recommendations and not actual behavior). On the other hand, Liang et al. (2014) make use of actual behavioral measures for the DV, introducing an element of reality into the experiment – making decisions to launch or not launch a new product development project – but presented a low experimental realism because it was conducted with MBA students and in a laboratory. Generalizing from these measures to strategy or policy implications still requires a bit of a leap of faith.

Given the resources needed to conduct field experiments, it is important to assure collaborations with partners in the industry for data collection, keeping in mind that smaller organizations usually provide faster and easier collaboration agreements (Gneezy, 2017). Choi, Sun, Liu, and Chen (2020), for instance, managed to collaborate with a small jewelry store in a local shopping mall to conduct a field experiment on how being promotion (vs. prevention) oriented can affect consumers' price choices. Another example can be found in Tsiros and Irmak (2020) where the authors got to collaborate with a lunch stand at a Farmers Market to investigate donations for a local elementary school with the purchase of every tuna bowl. Garrett and Gopalakrishna (2019) collaborated with an insurance company to conduct a field quasi-experiment for a period of 4 weeks, using life insurance as a reward for participants, in line with the collaborating company's products.

Methodological Considerations

Experiments present at least four methodological considerations to be addressed carefully, namely (1) the needed sample size; (2) the appropriate design to fit the research question (i.e., within, between or mixed); (3) manipulation checks; and (4) the moderation and mediation analyses to investigate mitigating factors and processes behind the effect, respectively.

Sample arrangements

The natural question that a marketing researcher would ask when designing an experimental study is: "*How many participants do I need for an experiment*?" The answer is not a magic number but depends on how large the treatment effect (i.e., how much the treatment moves the outcome) and on the standard deviation of the dependent variable.

The treatment effect is generally presented as an effect size, i.e., the extent of the difference of the variable object of investigation between groups. This measure is often presented also showing the outcome's standard deviation. The larger the treatment effect is, the fewer people are needed in the experiment. The way to know the treatment effect is to run a pilot study before running the actual experiment and comparing the mean outcome in the treatment group versus the control group (Viglia & Dolnicar, 2020).

Before presenting a tool to calculate the ideal sample size for a study, there is the need to summarize the probability of making errors in hypotheses testing. There are two main types of errors: Type 1, which refers to the risk of incorrectly rejecting the null hypothesis and claiming that the means of the treatment and control conditions are significantly different (when they are not), and Type 2, which refers to the risk of incorrectly not rejecting the null hypothesis and claiming that the means of the treatment and control conditions are not significantly different (when they are). The probability of making Type 1 error is referred to as α and probability of making Type 2 error is referred to as β. 1- β is the statistical power of the experiment.-Previous research has defined statistical power as "the probability that its null hypothesis (H0) will be rejected given that it is in fact false" (Faul, Erdfelder, Lang, & Buchner, 2007, p. 175).

Given that lack of statistical power decreases the possibility to reliably discriminate between the null hypothesis and the alternative hypothesis, ensuring that a study is sufficiently powered is of great importance (Faul et al., 2007). Here we present an empirical tool (among others available) used to calculate the necessary sample size for running an experiment. G*Power is a power analysis program and a standalone software developed by Erdfelder, Faul, and Buchner (1996) for statistical tests commonly used also in marketing behavioral research. The utility of G*Power is that it calculates the necessary statistical power based on a series of frequently used tests such as *t, F, z,* χ^2 or binomial reference distributions. G*Power[1] offers a free design-based approach where the researcher (1) selects the category of statistical test that one is interested in (e.g., correlations, means, proportions, variances, regressions), and (2) specifies the design characteristics (e.g., number of groups, whether the samples are independent or dependent, number of controls, expected level of power, etc.). At the specification of the design of the study, G*Power estimates a sample size that would provide the requested statistical power.

Hence, the necessary sample size depends on i) the suggested significance level and power, ii) the variance of outcomes and iii) the effect size between conditions. With the same treatment effect, a higher or lower sample dimension will make a null hypothesis rejected or not. When specifying the sample size, the researcher needs to take into consideration also common issues with sample arrangements: (1) *over-coverage in the sampling frame* (which occurs when the sample includes units that should not be part of the population, e.g., requesting only US participants on an online platform and users with VPN from non-US countries participating), (2) *under-coverage in the sampling frame* (which occurs when the sample does not include units that should be part of the population, e.g., failing to capture a representation of a portion of the population), and (3) *non-response bias* (which occurs when units that are supposed to be included, are not included for some reason, e.g., failing to get a portion of the sample because the invitation to take part in the study ends up in their spam;

[1] (https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html)

Mittal, 2019). These issues can be solved by including more demographic screening, different types of incentives, or by acknowledging these limitations in the study (Mittal, 2019).

Between, within or mixed design?

There are two main typologies of experimental designs: between-subjects design and within-subjects design. In the case of the between-subject design (e.g., Seggie, Griffith, & Jap, 2013, Study 2), participants are exposed to only one of the conditions (treatment or control). In the case of within-subject design, participants are exposed to all conditions. A between-subjects design compares differences between subjects who were exposed to different stimuli, while a within-subjects design focuses on changes among the same set of respondents from before and after the exposure to stimuli. Sometimes, these designs are combined, resulting in mixed designs (e.g., Dean, Griffith, & Calantone, 2016; Seggie et al., 2013, Study 1), where participants may be exposed to only one of the conditions of one factor (the between-subjects element), but all conditions of the other factor (the within-subjects element).

There are pros and cons behind these designs. While for within-subjects design the sample size required is smaller, and the probability of grasping true differences among conditions is greater, the setup and the analyses are easier for between-subjects designs instead. Moreover, employing a within-subjects design would give more statistical power to the study because we are not interested in the portion of error attributable to individual differences among participants in this type of design, but the risk is that exposing participants to all levels of the same factor may result in them understanding the purpose of the experiment and providing biased responses. A solution to this is the randomization of the order of the treatment conditions so that different participants are exposed to the conditions in a different order. For instance, An et al. (2020) have randomized the order of the scenarios that were shown within respondents. In employing a within-subjects design, the authors exposed every participating team to all three conditions (control condition, treatment 1, treatment 2) in a random order. This part of the experiment, given its characteristics, was a within-subject design. However, in this study, the authors combined a within-design experiment with a between-subject one. Specifically, in their research, the authors use a two factors design, with one of the factors being within-subject and the other one between. Specifically, the authors first randomly assign each of the teams to different simulations (team cooperation, inter-team competition, intra-team competition), which is the within-subject factor, and then randomly assign the teams to either high or low organizational identification scenario, which is the between-subject factor. Employing mixed-designs (e.g., Seggie et al., 2013, Study 1) provides benefits related both to within-subjects – greater statistical power – and to between-subjects – less learning and order effects.

Table 2. Advantages of different experimental designs

Experimental design	**Advantages**
Between-subject design	- Easier experimental setup - Simpler experimental data analysis - Lower risk of participants understanding the purpose of the experiment and providing biased responses - Shorter experimental sessions required
Within-subject design	- Smaller sample size required - Greater probability of grasping true differences among conditions (less noise) - Greater statistical power to the study - Greater alignment with most marketing theoretical mindsets
Mixed design	- Greater statistical power - Less learning effects - Less order effects

Manipulation checks

Manipulation checks are questions that are used to make sure that the treatment was perceived as intended. They are usually operated as quantitative questions asking for self-reported answers after the exposure of participants to the manipulation (Ejelöv & Luke, 2020). These answers (i.e., the manipulation checks) provide confidence that the effects on the dependent variables are due to the manipulation (e.g., Dean et al., 2016; Seggie et al., 2013; Taylor et al., 2020). If the majority of the respondents does not perceive the manipulation as intended, then the design of the study is flawed.

Ideally, researchers might look at manipulation checks previously validated in the literature, but it can also be easily possible to present a completely new manipulation check, especially when the variable to be manipulated is an objective variable.

While previous researchers have raised concerns regarding the effectiveness of manipulation checks from a quantitative point of view and regarding the possibility of creating demand effects among participants (Fiedler, McCaughey, & Prager, 2021), manipulation checks, if applied correctly, are highly beneficial to experimental researchers in understanding the effect of the independent variable on the dependent variable (i.e., presence of causality; Ejelöv & Luke, 2020; Fiedler et al., 2021).

Another aspect to consider related to manipulation checks is their positioning in the study, regarding (1) their inclusion in the main study (with the risk of overcomplicating the study) or in a pilot study, and (2) their inclusion before or after the dependent variable assessment (with the risk of creating demand effects; Kidd, 1976; Mills, 1969; Parro, & Hertel, 1999). One solution to the concerns regarding where to position manipulation checks may be to run separate pre-tests before conducting the main study (Hauser, Ellsworth, & Gonzalez, 2018; Kidd, 1976). However, the inclusion of manipulation checks in the main study remains crucial especially when there is the concern of confounds, as there is no guarantee that an experimental treatment will only manipulate the focal independent variable, without varying other variables too (Fiedler et al., 2021).

It is at the consideration of the researcher whether these questions are asked before or after the dependent variable, based on the theoretical expectation on the effects of manipulation checks over the study and on the dependent variable. Importantly, manipulation checks used in each study should always be reported, possibly be drawn from previous research – and if impossible, adapted – and should possibly be operationalized using multi-item measures (Pechmann, 2019).

Attention checks

While manipulation checks serve the purpose of understanding whether the manipulation worked as intended, instructional attention checks - often called attention checks (Paas, Dolnicar, & Karlsson, 2018) - help to see whether participants are paying attention to the provided instructions. Attention checks vary from reverse scaled questions, to more content-related questions, to timing response times in online-operated studies (Abbey & Meloy, 2017). Abbey and Meloy (2017) mainly classify attention checks into logical statements (e.g., preferring to eat fruit vs. paper), directed queries (e.g., For this query, mark X [insert X] and move on.), open-ended queries (e.g., Please summarize what was written in the scenario you just read), infrequency (e.g., asking several times the date of birth and checking for consistency), response time (e.g., evaluating efficiency and accuracy based on how long it took to the participant to answer), memory recall (e.g., please write down the name of the brand you were shown in the scenario before), and reverse scaling (e.g., changing the direction of the agreement from I agree to I do not agree with the statement). Researchers should consider using multiple attention checks especially if the study is lengthy (Abbey & Meloy, 2017).

For instance, Saab and Botelho (2020) used two questions to assess the attention of their respondents. First, to assess their attention in general, they asked their participants to indicate the product they had been informed about in the manipulation text. Moreover, as a further attention check, the authors asked participants to calculate the price difference of the two competing offerings for the products they read about in the manipulation text.

Again, it is of absolute importance to disclose the specific attention checks that were used in a study and if exclusion of participants based on attention checks significantly changes the results of the study (Abbey & Meloy, 2017). It is ideal to decide on the criteria for participants' exclusion before the study has been conducted, in order to diminish the risk of false positives (e.g., by preregistering a study, see Van't Veer & Giner-Sorolla, 2016).

To summarize, manipulation checks are to be treated differently from the attention check; the former one aims to assess the validity of the manipulation and the latter the attention that is being put into the study (Ejelöv & Luke, 2020; Sigall & Mills, 1998).

Analyzing experimental data in marketing

The first aim of researchers employing an experimental approach is that of comparing groups for differences. This allows them to suggest that a difference in the mean of the dependent variable between different groups can be a sign in the direction of causation claims. Hence, among the first analyses that researchers apply when investigating experimental data would be a comparison of means or frequencies between the experimental groups. For instance, a typical analysis is that of variance, also known as ANOVA, and defined as “a statistical technique utilizing an F ratio to determine if an independent variable has a statistically significant effect on a dependent variable" (Picardi & Masick, 2014, p. 105).[2]

There are two additional common aims of experiments, besides testing for main effects of the independent variable on the dependent variable: i) testing for moderation (also referred to as boundary conditions) and ii) testing for mediation. A moderator is a variable that strengthens or weakens an existing established relationship while a mediator is a variable that clarifies the mechanism behind that relationship. While the researcher may find empirically evidence for a mediator explaining why a relationship between two variable holds, this cannot rule out that there are no other possible explanations for why this relationship holds. One way to ensure that the effects have not been driven by confounds is the use of manipulation checks, as we have explained in the prior sections of this manuscript. Another solution is the measurement of those constructs that the researcher expects to possibly explain the relationship between the independent and the dependent variable. Measuring the constructs that present alternative explanations for why a relationship may hold will allow the researcher to test and eventually rule out their role as mechanisms in the predicted causality between the variables.

For instance, a study might want to measure if the effect of the manipulated independent variable is stronger or weaker depending on individual differences between respondents (e.g., personality traits, age, and gender) or some contextual factors (e.g., time of the day, brand positioning, etc.). This variable takes the name of moderator. Taylor et al. (2020) investigate the role of target decision legitimacy as a moderator in the relationship between target decision and activists' recommendation legitimacy, and in the relationship between activists' recommendation legitimacy and observer adoption. Dean et al. (2016) investigate the role of situation-specific factors relevant to new product introductions as

[2] For an overview of analyses of experiments with one or two factors (i.e., one-way and two-way ANOVAs), please refer to Appendix 2.

moderator instead. In other cases, the moderator can also be investigated as affecting another interaction effect, as it happens for instance in the study from Griffith, Hoppner, Lee and Schoenherr (2017).

To test for this statistically, we expect an interaction between the independent variable and the proposed moderator. For this reason, a moderator is often called an interaction effect (Das, J., & Dirienzo, 2010; Imbens & Wooldridge, 2009). The moderator can affect i) the main relationship between the independent and the dependent variable, ii) the relationship between the independent variable and the mediator, iii) the relationship between the mediator and the dependent variable, or iv) multiple relationships. The position of the moderator on the conceptual model should be guided by previous theoretical knowledge.

When both the independent variable and the moderator are categorical, the researcher can use ANOVAs to analyze the differences in means between conditions and the overall interaction effect. However, there are two additional analyses often applied when one of the variables in the interaction is continuous and the other is categorical: spotlight analysis and floodlight analysis. The spotlight analysis provides an estimate of the effect of one of the variables in the interaction – the categorical one – at specific values of the other variable in the interaction – the continuous one (Spiller, Fitzsimons, Lynch, & McClelland, 2013). This is particularly relevant when there are specific points in the continuous variable that are of interest to the researcher. The floodlight analysis provides an estimate of the effect of the categorical variable in the interaction at all values of the continuous variable (Spiller et al., 2013). This is relevant when the aim is to spot the areas where the interaction is significant and where it is not (for more details on spotlight and floodlight analyses, see Krishna, 2016; Spiller et al., 2013). In Table S1 in Appendix 1 and in Appendix 3, we provide guidance on how to run these analyses with STATA or SPSS.

A mediator is a concept that is different from that of the moderator. The mediator does not affect the strength of the relationship between two variables, it tries to explain why a change in the treatment may affect the outcome. Seggie et al. (2013) investigate the role of transaction cost as one potential mediator between opportunism form and satisfaction with the performance of the relationship. Pieters (2017), Rucker, Preacher, Tormala, and Petty (2011), and Zhao, Lynch, and Chen (2010) offer important insights and guidelines into meaningful mediation testing in experiments. The PROCESS macro for SPSS provides an efficient means to test for mediation effects using a bootstrapping approach (Hayes, 2017). In Appendix 4, we present guidelines on how to conduct mediation testing using SPSS. It is worth mentioning that mediation can take other forms, such as parallel mediation (when multiple mechanisms are tested contemporarily) or serial mediation (when one mechanism leads to another and so on).

An additional important aspect of the combination of mediator and moderator in a model is what Hayes (2013) refers to as conditional process analysis. This analysis is used when the researcher is investigating conditions under which the relation between the independent variable affects the dependent variable via a mediator. Hence, it links together into a single integrated analytical model both the mediator and the moderator. The author provides explanations with examples from past literature where the conditional effect (i.e., moderation) happens in the a-path of the model (i.e., between the independent variable and the mediator), in the b-path (i.e., between the mediator and the dependent variable), or both in the a-path and the b-path. The conditional effect can also occur in the direct effect between the independent and the dependent variable (i.e., c-path). In this case, the indirect effect is unconditional because, being it a product of the a-path and b-path unconditioned effects, it remains unconditional when the moderation occurs in the c-path. In the other cases (i.e., moderation occurring at least in the a-path, b-path, both a-path and b-path), we are dealing with a conditional indirect effect. If it occurs both in the c-path and any of the indirect links between the independent variable and the dependent one, the effect is considered to be conditional both at the indirect and direct effect (Hayes, 2013).

When analyzing moderation and mediation models, presenting the conceptual model helps the reader in understanding the relationships predicted and to be tested. This usually allows the researcher to also tell apart the moderator from the mediator, given that a moderator can still be sometimes confused with a mediator (Hayes, 2009). For more mediation models and guidelines, refer to Kirk (2013) and Hayes (2017).

General Discussion and Conclusion

Experiments remain one of the main methodological approaches in marketing and social sciences for testing causality between an independent and a (or many) dependent variable(s). To conclude that the treatment causes a specific variation in the dependent variable when applying an experimental approach, the researcher has to ensure that the requirements of the experiment are being carefully respected and attentively executed. While experiments have been greatly used in consumer-related research in marketing, their use in B2B remains behind and sometimes falls short of the applicability of this method. Most papers discuss the external validity and generalizability of their results, predicting real behavior, but they do so by basing their discussion on findings of surveys. However, both editors and reviewers of the B2B and Industrial Marketing Management (IMM) field, have been increasingly suggesting and requesting experimental approaches to enrich the internal validity of findings. That is because some type of questions – such as short-term and long-term implications of findings, boundary conditions, what are the effects of the treatment on the outcome and which players are the most affected – would have been hard to test in a hypothetical scenario with possible questions on the veracity of the findings also because of possible social desirability bias concerns.

In this article, we provide a short guide on how to conduct experiments in marketing research, borrowing from the procedures and guidelines provided from the other marketing areas. Specifically, we guide the researcher from the reasons behind choosing an experimental approach to investigate the conceptual model, to the types of available experiments and when and why each of them should be chosen. Moreover, we provide some guidelines on how to design experiments, explaining the main necessary criteria that should be respected when conducting experiments, and we suggest how the sample size should be determined. Finally, we provide some short guidelines on how to analyze different types of data resulting from experiments.

We see several benefits in applying experiments in the business marketing field. Combining experimental designs and approaches in a way to enhance behavioral realism will lead to greater robustness of findings and greater external validity of them too. Especially now, in the digital world, randomized experiments can be cheap and fast. It is possible to write a line of code to randomly assign participants to one group and it is not needed for users to fill surveys or even tell users that they are part of an experiment.

Combining data types for converging evidence and carefully considering the needed sample size also showing the effect size is a good way to proceed. Moreover, combining different experimental approaches in the empirical package can help to increase the extent of behavioral realism, documenting the phenomenon and some potential mechanism behind the effect in a laboratory or online experiment, and then enriching the external validity by running a field study (e.g., Steward, Narus, & Roehm, 2018).

Important empirical questions would ideally be addressed using at least a couple of toolboxes (e.g., a lab plus a field study; some natural data plus a lab study; a qualitative approach plus an experiment). Some possible examples in B2B can be found in Seggie et al. (2013) or in Steward and colleagues (2018), where the authors complement field experiments with other methods, such as in-depth interviews in Steward et al. (2018) or with longitudinal data in Seggie et al. (2013). In general, in the era of an enormous amount of available data, experiments are useful to draw causal relationships and understand “what is causing what”. Table 3 provides a short checklist for authors and reviewers.

To conclude, controlled experiments that are well-planned and executed have a practical bent. They help us learn interventions that work and that do not under different conditions, offering clear insights to business marketing researchers and practitioners.

Table 3. A checklist for authors and reviewers running or evaluating experimental studies in Marketing

	Question	**Check ✓**
RELEVANCE	Are authors addressing a real marketing problem?	
VALIDITY	Does the chosen sample allow for internal/external validity?	
RIGOR	Have authors come up with a clear design that rules out possible alternative explanations for the effects?	
REPLICABILITY	Is the design clearly explained so that it can be replicated?	
IMPACT	Is the dependent variable measuring actual behavior or - at the very least - showing some behavioral realism?	
INFORMATIVENESS	Have authors included detailed analyses (going beyond p-value by showing effect sizes)?	
ILLUSTRATIVENESS	Have authors included an adequate visual representation of the results?	

References

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53*, 63-70.

Allen, M. (2017). *The sage encyclopedia of communication research methods* (Vols. 1-4). Thousand Oaks, CA: SAGE Publications, Inc doi: 10.4135/9781483381411

An, D., Kreutzer, M., & Heidenreich, S. (2020). Always play against par? The effect of inter-team coopetition on individual team productivity. *Industrial Marketing Management*, *90*, 155-169.

Anderson, E. T., & Simester, D. (2011). A step-by-step guide to smart business experiments. *Harvard Business Review*, *89*(3), 98-105.

Barasch, A., Levine, E. E., Berman, J. Z., & Small, D. A. (2014). Selfish or selfless? On the signal value of emotion in altruistic behavior. *Journal of Personality and Social Psychology*, *107*(3), 393-413.

Bello, D., Leung, K., Radebaugh, L., Tung, R. L., & Van Witteloostuijn, A. (2009). From the editors: Student samples in international business research. *Journal of International Business Studies*, *40*(3), 361–364.

Bendixen, M., Bukasa, K. A., & Abratt, R. (2004). Brand equity in the business-to-business market. *Industrial Marketing Management*, *33*(5), 371-380.

Blake, J. (1999). Overcoming the ‘value-action gap’in environmental policy: Tensions between national policy and local experience. *Local Environment, 4*(3), 257-278.

Bruine de Bruin, W., & Ulqinaku, A. (2020). Effect of mortality salience on charitable donations: Evidence from a national sample. *Psychology and Aging*.

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

Carden, S. W., Holtzman, N. S., & Strube, M. J. (2017). CAHOST: An excel workbook for facilitating the Johnson-Neyman technique for two-way interactions in multiple regression. *Frontiers in Psychology*, *8*, 1293.

Carrington, M. J., Neville, B. A., & Whitwell, G. J. (2014). Lost in translation: Exploring the ethical consumer intention–behavior gap. *Journal of Business Research*, *67*(1), 2759-2767.

Choi, W. J., Sun, H., Liu, Y., & Chen, H. (2020). Guess Who Buys Cheap? The effect of consumers' goal orientation on product preference. *Journal of Consumer Psychology*, *30*(3), 506-514.

Cuervo-Cazurra, A., Andersson, U., Brannen, M. Y., Nielsen, B. B., & Reuber, A. R. (2016). From the editors: Can I trust your findings? Ruling out alternative explanations in international business research, *Journal of International Business Studies*, *47*, 881–897.

Das, J., & Dirienzo, C. (2010). Tourism competitiveness and corruption: A cross-country analysis. *Tourism Economics*, *16*(3), 477-492.

Dean, T., Griffith, D. A., & Calantone, R. J. (2016). New product creativity: Understanding contract specificity in new product introductions. *Journal of Marketing*, *80*(2), 39-58.

Ejelöv, E., & Luke, T. J. (2020). "Rarely safe to assume": Evaluating the use and interpretation of manipulation checks in experimental social psychology. *Journal of Experimental Social Psychology*, *87*, 103937.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 1-11.

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

Ferguson, J. L., & Mohan, M. (2020). Use of celebrity and non-celebrity persons in B2B advertisements: Effects on attention, recall, and hedonic and utilitarian attitudes. *Industrial Marketing Management*, *89*, 594-604.

Fiedler, K., McCaughey, L., & Prager, J. (2021). Quo Vadis, Methodology? The Key Role of Manipulation Checks for Validity Control and Quality of Science. *Perspectives on Psychological Science*, 1745691620970602.

Flere, S., & Lavrič, M. (2008). On the validity of cross-cultural social studies using student samples. *Field Methods*, *20*(4), 399-412.

Galizzi, M. M., & Navarro-Martínez, D. (2019). On the external validity of social preference games: a systematic lab-field study. *Management Science*, *65*(3), 976-1002.

Garrett, J., & Gopalakrishna, S. (2019). Sales team formation: The right team member helps performance. *Industrial Marketing Management*, *77*, 13-22.

Gneezy, A. (2017). Field experimentation in marketing research. *Journal of Marketing Research*, *54*(1), 140-143.

Griffith, D. A., Hoppner, J. J., Lee, H. S., & Schoenherr, T. (2017). The influence of the structure of interdependence on the response to inequity in buyer–supplier relationships. *Journal of Marketing Research*, *54*(1), 124-137.

Grimm, P. (2010). Social desirability bias. *Wiley international encyclopedia of marketing*.

Hada, M. (2021). Two Steps: A Primer on B2B Experiments. *Impact at JMR*, (January), Available at: https://www.ama.org/2021/01/26/two-steps-a-primer-on-b2b-experiments/

Hauser, D. J., Ellsworth, P. C., & Gonzalez, R. (2018). Are manipulation checks necessary?. *Frontiers in Psychology*, *9*, 998.

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

Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. Guilford Publications.

Hayes, A. F. (2017). Introduction to mediation, moderation, and conditional process analysis second edition: A regression-based approach.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5-86.

Jap, S. D., Robertson, D. C., Rindfleisch, A., & Hamilton, R. (2013). Low-stakes opportunism. *Journal of Marketing Research*, *50*(2), 216-227.

Kahneman, D., & Egan, P. (2011). *Thinking, fast and slow* (Vol. 1). New York: Farrar, Straus and Giroux.

Kidd, R. F. (1976). Manipulation checks: advantage or disadvantage? *Representative Research in Social Psychology, 7*, 160-165.

Kirk, R.E. (1995). *Experimental Design: Procedures for the Behavioral Sciences* (3rd edn.). Pacific Grove, CA: Brooks/Cole.

Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences*. Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781483384733

Krishna, A. (2016). A clearer spotlight on spotlight: Understanding, conducting and reporting. *Journal of Consumer Psychology*, *26*(3), 315-324.

Laursen, L. N., & Andersen, P. H. (2016). Supplier involvement in NPD: a quasi-experiment at Unilever. *Industrial Marketing Management*, *58*, 162-171.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world?. *Journal of Economic Perspectives*, *21*(2), 153-174.

Liang, B., Kale, S. H., & Cherian, J. (2014). Is the future static or dynamic? The role of culture on escalation of commitment in new product development. *Industrial Marketing Management*, *43*(1), 155-163.

McCoy, M., & Hargie, O. (2007). Effects of personalization and envelope color on response rate, speed and quality among a business population. *Industrial Marketing Management*, *36*(6), 799-809.

Mills, J. (1969). *Experimental Social Psychology*. Toronto: MacMillian.

Mittal, V. (2019). *How to get published in the best marketing journals*. Edward Elgar Publishing.

Morales, A. C., Amir, O., & Lee, L. (2017). Keeping it real in experimental research—Understanding when, where, and how to enhance realism and measure consumer behavior. *Journal of Consumer Research*, *44*(2), 465-476.

Oh, J. H., Peters, L. D., & Johnston, W. J. (2014). Who's acquiring whom?—Experimental evidence of firm size effect on B2B mergers and marketing/sales tasks. *Industrial Marketing Management*, *43*(6), 1035-1044.

Paas, L. J., Dolnicar, S., & Karlsson, L. (2018). Instructional Manipulation Checks: A longitudinal analysis with implications for MTurk. *International Journal of Research in Marketing*, *35*(2), 258-269.

Parro, W. G., & Hertel, P. (1999). Research methods in cognition and emotion. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 61-81). Chichester, England: Wiley.

Pechmann, C. (2019). *How to get published in the best marketing journals*. Edward Elgar Publishing.

Picardi, C. A., & Masick, K. D. (2014). Research methods. *Designing and conducting*. SAGE Publications.

Pieters, R. (2017). Meaningful mediation analysis: Plausible causal inference and informative communication. *Journal of Consumer Research*, *44*(3), 692-716.

Putnam-Farr, E., & Riis, J. (2016). "Yes/No/Not Right Now": Yes/No Response Formats Can Increase Response Rates Even in Non-Forced-Choice Settings. *Journal of Marketing Research*, *53*(3), 424-432.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, *5*(6), 359-371.

Ruiz, C. A. D., & Kowalkowski, C. (2014). Market representations in industrial marketing: Could representations influence strategy?. *Industrial Marketing Management*, *43*(6), 1026-1034.

Saab, A. B., & Botelho, D. (2020). Are organizational buyers rational? Using price heuristics in functional risk judgment. *Industrial Marketing Management*, *85*, 141-151.

Sarial-Abi, G., Vohs, K. D., Hamilton, R., & Ulqinaku, A. (2017). Stitching time: Vintage consumption connects the past, present, and future. *Journal of Consumer Psychology*, *27*(2), 182-194.

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

Seggie, S. H., Griffith, D. A., & Jap, S. D. (2013). Passive and active opportunism in interorganizational exchange. *Journal of Marketing*, *77*(6), 73-90.

Sigall, H., & Mills, J. (1998). Measures of independent variables and mediators are useful in social psychology experiments: But are they necessary? *Personality and Social Psychology Review*, *2*, 218-226.

Simester, D. (2017). Field experiments in marketing. In *Handbook of economic field experiments* (Vol. 1, pp. 465-497). North-Holland.

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

Steward, M. D., Narus, J. A., & Roehm, M. L. (2018). An exploratory study of business-to-business online customer reviews: external online professional communities and internal vendor scorecards. *Journal of the Academy of Marketing Science*, *46*(2), 173-189.

Summers, J. O. (2019). Guidelines for conducting research and publishing in marketing: From conceptualization through the review process. In *How to Get Published in the Best Marketing Journals*. Edward Elgar Publishing.

Tangpong, C., Li, J., & Hung, K. T. (2016). Dark side of reciprocity norm: Ethical compromise in business exchanges. *Industrial Marketing Management*, *55*, 83-96.

Taylor, K. M., Hajmohammad, S., & Vachon, S. (2020). Activist engagement and industry-level change: Adoption of new practices by observing firms. *Industrial Marketing Management*.

Tsiros, M., & Irmak, C. (2020). Lowering the Minimum Donation Amount Increases Consumer Purchase Likelihood of Products Associated with Cause-Related Marketing Campaigns. *Journal of Marketing Research*, 0022243720916463.

Ullah, S., Akhtar, P., & Zaefarian, G. (2018). Dealing with endogeneity bias: The generalized method of moments (GMM) for panel data. *Industrial Marketing Management*, *71*, 69-78.

Ulqinaku, A., Sarial-Abi, G., & Kinsella, E. L. (2020). Benefits of heroes to coping with mortality threats by providing perceptions of personal power and reducing unhealthy compensatory consumption. *Psychology & Marketing*, *37*(10), 1433-1445.

Valsesia, F., Proserpio, D., & Nunes, J. C. (2020). The Positive Effect of Not Following Others on Social Media. *Journal of Marketing Research*, 0022243720915467.

van Heerde, H. J., Moorman, C., Moreau, C. P., & Palmatier, R. W. (2021). Reality Check: Infusing Ecological Value into Academic Marketing Research. *Journal of Marketing*, *85*(2), 1–13. https://doi.org/10.1177/0022242921992383

Van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2-12.

Viglia, G., & Dolnicar, S. (2020). A review of experiments in tourism and hospitality. *Annals of Tourism Research*, *80*, 102858.

Zaefarian, G., Kadile, V., Henneberg, S. C., & Leischnig, A. (2017). Endogeneity bias in marketing research: Problem, causes and remedies. *Industrial Marketing Management*, *65*, 39-46.

Zellmer-Bruhn, M., Caligiuri, P., & Thomas, D. C. (2016). From the editors: Experimental designs in international business research, *Journal of International Business Studies*, *47*, 399–407.

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

Zimmer, M., Salonen, A., & Wangenheim, F. V. (2020). Business solutions as market signals that facilitate product sales. *Industrial Marketing Management*, *91*, 30-40.

Appendix 1

Table S1.

Generic STATA Codes for Experimental Testing

Tests	STATA Codes
One-Way ANOVA	oneway *DV IV*, tabulate
One-Way ANOVA (Post hoc testing)[1] mean comparison	pwmean *DV*, over[*IV*], mcompare(LSD) effects
One-Way ANOVA (Post hoc testing) contrast comparison	pwcompare *DV*, over[*IV*], mcompare(LSD) effects
One-Way MANOVA	manova *DV1 DV2 DV3 DV4 = IV* mvreg
Two-Way Between-Subjects ANOVA	anova *DV IV1##IV2*
Two-Way Between-Subjects ANOVA (contrasts)[2]	anova DV IV1##IV2 contrast ar.*IV1@IV2* contrast ar.*IV2@IV1*
Within-Subjects ANOVA[3]	anova *DV IV1 IV2*, repeated(*IV2*)
Spotlight analysis[4]	reg *DV IV1##IV2* margins, dydx(*IV1*) at(*IV2* =("*value 1 of IV2*" "*value 2 of IV2*")) atmeans
Floodlight analysis[5]	reg *DV IV1##IV2* margins, dydx(*IV1*) at(*IV2* =("*min value of IV2*" ("*unit change*") "*max value of IV2*")) atmeans

[1] *we have assumed LSD approach, but other approaches can be specified here, e.g., tukey, bonferroni, etc.*

[2] *here we are assuming IV1 is between-subjects and IV2 is repeated within-subjects*

[3] *first line: to compare differences in marginal means of all levels of IV1 at IV2*
second line: to compare differences in marginal means of all levels of IV2 at IV1

[4] *e.g., investigating the effect for moderator values 0 and 1, the 2nd part of the code would be:*
margins, dydx(IV1) at(IV2 =(0 1)) atmeans

[5] *e.g., investigating the effect for moderator that takes value 1 to 20, with 1 unit increase, the 2nd part of the code would be:*
margins, dydx(IV1) at(IV2 =(1 (1) 20)) atmeans

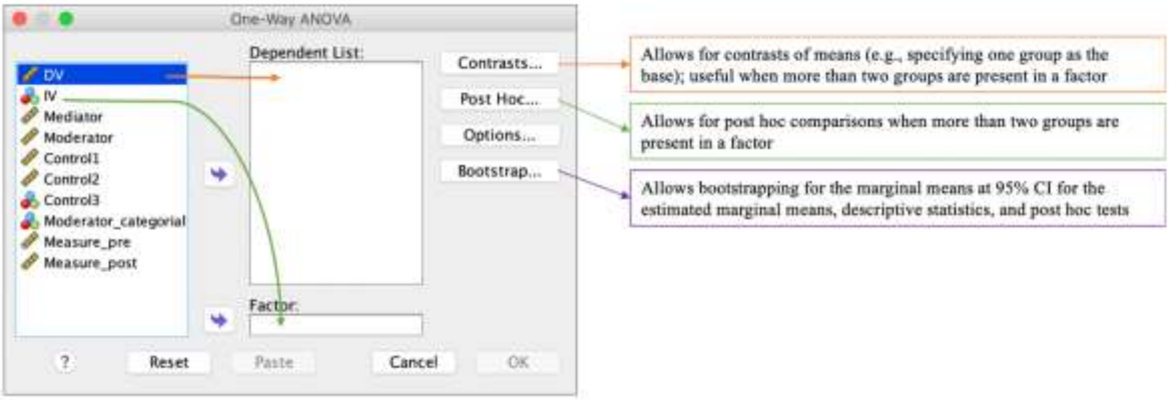
Appendix 2

How to analyze experimental data with one-way and two-way ANOVA

- **One-way ANOVA**

To test for mean comparison using a One-way ANOVA in SPSS the steps to follow are these:
Analyze – Compare Means – One-way ANOVA

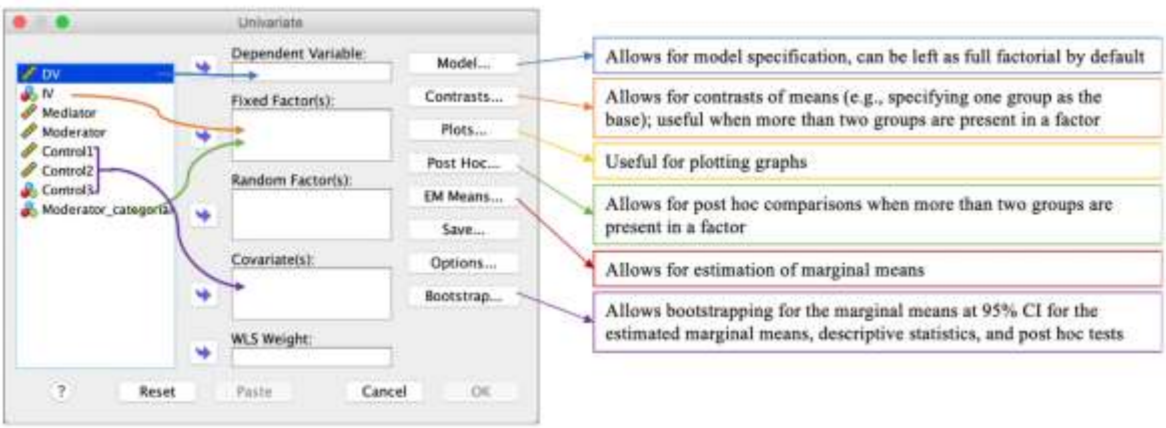
Step 1 **Step 2**



- **Two-way ANOVA**

To test a linear model for independent **between-subjects** factorial design in SPSS the steps to follow are these:
Analyze – General Linear Model – Univariate

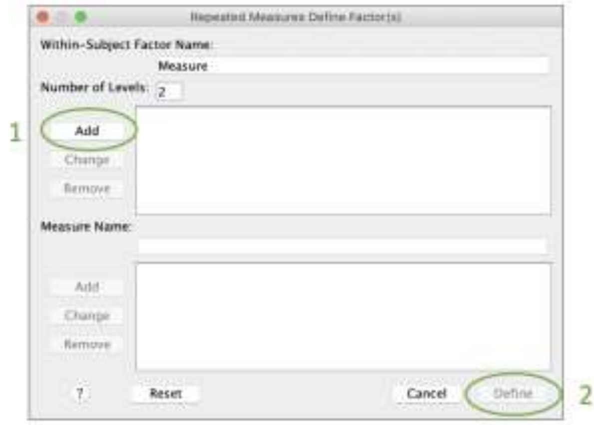
Step 1 **Step 2**



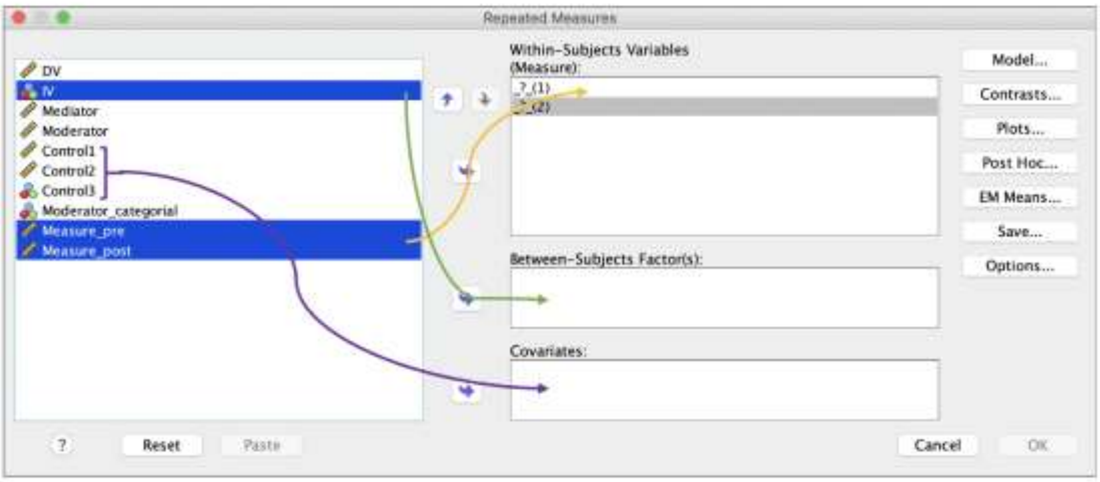
To test a model with a **repeated model** design in SPSS the steps to follow are these:
Analyze – General Linear Model – Repeated

We assume we have a repeated measure that we call Measure_pre and Measure_post, so it was measure before and after, for each participant. This variable has, hence, 2 levels (pre and post). Below is how we would set this up on SPSS:

Step 1



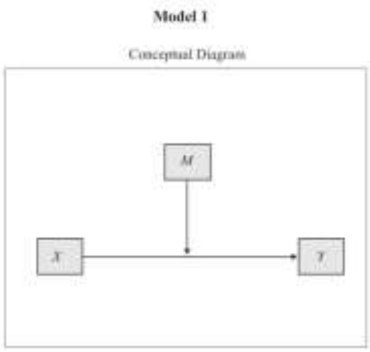
Step 2

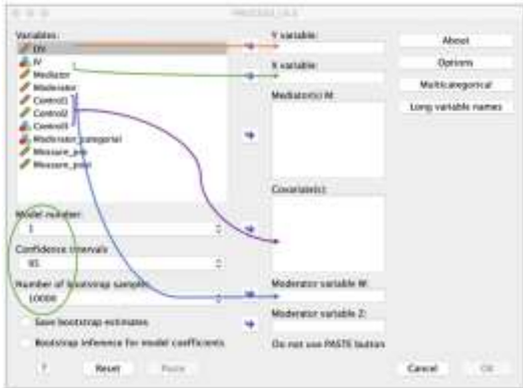


Appendix 3

- **Moderation: floodlight using PROCESS**

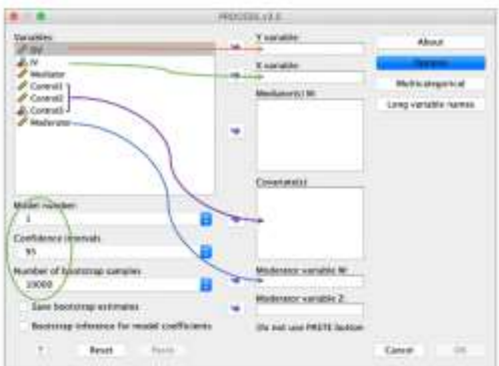
The PROCESS macro at SPSS allows for moderation testing both when variables are categorial, and for continuous and categorical variables combinations. For example, here, we've tested the Model 1 of the PROCESS macro, which predicts a simple moderation.



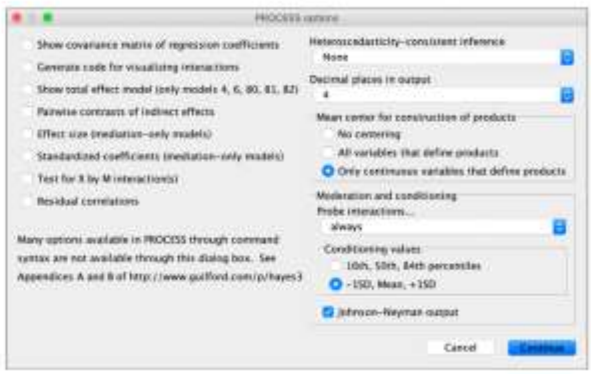


Moreover, the PROCESS macro at SPSS allows for the floodlight analysis when testing for any model that includes a moderation. For example, we've again tested Model 1 of the PROCESS macro, but it can be applied to any model including moderation (e.g., Model 7, 8, 14 of the PROCESS macro, etc.).

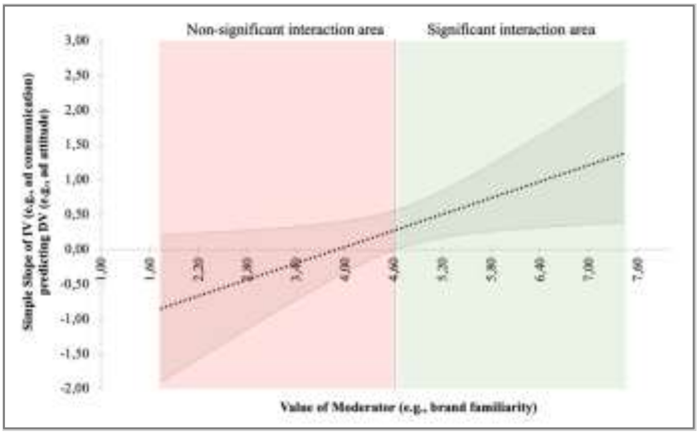
Step 1



Step 2



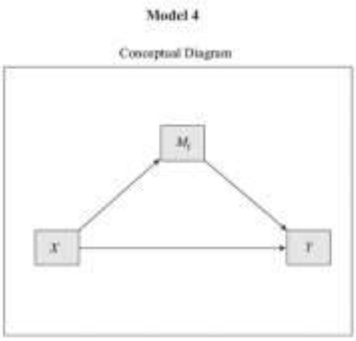
Here's an example of the presentation of the floodlight analysis results using the macro for excel provided by Carden, Holtzman, and Strube (2017):

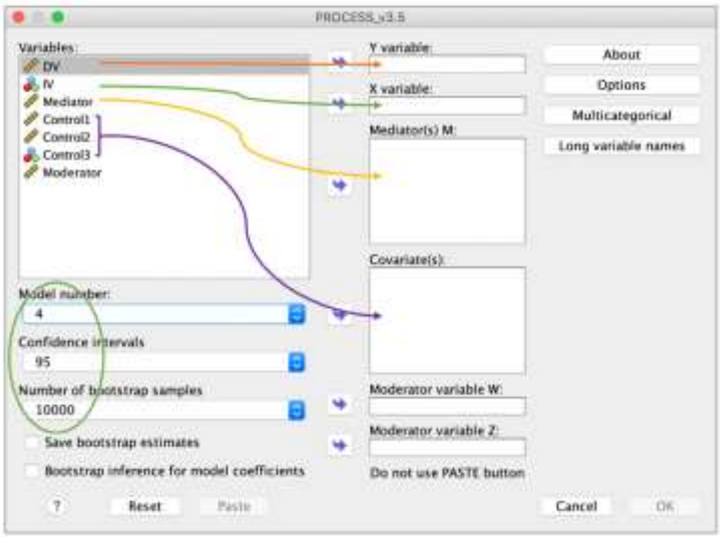


Appendix 4

- **Mediation: using PROCESS**

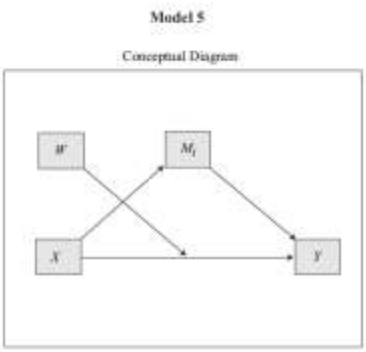
Instructions on how to install the PROCESS macro into SPSS and on the different model numbers can be found at:
https://www.processmacro.org/uploads/3/4/8/5/34858384/dialoginstall.pdf

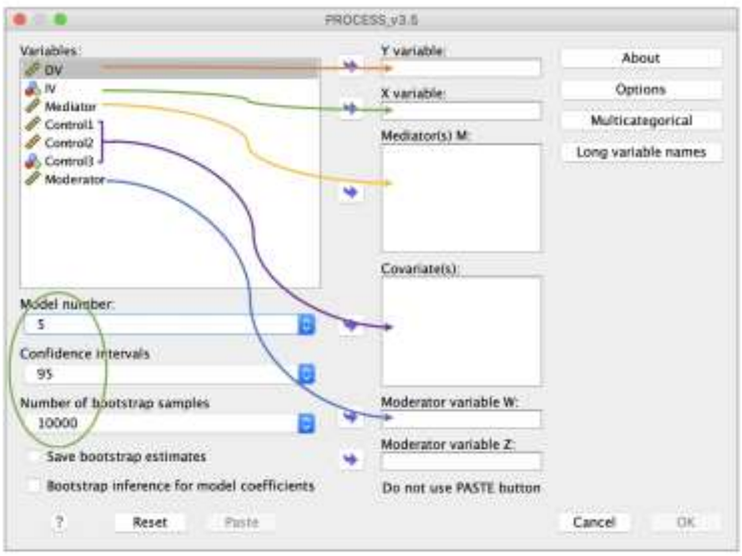




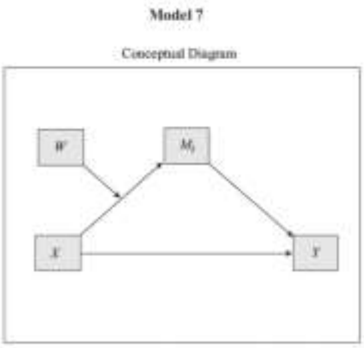
- **Moderated mediation: using PROCESS**

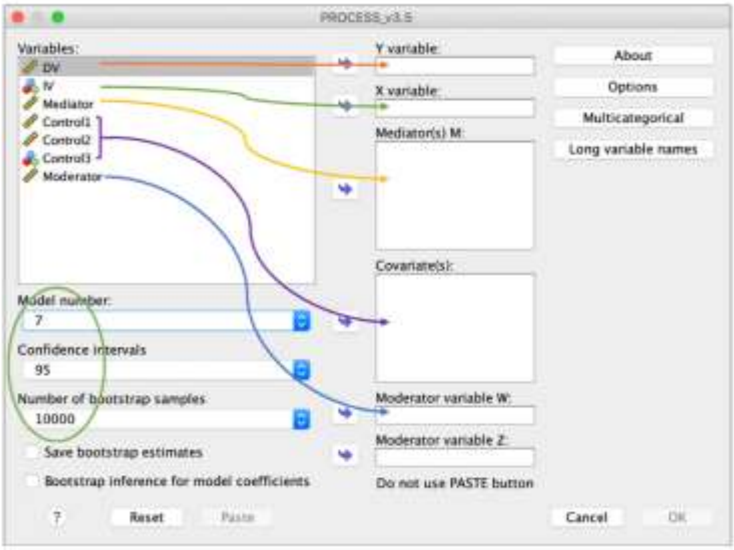
Theme a: When moderation is predicted in the c-path:



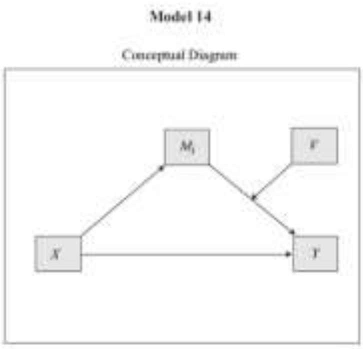


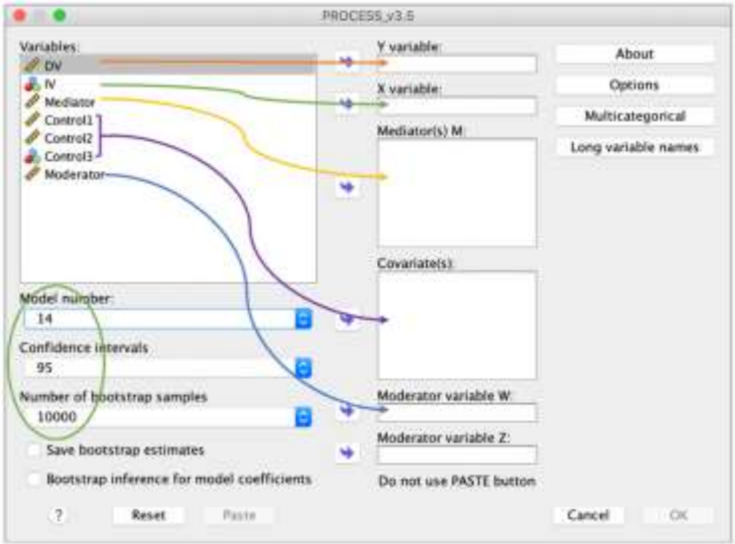
Theme b: When moderation is predicted in the a-path:





Theme c: When moderation is predicted in the b-path:





More combinations and more models, e.g., when the moderation is predicted on multiple paths, or when there are multiple mediators predicted, are available on Hayes (2012, 2013).