

RESEARCH METHODS

For Business and Marketing

GEORGE SELF

January, 2019 – Edition 1

George Self: *Research Methods*, For Business and Marketing, January, 2019

This work is licensed under a **Creative Commons** "Attribution-NonCommercial-ShareAlike 4.0 International" license.



FOREWORD

I have taught BASV 316, *Introductory Methods of Analysis*, on-line for the University of Arizona in Sierra Vista since 2010 and enjoy working with students on research methodology. I wanted a textbook that presented research in a practical way so students could use the lessons learned in their own research projects. I found an excellent book but over the years the cost of that book increased to the point that I felt like it was an unfair burden on students.

I began by looking for an acceptable “open source” book since authors make those available to students free of charge and I could modify the book to meet my own objectives. I could not find any that were focused on business research though I tried for several years—and keep looking to this day. I did, though, find a few open source books about research in the social and psychological sciences that were reasonably close to what I needed. So, I modified those books to emphasize business research and then provided my work to students free of charge.

Bhattacharjee[2], Blackstone[3], and Price[4] all released books about research that formed the major sources for this book. Those books are all open source and published under a Creative Commons license that permitted me to copy and modify them.

Three goals shaped the choices made about the topics covered by the text and how those topics are presented.

- The topics must have relevance for business students.
- Both qualitative and quantitative research methods are given roughly equal attention since both types of research are used in business.
- The text is engaging and readable.

While the book is useful in its current form, I will continually update it based on emerging trends in research.

This book is published under a Creative Commons **Attribution-NonCommercial-ShareAlike** license, just like the books that provided its foundation. The source is available at my GitHub account: <http://bit.ly/2xIjzXL>. It is my hope that students can use this book to learn about business research and other instructors can modify and use it for their own classes.

— George Self

BRIEF CONTENTS

I	BACKGROUND	1
II	QUANTITATIVE METHODS	3
1	EXPERIMENTAL RESEARCH	5
III	QUALITATIVE METHODS	19
IV	MIXED METHODS	21
V	REPORTING	23
VI	APPENDIX	25
2	APPENDIX	27
	GLOSSARY	29
	BIBLIOGRAPHY	37

CONTENTS

I	BACKGROUND	1
II	QUANTITATIVE METHODS	3
1	EXPERIMENTAL RESEARCH	5
1.1	Introduction	5
1.1.1	Basic Concepts	6
1.1.2	Two-Group Experimental Designs	8
1.1.3	Factorial Designs	10
1.1.4	Hybrid Experimental Designs	12
1.1.5	Quasi-Experimental Designs	14
1.1.6	Strengths and Weaknesses of Experimental Research	17
1.2	Summary	18
III	QUALITATIVE METHODS	19
IV	MIXED METHODS	21
V	REPORTING	23
VI	APPENDIX	25
2	APPENDIX	27
	GLOSSARY	29
	BIBLIOGRAPHY	37

LIST OF FIGURES

Figure 1	2X2 Factorial Design	11
----------	----------------------	----

LIST OF TABLES

Table 1	Pretest/Post-test Design	9
Table 2	Post-test Only Design	9
Table 3	Covariate Design	10
Table 4	2X2 Factorial Design	11
Table 5	Randomized Block Design	13
Table 6	Solomon Four-group Design	13
Table 7	Switched Replication Design	13
Table 8	Nonequivalent Groups Design	14
Table 9	Nonequivalent Switched Replication Design	14
Table 10	Regression-discontinuity Design	15
Table 11	Proxy Pretest Design	16
Table 12	Separate Pretest/Posttest Samples Design	16
Table 13	Nonequivalent Dependent Variable Design	17

LISTINGS

Part I

BACKGROUND

Research methods are grounded in philosophy, statistics, sociology, and many other disciplines. The chapters in this section introduce these background concepts.

Part II

QUANTITATIVE METHODS

Quantitative methods are based in the measurement of concepts and the statistical analysis of those measures. Quantitative methods include activities like sampling, surveys, and experimental research.

EXPERIMENTAL RESEARCH

1.1 INTRODUCTION

Experimental research, often considered to be the “gold standard” in research designs, is one of the most rigorous of all research designs. In this design, one or more independent variables are manipulated by the researcher (as treatments), subjects are randomly assigned to different treatment levels (random assignment), and the results of the treatments on outcomes (dependent variables) are observed. The unique strength of experimental research is its internal validity (causality) due to its ability to link cause and effect through treatment manipulation, while controlling for the spurious effect of extraneous variable.

One example of experimental research can be found in Alcott’s study of behavioral interventions[1]. In this study, the results of the *Opower* program were evaluated. Opower was a company that sent a “home energy report” to more than six million homes representing 85 utilities across the United States. The reports were designed to use social pressure to get homeowners to moderate their electricity use. For Alcott’s experiment, the three different sites were compared using pre- and post- intervention electricity usage statistics. He found that the initial home energy report caused high-frequency “action and backsliding,” but the cycles seem to attenuate over time. He also found that if the reports were discontinued the effect was still relatively persistent. Finally, he found that consumers are slow to habituate. While this study was focused on a single intervention involving energy conservation, it may be interesting to speculate if similar intervention efforts, in dieting, exercise, smoking cessation, etc., would experience similar results.

Experimental research is best suited for explanatory research rather than for descriptive or exploratory research, where the goal of the study is to examine cause-effect relationships. It also works well for research that involves a relatively limited and well-defined set of independent variables that can either be manipulated or controlled. Experimental research can be conducted in laboratory or field settings. Laboratory experiments, conducted in laboratory (artificial) settings, tend to be high in internal validity, but this comes at the cost of low external validity (generalizability), because the artificial (laboratory) setting in which the study is conducted may not reflect the real world. Field experiments, conducted in field settings such as in a real organization, and high in both internal and external validity. But such

experiments are relatively rare, because of the difficulties associated with manipulating treatments and controlling for extraneous effects in a field setting.

Experimental research can be grouped into two broad categories: true experimental designs and quasi-experimental designs. Both designs require treatment manipulation, but true experiments also require random assignment while quasi-experiments do not. Sometimes the literature refers to non-experimental research, which is only an all-inclusive term for research that does not employ treatment manipulation or random assignment, such as survey research.

1.1.1 *Basic Concepts*

TREATMENT AND CONTROL GROUPS. In experimental research, some subjects are administered one or more experimental stimulus called a treatment (the treatment group) while other subjects are not given such a stimulus (the control group). The treatment may be considered successful if subjects in the treatment group rate more favorably on outcome variables than control group subjects. Multiple levels of experimental stimulus may be administered, in which case, there may be more than one treatment group. For example, in order to test the effects of a new drug intended to treat a certain medical condition like dementia, if a sample of dementia patients is randomly divided into three groups, with the first group receiving a high dosage of the drug, the second group receiving a low dosage, and the third group receives a placebo such as a sugar pill (control group), then the first two groups are experimental groups and the third group is a control group. After administering the drug for a period of time, if the condition of the experimental group subjects improved significantly more than the control group subjects, we can say that the drug is effective. We can also compare the conditions of the high and low dosage experimental groups to determine if the high dose is more effective than the low dose.

TREATMENT MANIPULATION. Treatments are the unique feature of experimental research that sets this design apart from all other research methods. Treatment manipulation helps control for the “cause” in cause-effect relationships. Naturally, the validity of experimental research depends on how well the treatment was manipulated. Treatment manipulation must be checked using pretests and pilot tests prior to the experimental study. Any measurements conducted before the treatment is administered are called pretest measures, while those conducted after the treatment are post-test measures.

RANDOM SELECTION AND ASSIGNMENT. Random *selection* is the process of randomly drawing a sample from a population or a sam-

pling frame. This approach is typically employed in survey research, and assures that each unit in the population has an equal chance of being selected into the sample. Random *assignment* is the process of randomly assigning subjects to experimental or control groups. This is a standard practice in true experimental research to ensure that treatment groups are similar (equivalent) to each other and to the control group, prior to treatment administration. Random selection is related to sampling, and is therefore, more closely related to the external validity (generalizability) of findings. However, random assignment is related to design, and is therefore most related to internal validity. It is possible to have both random selection and random assignment in well-designed experimental research.

THREATS TO INTERNAL VALIDITY. Although experimental designs are considered more rigorous than other research methods in terms of the internal validity of their inferences (by virtue of their ability to control causes through treatment manipulation), they are not immune to internal validity threats. Some of these threats to internal validity are described below, within the context of a study of the impact of a special remedial math tutoring program for improving the math abilities of high school students.

- *History threat* is the possibility that the observed effects (dependent variables) are caused by extraneous or historical events rather than by the experimental treatment. For instance, students' post-remedial math score improvement may have been caused by their preparation for a math exam at their school, rather than the remedial math program.
- *Maturation threat* refers to the possibility that observed effects are caused by natural maturation of subjects (e.g., a general improvement in their intellectual ability to understand complex concepts) rather than the experimental treatment.
- *Testing threat* is a threat in pre/post designs where subjects' post-test responses are conditioned by their pretest responses. For instance, if students remember their answers from the pretest evaluation, they may tend to repeat them in the post-test exam. Not conducting a pretest can help avoid this threat.
- *Instrumentation threat*, which also occurs in pre/post designs, refers to the possibility that the difference between pretest and post-test scores is not due to the remedial math program, but due to changes in the administered test, such as the post-test having a higher or lower degree of difficulty than the pretest.
- *Mortality threat* refers to the possibility that subjects may be dropping out of the study at differential rates between the treatment and control groups due to a systematic reason, such that

the dropouts were mostly students who scored low on the pretest. If the low-performing students drop out, the results of the post-test will be artificially inflated by the preponderance of high-performing students.

- *Regression threat*, also called a regression to the mean, refers to the statistical tendency of a group's overall performance on a measure during a post-test to regress toward the mean of that measure rather than in the anticipated direction. For instance, if subjects scored high on a pretest, they will have a tendency to score lower on the post-test (closer to the mean) because their high scores (away from the mean) during the pretest was possibly a statistical aberration. This problem tends to be more prevalent in nonrandom samples and when the two measures are imperfectly correlated.

1.1.2 Two-Group Experimental Designs

The simplest true experimental designs are two group designs involving one treatment group and one control group, and are ideally suited for testing the effects of a single independent variable that can be manipulated as a treatment. The two basic two-group designs are the pretest/post-test control group design and the post-test-only control group design, while variations may include covariance designs. These designs are often depicted using a standardized design notation, where *R* represents random assignment of subjects to groups, *X* represents the treatment administered to the treatment group, and *O* represents pretest or post-test observations of the dependent variable (with different subscripts to distinguish between pretest and post-test observations of treatment and control groups).

PRETEST/POST-TEST CONTROL GROUP DESIGN. In this design, subjects are randomly assigned to treatment and control groups, subjected to an initial (pretest) measurement of the dependent variables of interest, the treatment group is administered a treatment (representing the independent variable of interest), and the dependent variables measured again (post-test). The notation of this design is shown in Table 1, where each group's process should be read from left-to-right. Thus, the Treatment Group (top line) is first assigned using a random process, then a pretest is administered, then a treatment of some sort is applied, finally, a post-test is administered. The Control Group (bottom line) follows the same process, except there is no treatment applied.

R	O ₁	X	O ₂	TREATMENT GROUP
R	O ₃		O ₄	CONTROL GROUP

Table 1: Pretest/Post-test Design

The effect, E , of the experimental treatment in the pretest post-test design is measured as the difference in the post-test and pretest scores between the treatment and control groups, as shown in Equation 1

$$E = (O_2 - O_1) - (O_4 - O_3) \quad (1)$$

Statistical analysis of this design involves a simple analysis of variance (ANOVA) between the treatment and control groups. The pretest post-test design mitigates several threats to internal validity, such as maturation, testing, and regression, since these threats can be expected to influence both treatment and control groups in a similar (random) manner. The selection threat is controlled via random assignment. However, additional threats to internal validity may exist. For instance, mortality can be a problem if there are differential dropout rates between the two groups, and the pretest measurement may bias the post-test measurement (especially if the pretest introduces unusual topics or content).

POST-TEST-ONLY CONTROL GROUP DESIGN. This design is a simpler version of the pretest/post-test design where pretest measurements are omitted. The design notation is shown in Table 2.

R	X	O ₁	TREATMENT GROUP
R		O ₂	CONTROL GROUP

Table 2: Post-test Only Design

The treatment effect is measured simply as the difference in the post-test scores between the two groups, as shown in Equation 2

$$E = (O_1 - O_2) \quad (2)$$

The appropriate statistical analysis of this design is also a two-group analysis of variance (ANOVA). The simplicity of this design makes it more attractive than the pretest/post-test design in terms of internal validity. This design controls for maturation, testing, regression, selection, and pretest/post-test interaction, though the mortality threat may continue to exist.

COVARIANCE DESIGNS. Sometimes, measures of dependent variables may be influenced by extraneous variables called *covariates*. Covariates are those variables that are not of central interest to an experimental study, but should nevertheless be controlled in order to eliminate their potential effect on the dependent variable and therefore allow for a more accurate detection of the effects of the independent variables of interest. The experimental designs discussed earlier did not control for such covariates. A covariance design (also called a “concomitant” variable design) is a special type of pretest post-test control group design where the pretest measure is essentially a measurement of the covariates of interest rather than that of the dependent variables. The design notation is shown in Table 3, where C represents the covariates:

R	C	X	O ₁	TREATMENT GROUP
R	C		O ₂	CONTROL GROUP

Table 3: Covariate Design

Because the pretest measure is not a measurement of the dependent variable, but rather a covariate, the treatment effect is measured as the difference in the post-test scores between the treatment and control groups as in Equation 3.

$$E = (O_1 - O_2) \quad (3)$$

Due to the presence of covariates, the right statistical analysis of this design is a two-group analysis of covariance (ANCOVA). This design has all the advantages of post-test only design, but with internal validity due to the controlling of covariates. Covariance designs can also be extended to pretest/post-test control group design.

1.1.3 Factorial Designs

Two-group designs are inadequate if the research project requires manipulation of two or more independent variables (treatments). In such cases, four or higher-group designs are required. Such designs, quite popular in experimental research, are commonly called factorial designs. Each independent variable in this design is called a factor, and each sub-division of a factor is called a level. Factorial designs enable researchers to examine not only the individual effect of each treatment on the dependent variables (called main effects), but also their joint effect (called interaction effects).

The most basic factorial design is a 2x2 factorial design, which consists of two treatments, each with two levels (such as high/low or present/absent). For instance, suppose a research project compares the

learning outcomes of two different types of instructional techniques (online and in-class) along with the time of instruction (1.5 or 3 hours per week). In this case, there are two factors: instructional type and instructional time; each with two levels (in-class and online for instructional type, and 1.5 and 3 hours/week for instructional time), as shown in Figure 1. If a third level of instructional time (maybe 6 hours/week) is desired then the second factor will consist of three levels and a 2x3 factorial design is required. On the other hand, if a third factor, such as group work (present versus absent), is desired, then a 2x2x2 factorial design is needed. In this notation, each number represents a factor, and the value of each factor represents the number of levels in that factor.

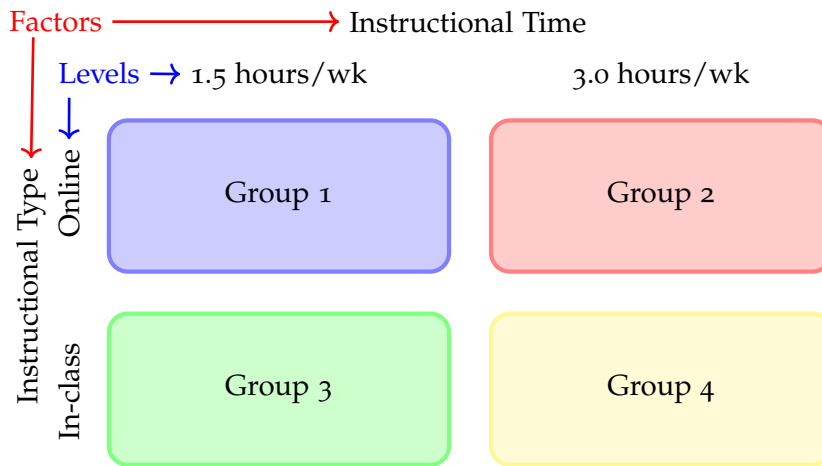


Figure 1: 2X2 Factorial Design

Factorial designs can also be depicted using a design notation, such as that shown in Table 4.

<i>R</i>	<i>X</i> ₁₁	<i>O</i> ₁	GROUP 1
<i>R</i>	<i>X</i> ₁₂	<i>O</i> ₂	GROUP 2
<i>R</i>	<i>X</i> ₂₁	<i>O</i> ₃	GROUP 3
<i>R</i>	<i>X</i> ₂₂	<i>O</i> ₄	GROUP 4

Table 4: 2X2 Factorial Design

R represents random assignment of subjects to treatment groups, *X* represents the treatment groups themselves (the subscripts of *X* represents the level of each factor), and *O* represent observations of the dependent variable. Notice that the 2x2 factorial design will have four treatment groups, corresponding to the four combinations of the two levels of each factor. Correspondingly, a 2x3 design will have

six treatment groups, and a $2 \times 2 \times 2$ design will have eight treatment groups. As a rule of thumb, each cell in a factorial design should have a minimum sample size of 20 (this estimate is derived from Cohen's power calculations based on medium effect sizes). So a $2 \times 2 \times 2$ factorial design requires a minimum total sample size of 160 subjects, with at least 20 subjects in each cell. It is obvious that the cost of data collection can increase substantially with more levels or factors in a factorial design. Sometimes, due to resource constraints, some cells in such factorial designs may not receive any treatment at all, which are called incomplete factorial designs, but incomplete designs decrease the ability to draw inferences about the those factors.

In a factorial design, a *main effect* is said to exist if the dependent variable shows a significant difference between multiple levels of one factor, at all levels of other factors. No change in the dependent variable across factor levels is the null case (baseline), from which main effects are evaluated. In the above example, perhaps a main effect of instructional type, instructional time, or both can be seen on learning outcomes. An *interaction effect* exists when the effect of differences in one factor depends upon the level of a second factor. In the above example, if the effect of instructional type on learning outcomes is greater for 3 hours/week of instructional time than for 1.5 hours/week, then there is an interaction effect between instructional type and instructional time on learning outcomes. Note that the presence of interaction effects dominate and make main effects irrelevant, and it is not meaningful to interpret main effects if interaction effects are significant.

1.1.4 Hybrid Experimental Designs

Hybrid designs are those that are formed by combining features of more established designs. Three such hybrid designs are randomized blocks design, Solomon four-group design, and switched replications design.

RANDOMIZED BLOCK DESIGN. This is a variation of the post-test-only or pretest/post-test control group design where the subject population can be grouped into relatively homogeneous subgroups (called blocks) within which the experiment is replicated. For instance, to replicate the same post-test-only design among university students and full-time working professionals (two homogeneous blocks), subjects in both blocks are randomly split between a treatment group (receiving the same treatment) or control group (see Table 5). The purpose of this design is to reduce the "noise" or variance in data that may be attributable to differences between the blocks so that the actual effect of interest can be detected more accurately.

UNIVERSITY STUDENTS	<i>R</i>	X	O ₁
UNIVERSITY STUDENTS	<i>R</i>		O ₂
PROFESSIONALS	<i>R</i>	X	O ₃
PROFESSIONALS	<i>R</i>		O ₄

Table 5: Randomized Block Design

SOLOMON FOUR-GROUP DESIGN. In this design, the sample is divided into two treatment groups and two control groups. One treatment group and one control group receive the pretest, and the other two groups do not. This design represents a combination of post-test-only and pretest/post-test control group design, and is intended to test for the potential biasing effect of pretest measurement on post-test measures that tends to occur in pretest/post-test designs but not in post-test only designs. The design notation is shown in Table 6.

<i>R</i>	O ₁	X	O ₂
<i>R</i>	O ₃		O ₄
<i>R</i>		X	O ₅
<i>R</i>			O ₆

Table 6: Solomon Four-group Design

SWITCHED REPLICATION DESIGN. This is a two-group design implemented in two phases with three waves of measurement. The treatment group in the first phase serves as the control group in the second phase, and the control group in the first phase becomes the treatment group in the second phase, as illustrated in Table 7. In other words, the original design is repeated or replicated temporally with treatment/control roles switched between the two groups. By the end of the study, all participants will have received the treatment either during the first or the second phase. This design is most feasible in organizational contexts where organizational programs (e.g., employee training) are implemented in a phased manner or are repeated at regular intervals.

<i>R</i>	O ₁	X	O ₂		O ₃
<i>R</i>	O ₄		O ₅	X	O ₆

Table 7: Switched Replication Design

1.1.5 Quasi-Experimental Designs

Quasi-experimental designs are almost identical to true experimental designs, but lacking one key ingredient: random assignment. For instance, one entire class section or one organization is used as the treatment group, while another section of the same class or a different organization in the same industry is used as the control group. This lack of random assignment potentially results in groups that are non-equivalent, such as one group possessing greater mastery of a certain content than the other group, say by virtue of having a better teacher in a previous semester, which introduces the possibility of selection bias. Quasi-experimental designs are therefore inferior to true experimental designs in interval validity due to the presence of a variety of selection related threats such as selection-maturation threat (the treatment and control groups maturing at different rates), selection-history threat (the treatment and control groups being differentially impact by extraneous or historical events), selection-regression threat (the treatment and control groups regressing toward the mean between pretest and post-test at different rates), selection-instrumentation threat (the treatment and control groups responding differently to the measurement), selection-testing (the treatment and control groups responding differently to the pretest), and selection mortality (the treatment and control groups demonstrating differential dropout rates). Given these selection threats, it is generally preferable to avoid quasi-experimental designs.

Many true experimental designs can be converted to quasi-experimental designs by omitting random assignment. For instance, the quasi-equivalent version of pretest/post-test control group design is called nonequivalent groups design (NEGD), as shown in Table 8, with random assignment R replaced by non-equivalent (non-random) assignment N . Likewise, the quasi-experimental version of switched replication design is called non-equivalent switched replication design (see Table 9).

N	O_1	X	O_2	TREATMENT GROUP
N	O_3		O_4	CONTROL GROUP

Table 8: Nonequivalent Groups Design

N	O_1	X	O_2		O_3
N	O_4		O_5	X	O_6

Table 9: Nonequivalent Switched Replication Design

In addition, there are quite a few unique non-equivalent designs without corresponding true experimental design cousins, including those listed below.

REGRESSION-DISCONTINUITY (RD) DESIGN. This is a non-equivalent pretest/post-test design where subjects are assigned to treatment or control group based on a cutoff score on a pre-program measure. For instance, patients who are severely ill may be assigned to a treatment group to test the efficacy of a new drug or treatment protocol and those who are mildly ill are assigned to the control group. In another example, students who are lagging behind on standardized test scores may be selected for a remedial curriculum program intended to improve their performance, while those who score high on such tests are not selected from the remedial program. The design notation can be represented as in Table 10, where C represents the cutoff score:

C	O_1	X	O_2	TREATMENT GROUP
C	O_3		O_4	CONTROL GROUP

Table 10: Regression-discontinuity Design

Because of the use of a cutoff score, it is possible that the observed results may be a function of the cutoff score rather than the treatment, which introduces a new threat to internal validity. However, using the cutoff score also ensures that limited or costly resources are distributed to people who need them the most rather than randomly across a population, while simultaneously allowing a quasi-experimental treatment. The control group scores in the *RD* design do not serve as a benchmark for comparing treatment group scores, given the systematic non-equivalence between the two groups. Rather, if there is no discontinuity between pretest and post-test scores in the control group, but such a discontinuity persists in the treatment group, then this discontinuity is viewed as evidence of the treatment effect.

PROXY PRETEST DESIGN. This design, shown in Table 11, looks very similar to the standard NEGD (pretest-post-test) design, with one critical difference: the pretest score is collected after the treatment is administered. A typical application of this design is when a researcher is brought in to test the efficacy of a program (e.g., an educational program) after the program has already started and pretest data is not available. Under such circumstances, the best option for the researcher is often to use a different prerecorded measure, such as students' grade point averages before the start of the program, as a proxy for pretest data. A variation of the proxy pretest design is to use

subjects' post-test recollection of pretest data, which may be subject to recall bias, but nevertheless may provide a measure of perceived gain or change in the dependent variable.

N	O_1	X	O_2	TREATMENT GROUP
N	O_3		O_4	CONTROL GROUP

Table 11: Proxy Pretest Design

SEPARATE PRETEST/POST-TEST SAMPLES DESIGN. This design is useful if it is not possible to collect pretest and post-test data from the same subjects for some reason. As shown in Table 12, there are four groups in this design, but two groups come from a single non-equivalent group, while the other two groups come from a different non-equivalent group. For instance, to test customer satisfaction with a new online service that is implemented in one city but not in another, customers in the first city serve as the treatment group and those in the second city constitute the control group. If it is not possible to obtain pretest and post-test measures from the same customers, you can measure customer satisfaction at one point in time, implement the new service program, and measure customer satisfaction (with a different set of customers) after the program is implemented. Customer satisfaction is also measured in the control group at the same times as in the treatment group, but without the new program implementation. The design is not particularly strong, because changes in any specific customer's satisfaction score before and after the implementation cannot be examined, only the average customer satisfaction scores. Despite the lower internal validity, this design may still be a useful way of collecting quasi-experimental data when pretest and post-test data are not available from the same subjects.

N_1	O_1		
N_1		X	O_2
N_2	O_3		
N_2			O_4

Table 12: Separate Pretest/Posttest Samples Design

NONEQUIVALENT DEPENDENT VARIABLE (NEDV) DESIGN. This is a single-group pre/post quasi-experimental design with two outcome measures, where one measure is theoretically expected to be influenced by the treatment and the other measure is not. For instance, if a new calculus curriculum is being designed for high school students, the curriculum is anticipated to influence students' post-test

calculus scores but not algebra scores. However, the post-test algebra scores may still vary due to extraneous factors such as history or maturation. Hence, the pre/post algebra scores can be used as a control measure, while the pre/post calculus scores can be treated as the treatment measure. The design notation, shown in Table 13, indicates the single group, N , followed by pretest O_1 and post-test O_2 for both calculus and algebra for the same group of students. This design is weak in internal validity, but its advantage lies in not having to use a separate control group.

An interesting variation of the NEDV design is a pattern matching NEDV design, which employs multiple outcome variables and a theory that explains how much each variable will be affected by the treatment. The researcher can then examine if the theoretical prediction is matched in actual observations. This pattern-matching technique, based on the degree of correspondence between theoretical and observed patterns is a powerful way of alleviating internal validity concerns in the original NEDV design.

N	O_1	X	O_2
	O_3		O_4

Table 13: Nonequivalent Dependent Variable Design

1.1.6 Strengths and Weaknesses of Experimental Research

A strength of the experimental method, particularly in cases where experiments are conducted in lab settings, is that the researcher has substantial control over the conditions to which participants are subjected. Experiments are also generally easier to replicate than are other methods of data collection, especially in cases where an experiment has been conducted in a lab setting.

On the other hand, experimental research is one of the most difficult of research designs, and should not be taken lightly. This type of research is often beset with a multitude of methodological problems.

- Though experimental research requires theories for framing hypotheses for testing, much of current experimental research is atheoretical. Without theories, the hypotheses being tested tend to be ad hoc, possibly illogical, and meaningless.
- Many of the measurement instruments used in experimental research are not tested for reliability and validity and are incomparable across studies. Consequently, results generated using such instruments are also incomparable.
- Many experimental research use inappropriate research designs, such as irrelevant dependent variables, no interaction effects, no

experimental controls, and nonequivalent stimulus across treatment groups. Findings from such studies tend to lack internal validity and are highly suspect.

- The treatments (tasks) used in experimental research may be diverse, incomparable, and inconsistent across studies and sometimes inappropriate for the subject population. For instance, undergraduate student subjects are often asked to pretend that they are marketing managers and asked to perform a complex budget allocation task in which they have no experience or expertise. The use of such inappropriate tasks introduces new threats to internal validity (i.e., subject's performance may be an artifact of the content or difficulty of the task setting), generates findings that are non-interpretable and meaningless, and makes integration of findings across studies impossible.

The design of proper experimental treatments is a very important task in experimental design, because the treatment is the *raison d'être* of the experimental method, and must never be rushed or neglected. To design an adequate and appropriate task, researchers should use prevalidated tasks if available, conduct treatment manipulation checks to check for the adequacy of such tasks (by debriefing subjects after performing the assigned task), conduct pilot tests (repeatedly, if necessary), and if there is doubt, using tasks that are simpler and familiar for the respondent sample than tasks that are complex or unfamiliar.

Time, other resources such as funding, and even one's topic may limit a researcher's ability to conduct an experiment. For researchers in the medical and health sciences, conducting an experiment could require denying needed treatment to patients, which is a clear ethical violation. Even those whose research may not involve the administration of needed medications or treatments may be limited in their ability to conduct a classic experiment. In social scientific experiments, for example, it might not be equitable or ethical to provide a large financial or other reward only to members of the experimental group.

1.2 SUMMARY

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et

Part III

QUALITATIVE METHODS

Qualitative methods are based in the evaluation of non-numeric data, like photographs and text documents. These methods include activities like field work, unobtrusive, and interpretive research methods.

Part IV

MIXED METHODS

All quantitative and qualitative research methods have certain strengths and weaknesses. Mixed methods are an attempt to use more than one research method on a given project to utilize the strengths of each method while mitigating their weaknesses.

Part V

REPORTING

After a research project is completed, the investigator must report the results of the project, often in both written and oral forms. This chapter concerns the reporting process.

Part VI

APPENDIX

APPENDIX

No appendix content yet.

GLOSSARY

applied research	Research that is intended to be applied to a situation rather than further the knowledge of some topic. For example, if a researcher completes a project designed to increase the sales of bottled water in a small town it would be considered applied research. <i>see</i>
basic research	Research that is intended to be further the knowledge of some topic rather than be applied to a specific situation. For example, if a researcher completes a project designed to refine some aspect of the Law of Supply and Demand it would be considered basic research. <i>see</i>
bivariate	A type of analysis involving two variables. Examples of bivariate analysis include finding a correlation and regression. <i>see</i> univariate
concurrent validity	The degree that a measure relates to an outcome that is presumed to occur simultaneously. For example, the results of a new employee attitude test would be the same as an older test if those tests have high concurrent validity. <i>see</i> predictive validity
construct validity	The degree to which a test measures what it claims to measure. For example, if a research project purports to investigate some aspect of local farmers' markets, does the project actually research that aspect? Construct validity is sometimes thought to be the overarching type of validity since research projects that do not address the construct of interest can have no other validity. <i>see</i> validity

content validity	A determination of whether a measure correctly assesses the construct's content. For example, if a research project is attempting to determine the drivers for total sales in a store but only measured the price of the merchandise being sold then ignoring factors like advertising, competition, and even the general economy of the region would call into question the content validity of the study. <i>see</i> validity
continuous data	Continuous data are a type of quantitative data that can represent any measured value, including fractions and decimals. In mathematics terms, continuous data are members of the real number system. <i>see</i> quantitative data
convergent validity	The closeness that two measures relate to, "converge on," a single construct. For example, if a research project measures the amount of sales of carbonated drinks, fruit juices, and bottled water in a store it would be expected that those would converge on a construct of "drink sales." <i>see</i> discriminant validity
criterion validity	The degree to which a measure is related to an outcome. <i>see</i> validity
dependent variable	Dependent variables are the outcomes for an observation. For example, if a medical researcher conducts an experiment where a drug is administered and then the patient's blood pressure is measured, the blood pressure reading is the dependent variable; that is, the blood pressure depends on the drug being administered. <i>see</i> independent variable
descriptive research	Research that is designed to describe observed phenomena. The goal is to improve understanding rather than explore new ideas. <i>see</i> exploratory research
discrete data	Discrete data are a type of quantitative data that can be counted with integers. In mathematics terms, discrete data are integers, though negative values are rather rare. <i>see</i> quantitative data

discriminant validity	The degree that a measure does not measure, “discriminates between,” one of two competing constructs. For example, a measure of the sale of toiletries in a department store would not be related to the construct of “drink sales.” <i>see</i> convergent validity
explanatory research	Research that is designed to explain an observed phenomena or process. <i>see</i> exploratory research
external validity	The degree to which a research project’s results can be applied outside the context of the study. For example, if the results of a research project that studied manufacturing firms in the mid-west could be applied to firms in the south then that study would have high external validity. <i>see</i> validity
face validity	A determination of whether an indicator is a reasonable measure of an underlying construct “on its face.” For example, is the amount of money spent on live theater tickets a measure social class? <i>see</i> validity
independent variable	Independent variables are those that create an observed effect. For example, if a farmer conducts an experiment where different types of fertilizer are applied to two fields in order to see which is more effective then the type of fertilizer is the independent variable; that is, the type of fertilizer is the variable that is creating the observed effect. <i>see</i> dependent variable
inductive research	A research methodology that works from specific observations to a general theory. This is sometimes called the “theory-building” form of research. 12

internal validity	The degree to which a research project avoids confounding multiple variables within the study. A project with high internal validity facilitates selecting one explanation over an alternate since the number of confounding variables are controlled. <i>see</i> validity
interval data	Interval data are a type of quantitative data that are measured along a scale where each point is an equal distance from the next. It is possible to compare the distance between two points on an interval scale; for example, the difference between 90 and 100 degrees is the same as the difference between 40 and 50 degrees. However, since an interval scale does not have a zero point, stating 100 degrees is twice as hot as 50 is not possible. <i>see</i> quantitative data
nominal data	Nominal data are a type of qualitative data that are grouped but with no order implied in the grouping. As an example, the gender of survey respondents is nominal data. <i>see</i> qualitative data
nonparametric	Nonparametric data are data that do not conform to a distribution, are skewed, or are qualitative in nature. Statistical tests that work with nonparametric data are generally less powerful and predictive than tests that work with parametric data. <i>see</i> nonparametric
ordinal data	Ordinal data are a type of qualitative data that are grouped where the groupings have an implied order. As an example, the “satisfaction” rating on a customer survey typically permits respondents to choose from several levels where one level is somehow better than another. <i>see</i> qualitative data
paradigm	A pattern or model of how things work in the world. <i>see</i> theory

parametric	Parametric data are data that conform to a distribution, usually a normal distribution. Statistical tests that work with parametric data are generally much more powerful and predictive than tests that work with nonparametric data. <i>see</i> nonparametric
positivist	A researcher who uses positivist techniques on research projects. <i>see</i> positivism
predictive validity	The degree to which a measure predicts an outcome. For example, does increasing beer sales (a measure) predict increasing potato chip sales? <i>see</i> concurrent validity
qualitative data	Qualitative data approximates or describes attributes that cannot be directly measured, like employee morale, customer relationships, and management effectiveness. Typically, qualitative data attempt to answer questions like “why” and “how come.” <i>see</i> quantitative data
qualitative research	Qualitative research typically intends to explore observed phenomena with a goal of developing hypotheses and dive deep into a problem. Qualitative data collection involves semi-structured activities like focus groups and ethnographies. 7, <i>see</i> quantitative research
quantitative data	Quantitative data are numeric measurements of attributes, like the number of employees, the median value of housing, and total revenue. Quantitative data are gathered and analyzed using statistical methods. <i>see</i> qualitative data
quantitative research	Quantitative research typically uses numerical data and statistical analysis to find patterns and generalize results to a large population. Quantitative data collection involves structured activities like surveys, interviews, and systematic observations. <i>see</i> qualitative research

ratio data	Ratio data are a type of quantitative data that are measured along a scale where each point is an equal distance from the next and there is a zero point. An example of ratio data is people's heights, which is measured along a uniform scale, e.g. inches or centimeters. Because there is a true zero point, it is possible to determine that one person is twice as tall as another. <i>see</i> quantitative data
reliability	A descriptor for the consistency of a concept's measure. It is desirable to achieve the same, or nearly same, values for each sampling. For example, if the mean age of the people in one sample is 30 while in another is 50 then this would indicate a problem with reliability of the data. <i>see</i> validity
statistical conclusion validity	The degree to which the conclusions found in a research project are correct. Studies with high statistical conclusion validity minimize the two types of statistical errors: Type I (finding a correlation when there is none) and Type II (failing to find a correlation when one exists). <i>see</i> validity
theory	A system of ideas that is intended to explain phenomena. Theories that are accepted by scientists have been repeatedly tested and can be used to make accurate predictions. Unlike common usage, a scientific theory is a tested, falsifiable explanation for phenomena. <i>see</i> paradigm
translational validity	The degree to which a construct has been measured by a research project. Translational validity is divided into two types: face and content. <i>see</i> validity
univariate	A type of analysis involving a single variable. Univariate analysis findings include the central measure, standard deviation, and frequency distributions. Graphic tools include box plots for continuous data and bar plots for discrete data. <i>see</i> bivariate

validity

A descriptor of whether a research project is measuring the variable under question. For example, if a project hypothesis is that older men tend to tip more than younger men then the validity of the study would be in question if the researcher only sampled men under the age of 40. *see* [reliability](#)

BIBLIOGRAPHY

- [1] Hunt Allcott and Todd Rogers. "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation." In: *American Economic Review* 104.10 (2014), pp. 3003–37.
- [2] Anol Bhattacharjee. "Social science research: Principles, methods, and practices." In: (2012).
- [3] Amy Blackstone. *Principles of sociological inquiry—qualitative and quantitative methods*. The Saylor Foundation, 2012.
- [4] Paul Christopher Price, I-Chant A Chiang, and Rajiv Jhangiani. *Research methods in psychology*. BCcampus, BC Open Textbook Project, 2015.

COLOPHON

This document was typeset using the typographical look-and-feel classicthesis developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". classicthesis is available for both L^AT_EX and L^yX:

<https://bitbucket.org/amiede/classicthesis/>

Happy users of classicthesis usually send a real postcard to the author, a collection of postcards received so far is featured here:

<http://postcards.miede.de/>