Annals of Tourism Research 80 (2020) 102858

Contents lists available at ScienceDirect

Annals of Tourism Research

journal homepage: www.elsevier.com/locate/annals

A review of experiments in tourism and hospitality

Giampaolo Viglia[a,*], Sara Dolnicar[b]

[a] University of Portsmouth, Department of Marketing and Sales, Portland Street, PO13DE, United Kingdom
[b] The University of Queensland, Blair Drive, St Lucia, Queensland 4067, Australia

ARTICLE INFO

Associate editor: Scott McCabe

Keywords:
Experiment
Validity
Research design
Sample size
Moderator
Mediator

ABSTRACT

Well-designed and executed experiments prove cause-and-effect relationships. The ability to draw causal conclusions is critical to knowledge development in any field of research. In this article, we discuss the benefits of experimental designs over alternative research approaches for the social sciences, discuss advantages and disadvantages of different types of experiments, review existing experimental studies specific to tourism and hospitality, and offer guidance to researchers who wish to conduct such studies. Properly executed experiments using actual behaviour of real stakeholders as a dependent variable lead to conclusions with high external validity. Our discussion of practical implementation issues culminates in a checklist for researchers. The article launches the *Annals of Tourism Research* Curated Collection on experimental research in tourism and hospitality.

Follow the data, what people do and not what they say. Replace guesses and shoddy correlations with what actually works – causally.
Stephens-Davidowitz, Everybody lies (2017, pp. 240, 284)

Introduction

Knowledge development is the aim of any academic field of research. The Oxford dictionary defines knowledge as the "theoretical or practical understanding of a subject" (Oxford Dictionary, 2018). John Rossiter (2001, 2002) distinguishes between three forms of knowledge: first order knowledge results from the description and naming of constructs. First order knowledge does not provide insights into associations between constructs. Second order knowledge is the understanding of non-causal relationships between constructs, and typically results from association studies, from studies investigating correlations between constructs. This can be illustrated using a research question frequently studied in tourism and hospitality: the association of tourist satisfaction, loyalty and intention to revisit (Dolnicar, Coltman, & Sharma, 2015). To investigate this question, researchers tend to use one-off cross sectional survey study designs, asking tourists to provide self-reports on these three constructs. Such research designs allow conclusions about *associations* between constructs: tourists who are more satisfied, are also more loyal to the destination; tourists who are more loyal to the destination, express a higher intention to revisit the destination. These insights represent second order knowledge; they *describe* relationships, but they *cannot explain* them. Because all empirical measures are taken at the same point in time, it remains unclear if tourist satisfaction *causes* loyalty, or if loyalty *causes* re-visitation intention.

While correlations can point to possible causal relationships, they cannot prove which construct drives the other construct(s); they cannot prove cause-and-effect relationships. If such proof is required, third order knowledge has to be created. It allows conclusions about cause-and-effect relationships and statements of the following type: if we do X, the consequence will be Y. This kind of

* Corresponding author.
E-mail address: giampaolo.viglia@port.ac.uk (G. Viglia).

https://doi.org/10.1016/j.annals.2020.102858
Received 6 August 2019; Received in revised form 24 December 2019; Accepted 29 December 2019
0160-7383/

knowledge pushes a field of research forward and, at the same time, is most useful in providing practical recommendations. For example: if you want to mitigate theme park crowding, send real-time information coupled with incentives on tourists' mobile phones (Brown, Kappes, & Marks, 2013). If you want to reduce both the negative environmental impact and cost of daily room cleaning in high quality hotels, share the savings with guests by buying them a drink (Dolnicar, Knezevic Cvelbar, & Grün, 2018). If you want to reduce plate waste at a buffet, reduce plate size (Kallbekken & Sælen, 2013). If you want to elicit mental imagery and sense of presence prior to a hotel experience, use virtual reality instead of images or 360′ tours (Bogicevic, Seo, Kandampully, Liu, & Rudd, 2019).

A review of tourism marketing research concludes that the dominant form of knowledge currently being created in the field of tourism is second order knowledge (Dolnicar & Ring, 2014): 87% of reviewed studies used research designs allowing conclusions about associations, but not about cause-and-effect relationships. Beyond the narrow area of tourism marketing, we see a similar picture: experiments, especially field experiments, represent a tiny fraction of studies in tourism and hospitality (Fong, Law, Tang, & Yap, 2016). Of 88 articles published in *Annals of Tourism Research* in 2018, for example, only seven (less than 8%) were based on experiments.

The aim of this article is to stimulate uptake of well-designed and well-executed experimental research among tourism and hospitality researchers (1) by outlining clearly the key benefits of experiments over alternative research approaches commonly used in the social sciences; (2) by discussing experimental research in tourism and hospitality conducted to date, highlighting common pitfalls and key design elements to consider; and (3) by providing practical guidance to researchers on how to conduct experimental studies to ensure valid conclusions. A shift towards well-designed and well-executed experimental research in tourism and hospitality will lead to a quantum leap in both knowledge creation, and in the practical usefulness of such knowledge to industry.

The experiment

An experiment assesses the causal effect of an intervention X (the independent variable) on an outcome Y (the dependent variable), as illustrated in Fig. 1. To generate valid third order knowledge, an experiment has to be well designed and executed. To be able to validly determine if X affects Y, the intervention X has to occur *before* the outcome Y. Without intervention the value of the outcome variable Y would be the same at time 0 and time 1, assuming that nothing else has changed which may affect Y. The assumption of nothing else changing is referred to as the ceteris paribus condition, meaning “other things being equal” in Latin. The top row in Fig. 1 illustrates this case: the outcome variable Y has a value of Y_0 at time 0. If time passes and no external factors affect Y, the outcome variable takes a value of Y_1, which is identical to Y_0. The purpose of this so-called *control group* is to make sure that the passing of time, or anything external to the experiment occurring during this time, does not affect the value of the dependent variable.

The bottom row in Fig. 1 illustrates the change in the outcome variable Y as a consequence of the intervention X. The dependent variable Y initially has a value of Y_0. Then time passes *and* an intervention occurs. If the dependent variable Y_1 is still equal to Y_0, ceteris paribus, the independent variable X did not affect Y. If, however, Y_1 is different from Y_0, we can conclude that X caused the change in Y.

For this conclusion to be valid, however, we must test both situations: the top and the bottom case in Fig. 1 assign participants randomly to the two groups. For example, in a quasi experiment on the effect of price changes (X) on one outdoor swimming pool at a tourist destination, the weather may be warm and sunny during the first measurement with the lower price (Y_0), but cold and rainy for the second measurement with the higher price (Y_1). Almost certainly, fewer people will use the outdoor pool at the second point in time (Y_1). But this change in Y may not be caused by the change in price (X). Rather, it may simply reflect weather that is more or less conducible for swimming in an outdoor pool. We need the control group to capture changes in Y not caused by the independent variable X. The control group tells us how much change in Y we need to expect in the experimental group, even if X has no effect on Y at all.

The bottom case in Fig. 1 is the *experimental group*. The purpose of the experimental group is to determine whether X leads to a change in Y. For example, if we increase the entrance fee to the outdoor swimming pool at the destination, it is likely that fewer tourists will come for a swim (Y_1 will decrease). Once we have corrected for the decrease in Y_1 due to the bad weather, any additional decrease in Y_1 is caused by the price increase, ceteris paribus.

Weather may not be the only external factor to affect the change in Y in field experiment where randomization of study participants is not always possible to achieve (which is why they are technically quasi experiments and not real experiments). For example, in the case of the swimming pool experiment, the first measurement (Y_0) may have fallen on a day when local schools have lessons at the local pool. The second measurement (Y_1) may have fallen on a day where this is not the case. This drop in pool

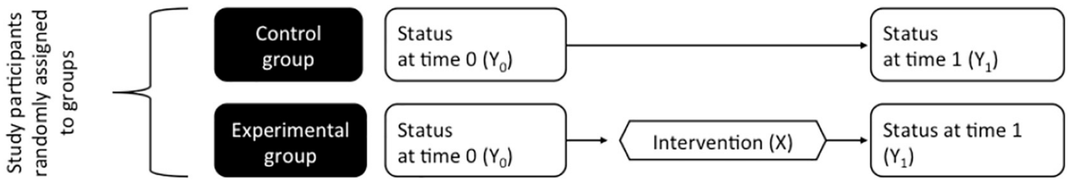

Fig. 1. Example of an experimental design.

visitation would be incorrectly interpreted as being due to the independent variable X. Or the first measurement (Y_0) may have been on a normal day, while the second measurement (Y_1) took place during a major local festival, diverting people from the pool to the festival site. The fact that all other things are equal cannot just be assumed. It requires careful planning and anticipation of any possible factors that could influence Y. Such external factors, affecting the result of the research in an unwanted way, are called *confounding variables.* A confounding variable is related to the dependent variable Y, and explains all or part of the association between X and Y (Fritz & Lester, 2016). Optimally, we want to avoid confounding variables. In our pool experiment we can do that by not running the experiment during the festival. Sometimes we cannot avoid the interference of confounding variables with our research design. The weather, for example, cannot be kept constant. In such instances we need to measure the values of the confounding variables at time 0 and time 1. In general, randomization allows reducing the likelihood that changes in Y are due to factors other than changes in X.

Several non-tourism sources offer clear insights on how to design experiments. Focusing on specific textbooks entirely devoted on experimental research, Box, Hunter, and Hunter (2005) offer a detailed overview on the necessary statistical knowledge to design experiments, while Lawson (2014) provides actionable insights on how to analyse experimental data in R.

For an experiment to be classified as a *true* experiment, study participants need to be randomly assigned to control and experimental group(s). If the assignment is not random, it is difficult to argue the ceteris paribus condition – the composition of study participants in each of the groups may affect the result (Seltman, 2012). If we do not randomly assign people in our pool experiment to the control and the experimental group, for example, our control group may contain more elderly people who attend the weekly aquarobics class, and the experimental group may contain more tourists. The aquarobics participants will still come to attend their class even if the weather is bad, but the tourists will not. The non-random assignment in such a design could lead to the incorrect conclusion that the price increase had a huge impact on visitation of the pool, when much of the cause was in fact the difference in control and experimental group composition. In general, without randomization, participants will self-select into the control and the experimental group, biasing the results.

Experiments can be implemented using *between-subject* or *within-subject designs*. In a between-subjects design, each study participant is part of one single group, either the control group or an experimental group. As a consequence, study participants in experimental groups are only exposed to one intervention X. The effect of intervention X on the outcome Y is tested *between* people. An example of a between-subject design in tourism is the study by Araña and León (2016) who measure overall CO_2 emission levels depending on whether subjects were exposed to market vs non-market-based sustainability policies. Another example is Cornelis' (2010) study in which one group of participants is exposed to a co-branding option only, and a second group to a single-branding option only. Lin, Yang, and Wan (2015) test the effect of discounts by comparing willingness to repurchase for a group receiving a small discount and a group receiving a major discount.

In within-subject designs, each study participant is exposed to all interventions X. The effect of X on Y is tested for the same study participant (*within* the person) before and after exposure to X. Tessitore, Pandelaere, and Van Kerckhove (2014) look at how the presence of a reality television affects destination image and travel intentions by measuring the behaviour of participants before and after the reality. Kim, Kim, and Bolls (2014) expose the same study participants to a touristic image and to a video, concluding that videos cause a stronger emotional reaction. Okazaki and Hirose (2009) investigate how people react to different media in travel information search, finding consistent different patterns across gender. When comparing multiple treatments, between subjects designs can provide greater confidence than within-subject designs because learning effects can arise when participants are exposed multiple times to a different treatment. But between-subject designs require more participants, increasing cost for researchers.

Validity, reliability and manipulation checks

A good research design produces valid and reliable results. *Validity* indicates that a measure "measures what it purports to measure" (Carmines & Zeller, 1979, page 4). *Reliability* indicates the "degree to which results are consistent across repeated measurements" (Carmines & Zeller, 1979, page 7). In the context of experimental research, we differentiate between two types of validity. *Internal validity* refers to the degree to which the results are attributable to the intervention X. Internal validity can be increased by controlling for any possible confounding factors. *External validity* refers to the generalisability of findings. External validity can be increased by conducting experiments in contexts in which the intervention will be used. Internal and external validity compete with one another in experimental designs (Schram, 2005). For example, if we want to avoid potential biases caused by weather, aquarobics classes or school groups, we need to conduct our swimming pool entry price experiment in a laboratory using a hypothetical scenario. But if we want to maximize generalizability, we need to conduct the experiment at the pool because the decisions respondents make in a laboratory are unlikely to be the same as the decisions they would make when faced with a higher ticket price at the pool. Sometimes internal validity is more important; sometimes external validity is more important. As a consequence, no single experimental design is optimal under all circumstances. Rather, the design has to be carefully selected in view of the research question.

Reliability means that the study produces the same results if repeated under the same conditions. Reliability in experimental research is best ensured through a clean, well-documented experimental design, which accounts for all potential confounding variables. Documentation is critical to enabling replication studies and, with it, the future assessment of reliability (Seltman, 2012).

Manipulation checks are a crucial element of experimental research. The purpose of manipulation checks is to ensure that the intervention (the independent variable) X has the intended effect. The manipulation check is the experimental researcher's life insurance: it proves that the intervention does what it was developed to do before it is used in the actual experiment. If the manipulation check fails, running the full experiment is a waste of time, effort and money. The manipulation check must be reported alongside the full experiment to give readers the confidence about X working as intended. Sparks, Perkins, and Buckley (2013)

Table 1
Types of experiments.

	Advantages	Disadvantages	Sub-types	Illustrative tourism examples
True experiments				
Laboratory experiment	• high internal validity • researcher has full control	• low external validity	physical vs. online	Babakhani et al. (2017) Ert et al. (2016) Huang et al. (2016) Hwang and Mattila (2018) Jun and Vogt (2013) Pera et al. (2019) Tassiello et al. (2018) Wu et al. (2017)
Field experiment	• high external validity • in the real life context that matters • measuring actual behaviour	• low internal validity • researcher does not have full control • many possible confounding variables • difficult to randomly assign participants to groups	framed vs. natural	Baca-Motes et al. (2012) Dolnicar, Knezevic Cvelbar, and Grün (2017a) Goldstein, Cialdini, and Griskevicius (2008) Grazzini, Rodrigo, Aiello, and Viglia (2018) Jacob and Guéguen (2012) Kallbekken and Sælen (2013) Nguyen (2016) Reynolds, Merritt, and Pinckney (2005) Rong-Da Liang (2017) Viglia, Maras, Schumann, and Navarro-Martinez (2019)
Other types of experiments				
Natural and quasi experiment	• high external validity • in the real life context that matters	• low internal validity • X not controlled by researcher (only in natural experiments) • participants not randomly assigned • many possible confounding variables	natural vs. person-by-person treatment	Araña and León (2008) Chen, Lehto, and Cai (2013) Dolnicar et al. (2017b) Dolnicar et al. (2018) Karlsson and Dolnicar (2016) Pabel and Pearce (2016) Zavattaro and Fay (2019)
Discrete choice experiment	• high internal validity if designed well	• low external validity	stated vs. revealed	Albaladejo-Pina and Díaz-Delfa (2009) Chen, Masiero, and Hsu (2018) Karlsson, Kemperman, & Dolnicar, 2017 Kim and Park (2017) Masiero and Nicolau (2012) Nicolau and Mas (2006)

provide an excellent example of a manipulation check used in tourism research. The intervention X they use is the content type of online reviews, which can be either vague or specific. In the manipulation check study participants answer the following questions: "I think the reviews provided specific information" and "Overall, I felt most of the reviews were a bit vague." If the majority of study participants do not perceive the vague reviews as vague, and the specific reviews as specific, the manipulation check fails and there is no point in running the experiment.

Types of experiments

Table 1 lists different types of experiments (as classified by Harrison and List (2004) and Seltman (2012)), states advantages and disadvantages of different designs, and points to examples of each of those types of designs in tourism research. The two main categories are laboratory experiments and field experiments. When neither of those can be implemented, quasi experiments and choice experiments are available as an alternative. Additionally, researchers can sometimes make use of naturally occurring events to measure the impact of one variable on another one (natural experiments).

Laboratory experiments

Every intervention X is designed to change an outcome Y in a specific real life setting. Laboratory experiments do not test the intervention X in that real life context. Instead, they take place in a space where the behaviours under study do not usually occur: in a laboratory. The laboratory can be a room equipped with a computer, an eye tracking device, a skin conductance instrument and an EEG; or it can be a space that is set up to mimic as closely as possible the real life context, such as a space set up like a real travel agency.

The main advantage of the laboratory setup is that the researcher is in full control of the environment. The researcher can ensure that the ceteris paribus condition is met: that everything except the intervention X is the same for all study participants. For that reason, laboratory experiments have high internal validity: changes in the outcome Y are caused only by changes in the intervention X. The disadvantage of the laboratory experiment is that it is artificial: it does not study people in the real life context for which intervention X is designed. People behave differently in a laboratory setting, reducing the external validity of laboratory experiments. In sum, laboratory experiments tell us that intervention X affects outcome Y, but we cannot be sure that this happens in different real life contexts.

Much tourism and hospitality knowledge has resulted from laboratory experiments: Jun and Vogt (2013) manipulated consumer involvement, the strength of textual information and the availability of a picture of the product advertised to test how it affects stated intention to purchase; Ert, Fleischer, and Magen (2016) determined how photos of Airbnb hosts affect stated bookings by Airbnb guests; Babakhani, Ritchie, and Dolnicar (2017) tested the effect of different carbon offsetting appeals presented on airline booking webpages on passengers' stated intention to pay for voluntary carbon offsets; and Huang, Cheng, Chuang, and Kuo (2016) tested how different types of framing a pro-environmental communication message (positive and negative) affects stated pro-environmental behaviour of people with high or low levels of environmental concern.

Laboratory experiments are particularly useful for the identification of psychological processes causing the change in the outcome Y. Hwang and Mattila (2018) tested the effect of different reward types on stated behavioural loyalty, concluding that self-construal explains whether a luck-based reward or a loyalty-based reward is more effective. Similarly, Tassiello, Viglia, and Mattila (2018) show that better ratings given by hotel guests in handwritten feedback is explained by the activation of empathy. Isolating the psychological role of empathy in a field experiment would have been impossible (see also Pera, Viglia, Grazzini, & Dalli, 2019). Laboratory experiments are also unavoidable when the intervention X is difficult to manipulate in the field. Wu, Shen, Fan, and Mattila (2017), for example, test the effect of using factual versus figurative language when writing reviews on consumers' attitudes and stated purchase intentions.

Increasingly, laboratory experiments are conducted using online platforms such as MTurk, Qualtrics and Prolific Academic (Peer, Brandimarte, Samat, & Acquisti, 2017). Qualtrics is an online survey platform. MTurk provides access to crowdsourced individuals to complete tasks. Both Qualtrics and Prolific Academic provide access to survey panels (individuals that are willing to complete surveys). While using these platforms is appropriate when testing general theories and principles, testing specific effects requires a sample that resembles the target population (real tourists in the field). Because study participants cannot be monitored on such platforms, data quality may be low (Goodman & Paolacci, 2017). Not surprisingly, therefore, Crump, McDonnell, and Gureckis (2013) find evidence that actual laboratory studies are more accurate in predicting behaviour than online survey experiments.

Field experiments

Field experiments are conducted in the real life settings for which interventions X are designed (Gerber & Green, 2012). For example, if the intervention is the reduction of plate size intended to help hotels reduce plate waste at buffets (as in Kallbekken & Sælen, 2013), a field experiment has to be conducted at a hotel buffet. As opposed to laboratory experiments, field experiments have high external validity. This means: if the intervention X affects the outcome Y, we know that the intervention works in the real life context. We can immediately derive a practical recommendation for that context. The disadvantage of field experiments is that the real life context is not easily controlled. Many external factors (confounding variables) other than the intervention X can affect the outcome Y. It is therefore critically important to ensure that all possible external factors that may change Y are identified and either excluded or measured during the experiment. That way potential effects of confounding variables can be avoided or accounted for during data analysis.

A number of field experiments have contributed to tourism and hospitality knowledge: Baca-Motes et al. (2012) increased hotel towel reuse by over 40% among guests wearing a publicly visible pin indicating their commitment to towel reuse. Their work highlights “how a small, carefully planned intervention can have a significant impact” (p. 1070). Kallbekken and Sælen (2013) – inspired by obesity prevention research (Freedman & Brochado, 2010) – reduced buffet food waste by 20% with smaller plates and by 21% with a sign encouraging people to return to the buffet. Filimonau, Lemmer, Marshall, and Bejjani (2017) effectively improved sustainable consumer choices by re-designing a restaurant menu. Grazzini et al. (2018) significantly increased guests' recycling in hotels using message framing, and Mair and Bergin-Seers (2010) increased towel reuse by 4% through interventions combining information, norms and incentives.

Reese, Loew, and Steffgen (2014) show that normative appeals – which express a value judgment about whether a situation is desirable or undesirable – are more powerful than standard environmental messages in decreasing towel use. Gössling, Araña, and Aguiar-Quintana (2019) confirm these findings, showing that, compared to a normal message, a normative message leads to a 6.8% decline in towel and a 1.2% decline in bed linen use.

Results from field experiments have often challenged previous evidence. For instance, pro-environmental values, assumed to be effective in changing behaviour, have failed to significantly increase tourists' hotel towel reuse and decrease room electricity consumption, suggesting that interventions in hedonic contexts—such as tourism—may require the use of more tangible benefits in order to change behaviour (Dolnicar et al., 2017a). Similarly, Reynolds et al. (2005) have implemented two menu layout conditions in restaurant and showed that menu labelling does not affect sales directly. Work by Rong-Da Liang (2017) shows that agritourism activities with a service-dominant-logic design, contrarily to expectations, do not comprehensively enhance the experience of tourists' learning agricultural knowledge. Araña and León (2016) show how an interplay of emotions (sadness and empathy) reduces the importance of price, and increases the importance of low CO_2 emissions during travel decision making. With respect to framing of messages, Araña and León (2013) demonstrate that proposing carbon-offsetting policies as default – thus asking tourists to opt-out if they do not wish to voluntarily purchase offsets – substantially increases sustainable choices compared to opt-in options. The work of Nguyen (2016) shows that loss averse tourists are more likely to overspend, providing empirical support for prospect theory-based approaches as an alternative to the more commonly adopted expected theory-based approaches.

Field experiments also provide a deeper understanding of tourist experiences. For instance, Kang and Gretzel's (2012) study podcast tours, concluding that the human voice – conveyed via audio-only media – creates positive tourist experiences. Engeset and Elvekrok (2015) show how authenticity increases tourist satisfaction. And Antón, Camarero, and Garrido (2018) identify factors creating memorable experiences in museums.

In some cases, the nature of one's research question directly dictates the experimental context. For example, research examining the effect on people's attitudes and behaviour of "Facebook liking" needs to be conducted on Facebook, or in a laboratory simulation of Facebook (John, Emrich, Gupta, & Norton, 2017). When investigating more general questions, researchers have some flexibility in choosing the study setting, allowing field experiments in different contexts. Sometimes participants know they are part of an experiment (*frame*d field experiment), sometimes they do not (*natural* field experiment, Harrison & List, 2004).

Field experiments are not as powerful as laboratory experiments when it comes to understanding the reasons for the intervention X affecting the outcome Y. This is because researchers have less control over all aspects of the experiment (Gneezy, 2017). The only way to mitigate this weakness is to identify potential confounding variables in advance, and either vary or measure them, and include them in the data analysis.

Importantly, researchers do not have to choose between laboratory and field experiments. Sometimes the most powerful knowledge results from a sequence of both approaches. We can start with a laboratory experiment to systematically test the reasons for certain behaviours. Then we can conduct a field experiment to ensure that the effects observed in the laboratory generalise to the real life context. Or we can conduct a field experiment to test if an intervention works, and then follow up with a laboratory experiment to understand why exactly it works. The ultimate aim is understanding the causal process underlying a phenomenon, and establishing moderators or boundary conditions of the existing phenomenon.

While laboratory and online experiments are relatively easy to set up, field experiments are time consuming, logistically difficult, disruptive and costly. Identifying potential tourism partners and obtaining their consent, in addition to planning and collecting data, is key to successful implementation. An effective way to address this concern is to design experiments that integrate into what is already occurring. Tourism organizations are often familiar and comfortable with the idea of pilot programs, especially when we are not sure yet if something works. Proactively presenting potential caveats and discussing ways to address them is of paramount importance. Generally: the smaller the organization, the faster an agreement can be reached (Gneezy, 2017). Engeset and Elvekrok (2015), for instance, agreed with the Norwegian Trekking Association to run field experiments on authenticity with a selection of mountain lodges. Jacob and Guéguen (2012) came to an agreement with two restaurants in medium-size cities in France to investigate the relationship between exposure to altruistic quotes and helping behaviour. Viglia et al. (2019) implemented a field experiment on Pay-What-You-Want starting from a need of the restaurant owner: increasing the awareness of the place. The research program on reducing the environmental harm done by tourists in hotel (Dolnicar et al., 2017, 2017a, 2018; Juvan, Grün, & Dolnicar, 2018) was only possible because of the commitment and ongoing participation in field experimentation by Slovenian hotels.

Natural experiments and quasi experiments

Alasuutari, Bickman, and Brannen (2008) define natural experiments as a form of experiments where treatment occurs naturally (or is unplanned), while quasi experiments as a form of experiment where the treatment is intentional or planned. Compared to the other forms of experiments presented, membership of study participants is not random. As a consequence, neither natural experiments not quasi experiments comply with both key criteria of the true experiment: the researcher being in control of the intervention X, and the researcher randomly assigning study participants to control and experimental conditions (Hyman, 1982). For example, we may want to test whether the number of tourists to the UK changes after Brexit. The researcher cannot determine at which point in time the UK leaves the EU, and the researcher cannot ensure that the same kinds of tourists consider travelling to the UK before and after Brexit. The researcher can only test the difference, and try to correct for all possible confounding variables during data analysis. Because natural experiments violate the strict condition of true experiments, they have lower internal validity. Araña and León (2008), for example, investigate the effect of terrorism on tourism demand using as a cut-off date the moment of the terrorism attack. In quasi experiments, person-by-person treatment information can be leveraged to see the effect of the intervention X on the outcome Y. Chen et al. (2013) test the effect of vacations on subjective well-being. They use a longitudinal design, measuring subjective well-being before and after the vacation for each study participant. The design does not allow random assignment of study participants to experimental (vacation / no vacation) conditions.

Similarly, Dolnicar et al. (2018) test if replacing cotton serviettes at breakfast buffets with recycled paper serviettes (while still making cotton serviettes available on a self-service basis) reduces the use of cotton serviettes. Again, this design does not permit random allocation of study participants (hotel guests) to the control and experimental condition. In this case the researchers took two measures to reduce contamination of findings by differences in guest composition: they chose a period of time where the guest mix in the hotel is similar, and they collected guest mix data and accounted for differences in guest mix during data analysis.

In other examples of quasi experiments in tourism, Pabel and Pearce (2016) tested the effect of humour in tour guides' presentations on tourist comfort during their vacation; Zavattaro and Fay (2019) tested the effect of using Brand USA to increase return of investment for the United States. By comparing areas where the brand was used versus areas where the brand was not used, they show that – when controlling for confounding variables – Brand USA did not increase return of investment. Karlsson and Dolnicar (2016) tested whether tour boat eco-certification affects tourists' boat choices, concluding that – while 60% of boat passengers stated that they considered the environment when selecting one of the two available boat tours – only 14% were able to correctly answer the question whether or not the boat they ultimately boarded was eco-certified. This suggests that any effect, if it exists at all, is small. Reiser and Simmons (2005) tested the effectiveness of eco-label promotion on tourist behaviour, showing that attitudes towards ecolabels are an unreliable predictor of responsible environmental tourist behaviour.

Quasi-experiments have also been adopted to test whether flagship urban projects contribute to city image change (Smith, 2006), and to assess the effect of co-branding strategies on hotel and restaurant brands (Tasci & Guillet, 2011). Hahm and Wang (2011) examined the impact of a film (Lost in Translation) on the featured destination's (Japan's) image, and people's stated intentions to travel to Japan, using a one-group prefilm-postfilm quasi-experimental design. Becken and Wilson (2007) used a quasi-experiment to understand whether information about regional attractions influences tourists' itineraries. March and Woodside (2005) explored the relationship between planned and reported consumption behaviours, concluding that implemented tourism strategy varies systematically from planned strategy. Quasi-experiments are also useful for measuring the impacts of advertising and marketing programs (Woodside, 2010): Woodside III, T, and MacDonald (1997) tested the efficacy of a free 130-page visitor information guide on changing destination behaviours, and increasing the expenditures of visitors to Prince Edward Island.

Choice experiments

Discrete choice experiments – often referred to as discrete choice modelling – allow researchers to carefully design product or behavioural alternatives, and ask study participants to indicate which they would choose. Choice experiments are typically implemented as survey studies: for example, survey respondents may see ten pairs of products with specific product features. Respondents indicate –for each pair – which product they would choose. Their choices allow data analysts to determine the importance of each product feature. The random utility theory framework – which postulates that a person will choose the option that maximizes their utility (McFadden, 1980) – underlies choice modelling. Discrete choice modelling is used across many different fields of research, including marketing, transportation, housing, and environmental economics. Compared to field experiments and quasi experiments, discrete choice models have low external validity but, if designed well, have high internal validity. The key advantage of choice experiments is that one can test products which do not actually exist. Discrete choice experiments can be of two types: stated or revealed (Louviere, Hensher, & Swait, 2000). In stated choice experiments participants are asked to state their choice. In revealed choice experiments preferences are elicited from actual behaviour.

Tourism researchers have embraced choice modelling, and use it to understand destination choice (Masiero & Nicolau, 2012; Nicolau & Mas, 2006); rural accommodation selection (Albaladejo-Pina & Díaz-Delfa, 2009; Chaminuka, Groeneveld, Selomane, & Van Ierland, 2012); hotel selection (Huertas-Garcia, Laguna García, & Consolación, 2014; Kim & Park, 2017); and approval of guest inquiries by Airbnb hosts (Karlsson et al., 2017). Choice experiments also help understand factors that influence destination choice, domestically (Huybers, 2003a, 2003b) and internationally (Morley, 1994). Huybers and Bennett (2000) assess the relative importance of the natural environment; Brau (2008) investigates the relationship between natural and man-made attractions in location choice; Huybers (2003a, 2003b) identifies key attributes (such as crowdedness, nightlife, season and alike) affecting destination choice by Sydney resident; and Hsieh et al. (1993) explain the influence of socio-demographics, travel characteristics and psychographic variables on travel mode choice, pointing to the central role of psychographic characteristics. More recently, choice experiments provided insights into how sequential exposure to attributes affects destination choice (Oppewal, Huybers, & Crouch, 2015). Rashidi and Koo (2016) find evidence that travel party choices, travel mode choices, and expenditure decisions are interrelated.

Discrete choice experiments have also helped understand how tourism affects local residents (Figini, Castellani, & Vici, 2009), vacation length (Grigolon, Borgers, Kemperman, & Timmermans, 2014), spending allocation decisions among vacations (domestic and overseas) and other categories of discretionary expenditure (Crouch et al., 2007). Crouch, Del Chiappa, and Perdue (2019) explored the factors that determined the choice of a host city for international conventions. Chen et al. (2018) use discrete choice experiments to identify sources of preference heterogeneity for Chinese outbound tourists.

Tourism researchers have also adopted agent-based modelling, which represents complex systems of autonomous agents or actors. By simulating the many possible outcomes of agent behaviours it is possible to explore the importance of different choices (Nicholls, Amelung, & Student, 2017). For instance, Boavida-Portugal, Ferreira, and Rocha (2017) use agent-based modelling to study tourist decision making when choosing a holiday destination.

Study participants

The validity of any research depends on the study participants. Many researchers are tempted to use *convenience samples*: samples of study participants that are easy to access, such as students or members of online survey panels. The problem with convenience sampling is that the people participating in the study may not actually be behaving in the same way as the people we are trying to understand. For example, if we are interested in the accommodation choice behaviour of business travellers, using university students as study participants is not a good design choice because it is unlikely that the accommodation choices of university students resemble those of business travellers. In other instances, it is permissible to use student sample: if we are investigating the accommodation choice behaviour of university students. Or if we are investigating very fundamental human behaviour. For example, when we are presented with a tourism advertisement we may want to understand which sections of the advertisement attract the most attention. The cues attracting attention – such as bright colours – can be assumed to be of general nature, making the use of student samples permissible (Calder, Phillips, & Tybout, 1981; Šerić & Praničević, 2018). Justifying the sampling approach and explaining the advantages and disadvantages associated with this choice is critically important.

In addition, for an experiment to be a *true* experiment, the assignment of study participants to experimental conditions must be random. While easy to achieve in laboratory experiments (Calder et al., 1981), random assignment is difficult to implement in field experiments with real consumers being studied (Juvan et al., 2018).

The best we can do in field experiments to avoid differences in the samples affecting the results is to (1) select the timing of our study in a way that ensures – as much as possible – that the composition of tourists will be the same across all experimental groups, and (2) collect data on personal characteristics of the tourists to allow checking at the stage of data analysis whether the composition of study participants varied across experimental groups and whether this variation affected the outcome Y. This problem is well illustrated by the study on replacing thick cotton serviettes in hotel breakfast buffet dining rooms with more environmentally friendly recycled paper serviettes (Dolnicar et al., 2018). The intervention X was the change from providing cotton serviettes on the tables to providing recycled paper serviettes (while still making cotton serviettes available at the buffet). The outcome Y of interest was the number of cotton serviettes used. The two experimental conditions were implemented sequentially in a between-subjects design. An analysis of the guest mix showed that – despite timing the experiment such that no differences in guest mix were expected – tourists in each of the two experimental groups were not the same (they differed in length of stay, purpose of the trip, country of origin, and room type booked). As a consequence, a number of potentially confounding variables (check-in date, check-out date, number of adults in the room, number of children in the room, type of guest (leisure, business), room type (standard, superior, other) and country of origin) had to be included in the data analysis.

The dependent variable Y

People are bad at reporting their past and predicting their future behaviour. Reports of past behaviour tend to be inaccurate for a range of reasons, including difficulties recalling the actual behaviour (for example, exact travel routing), not wanting to admit to the actual behaviour (for example, littering in National Parks), or constructing ex-post explanations for behaviour that are not the real causes of their actions (such as taking the train for environmental reasons, when really it was the cheapest option; Kahneman, 2011). GPS studies conclude that stated behaviour under-reports actual behaviour by as much as 60% (Stopher & Greaves, 2009).

Accurately stating behavioural intentions is even more difficult because intentions are, by definition, hypothetical. The more hypothetical questions about behavioural intentions are ("Do you intend to engage in space travel when it is commercially viable?"), the lower the likelihood of the stated intentions being predictive of actual behaviour. An extensive body of empirical work has investigated and re-investigated the link between behavioural intentions and actual behaviour, leading to the overwhelming conclusion that predictive validity of intentions is rather low. A recent empirical study conducted in the context of boat tours and studying the stated and actual consideration of pro-environmental credentials of otherwise identical boat tours concludes a gap of 46% (Karlsson & Dolnicar, 2016).

Attitude is another construct that is frequently treated as a proxy for behaviour. Theories postulating attitudes as key antecedents of behaviour (such as the theory of planned behaviour; Ajzen, 1991) are the likely reason for this. Empirical evidence shows that, while a certain behaviour might be driven by a specific attitude, a specific attitude does not necessarily cause the behaviour. This difference "*between what people say and what people do*" (Blake, 1999, p. 275) is referred to as the attitude-behaviour gap (see Blake, 1999; Carrington, Neville, & Whitwell, 2014; and in tourism Juvan & Dolnicar, 2014).

Because past behaviour, stated behavioural intentions, and attitudes are not highly predictive of actual behaviour, the best choice of an outcome Y in experimental research is to measure actual behaviour. Measuring actual behaviour dramatically increases the validity of conclusions drawn (Morales, Amir, & Lee, 2017). Although measuring actual behaviour typically requires more effort, many behaviours of interest in tourism and hospitality research are readily accessible to empirical measurement, including: eye movements when inspecting a restaurant menu (Yang, 2012), skin conductance when viewing a destination advertisement (e.g. Li, Walters, Packer, & Scott, 2017), and voluntarily waiving the daily room clean in a hotel (e.g. Dolnicar et al., 2017). Other behaviours that can be measured are the reuse of hotel towels (e.g. Baca-Motes, Brown, Gneezy, Keenan, & Nelson, 2012; Goldstein et al., 2008; Mair & Bergin-Seers, 2010), repeated use of the same airline (e.g. Dolnicar, Grabler, Grün & Kulnig, 2011), online booking conversion rates (Di Fatta, Patton, & Viglia, 2018), or electricity use in the hotel room (e.g. Dolnicar et al., 2017a), just to offer a few examples.

Fig. 2 summarizes the two features of experiments discussed so far (the nature of study participants, and the nature of the outcome Y), pointing to the strengths and weaknesses of each of the approaches in terms of external validity (Fig. 2).

Tourism studies fall into all four quadrants of Fig. 2. For example, Nicolau and Sellers (2012) and Viglia, Mauri, and Carricano (2016) test the reaction of travellers to realistic hotel price changes in real hotels using the actual target population, but fail to measure actual behaviour, relying instead on stated behavioural intentions only. This is an example of quadrant 4 in Fig. 2. An example for quadrant 1 is provided by Ding, Grewal, & Liechty, 2005 who measure actual behaviour – actual Chinese dinner purchases – but use a convenience sample.

Drawing firm conclusions and making recommendations to practitioners about how they should act on the basis of research designs falling in quadrants 1, 3 and 4 still requires a bit of a leap of faith.

Sample size

There is no single optimal sample size for experiments. Two factors drive sample size requirements: (1) the expected effect of the independent variable X on the dependent variable Y (this effect is referred to as *treatment effect*), with larger treatment effects requiring lower samples, and (2) the standard deviation of the dependent variable Y.

	Study participants are NOT from the population of interest	Study participants are from the population of interest
Outcome Y is a measure of actual behaviour	QUADRANT 1 Effect of X on Y for population of interest questionable	QUADRANT 2 Effect of X on Y can be determined for this population
Outcome Y is NOT a measure of actual behaviour	QUADRANT 3 Effect of X on actual behaviour Y is VERY questionable (i.e., it cannot be determined)	QUADRANT 4 Effect of X on actual behaviour Y for population of interest questionable

Fig. 2. The impact of study participants and outcome variable Y on external validity.

Sample size requirements are closely linked to the concept of the probability of making errors in hypotheses testing. Type 1 error is the probability of incorrectly rejecting the null hypothesis, and concluding that means are different when they are not. In the context of experimental research we would make a Type 1 error when concluding a difference in the effect of X on Y between experimental and control group when in reality there is no difference. We use α to denote the probability of making a Type 1 error.

Type 2 error is the probability of incorrectly failing to reject the null hypothesis, concluding that means are not different when they are. In the context of experimental research we would make a Type 2 error when concluding no difference in the effect of X on Y between experimental and control group when in reality there is a difference. We use β to denote the probability of making a Type 2 error. The power of the experiment is 1-β. The statistical power is the likelihood that a study will detect an effect when an effect actually exists. If statistical power is high, the probability of making a Type 2 error drops. A typical power selection is = 0.8 (0.2 probability of Type 2 error).

Assuming an equal variance across conditions, $\sigma_1^2 = \sigma_2^2$ we can write that:

$$n_0^* = n_1^* = n^* = 2(t_{\alpha/2} + t_{\beta})^2\left(\frac{\sigma}{\delta}\right)^2 \tag{1}$$

where $n_0^* = n_1^*$ is the sample of each experimental condition and the t are the values of the T-student statistics based on the probabilities of α and β. The optimal sample size depends on the ratio of the effect size σ to the standard deviation δ. Hence, effect sizes can just as easily be expressed in standard deviations. The necessary sample size increases with the desired significance level. Power increases proportionally with variance of the dependent variable, and decreases inversely proportionally with the square of the minimum detectable effect size. Given that the standard (Raudenbush & Liu, 2000) is to use $\alpha = 0.05$ and to have power of 0.80 ($\beta = 0.20$), if we want to detect a one-standard deviation change using the standard approach, we would need.

$n^* = 2(1.96 + 0.84)^2 * (1)^2 = 15.68$ observations for each group. In many tourism studies, however, is pretty uncommon to find a one-standard deviation change in the outcome variable. Assuming a 1/3 standard deviation change, the optimal sample would be$n^* = 2(1.96 + 0.84)^2 * (3)^2 = 15.68 * 9 \sim 141$ observations for each group. This means that if we assign 141 people to the experimental group, and 141 people to the control group, and if the true treatment effect is 1/3 of a standard deviation, then there is an 80% probability that – when we compare the mean Y for the experimental condition with the mean Y in the control condition – the difference will be statistically significant at the 5% level (using a two-sided *t*-test).

With the same treatment effect, a higher or lower sample can lead to the null hypothesis being rejected or not. For this reason, it is good practice to complement the results with an effect size measure, such as Cohen's d (Cohen, 1988).

Between, within and mixed designs

When looking at the statistical differences between a within-subject design and a between-subject design we need to look at the decomposition of the errors in the differences across experimental groups and within group. The total sum of squared differences between the individual values and the mean can be written as $SS_T = \Sigma\ (value - mean)^2$, where $SS_T = SS_{EFFECT} + SS_{ERROR}$.

The observed variation across all study participants (SS_T) has two causes: the effect of the intervention X (the treatment effect), and individual differences between study participants. The variation between experimental conditions (SS_{EFFECT}) is due to the treatment effect. The variation between study participants assigned to the same condition (SS_{ERROR}) is due to individual differences. In a within-subject design, because it is the same person, the difference between the outcome variable Y across two conditions cannot be due to age, personality, or any other individual difference. Therefore, given that all of those sources of error variance are removed,SS_{ERROR} is smaller in a within-subject design compared to a between-subject design.

Within-subjects designs have more statistical power, but they have a major limitation: by exposing each study participant to all interventions X, there is a risk of learning and order effects. That means that study participants may modify their behaviour when confronted with the second intervention based on their experience with and learning from the first intervention. An example from the tourism literature is the study by Nguyen (2016) who analyses risk and time preferences of Singapore tourists. Being exposed to all the interventions more than once made participants learn more how the outcome variable is affected by risk and time. One way of controlling these effects in within subject-designs is to randomize the order of the interventions X.

The combination of a between-subject design and a within-subject design is referred to as a *mixed design*. Mixed designs combine the advantages of within-subjects and between-subjects designs. The within-subjects designs in a mixed design add statistical power, while the between-subjects designs help to rule out learning and order effects. An example of a mixed design experiment in tourism and hospitality is the study by Viglia and Abrate (2014). The authors showed study participants a series of past hotel prices. Each participant saw a complete series of prices (within-subjects design). The authors also manipulated the source of information of these prices – the Internet or personal interaction – between subjects.

We can conclude that between-subjects designs can provide greater confidence than within-subject designs if one is comparing multiple treatments, but there are implications with respect to number of participants required (and thus also cost issues).

Moderators and mediators

A *moderator* is a variable that explains under what conditions X causes Y (Baron & Kenny, 1986). In other words, a moderator clarifies the boundaries within which X influences Y (Seltman, 2012). Contrarily to structural equation models where moderators are generally self-reported variables prone to social desirability bias (Tourangeau, Rips, & Rasinski, 2000), experimental researchers manipulate the levels of the moderator. For instance, Gneezy et al. (2010) – studying the relationship between price and souvenir photo purchase – use the presence (or absence) of charity contributions as a moderator of the main X ➔ Y effect. Rodríguez-Molina, Frías-Jamilena, and Castañeda-García (2015) find that the strength of the relationship between the number of alternatives presented and destination image is moderated by the degree of involvement. They use as a moderator either a high involvement task (booking a trip) or a low involvement task (exploring a website). Statistically, a moderator is the coefficient between the intervention X and the outcome Y. In tourism economics, a moderator is often referred to as an *interaction effect* given that, statistically speaking, a moderation is the coefficient of the interaction between the intervention X and the proposed moderator (Das & Dirienzo, 2010; Imbens & Wooldridge, 2009).

A *mediator* is a third factor that explains the relationship between X and Y. As opposed to a moderator, a mediator does not affect the strength of a proposed X-Y relationship. Instead, the mediator explains why X is associated with Y. In their study on the role of handwriting versus typewriting on the subsequent online review evaluation by hotel guests, Tassiello et al. (2018) demonstrate that empathy is a mediator of the X-Y relationship. Handwriting triggers empathy. Empathy, in turn, has a positive effect on the online review such that, when controlling for it, the main X-Y relationship is no longer significant. Statistically, mediation analysis requires more theoretical and implementing steps than moderation analysis (see Baron & Kenny, 1986; Preacher & Hayes, 2008). First, to claim the presence of mediation, we expect that X affects Y. We also expect that X has an effect on the proposed mediator. The mediator must influence Y when both the mediator and the X are used as predictors. The coefficient relating X to Y must be larger in the original model compared to the coefficient in the third step. As clarified in Rucker et al. (2011), if the main X ➔ Y relationship is not significant in the third step (when controlling for the mediator), the result is called full mediation. Alternatively, if the main relationship is weaker but still significant, the result is defined as partial mediation. It is important to recognize an ongoing discussion on “mediation-only” models (see Zhao et al., 2010), in which the researcher is unable to demonstrate a X-Y main effect, as there may be an indirect significant effect. An example is Kim and Lakshmanan's (2015) study, where the authors show a mediation model on novelty perceptions with no initial link between X and Y. When using secondary data and assuming full mediation, a mediation analysis is comparable to an instrumental variable estimation in tourism economics (Otter, Pachali, Mayer, & Landwehr, 2018).

In sum, 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.

For mediation and moderation models, the statistical software G-Power and the PROCESS add-on to SPSS (Hayes, 2017) are examples of practical tools to perform these analyses without programming skills. It is important to present the proposed relationships graphically to facilitate the reader's understanding of the model to be tested. Fig. 3 presents a possible way to show the main relationship between the intervention X and the outcome Y, a moderator and a mediator graphically. In this case it is assumed that

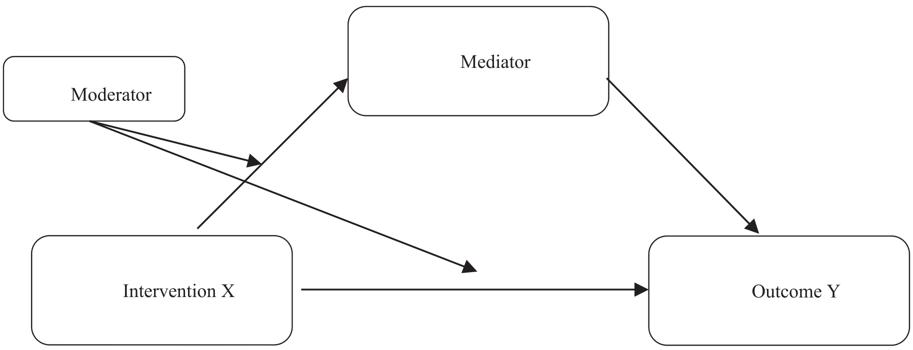

Fig. 3. Graphical representation of an experimental moderated mediation.

Table 2
The benefits of experiments.

	Experiment	Observation	Survey
Form of knowledge created	3rd order knowledge	1st order knowledge 2nd order knowledge	1st order knowledge 2nd order knowledge
Permissible conclusions	Explanation (cause-and-effect)	Description (of concepts or associations)	Description (of concepts or associations)
Ability to randomize the sample	High	Low	Low
Ability to manipulate the intervention and the moderator	Yes	No	Partial (possible to assign different Interventions, but no control over them)
Ability to eliminate confounding factors	High	Low	Partial (it is sometimes possible to control for third factors)
Ability to recommend to industry measures that are effective in changing consumer behavior	Yes	No	No

the moderator has an impact on the main relationship (X ➔ Y) and on the relationship between the X and the mediator.

Moderators are sometimes confused with a mediator (Hayes, 2009). Imagine investigating the impact of co-creation on tourists' stated willingness to pay for a product (Tu, Neuhofer, & Viglia, 2018). Age is an exogenous factor that might strengthen or reduce the effect of co-creation on stated willingness to pay and thus can be a moderator. But age cannot be a mediator because it is not affected by the X (co-creation). Age also does not affect Y. A possible mediator could be customer engagement: co-creation might trigger customer engagement, and customer engagement might drive willingness to pay, thus acting as the mechanism behind the main X ➔ Y relationship. In this model, age affects both the main X ➔ Y relationship and the relationship between X and the mediator (i.e., the model, conceptually, is the one presented in Fig. 3). It is not always trivial to understand the role of a moderator in a mediation model because the moderator can affect: the X ➔ Y relationship only, the link between X and the mediator, or both paths (as in the case above). Always guided by the theoretical knowledge on the tourism phenomena, the researcher might have to test all possible moderation effects to really understand how the overall model works better empirically. For instance, stemming from construal level theory, Li, McCabe, and Xu (2019) explore the moderation effect of mind-set on tourism preference shift over time.

Conclusion

Well-designed and executed experiments overcome most of the limitations associated with one-off cross-sectional survey research with no randomization, which rely solely on self-report measures. Experiments create third order knowledge about cause-and-effect relationships between constructs and, as a consequence, allow tangible practical recommendations that can be derived with confidence. Table 2 summarizes the key benefits of experimental research in the social sciences. Appendix A provides a checklist for tourism and hospitality researchers.

Insights gained from experiments can further be enriched with complementary research methods, combining, for example, a laboratory and a field experiment; secondary data analysis and a laboratory experiment; or a qualitative study and an experiment. Examples in tourism include: Kim and Jang (2014) who complement their laboratory experiment with a field study, Walters, Wallin, and Hartley (2019) who run a choice experiment embedded within a classic between experimental design, and Chan et al. (2015) who offers evidence from a qualitative study and an experiment. The nature of participants and the degree of behavioural measures may differ across the empirical package. For instance, one can have a series of laboratory experiments to document a phenomenon, and then follow these with a field experiment to strengthen the external validity of the claims. Take the case of Lin et al. (2015) where the authors first run a laboratory experiment, and ensure the generalizability and representativeness of their findings in a field experiment. Alternatively, given a main effect that was previously found in a field study (Araña, León, Moreno-Gil, & Zubiaurre, 2013; Choi, Mattila, & Upneja, 2018; Tassiello et al., 2018), we can learn a lot from lab-like studies investigating the underlying mechanisms. When different methods yield different results, trying to reconcile them enhances our understanding, giving robustness to our results.

Most importantly: choosing a research design is not a matter of personal preference. It should not be driven by the research design predominantly used in a certain field at a certain point in history, or by the pressure to minimize implementation effort to increase speed of publication. Rather, the key criterion is the suitability of the research design to answer the question asked. If the question is to explain the cause of something, a survey or observation will typically not be a suitable approach. For other question, of course, it may be. For example, a survey study is perfectly suitable to determine how tourists perceive a destination. And an observation is suitable for describing a specific behaviour. Ultimately, however, the suitability of the research design, and the quality of its implementation dictates whether valid and reliable conclusions can be drawn from any given study.

Appendix A. Experimental design checklist

Task	Examples	Things to consider (not necessarily in this order)	Check when done
Determine your dependent variable Y	The number of times a hotel room is booked. The number of clicks on a web link. The physiological activation of a person. The number of seconds a tourist looks at a piece of information.	Actual behaviour is a stronger dependent variable that any self-report measure. Self-reported measures can also suffer from social desirability bias.	☐
Determine your independent variable X	An advertisement for Brisbane (Australia). A price promotion. A pro-environmental message.		☐
Obtain ethics approval	Submit an ethics application to the relevant ethics authority. Human ethics requirements vary greatly by country and by university. To protect the researcher and the study participants, we recommend adhering to the strictest available guidelines.	In many countries any research involving humans needs to be approved *before* study participants are contacted.	☐
Prove that X has the intended effect (manipulation check) which is hypothesized to cause a change in Y	If an advertisement is meant to increase people's perception of a city as friendly, do people who have seen the advertisement rate that city as more friendly? If the pro-environmental message is intended to explain harm done to the environment by a specific action, the manipulation check must prove that the message indeed achieves this.	Use the same types of participants that are used in the experiment. If the intervention does not work as intended, revise until you can prove it does, or abandon the experiment. Do not run the experiment with an independent variable (intervention) that has not been proven to work in a manipulation check.	☐
Determine possible confounding variables	Age may affect technology adoption. Income may affect sensitivity towards price promotions.	Make sure that you consider anything that could affect Y other than X. You will need to collect data on possible confounding variables, and include them in the data analysis to ensure you are not drawing incorrect conclusions about the effect of X on Y.	☐
Determine the experimental setting (field experiment or laboratory experiment)	Field experiment • A hotel • A tourist destination • An airline's flight booking webpage Laboratory experiment • An eye-tracking laboratory • A mock up travel agency • A series of price quality options to pick from	You have more control over an experiment in the laboratory. This means the internal validity is higher because there are fewer external factors affecting the result. This internal validity comes at the expense of lower external validity or generalizability. Field experiments are "messier" because not everything can be controlled for, but results have higher external validity (generalizability).	☐
Determine who the study participants will be (not required for field experiments)	Students Tourists Managers of tourism businesses Employees of tourism businesses	For field experiments study participants, by definition, are people who are usually present in this environment. For laboratory experiments, researchers select participants. The profile of participants should be as close as possible to the profile of the people the intervention is targeted at to increase external validity.	☐
Determine if study participants can be randomly assigned to experimental conditions	Assign two different sustainability types of communications randomly across tourists. Assigning randomly participants to different price conditions.	If it is not possible to assign participants randomly, consider whether to run quasi experiments or discrete choice experiments.	☐
Determine how many observations are needed	Run a pilot study before running the experiment	The required sample size depends on the expected effect size (how much X is expected to affect Y). The smaller the expected effect size, the higher the sample requirements.	☐
Determine how many groups are needed	One control and one treatment group or several treatment groups	The more the groups, the higher the number of the participants.	☐
Conduct the experiment and analyse the data	The experiment should integrate easily into what is already occurring. Look for natural opportunities to experiment (pilots, new program, gradual rollout, etc.). Analyse the data with a statistical software (R, SPSS, SAS, STATA etc.).	It is critically important that all potential confounding variables are included. These may include less obvious things as weather, the number of hours of daylight, etc.	☐

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211.
Alasuutari, P., Bickman, L., & Brannen, J. (Eds.). (2008). *The SAGE handbook of social research methods*. New York: Sage.
Albaladejo-Pina, I. P., & Díaz-Delfa, M. T. (2009). Tourist preferences for rural house stays: Evidence from discrete choice modelling in Spain. *Tourism Management, 30*(6), 805–811.
Antón, C., Camarero, C., & Garrido, M. J. (2018). A journey through the museum: Visit factors that prevent or further visitor satiation. *Annals of Tourism Research, 73*,

48–61.
Araña, J. E., & León, C. J. (2008). The impact of terrorism on tourism demand. *Annals of Tourism Research, 35*(2), 299–315.
Araña, J. E., & León, C. J. (2013). Can defaults save the climate? Evidence from a field experiment on carbon offsetting programs. *Environmental and Resource Economics, 54*(4), 613–626.
Araña, J. E., & León, C. J. (2016). Are tourists animal spirits? Evidence from a field experiment exploring the use of non-market based interventions advocating sustainable tourism. *Journal of Sustainable Tourism, 24*(3), 430–445.
Araña, J. E., León, C. J., Moreno-Gil, S., & Zubiaurre, A. R. (2013). A comparison of tourists' valuation of climate change policy using different pricing frames. *Journal of Travel Research, 52*(1), 82–92.
Babakhani, N., Ritchie, B. W., & Dolnicar, S. (2017). Improving carbon offsetting appeals in online airplane ticket purchasing: Testing new messages, and using new test methods. *Journal of Sustainable Tourism, 25*(7), 955–969.
Baca-Motes, K., Brown, A., Gneezy, A., Keenan, E. A., & Nelson, L. D. (2012). Commitment and behavior change: Evidence from the field. *Journal of Consumer Research, 39*(5), 1070–1084.
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.
Becken, S., & Wilson, J. (2007). Trip planning and decision making of self-drive tourists— quasi-experimental approach. *Journal of Travel & Tourism Marketing, 20*(3-4), 47–62.
Blake, J. (1999). Overcoming the "value-action gap" in environmental policy: Tensions between national policy and local experience. *Local Environment, 4*(3), 257–278.
Boavida-Portugal, I., Ferreira, C. C., & Rocha, J. (2017). Where to vacation? An agent-based approach to modelling tourist decision-making process. *Current Issues in Tourism, 20*(15), 1557–1574.
Bogicevic, V., Seo, S., Kandampully, J. A., Liu, S. Q., & Rudd, N. A. (2019). Virtual reality presence as a preamble of tourism experience: The role of mental imagery. *Tourism Management, 74*, 55–64.
Box, G. E., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for experimenters. In Wiley Series in Probability and Statistics.* NJ, USA: Wiley Hoboken.
Brau, R. (2008). Demand-driven sustainable tourism? A choice modelling analysis. *Tourism Economics, 14*(4), 691–708.
Brown, A., Kappes, J., & Marks, J. (2013). Mitigating theme park crowding with incentives and information on mobile devices. *Journal of Travel Research, 52*(4), 426–436.
Calder, B. J., Phillips, L. W., & Tybout, A. M. (1981). Designing research for application. *Journal of Consumer Research, 8*(2), 197–207.
Carmines, E. G., & Zeller, R. A. (1979). *Reliability and Validity Assessment.* Newbury Park: Sage.
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.
Chaminuka, P., Groeneveld, R. A., Selomane, A. O., & Van Ierland, E. C. (2012). Tourist preferences for ecotourism in rural communities adjacent to Kruger National Park: A choice experiment approach. *Tourism Management, 33*(1), 168–176.
Chan, W., Lee, S. C., Hon, A., Liu, L., Li, D., & Zhu, N. (2015). Management learning from air purifier tests in hotels: Experiment and action research. *International Journal of Hospitality Management, 44*, 70–76.
Chen, N., Masiero, L., & Hsu, C. H. (2018). Chinese outbound tourist preferences for all-inclusive group package tours: A latent class choice model. *Journal of Travel Research, 47287518794331*.
Chen, Y., Lehto, X. Y., & Cai, L. (2013). Vacation and well-being: a study of Chinese tourists. *Annals of Tourism Research, 42*, 284–310.
Choi, C., Mattila, A. S., & Upneja, A. (2018). The effect of assortment pricing on choice and satisfaction: The moderating role of consumer characteristics. *Cornell Hospitality Quarterly, 59*(1), 6–14.
Cohen, J. (1988). The effect size index: d. Statistical Power Analysis for the Behavioral. *Sciences, 2*, 284–288.
Cornelis, P. C. (2010). Effects of co-branding in the theme park industry: a preliminary study. *International Journal of Contemporary Hospitality Management, 22*(6), 775–796.
Crouch, G. I., Oppewal, H., Huybers, T., Dolnicar, S., Louviere, J. J., & Devinney, T. (2007). Discretionary expenditure and tourism consumption: Insights from a choice experiment. *Journal of Travel Research, 45*(3), 247–258.
Crouch, G. I., Del Chiappa, G., & Perdue, R. R. (2019). International convention tourism: A choice modelling experiment of host city competition. *Tourism Management, 71*, 530–542.
Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS one, 8*(3), e57410.
Das, J., & Dirienzo, C. (2010). Tourism competitiveness and corruption: A cross-country analysis. *Tourism Economics, 16*(3), 477–492.
Di Fatta, D., Patton, D., & Viglia, G. (2018). The determinants of conversion rates in SME e-commerce websites. *Journal of Retailing and Consumer Services, 41*, 161–168.
Ding, M., Grewal, R., & Liechty, J. (2005). Incentive-aligned conjoint analysis. *Journal of Marketing Research, 42*(1), 67–82.
Dolnicar, S., Grabler, K., Grün, B., & Kulnig, A. (2011). Key drivers of airline loyalty. *Tourism Management, 32*(5), 1020–1026.
Dolnicar, S., & Ring, A. (2014). Tourism marketing research – past, present and future. *Annals of Tourism Research, 47*, 31–47.
Dolnicar, S., Coltman, T., & Sharma, R. (2015). Do satisfied tourists really intend to come back? Three concerns with empirical studies linking satisfaction to behavioral intentions. *Journal of Travel Research, 54*(2), 152–178.
Dolnicar, S., Knezevic Cvelbar, L., & Grün, B. (2017). A sharing-based approach to enticing tourists to behave more environmentally friendly. *Journal of Travel Research, 0047287517746013*.
Dolnicar, S., Knezevic Cvelbar, L., & Grün, B. (2017a). Do pro-environmental appeals trigger pro-environmental behavior in hotel guests? *Journal of Travel Research, 56*(8), 988–997.
Dolnicar, S., Knezevic Cvelbar, L., & Grün, B. (2018). Changing service settings for the environment: How to reduce negative environmental impacts without sacrificing tourist satisfaction. *Annals of Tourism Research, 76*(C), 301–304.
Engeset, M. G., & Elvekrok, I. (2015). Authentic concepts: Effects on tourist satisfaction. *Journal of Travel Research, 54*(4), 456–466.
Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management, 55*, 62–73.
Figini, P., Castellani, M., & Vici, L. (2009). Estimating tourism effects on residents: A choice modelling approach to the case of Rimini. *Advances in tourism economics* (pp. 145–164). Physica-Verlag HD.
Filimonau, V., Lemmer, C., Marshall, D., & Bejjani, G. (2017). 'Nudging' as an architect of more responsible consumer choice in food service provision: The role of restaurant menu design. *Journal of Cleaner Production, 144*, 161–170.
Fong, L. H. N., Law, R., Tang, C. M. F., & Yap, M. H. T. (2016). Experimental research in hospitality and tourism: a critical review. *International Journal of Contemporary Hospitality Management, 28*(2), 246–266.
Freedman, M. R., & Brochado, C. (2010). Reducing portion size reduces food intake and plate waste. *Obesity, 18*(9), 1864–1866.
Fritz, M. S., & Lester, H. F. (2016). *Mediator Variables.* Oxford (UK): In Oxford Research Encyclopedia of Psychology.
Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation.* New York: WW Norton.
Gneezy, A., Gneezy, U., Nelson, L. D., & Brown, A. (2010). Shared social responsibility: A field experiment in pay-what-you-want pricing and charitable giving. *Science, 329*(5989), 325–327.
Gneezy, A. (2017). Field experimentation in marketing research. *Journal of Marketing Research, 54*(1), 140–143.
Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research, 35*(3), 472–482.
Goodman, J. K., & Paolacci, G. (2017). Crowdsourcing consumer research. *Journal of Consumer Research, 44*(1), 196–210.
Gössling, S., Araña, J. E., & Aguiar-Quintana, J. T. (2019). Towel reuse in hotels: Importance of normative appeal designs. *Tourism Management, 70*, 273–283.
Grazzini, L., Rodrigo, P., Aiello, G., & Viglia, G. (2018). Loss or gain? The role of message framing in hotel guests' recycling behaviour. *Journal of Sustainable Tourism,*

26(11), 1944–1966.
Grigolon, A. B., Borgers, A. W., Kemperman, A. D., & Timmermans, H. J. (2014). Vacation length choice: A dynamic mixed multinomial logit model. *Tourism Management, 41*, 158–167.
Hahm, J., & Wang, Y. (2011). Film-induced tourism as a vehicle for destination marketing: Is it worth the efforts? *Journal of Travel & Tourism Marketing, 28*(2), 165–179.
Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature, 42*(4), 1009–1055.
Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication monographs, 76*(4), 408–420.
Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis second edition: A regression-based approach.* New York: Guilford Publications.
Hsieh, C. C., Babcock, R. W., Jr, & Stenstrom, M. K. (1993). Estimating emissions of 20 VOCs. II: Diffused aeration. *Journal of Environmental Engineering, 119*(6), 1099–1118.
Huang, M. C. J., Cheng, Y. H., Chuang, S. C., & Kuo, K. S. (2016). Framing makes tourists more environmentally conservative. *Annals of Tourism Research, 61*(C), 242–244.
Huertas-Garcia, R., Laguna García, M., & Consolación, C. (2014). Conjoint analysis of tourist choice of hotel attributes presented in travel agent brochures. *International Journal of Tourism Research, 16*(1), 65–75.
Huybers, T. (2003a). Modelling short-break holiday destination choices. *Tourism Economics, 9*(4), 389–405.
Huybers, T. (2003b). Domestic tourism destination choices—a choice modelling analysis. *International Journal of Tourism Research, 5*(6), 445–459.
Huybers, T., & Bennett, J. (2000). Impact of the environment on holiday destination choices of prospective UK tourists: implications for Tropical North Queensland. *Tourism Economics, 6*(1), 21–46.
Hwang, Y., & Mattila, A. S. (2018). Is it my luck or loyalty? The role of culture on customer preferences for loyalty reward types. *Journal of Travel Research, 57*(6), 769–778.
Hyman, R. (1982). Quasi-experimentation: Design and analysis issues for field settings. *Journal of Personality Assessment, 46*(1), 96–97.
Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5–86.
Jacob, C., & Guéguen, N. (2012). Exposition to altruism quotes and helping behavior: a field experiment on tipping in a restaurant. *Annals of Tourism Research, 39*(3), 1694–1698.
John, L. K., Emrich, O., Gupta, S., & Norton, M. I. (2017). Does "liking" lead to loving? The impact of joining a brand's social network on marketing outcomes. *Journal of Marketing Research, 54*(1), 144–155.
Jun, S. H., & Vogt, C. (2013). Travel information processing applying a dual-process model. *Annals of Tourism Research, 40*, 191–212.
Juvan, E., & Dolnicar, S. (2014). The attitude–behaviour gap in sustainable tourism. *Annals of Tourism Research, 48*, 76–95.
Juvan, E., Grün, B., & Dolnicar, S. (2018). Biting off more than they can chew: food waste at hotel breakfast buffets. *Journal of Travel Research, 57*(2), 232–242.
Kahneman, D. (2011). *Thinking, fast and slow.* New York: Macmillan.
Kallbekken, S., & Sælen, H. (2013). 'Nudging' hotel guests to reduce food waste as a win–win environmental measure. *Economics Letters, 119*(3), 325–327.
Kang, M., & Gretzel, U. (2012). Effects of podcast tours on tourist experiences in a national park. *Tourism Management, 33*(2), 440–455.
Karlsson, L., & Dolnicar, S. (2016). Does eco certification sell tourism services? Evidence from a quasi-experimental observation study in Iceland. *Journal of Sustainable Tourism, 24*(5), 694–714.
Karlsson, L., Kemperman, A., & Dolnicar, S. (2017). May I sleep in your bed? Getting permission to book. *Annals of Tourism Research, 62*, 1–12.
Kim, D., & Park, B. J. R. (2017). The moderating role of context in the effects of choice attributes on hotel choice: A discrete choice experiment. *Tourism Management, 63*, 439–451.
Kim, J., & Lakshmanan, A. (2015). How kinetic property shapes novelty perceptions. *Journal of Marketing, 79*(6), 94–111.
Kim, J. H., & Jang, S. S. (2014). A scenario-based experiment and a field study: A comparative examination for service failure and recovery. *International Journal of Hospitality Management, 41*, 125–132.
Kim, S. B., Kim, D. Y., & Bolls, P. (2014). Tourist mental-imagery processing: Attention and arousal. *Annals of Tourism Research, 45*, 63–76.
Lawson, J. (2014). *Design and Analysis of Experiments with R.* London: Chapman and Hall/CRC.
Li, Q., McCabe, S., & Xu, H. (2019). Always best or good enough? The effect of 'mind-set' on preference consistency over time in tourist decision making. *Annals of Tourism Research, 75*, 186–201.
Li, S., Walters, G., Packer, J., & Scott, N. (2017). A comparative analysis of self-report and psychophysiological measures of emotion in the context of tourism advertising. *Journal of Travel Research, 0047287517733555*.
Lin, M. L., Yang, J. T., & Wan, C. S. (2015). Effect of restaurant discount coupon depth on re-consumption willingness: A moderating role of brand image. *Tourism and Hospitality Research, 15*(3), 193–205.
Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: analysis and applications.* Cambridge (UK): Cambridge University Press.
Mair, J., & Bergin-Seers, S. (2010). The effect of interventions on the environmental behaviour of Australian motel guests. *Tourism and Hospitality Research, 10*(4), 255–268.
March, R., & Woodside, A. G. (2005). Testing theory of planned versus realized tourism behavior. *Annals of Tourism Research, 32*(4), 905–924.
Masiero, L., & Nicolau, J. L. (2012). Tourism market segmentation based on price sensitivity: Finding similar price preferences on tourism activities. *Journal of Travel Research, 51*(4), 426–435.
McFadden, D. (1980). Econometric models for probabilistic choice among products. *Journal of Business,* S13–S29.
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.
Morley, C. L. (1994). Experimental destination choice analysis. *Annals of Tourism Research, 21*(4), 780–791.
Nguyen, Q. (2016). Linking loss aversion and present bias with overspending behavior of tourists: Insights from a lab-in-the-field experiment. *Tourism Management, 54*, 152–159.
Nicholls, S., Amelung, B., & Student, J. (2017). Agent-based modeling: A powerful tool for tourism researchers. *Journal of Travel Research, 56*(1), 3–15.
Nicolau, J. L., & Mas, F. J. (2006). The influence of distance and prices on the choice of tourist destinations: The moderating role of motivations. *Tourism Management, 27*(5), 982–996.
Nicolau, J. L., & Sellers, R. (2012). The free breakfast effect: an experimental approach to the zero price model in tourism. *Journal of Travel Research, 51*(3), 243–249.
Okazaki, S., & Hirose, M. (2009). Does gender affect media choice in travel information search? On the use of mobile Internet. *Tourism Management, 30*(6), 794–804.
Oppewal, H., Huybers, T., & Crouch, G. I. (2015). Tourist destination and experience choice: A choice experimental analysis of decision sequence effects. *Tourism Management, 48*, 467–476.
Otter, T., Pachali, M. J., Mayer, S., & Landwehr, J. (2018). Causal inference using mediation analysis or instrumental variables-full mediation in the absence of conditional independence. Available at http://dx.doi.org/10.2139/ssrn.3135313, last accessed 21.12.2019.
Pabel, A., & Pearce, P. L. (2016). Tourists' responses to humour. *Annals of Tourism Research, 57*, 190–205.
Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163.
Pera, R., Viglia, G., Grazzini, L., & Dalli, D. (2019). When empathy prevents negative reviewing behavior. *Annals of Tourism Research, 75*, 265–278.
Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891.
Rashidi, T. H., & Koo, T. T. (2016). An analysis on travel party composition and expenditure: a discrete-continuous model. *Annals of Tourism Research, 56*, 48–64.
Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods, 5*(2), 199–213.
Reese, G., Loew, K., & Steffgen, G. (2014). A towel less: Social norms enhance pro-environmental behavior in hotels. *The Journal of Social Psychology, 154*(2), 97–100.
Reiser, A., & Simmons, D. G. (2005). A quasi-experimental method for testing the effectiveness of ecolabel promotion. *Journal of Sustainable Tourism, 13*(6), 590–616.

Reynolds, D., Merritt, E. A., & Pinckney, S. (2005). Understanding menu psychology: An empirical investigation of menu design and consumer response. *International Journal of Hospitality & Tourism Administration, 6*(1), 1–9.
Rodríguez-Molina, M. A., Frías-Jamilena, D. M., & Castañeda-García, J. A. (2015). The contribution of website design to the generation of tourist destination image: The moderating effect of involvement. *Tourism Management, 47*, 303–317.
Rong-Da Liang, A. (2017). Considering the role of agritourism co-creation from a service-dominant logic perspective. *Tourism Management, 61*, 354–367.
Rossiter, J. R. (2001). What is marketing knowledge? Stage I: forms of marketing knowledge. *Marketing Theory, 1*, 9–26.
Rossiter, J. R. (2002). The five forms of transmissible, usable marketing knowledge. *Marketing Theory, 2*, 369–380.
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.
Schram, A. (2005). Artificiality: The tension between internal and external validity in economic experiments. *Journal of Economic Methodology, 12*(2), 225–237.
Seltman, H. J. (2012). Experimental design and analysis. Online at: http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf, last accessed 21.12.2019.
Šerić, M., & Praničević, D. G. (2018). Consumer-generated reviews on social media and brand relationship outcomes in the fast-food chain industry. *Journal of Hospitality Marketing & Management, 27*(2), 218–238.
Smith, A. (2006). Assessing the contribution of flagship projects to city image change: a quasi-experimental technique. *International Journal of Tourism Research, 8*(6), 391–404.
Sparks, B. A., Perkins, H. E., & Buckley, R. (2013). Online travel reviews as persuasive communication: The effects of content type, source, and certification logos on consumer behavior. *Tourism Management, 39*, 1–9.
Stopher, P., & Greaves, S. (2009). Missing and inaccurate information from travel surveys – pilot results. 32nd Australasian Transport Research Forum (ATRF) 2009, Auckland, New Zealand, 1st October 2009.
Tasci, A. D., & Guillet, B. D. (2011). It affects, it affects not: A quasi-experiment on the transfer effect of co-branding on consumer-based brand equity of hospitality products. *International Journal of Hospitality Management, 30*(4), 774–782.
Tassiello, V., Viglia, G., & Mattila, A. S. (2018). How handwriting reduces negative online ratings. *Annals of Tourism Research, 73*, 171–179.
Tessitore, T., Pandelaere, M., & Van Kerckhove, A. (2014). The amazing race to India: Prominence in reality television affects destination image and travel intentions. *Tourism Management, 42*, 3–12.
Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* Cambridge: Cambridge University Press.
Tu, Y., Neuhofer, B., & Viglia, G. (2018). When co-creation pays: stimulating engagement to increase revenues. *International Journal of Contemporary Hospitality Management, 30*(4), 2093–2111.
Viglia, G., & Abrate, G. (2014). How social comparison influences reference price formation in a service context. *Journal of Economic Psychology, 45*, 168–180.
Viglia, G., Mauri, A., & Carricano, M. (2016). The exploration of hotel reference prices under dynamic pricing scenarios and different forms of competition. *International Journal of Hospitality Management, 52*, 46–55.
Viglia, G., Maras, M., Schumann, J., & Navarro-Martinez, D. (2019). Paying before or paying after? Timing and uncertainty in pay-what-you-want pricing. *Journal of Service Research, 1094670519835308*.
Walters, G., Wallin, A., & Hartley, N. (2019). The threat of terrorism and tourist choice behavior. *Journal of Travel Research, 58*(3), 370–382.
Woodside, A. G. (2010). Tourism advertising and marketing performance metrics. In: Tourism-marketing performance metrics and usefulness auditing of destination websites (pp. 1-14). Emerald Group Publishing Limited. Bingley (UK).
Woodside, A. G., III, T, R. J., & MacDonald, R. (1997). Measuring linkage-advertising effects on customer behaviour and net revenue: Using quasi-experiments of advertising treatments with novice and experienced product-service users. Canadian. *Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration, 14*(2), 214–228.
Wu, L., Shen, H., Fan, A., & Mattila, A. S. (2017). The impact of language style on consumers' reactions to online reviews. *Tourism Management, 59*, 590–596.
Yang, S. S. (2012). Eye movements on restaurant menus: A revisitation on gaze motion and consumer scanpaths. *International Journal of Hospitality Management, 31*(3), 1021–1029.
Zavattaro, S. M., & Fay, D. L. (2019). Brand USA: A natural quasi-experiment evaluating the success of a national marketing campaign. *Tourism Management, 70*, 42–48.
Zhao, X., Lynch, J. G., Jr, & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of consumer research, 37*(2), 197–206.