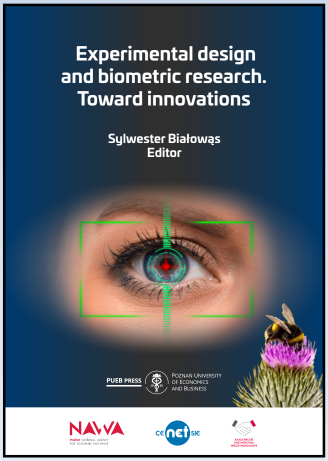

Experimental design and biometric research. Toward innovations

Sylwester Białowąs
Editor

POLISH NATIONAL AGENCY FOR ACADEMIC EXCHANGE

AKADEMICKIE PARTNERSTWA MIĘDZYNARODOWE

eISBN 978-83-8211-079-1

https://doi.org/10.18559/978-83-8211-079-1

PUEB PRESS

Poznań University of Economics and Business

Experimental design and biometric research. Toward innovations, pp. 9-35
https://doi.org/10.18559/978-83-8211-079-1/I1

1.

EXPERIMENTAL DESIGN

Sylwester Białowąs, Atanaska Reshetkova, Adrianna Szyszka
Poznań University of Economics and Business

Atanaska Reshetkova
D.A. Tsenov Academy of Economics

Adrianna Szyszka
Poznań University of Economics and Business

Abstract: Experiment is a research method appropriate to examine causal relationships, also in relation to the current problems of science, including sustainable development.

Conducting experiments can take place in laboratory conditions, but also in natural environments. The main objective of an experiment is always to test what the researcher actually wants and to obtain results that can be generalised to the entire population. In other words, planning experiments requires considering many aspects related to their internal and external validity. The key aspect that needs to be considered in conducting experiments is proper problem defining, as well as the concepts of causality, manipulation or null and alternative hypotheses. It is also worth bearing in mind that in social sciences, when engaging participants in research, caution must be exercised. Depending on whether each participant of the experiment is exposed to all conditions or different people test different ones, the classification of experiments is distinguished into within-subjects and between-subjects design. In this chapter, the most commonly used experimental designs in this division are presented. However, the experimental method offers more complex schemes such as randomised block design or Latin square design. Finally, the obtained findings should be properly presented—in the form of a report following APA standards.

Keywords: experiment, experimental design, randomisation, validity.

Sylwester Białowąs, Atanaska Reshetkova, Adrianna Szyszka

1.1. Introduction to the experimental method

1.1.1. The definition of experiment

The experimental model is the best way to check research hypotheses about cause and effect relationships between variables. Experiments allow a researcher to observe and influence a specific phenomenon. Conducting an experiment requires a precise definition of the problem under investigation, as well as analysis of the conditions related to the phenomenon (Stachak, 1997, p. 146). Thus, an experiment should be defined as a conscious, purposeful and planned evocation of specific states and processes. While performing the experiment, the researcher deliberately changes certain factors in the examined situation and, at the same time, controls other factors in order to learn what the effects of the undergoing change are in the course of observation (Sułek, 1979).

An experiment depends on manipulating one or more independent variables and measuring their effect on one or more dependent variables while simultaneously controlling the effect of extraneous variables (Burns, Veeck, & Bush, 2017; Malhotra, Nunan, & Birks, 2017). In other words, experiments help discover causal relationships and ensure that the observed effect concerning the dependent variable is because of the independent variable and not due to other aspects (extraneous variables) (Burns et al., 2017). For example, if checking the effectiveness of fertiliser for plants, in the experiment, two flowers are planted separately and their growth is observed for 60 days, with one of the plants being additionally fertilised.

The three key concepts related to experimental research are:

- independent variable **manipulation** (in the example, one of the flowers will be additionally fertilised);
- **control** of extraneous variables, which may be relevant for the independent variable (in this example, it must be certain that the two flowers grow under similar conditions, i.e. the flowers are planted close together, have the same lighting parameters and get the same amount of water every day);
- variability **measurement** of a dependent variable resulting from the researcher's influence on this variable using independent variables (Brzeziński, 1999, p. 282).

An important distinction is the division of experiments into those performed at a laboratory and in the field. In the case of laboratory experiments, the researcher creates an artificial environment meeting conditions for the tested problem. Field experiments are carried out under real market conditions (Malhotra et al., 2017). Conducting experiments in natural settings creates a more realistic environment but is more expensive and time-consuming in comparison to laboratory experiments (Hair, Bush, & Ortinau, 2003).

1.1.2. Experiments and other methods of scientific research

There are three types of research designs, i.e. exploratory, descriptive and causal. Together, descriptive and causal methods are called conclusive (Malhotra et al., 2017). The main difference between these categories regards their goals—in exploratory studies, the researchers aim to understand the nature of the problem, while in those conclusive—measuring the phenomenon and examining dependencies are of concern (Malhotra et al., 2017). Exploratory research designs are used to clarify and define the problem, obtain additional insights and formulate research objectives. It is especially helpful when little is known about the investigated phenomenon (Burns et al., 2017). Exploratory methods find their application when the problem is difficult to be measured quantitatively (Malhotra et al., 2017).

Descriptive studies provide information about certain aspects of the problem: who, what, where, when and how. This research design allows researchers to describe and measure the phenomena (Burns et al., 2017). This process should be preceded by formulating a hypothesis and defining a problem. This is usually the description of market characteristics or functions that are planned, structured and based on a representative sample (Malhotra et al., 2017).

The last category—causal research designs, enable the measurement of causality in relationships which can be observed when one (or more) variables affect one (or more) variables (Burns et al., 2017). Experiments are the primary method among causal research designs (Malhotra et al., 2017), providing the researcher with the ability to answer the question as to why something occurs and why it may be observed under specific conditions. Examining cause and effect dependencies further allows the researchers to make predictions about various phenomena occurring on the market (Hair et al., 2003). The experiments are considered as research designs that measure the causes and effects of the variables most accurately. Non-experimental studies that are also used for examining cause and effect relationships sometimes do not fulfil all the aspects of causality (Malhotra et al., 2017).

Advantages of the experimental method include:

- enabling the verification of cause-effect relationships relatively easily in comparison to other methods (Moore, McCabe, Alwan, Craig, & Duckworth, 2011);
- helping the researcher to control the experimental conditions and factors that are not significant for the study (Moore et al., 2011);
- easy replication (experiments are repeated more often than other methods), proving the experiment's accuracy (Sułek, 1979);
- making it possible to study the simultaneous influence of more than one factor—separately, the variables may affect a dependent variable in a different way than their interaction (Moore et al., 2011).

Limitations of the experimental method include:

- the researcher potentially not being able to control extraneous variables, particularly in field experiments (Malhotra et al., 2017);
- being time-consuming—especially when the effects of the manipulation are examined in the long run (e.g. effectiveness of an advertising campaign) (Malhotra et al., 2017);
- conducting the experiment on many occasions being relatively expensive (Malhotra et al., 2017);
- potential ethical implications of the conducted experiments (Burns et al., 2017);
- the experiment's results potentially being affected by the artificiality of an experimental situation (Moore et al., 2011).

1.1.3. Research design: type of data

In research design, there are two types of data. The first category refers to the kind that is not obtained by the researcher in the research project or that has been collected for other purposes. This group of data sources refers to surveys and records that are prepared by different companies or organisations. If this data is publicly available, the researcher may use it for research purposes (Burns et al., 2017). It is usually in the form of written documents and is referred to as "desk research". Secondary data is helpful both in defining the objectives of a study and confronting the obtained results. The fact that the data was collected previously by someone else makes it relatively inexpensive and easy to access. On the other hand, since it has been collected for different purposes, the secondary data may cover issues that do not fit perfectly with the research objectives. There is also a risk that this information will be out of date. The secondary data may be obtained from government sources represented by statistical departments. Information is also provided by academic sources and company documents or annual reports. Secondary data come from market research publishers, organisation websites and even private citizens (Bridley, 2013).

Primary data is collected intentionally for a specific purpose. The researcher obtains this kind of data while conducting the research project (Bridley, 2013; Burns et al., 2017). She/he has various possibilities to contact participants and gain the information. This may take place by phone, e-mail, post and/or in person. Among the forms of this kind of data, we can distinguish interviewing and self-completion methods (Bridley, 2013). Experimentation also belongs to this category as a form of gathering primary quantitative data (Malhotra et al., 2017).

1.1.4. Application in economics and management

The experimental method (including field experiments) is commonly used in marketing research, especially in the aspects of communications and advertising (Malhotra et al., 2017). In this area, a popular practice is test marketing which is a type of field experiment. The researchers mostly use test marketing for two purposes—to evaluate the sales potential of a product or service or to test elements of the marketing mix. Furthermore, test markets are used to assess media usage, prices or sales promotions. Although this practice may be expensive for the company, it enables in-advance testing if the product may succeed in the market. There are four main types of test markets—standard, controlled, electronic and stimulated. Standard test markets may provide reliable results because they are conducted in real settings, i.e. using the regular distribution channels of the company. In a controlled model, the experiments are conducted by out-company research firms that test the adjusted distribution channels (Burns et al., 2017). Electronic test markets depend on gathering data from consumers who use an identification card that registers the purchase of goods or services. In simulated test markets, researchers interview selected participants and observe their purchasing behaviours as well as attitude towards the product (Hair et al., 2003).

Experiments are successfully implemented in the area of organisational research. For example, in examining issues related to work efficiency, the profitability of an organisation, the level of task performance or the attitudes and satisfaction of employees (Stachak, 1997). Another field in which companies use experiments is consumer behaviour.

1.2. Key concepts prior to planning an experiment

1.2.1. Causality

Identifying causal relationships is one of the most interesting yet challenging research goals in any scientific field. The concept of causality is rather simple to understand: when one phenomenon is the reason why another manifests, then we have a cause and its effect. How to validate this assumed causal relationship is a question requiring more attention. As in any other research methodology, the experimental method has certain significant criteria that need to be met in order to conclude that a causal relationship really exists.

In any book of statistics and/or research methodology, it is said that correlation does not necessarily mean causation. Two variables can be associated and this still might not mean that one of them is responsible for the changes in the values (levels) of the other. There may be, for example, a third variable influencing both of them, and thus—making them seem like a cause and its effect, when they are actually both an effect of another cause variable. The pollution of oceans and of air are correlated, but that does not mean that one is a cause of the other. However, the correlation between two variables is the first criterion to pronounce their relationship for causation. In other words, the researcher must be sure that there is an observed association between the cause variable (also called independent variable), and the effect variable (dependent variable).

Another important condition is that the cause must precede the effect. Only this way can we assume that the variation in the independent variable is the cause for variation in the dependent one. For example, suppose we are trying to prove that bad marks at school make admission to college harder. The marks should be received before students send their application to a given college—any other way around contradicts common logic. There are many cases in which it is difficult to decide which came first which makes pointing to the cause and effect variables difficult.

To conclude that there is a causal relationship between any two phenomena, it is necessary to manipulate the impact of the influencing factor, to control all other factors that may influence the test subjects and to compromise the validity of the results. Of course, this is possible only if we conduct an experimental study to conclude causality. The adoption of any other research strategy may lead to the assumption of some correlational degree between variables, which solely, cannot serve as an argument for causality. The choice of experimental design plays a critical role in drawing a conclusion about the causality.

Finally, testing for causality requires the influencing factor to have at least two levels to compare their effects on the response variable.

1.2.2. Independent and dependent variables

As previously stated, in experimental data we can identify two types of variables: dependent and independent. The independent variables are those manipulated by the researcher with the expectation that they will cause some effect on the experimental subjects. If no effect is observed, the reason could be either that there is no causal relationship or that the manipulation of the independent variable was not done properly. For example, the researcher may choose to experiment with only two levels of the variable while there are three or more that can be tested. The dependent variables represent the response and their values are expected to be a result of independent variable manipulation. In social experiments, dependent variables may measure, for example, the actual or intended behaviour of participants, or different psychological processes. Choosing the right dependent variables that can actually capture the supposed effect is just as important as the right manipulation of the independent variables. Sometimes more than one dependent variable can be used in order to ensure accurate measurement. The manipulation and application of the independent variable on the participants is often called 'treatment'.

The goal of the experiment is to determine the function *f* that relates the dependent variable *y* to the independent variables $x_1, x_2, x_3 \ldots, x_k$, i.e., the cause and effect:

$$y = f(x_1, x_2, x_3 \ldots, x_k)$$

The independent variables are also called factors. There are different types of factors:

- Continuous factors—these are variables that can assume any value in a given interval. Values taken on by continuous factors are therefore represented by continuous numbers.
- Discrete factors—these can assume only a limited number of values. Values taken on by discrete factors can be names or words. Numbers are usually used as codes or labels, and not to denote quantity. For example, the type of labelling used on a product—bio, natural, eco—represents the possible values (levels) that a discrete factor can assume.
- Ordinal factors—these are discrete factors that can be put in a logical order. For example, the ranking of some objects as first, second and third is an ordinal factor. Size defined as small, medium and large is also this type of factor.

It should be noted that some continuous factors can be transformed into discrete or ordinal ones by creating two or more categories. For example, age is a continuous factor that can be transformed into an ordinal one with three levels: young adults, adults, seniors.

Very often, researchers are interested in testing the effect of more than one independent variable at the same time. In experiments with two or more factors, an interaction between these factors can be observed. If there is an interaction between two independent variables, the effect produced by one of them is different at each level of the other one. For example, let us suppose that we want to compare the effect of different product packaging on consumer perception of its healthiness. We decide to manipulate two independent variables of the packaging—labelling (bio, natural, eco) and colour (green and blue). If the labelling has a different effect on consumers' perceptions when the package is green and when it is blue, then this means that the two factors are in interaction. The mean values regarding the stated perceptions of the product healthiness across different levels of independent variables are presented in Figure 1 (a).

When the package is green and the product is labelled as “bio”, it is perceived as healthier than when the labelling is “natural” or “eco”. But when the package is blue, the product is perceived as healthiest if it is labelled as “eco”. We can conclude that there is an interaction between these two factors. It should be borne in mind that when there is a significant interaction between the independent variables, it is only meaningful to interpret the interaction effects on the dependent variables.

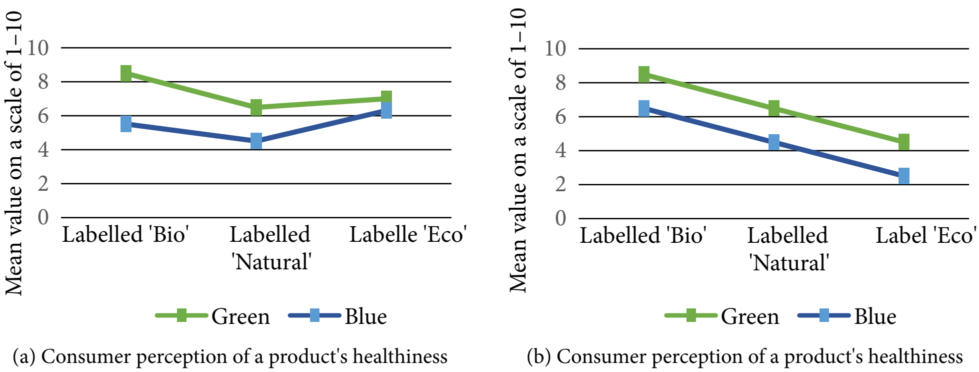

Figure 1. Interaction between independent variables (a). No interaction between independent variables (b)

Source: Own elaboration.

If no interaction exists, then each independent variable effect is interpreted. With regard to the previous example, if there is no interaction between the factors, the results would look like those presented in Figure 1 (b). The product is perceived as healthiest when it is labelled “bio”, followed by “natural” and “eco”. This does not change when the colour of the packaging changes from green to blue. When the packaging is green, the product is perceived as healthier than when it is blue, regardless of the labelling. These independent effects of each factors are called main effects.

1.2.3. Experimental and control groups

In order to test for a causal relationship between two variables, the one assumed as the cause should be manipulated and the participants exposed to its impact. Then, the response variable should be measured. Following, it needs to be concluded whether a significant effect is observed. There are different ways to organise the experiment depending on its goals and the studied phenomena. Sometimes, all participants are subjected to the same levels of the independent variables and other times—they are divided into groups, each group subjected to a different level of the independent variable. The specific way of dividing participants into groups and applying treatments is discussed later in this chapter. However, participants in every experiment can be distributed either in an experimental or in a control group. Every experiment has at least one experimental group which is exposed to one or more factor levels. A true experiment includes a control group as well. This is a group of participants not subject to the impact of any factor, thus, providing a baseline for comparison of the dependent variables' mean values.

1.2.4. Selecting research participants

While conducting research, the main purpose is to draw conclusions about the population that is the entire group under the study. However, markets sometimes consist of millions of individuals and the researcher is not able to carry out the experiment on the entire group. Conducting the experiment on the whole population would be time-consuming, expensive and also ineffective. Fortunately, in order to achieve research objectives, the researcher may use a sample (Burns et al., 2017). A sample is a subgroup of the population that represents the entire group and is selected for participation in the experiment. A representation should be suitable since the sample is designed to accurately reflect the characteristics of the group (Burns et al., 2017).

Sample sizes differ across different studies. In order to choose the number of elements to participate in a study, the researcher should take several aspects into account. It is crucial to consider the significance of the study, the number of variables, the sample sizes used in similar experiments and the available resources to conduct the study (Malhotra et al., 2017).

The way of assigning participants to samples induces the division into probability and non-probability sampling. In probability sampling, each unit has a specific chance of being assigned to the sample—the general probability is known. On the other hand, in the non-probability designs, the selection of a sample depends on subjective criteria. In this case, the researcher may correctly assign units to the sample based on his/her expertise, but, at the same time, there is no way to ensure that this selection will be free from bias (Mazzochi, 2008). Therefore, the main difference among both schemes depends on the intervention of the researcher. In probability sampling, the selection of participants is determined by the applied method (Burns et al., 2017). In this chapter, focus will be mostly on probability sampling methods. In this group, the following methods of selection for the sample can be distinguished: simple random, systematic, cluster and stratified sampling types. In the most basic of these methods—simple random sampling—participants have the same chances of being selected (the selection hinges on luck and probability). A similar procedure is systematic sampling, for which the participants are listed by the researcher who randomly selects only the first assigning number for the first unit in the sample. The rest of the participants is extracted consecutively following the selected starting point. In cluster sampling, the population is divided into complementary groups—each of them should reflect the population. The random selection applies to the clusters. This type of sampling is often an initial step in a more advanced procedure and is useful in relation to electronic databases (e.g. people whose name starts with the letter A, B, C, etc.) or geographical areas (cities, neighbourhoods). The method depending on dividing the population into groups is also stratified sampling. However, in the case of this sampling procedure, the groups are distinguished on the basis of the common characteristic so that the units are similar inside groups and heterogeneous among different strata. It is especially helpful when the distribution of the population is not normal.

1.3. Planning an experiment

1.3.1. Defining the problem and research questions

The key part of the research process is to properly define the problem. It is important because defining the problem influences the research questions, hypotheses and research procedure. If this part is not carried out correctly, there is a risk that we will not get answers to the issues that we want to examine.

In economic practice, a problem is often defined after failure to achieve a goal or after an opportunity has been identified. After that, managers aim to understand the background of the problem, define what decisions should be made and learn about additional sources of information to fully understand it (Burns et al., 2017). This may constitute a basis for scientific exploration. Therefore, the first step in the research process is initial observation and identification of an issue that needs explaining. After finding a lack of knowledge and solutions in some areas, exploratory research should be conducted. It will help the researcher to better structure the problem and clarify the scope of the investigation. The common practice is to investigate previous studies regarding the topic. This step mainly involves conducting a review of literature—before further exploring the problem it is essential to examine scientific publications, books and articles relevant to the issue. On this basis, specific research questions should be identified, and then—the resulting hypotheses (Zikmund, Babin, Carr, & Griffin, 2010).

1.3.2. Null and alternative hypotheses as well as significance

By conducting an experiment, the researcher tries to verify whether the empirical evidence is consistent with the assumed hypothesis. In statistics, it is impossible to show that any statement is undeniably true. However, there are ways to show that some dependencies are not true. In this part of the chapter, the main assumptions of statistical hypothesis testing are presented. Here, the null hypotheses can be distinguished according to whether there is no difference between the groups that are compared in the study. The null hypothesis allows to suggest that no effect is observed but the researcher usually wants to demonstrate that there are dependencies. Thus, there is an alternative hypothesis which predicts that there is a significant difference between the groups under study. This hypotheses is mainly the one that the researcher aims to support (Jackson, 2008). While conducting the experimental procedure, we want to reject the null hypothesis, which would indicate that the findings are consistent with the alternative hypothesis.

The concept of statistical significance is crucial in testing statistical hypotheses. If some difference is statistically significant, this means that it does not happen by chance. In social sciences, the usually chosen level of statistical significance (alpha level) is 5%. It allows the researcher to reject the null hypothesis and indicates that the probability that the tested dependence is due to chance is 5 in 100 (Jackson, 2008).

1.3.3. Data presentation and report structure (APA standards)

After planning, conducting the experiment and data analysis, the results should be properly presented. In this section, the principles of presenting and reporting results will be discussed. Creating the report follows the standards of scientific papers. The research should be fully clear for readers, the conclusions thoroughly explained and presented in a way that allows them to be compared to other studies. This is why the comprehensive standards of reporting are indispensable. The main rule is that all information relevant to the experiment should be included in the report.

The structure of a typical report follows the structure of scientific articles and is presented below:

- Introduction (literature review, main hypotheses)
- Method (design, participants, procedure)
- Results
- Discussion (interpretation, limitations)

The document should also follow the formal structure including: title, abstract, keywords, references and appendix.

Introduction

The first part of a report is the 'Introduction', in which the importance of the problem under study is shown. In this part, the 'Literature review' should also be presented—it is suggested to define the scope of the problem, its theoretical and practical aspects and to indicate what was the subject of research earlier and what remains unexplained. The main hypothesis should be formulated on the basis of the analysed theories. Thus, the introduction involves the description of the study goals as well.

Method

In the next part of the report, the implemented method(s) should be described. The 'Method' section should contain a description of the study participants, including information about their demographic characteristics (e.g. age, nationality, level of education), as well as aspects relevant to the study. Here, the procedures for selecting participants should be presented—the sampling method, time and place of collecting the data, agreements with participants and ethical and safety considerations. In the report, the number of participants taking part in the experiment, the number of participants in experimental and control groups as well as the number of participants that did not complete the experiment should be shown. The 'Method' section involves the inclusion and exclusion criteria for participants. Then, there should be a description of the sample—the number of participants in the study and the planned sample size. If such procedures were used, the methodological part of the report should include information on masking the purpose of the study, training to which collectors were subjected or additional methods. In this section, the research design (whether the between-subjects or within-subjects procedure was applied), the conditions of the study (natural or manipulated) and the assignment to different conditions (if applicable) are described. If the experiment includes manipulations/interventions, it should be precisely described what they consisted of and how they were applied—settings, the duration of exposure and the number of manipulations.

Results

The next section of the report focuses on the 'Results' section of the experiment. An accurate and impartial presentation of the results is the crucial part of the report. All the important results of the study should be presented with attention to detail and as clearly as possible. In the report, data that are not consistent with the assumed hypotheses should not be omitted—the insignificant dependencies and small effect sizes should be mentioned as well. Raw data and additional materials may be included in the 'Appendix'. When reporting the results, it is recommended to reflect the sequence of the hypotheses presented earlier. When it comes to statistical tests, reporting involves a sufficient set of statistics that are indispensable to understand the outcome. The description should include the value of the test statistic, the degrees of freedom, the *p* value and the magnitude of the effect. The measures of effect size may also be added to this section.

Discussion

The next part of the report regards the 'Discussion' section. The next step, after presenting the results, is to interpret them and draw conclusions from the conducted experiment. It is important to keep this section consistent with the previous one regarding the results. In this section of the report, it should be indicated whether the findings support or do not support the hypotheses. If contradictory or unclear results are obtained, possible causes need to be indicated. Moreover, in the report, the results obtained in relation to the studies of other researchers are presented and the observed differences and similarities are explained. In general, the main implications of the study should be emphasized. In this section, the limitations and strengths of the study are given.

Example

Perception time in forming attitudes towards art

Abstract: In the study, it is examined whether an extremely short exposure to stimuli enables the formulation of aesthetic judgments. In order to determine the time of aesthetic experience formation, an experiment has been conducted in which 12 paintings were displayed during 40 ms. In the previous study, 40 ms was assessed as the minimum exposure duration to process the visual stimuli. The initial judgments were confronted with the judgments formed after longer exposure (10 s). By comparing long- and short-term exposure, it is possible to establish consistency of the observed judgments. The database comprises pairs of works of art by the same artists with a similar composition and auctioned at similar prices, which makes it possible to assess the consistency of judgments with regard to a particular style. The experiment was conducted on a sample of 30 participants. The main findings allow to indicate that 40 ms is a sufficient time to formulate aesthetic judgment.

Keywords: art perception, formulating aesthetic judgments.

Introduction

When thinking of an aesthetic judgment, it must be considered how well a work of art expresses and influences others with feelings and emotion. The processes underlying the aesthetic experience have been described from both perceptual/cognitive and motivational viewpoints.

In previous research, it has been confirmed that ultra-short exposures (below 1 s) may be sufficient to formulate aesthetic judgements and attitudes. Cupchik and Berlyne (1979) assessed whether people are able to distinguish collative properties with presentation times of 50 ms. They have confirmed that this time allowed the participants to obtain relevant visual information. Locher, Krupinski, Mello-Thoms and Nodine (2007) noted that the time needed to form a significant holistic impression of the painting is about 100 ms.

The most extreme time range was tested in the study by Augustin, Leder, Hutzler and Carbon (2008). They found that 10-ms exposure may be enough to find traces of visual processing effects. In the same study, they confirmed such a significant effect after the presentation of 50 ms.

Main hypotheses

The previous study allows us to state that within the range of 50 to 100 ms, people are able to process visual stimuli and formulate judgment. We aimed to test if the shorter presentation time could be sufficient for similar effects to be observed.

The main hypothesis allows to indicate that a presentation time of 40 ms is sufficient to formulate aesthetic judgments.

Method

In the study the within-subjects, one-group pretest–posttest design was used. There was one independent variable (exposure time) with two levels (40 ms, 10 s). The dependent variable was the aesthetic pleasure measured as a self-reported assessment on the interval scale of 0 (not at all) to 10 (extremely pleasing).

Participants

The recruited participants were students from PUEB. The sample included 52 participants who were not selected randomly. There were 35 women and 17 men between the age of 18 and 31.

Procedure (including technical aspects)

The study took place under manipulated conditions—in the laboratory at the university. We have displayed the stimuli on a 75-inch screen in constant and dimmed light conditions. In the first stage of the experiment, each participant was shown 9 images for 40 ms each. After every stimulus, the participants evaluated their experience by answering the question as to whether the image was pleasing. Each image was preceded by one second of a grey screen with a cross sign in a circle (attention focusing point assuring same visual range for each picture). The second stage was a series of tasks not related to the experiment which was intended to clear the short-term memory of the previously seen stimuli. The third stage was conducted in the same manner as the first one, but the exposure time for each picture was 10 s.

Results

In order to test aesthetic judgment, it was decided to test if there was a difference between scores obtained for short- and long-time exposures. Due to the lack of normal distribution of differences, we decided to apply the Wilcoxon signed-rank test performed 12 times for each picture separately. In 11 cases, it was found that there were no differences between scores (see appendix)—p value was higher than the assumed alpha level ($p > .05$).

In one example, it was found that the evaluation of experience (whether the image was pleasing) changed significantly $Z(52) = -2.54$, $p = .011$. The evaluation that the image was pleasing for the 40-ms exposure time was higher ($Mdn = 5$) compared to the 10-s exposure ($Mdn = 4$). However, the effect size was rather small ($r = .25$).

Discussion

The findings are consistent with the assumed hypotheses that the visual exposure of 40 ms can be sufficient to formulate aesthetic judgment. This means that the obtained results are consistent with the previous findings.

In our study, it was shown that aesthetic judgment may be formed even in a shorter time (40 ms) than expected by other researchers.

Limitations

It may be considered whether the second stage sufficiently separated both experiences of processing visual stimuli. For future research, random sampling could also be considered. In further research, an even shorter time for the initial exposure could be applied—our results do not ensure that 40 ms is the limit of processing visual information enabling the formulation of aesthetic judgment.

References and appendix

Here, tables with detailed results should be presented. Due to the space limitations in this sample report, they have not been included.

References

Augustin, D. M., Leder, H., Hutzler, F., & Carbon, C.-C. (2008). Style follows content: on the microgenesis of art perception. *ActaPsychologica*, *128*(1), 127–138. https://doi.org/10.1016/j.actpsy.2007.11.006

Cupchik, G. C., & Berlyne, D. E. (1979). The perception of collative properties in visual stimuli. *Scandinavian Journal of Psychology*, *20*(2), 93–104. https://doi.org/10.1111/j.1467-9450.1979.tb00688.x

Locher, P., Krupinski, E. A., Mello-Thoms, C., & Nodine, C. F. (2007). Visual interest in pictorial art during an aesthetic experience. *Spatial Vision*, *21*(1–2), 55–77. https://doi.org/10.1163/156856807782753868

1.4. Types of experimental research design

In this chapter, we will focus mainly on true experimental designs. However, we present other types of experiments as well to properly adjust the design to the goal of the study.

1.4.1. Within-subjects and between-subjects experimental designs

Conducting an experiment involves deciding what specific experimental design is to be used. Only after this decision is made, we can determine other important elements of the experimental procedure. Experimental design refers to the way of organising the tested subjects and intervening factors so as to minimise the uncontrolled variation in the effect variable. Of course, the choice of experimental design depends mainly on the purpose of the experiment and the nature of the studied phenomena, but also requires balance between the ability to correctly detect an existing causal effect and the precision with which this effect can be measured (Bellemare, Bissonnette, & Kröger, 2014). However, this is not the only consideration that makes the choice of an experimental design critical for the success of the study: the results can vary considerably depending on the chosen design. In social sciences, there are two major types of experimental designs: within-subjects and between-subjects design.

In the within-subjects experimental design, each participant (or subject) is exposed to all factors levels. In other words, all levels of the independent variable are administered in a consecutive manner on the same group of participants. Because we do not need to assign subjects to different groups, this experimental design requires fewer participants. From a practical point of view, the design is preferred when participants have to fulfil specific conditions to be recruited for the experiment and generally, it is hard to find enough people who are willing to participate.

In the between-subjects experimental design, the participants are divided into separate groups and each group is subjected to only one factor level. To describe the impact of the factor, the differences in the mean values of the effect variable are observed.

The main types of within-subjects and between-subjects designs are presented in Table 1. The experimental groups are indicated by an 'E', a factor is indicated with an 'X' (but not its levels), observations are indicated by an 'O', and control groups are indicated with a 'C'. In the between-subjects design, the indication 'A' represents the sample that is divided into equivalent groups prior to experimental treatment.[1]

[1] The sequence of letters in each row represents the order in which particular actions are taken. For example, in the first design (pretest / posttest), there is only one experimental group –E_1. The dependent variable is measured before the group is exposed to the treatment (O_1). Then, treatment X takes place, and the dependent variable is measured once more (O_2). Parallel rows represent independent, parallel testing of another experimental group.

Table 1. Types of within-subjects and between-subjects designs

Experimental design			Type
	Pretest / Posttest	(1)	$E_1\ O_1\ X\ O_2$
Within-subjects	Pretest / Posttest with control group	(2)	$E_1\ O_1\ X\ O_2$ $C_1\ O_3\ O_4$
	Pretest / Posttest: four-group design	(3)	$E_1\ O_1\ X\ O_2$ $C_1\ O_3\ O_4$ $E_2\ X\ O_5$ $C_2\ O_6$
Between-subjects	Posttest with control group	(4)	$A\ X\ O_1$ $A\ O_2$

Source: (Campbell, 1957; Dimitrov & Rumrill, 2003).

The simplest type of within-subjects design is the pretest / posttest with a control group. The experimental group is observed at least twice: prior to and after testing. The number of observations increases with the number of factor levels that are to be tested. Then the mean values of the dependent variable at each observation are compared for significant differences. Because different external factors that are outside the researcher's control can cause the dependent variable's mean to change between the observations (such as history or maturation), a control group is also included in the experiment. Measuring the dependent variable in the control group in parallel to the experimental group, but without applying any treatment, can help determine whether the observed effect can be attributed to the applied treatment or other external factors are contributing.

The pretest / posttest: four-group design, also known as the Solomon four group design, is somehow an extension of the previous design. It includes two more groups—one experimental and one control group, tested in parallel. Only the post-test is performed in the second experimental group. By omitting the pre-test, the researcher aims to avoid some threats to internal validity—performing the same testing twice can lead to the occurrence of the testing effect. The same logic applies to the second control group, which is only tested once.

The between-subjects design aims to overcome the weaknesses of the described designs by dividing the participants into different groups that are exposed to only one treatment. There is no pretest in this type of experimental design, thus, there is no threat to the internal validity because of the testing effect. Each experimental group is exposed to different conditions and this is why participants' behaviour cannot be influenced by more than one combination of factor levels.

Considerations when choosing an experimental design

True experimental designs are associated with different degrees of external factor control, leading to the occurrence of systematic error, and with a different statistical power of the tests used. The researcher must decide whether to accept a certain decrease of internal validity at the expense of increasing statistical power, or vice versa; what possible measures can be taken to address the limitations of the preferred experimental design. Comparing the within-subjects and between-subjects experimental designs can be done within two aspects: the potential to provide internal validity and statistical power.

Statistical power of tests

The interference theory poses that the null hypothesis expresses the lack of difference or effect in the observed means (Sawyer & Ball, 1981, p. 275). In most cases, the researcher seeks to reject the null hypothesis in order to accept that there is a significant effect of tested factor levels. Statistical power expresses the probability that the applied statistical test will lead to correct rejection of the null hypothesis[2], therefore, the higher the power of the test, the greater the probability that the conclusion made about the existence of a causal relationship is correct. Statistical power is a function of the test's significance level, the sample size and effect size, thus increasing the sample will increase the statistical power when fixing the other two components (Chase & Chase, 1976, p. 234).

Determining the desired statistical power level before conducting the experiment is important both for the correct definition of the required sample size and for assessing the appropriateness of the study as a whole.[3] However, choosing an adequate power level can be difficult when the estimated size of the effect is unknown. The empirical level of significance should not be used as a measure of the effect size, since both the statistical significance for a particular level of α and the size of the effect are a function of the sample size: even small effects will almost certainly be significant in large samples, while large effects may not be considered significant if the sample is small (Sawyer & Ball, 1981, p. 281). Overcoming the problem of the lack of a preliminary idea of the effect size can be done by conducting a pilot study.

The link between statistical power and the significance level of the test presupposes its relatively lower values when applying conservative tests.[4] Because the between-subjects design is generally more conservative (Charness, Gneezy, & Imas, 2012, p. 2), it is characterised by lower statistical power. Its conservatism stems from the need to apply post hoc contrast tests, some of which are particularly conservative when it comes to comparing more than three pairs of groups (Privitera, 2015,

[2] Statistical power is equal to $1 - \beta$, where β is the probability of failure to reject the null hypothesis when it should be rejected.

[3] For example, a limited budget may force the researcher to change the research design in order to achieve a satisfactory level of statistical power when a relatively small effect is of interest and it is necessary to experiment with a larger sample (which will be more costly).

[4] A statistical test is conservative when the α level is reduced and as a result, the level of β increases.

p. 374), although the researcher may choose to apply a more liberal test. Dividing participants into separate groups, which should be treated with different factor levels, results in a smaller number of subjects in each group, compared to the within-subject design where all participants are in one group. Repeated measurements in the within-subjects design provide more observations from one subject. This allows the researcher to work with a larger sample size and therefore, the statistical power of the tests is higher. Achieving an acceptable level of statistical power in between-subject experiments may require up to four times more subjects than in within-subject design experiments when the number of experimental sessions is small (Bellemare, Bissonnette, & Kröger, 2014, p. 3). This problem becomes more serious with the inclusion of additional factors or factor levels because if the researcher is unable to recruit more participants, this can lead to forming more groups of even smaller size. In such cases, only the recruitment of additional participants could increase the statistical power of the between-subjects experiment.

1.4.2. Different types of experimental designs

The choice of an appropriate experimental design is essential to precisely explore the aim of research. The previous part of this chapter focuses mainly on true experimental designs. However, different categories of experimental designs can be distinguished as well.

In true experimental designs, the researcher employs randomisation in order to assign participants to groups. On the other hand, the main characteristic of pre-experimental design is the lack of randomisation. Here, the most simple design is a one-shot case study which involves only a single measurement after the treatment applied to one group (Malhotra et al., 2017). Sometimes, while conducting experiments, the researcher is unable to control when the procedure is applied and how the participants are assigned to groups. Such a category of experimental designs is called quasi-experimental. Measurements are made at various time intervals—some of them are taken before the treatment and some after. This design can be helpful for practical reasons, particularly when the researcher wants to observe the effects over a long time period (Field & Hole, 2013). There can be a simple time series with one group and a multiple time series with another group that plays the role of a control group.

Occasionally, the researcher seeks to examine the influence of more than one independent variable. The research design may also intend to control the nuisance factors. A more advanced experimental design that enables the researcher to exert control over an extraneous variable is the randomised block design. In this design, the blocking technique is applied, which depends on dividing the participants into the groups on the basis of a similar variable level. This type of experiment is used when the external factor that may influence the performance has been identified and may be controlled. For instance, a company introduces a new line of environmentally-friendly cosmetic products. The company aims to popularise the product by organising public campaigns. After creating three public campaigns (E, F, G) that differ from each other with regard to the content of information about the product, the company aims to examine their effectiveness. At the same time, the managers assume that the reception of the campaign may vary depending on the usage of similar products before the experiment. Therefore, the information about using a similar product in the past is considered as a blocking variable. Participants are classified into groups based on the assumption that they have often, rarely or never before used other environmentally-friendly cosmetic products in the past. The random assignment occurs at two stages—in selecting participants for the experiment and in assigning them to the types of public campaign (treatment groups). This design is presented in Table 2.

Table 2. Random block design—example

Number of block group	Using environmentally-friendly cosmetic products before	Type of public campaign		
		High amount of information (H)	Medium amount of information (M)	Low amount of information (L)
1	Often	H	M	L
2	Rarely	H	M	L
3	Never	H	M	L

Source: Own elaboration.

A different kind of experimental design in which the blocking technique also finds its application is the Latin-square design. This kind of practice allows the researcher to control two external factors. Again, blocking intends to reduce the additional source of variability. The main rule is that for both variables, the numbers of levels are the same—the scheme of experimental design that is presented in the table has the same number of rows and columns (Montgomery, 2001).

1.5. Conducting experiments

1.5.1. Internal and external validity

The results of a properly planned and conducted experiment should be valid. In general, the concept of validity is about the "extent to which the conclusions drawn from the experiment are true" (Hair et al., 2003). Specifically, this refers to two aspects—the certainty that the effect on the dependent variable is due to the independent variable and that the results may be applied to a larger population of interest in a real-world context (Burns et al., 2017; Malhotra et al., 2017). Two forms of validity can be distinguished—that is, internal and external validity. Internal validity concerns the accuracy of the experiment. It is the "extent to which the results of the experiment are attributed to the manipulation of the independent variable" (Jackson, 2008; Malhotra et al., 2017). In other words, the researcher needs to be sure that in the study, the causal relationships that may be explained by the experimental treatment (and not by other reasons) are accurately examined (Hair et al., 2003). Thus, if high internal validity of the experiment is to be ensured, there should not be any confounds, i.e. extraneous variables or flaws of the experiment that are not controlled by the researcher (Jackson, 2008). What this means is that controlling variables other than the treatment is an indispensable condition for internal validity (Malhotra et al., 2017).

External validity, on the other hand, indicates whether the observed relationships between independent and dependent variables can be generalised. This means whether the obtained results can be projected onto conditions beyond the experimental situation and if they are true for the entire population to which the study applies (Hair et al., 2003; Malhotra et al., 2017). While planning the experiment, the researcher needs to be cautious and include all aspects that may be relevant in real-world settings in the experiment. The truth is, considering both internal and external validity in designing a study is a challenge for the researcher. In order to ensure a satisfactory level of these two forms of validity, the need to compromise may sometimes occur. The laboratory experiment can be conducted in the case of wanting to control the extraneous variables which increase the internal validity of the study. However, the laboratory conditions differ from the real ones, which may reduce the external validity (Malhotra et al., 2017).

While planning and conducting experiments, the researcher may encounter several threats, both to internal and external validity. Enlisting those threats and errors may help in avoiding some of them.

1.5.2. Experimental errors (threats to validity)

The following threats regarding the internal validity can be enumerated:

History effect: The history effect occurs when the specific event that is outside the experimental situation may violate the results. What should be noted is that this does not mean that the event occurred in the past, before the experiment. Contrarily, it applies to the factors taking place during the time of the study (Jackson, 2008; Malhotra et al., 2017).

Maturation effect: Another aspect that may limit experiment results is maturation. In the case of the experiments that last over a period of time, the participants may naturally grow and develop (e.g. became older, more tired or interested). These changes are not caused by a specific event but involve the participants and occur with the passage of time. If people change during the period of the study, the effect on the dependent variable may be caused by this instead of the independent variable. How can we deal with the maturation effect? Implementing a control group in the study may help (Jackson, 2008; Malhotra et al., 2017).

Testing effect: Testing effect refers to the process of experimentation. In some studies, pre- and posttest measures are included, and some are conducted on a daily, weekly or monthly basis. These repeated measures may affect the experiment results in two ways—when the prior measure has impact the later one (the main testing effect) or when the prior measure changes the participant's reaction towards the independent variable (the interactive testing effect). Being tested numerous times itself may well influence the dependent variable, decreasing internal validity. Here, such effects as the practice effect—when participants take some tests several times and learn how to perform it better—can also be mentioned. Also, the fatigue effect may occurs when participants become tired of repeating the same procedure and then get lower scores (Jackson, 2008; Malhotra et al., 2017).

Regression: This effect means that in the course of the study, the extreme scores tend to move closer to the mean. It happens when participants are initially chosen on the basis of their extreme scores and then, with several more tests, their scores regress to the average values. It may also distort the internal validity because the observed change may be caused by the changes in scores instead of the independent variable.

Instrumentation: Another threat to internal validity is directly related to the measuring instrument, observation techniques and measurement processes (Hair et al., 2003). The measurement sometimes takes place through observation and the observers may become tired, bored or may lose their focus during the experiment. This also happens because of the lower accuracy of scorers or due to changes in administration procedures (Hair et al., 2003; Jackson, 2008). This threat occurs especially when there is a pretest and posttest study (Malhotra et al., 2017).

Selection bias: The threat to internal validity that depends on inappropriate selection or assignment of the participants to treatment groups is called selection bias (Hair et al., 2003). In this case, the changes in the dependent variable would be impossible to compare because the group may differ initially. This happens when the researchers select participants on the basis of their subjective judgement or when they let the participants assign themselves to groups on their own (Malhotra et al., 2017).

Mortality (attrition): In the case of experiments with experimental and control groups, there is a risk that along with the course of the study, the number of participants will change. As a result, this inequality between both groups may lead to distortion of internal validity. The differences in the group sizes occur due to many reasons—sometimes people just refuse to take part in the experiment. Thus, it may not be known whether those who participate in the study would react to the treatment in the similar way to those who resigned from participation (Jackson, 2008; Malhotra et al., 2017).

Diffusion of treatment: Another effect that is particularly risky in maintaining internal validity of the experiment is diffusion of treatment. This matches the relationships between the participants of the study who may react differently to the treatment because of the information which they exchange. It may sometimes occur when students are taking part in the experiment. They may know each other and discuss the study during its course or talk about it with the students who have not yet participated in the study. Sharing information about the experiment may influence the reaction to treatment. The researcher should ask the participants not to communicate during the study. The threat may be limited by conducting each part of the experiment in the shortest time possible (Jackson, 2008).

Experimenter effect: The results of the experiment may well be violated by the experimenter. The experimenter is responsible for designing the study and puts a lot of time and effort into this process. Occasionally, the researcher may unintentionally encourage participants to react in a way desired for the purpose of the study. This may be done by body language or mimics. The possible solution could be using the method of blinding, in which the researcher interacting with the participants does not know the details of the treatment (Jackson, 2008).

There are several threats also connected with external validity. The risk of limiting the possibility of result generalisation may be violated by involving mostly student participants in the study. This is a common practice due to accessibility, low cost and time. However, the researcher should be careful with including students mostly in the experiment as they sometimes may not be representative of the target population (Zikmund et al., 2010). The selection may also be crucial in different aspects. For example, if the study demands a lot of time, the participants who are involved are only those who have the motivation and time to take part in it, which also, may not be a good reference to the whole population. Apart of the inaccurate selection of participants who cannot be applied to the whole population, another issue with generalisation refers to the setting of the experiment. This involves conducing laboratory experiments in which there is a risk that the participants' reactions to the treatment may differ compared to those present in the natural environment. Some other aspects that may affect the external validity are timing of the study or exposure to pre-measurements, which can change the participant's reaction to the treatment.

Limitations of validity in between-subject and within-subject designs

It can be pointed out that in experimental studies, the external validity is unattainable for the whole population, but for its subgroups, it is formed on the basis of the characteristics of the participating subjects. While controlling the adverse effects of external factors ensures internal validity, it could be an obstacle in achieving external validity for laboratory experiments, as conditions may deviate too much from the actual environment of the studied phenomena.

In within-subjects experiments, at least two observations are performed—one before and one after the treatment. Due to the fact that the group should be subjected to all levels of treatment, increasing the number of factors and/or their levels leads to multiple testing. This allows the comparison of each subject to him/herself from before and after the treatment, but it opens space to some unwanted variation caused by re-testing. In the time period between two observations, various side events may occur, or systematic effects may appear (such as fatigue, distraction, anxiety, boredom). Each subsequent test may result in increased experience of the subjects with the experimental procedure. However, some balancing techniques can be used to deal with these effects. Adding an untreated control group to the design allows generalisation of the results to any other equivalent and pretested group (Campbell, 1957, p. 302). It is assumed that the impact of external factors on the experimental and control groups is relatively the same.

Between-subjects experiments do not suffer from the shortcomings described above, as the observed values of the effect variable in at least two experimental groups are compared, and each group is tested only once. It is assumed that any differences in the group means of the dependent variable would be due to the different level of treatment in each group. However, when conducting a between-subjects experiment, reasonable doubts may arise as to whether the observed differences are caused by different personal characteristics of the subjects in the groups. Application of a randomisation technique and ensuring equivalence of the groups is mandatory to eliminate the possibility that differences in participants' characteristics are the reason for the observed difference.

In this section, the aspects that may help the experimenter to control for extraneous variables and enhance validity of experiments are presented. Among the crucial aspects, the role of randomisation should be distinguished. Randomisation is the procedure of assigning participants to groups randomly, which helps ensure that the groups are equal and comparable (Hair et al., 2003). Participants chosen for the study should also be randomly selected for the experimental and control groups. Thanks to this technique, the researcher may assume that the confounding factors will be displayed in the whole group equally. For this to be ensured, the sample should be reasonably large. Sometimes, for the purposes of the study, the researcher may need to select participants with certain characteristics. This procedure—known as matching—is then a step prior to group assignment (Malhotra et al., 2017). The advisable practice, if possible, is also including a control group in the experimental design, which may help deal with the maturation, history, testing or instrumentation effects (Jackson, 2008).

1.5.3. Ethics in experimentation

When conducting experiments involving people, the researchers need to take care of the participants' well-being and morals during the study. In many studies, especially those demanding active participation, informed consent is indispensable. The participants need to know in what activity they will be involved. The participants should be assured that the confidential information obtained during the experiment is not to be disclosed. What should also be worth noting is that the question of confidentiality is crucial, not only for the participants, but also for the companies that are commissioning or sponsoring the research.

The role of researcher is to precisely explain all procedures to the participants. They should be informed about the extent to which they will be involved. Every participant has the right to leave or resign from the experiment at any time.

For some participants, taking part in an experiment may be a totally new situation, thus, they feel stressed. A helpful practice in dealing with this is called debriefing—this consists in informing the participants about the main objectives of the experiment and its hypotheses. This also creates the chance for participants to ask any questions they may have.

References

Bellemare, C., Bissonnette, L., & Kröger, S. (2014). *Statistical power of within and between-subjects designs in economic experiments* (The IZA Discussion Paper series).

Bridley, N. (2013). *Marketing research*. Oxford University Press.

Brzeziński, J. (1999). *Metodologia badań psychologicznych*. Wydawnictwo Naukowe PWN.

Burns, A. C., Veeck, A., & Bush, R. F. (2017). *Marketing research. Global edition*. Pearson Education Limited.

Campbell, D. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, *54*(4), 297-312.

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

Chase, L., & Chase, R. (1976). A statistical power analysis of applied psychological research. *Journal of Applied Psychology*, *61*(2), 234-237. https://doi.org/10.1037/0021-9010.61.2.234

Dimitrov, D., & Rumrill, P. (2003). Pretest-posttest designs and measurement of change. *Work*, *20*(2), 159-165.

Field, A., & Hole, G. (2013). *How to design and report experiments*. Sage Publications Ltd.

Hair, J. F. J., Bush, R. P., & Ortinau, D. J. (2003). *Marketing research: Within a changing information environment*. McGraw-Hill/Irwin.

Jackson, S. L. (2008). *Research methods and statistics: A critical thinking approach* (3rd ed.). Wadsworth Cengage Learning.

Malhotra, N. K., Nunan, D., & Birks, D. F. (2017). *Marketing research: An applied approach* (5th ed.). Pearson Education Limited.

Mazzochi, M. (2008). *Statistics for marketing and consumer research*. Sage Publications Ltd.

Montgomery, D. C. (2001). *Design and analysis of experiments* (5th ed.). John Wiley & Sons, Inc.
Moore, D. S., McCabe, G. P., Alwan, L. C., Craig, B. A., & Duckworth, W. M. (2011). *The practice of statistics for business and economics* (3rd ed.). W. H. Freeman and Company.
Privitera, G. (2015). *Statistics for the behavioral sciences* (2nd ed.). Sage Publications, Inc.
Sawyer, A., & Ball, D. (1981). Statistical power and effect size in marketing research. *Journal of Marketing Research*, *18*(3), 275-290.
Stachak, S. (1997). *Wstęp do metodologii nauk ekonomicznych*. Książka i Wiedza.
Sułek, A. (1979). *Eksperyment w badaniach społecznych*. Wydawnictwo Naukowe PWN.
Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2010). *Business research methods* (8th ed.). South-Western Cengage Learning.