University of Groningen

Research designs for environmental issues

Abrahamse, Wokje; Schultz, P. Wesley; Steg, Linda

Published in:
Research Methods for Environmental Psychology

DOI:
[10.1002/9781119162124.ch4](https://doi.org/10.1002/9781119162124.ch4)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Abrahamse, W., Schultz, P. W., & Steg, L. (2016). Research designs for environmental issues. In R. Grifford (Ed.), *Research Methods for Environmental Psychology* (pp. 53-70). Wiley-Blackwell.
<https://doi.org/10.1002/9781119162124.ch4>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Research Designs for Environmental Issues

Wokje Abrahamse¹, P. Wesley Schultz², and Linda Steg³

¹*Victoria University of Wellington, New Zealand*

²*California State University, San Marcos, CA, United States*

³*University of Groningen, The Netherlands*

“Hey, guys, listen to this!” Ethan and Gabriel look up from their bowls of cereal. Maria is scanning the morning headlines. “Green space linked to obesity in city dwellers.” Ethan has a sip of coffee and ponders this for a moment. “That doesn’t make any sense,” he says. “You’d expect it to be the exact opposite, green space linked to slim waists in city dwellers.” Gabriel just grunts. He’s not fully awake yet.

Maria continues reading: “The researchers collected data on more than 10,000 people living within walking distance of a park and found unexpectedly high levels of obesity.” “Did they compare this to people living further away from a park?” asks Ethan. “The article doesn’t say,” Maria responds. “Well, there you have it. No comparison group.” “Hang on a minute.” Gabriel’s coffee has finally kicked in. “Surely there are other factors that play a role in the prevalence of obesity. In my social influence class, our prof was telling us about the role of social networks in the spread of obesity. If your friend is obese, your likelihood of also being obese significantly increases. This could help explain the results.”

“Either way,” Maria says “the results are based on correlational data, so there is no way they can claim cause and effect. They cannot rule out the possibility that another variable is influencing both obesity and walking distance to green space.” “Maybe that’s just the headline though. You know how the media loves sound bites. Is this FOX News?” Maria nods. “My point exactly.”

Environmental psychology is an empirical science. This means that the discipline is grounded in research in which hypotheses are tested using observations. Research questions and theories are used to guide predictions about human beliefs, emotions, behavior, or performance, and these predictions are tested through various research designs and measurement techniques. Research studies in environmental psychology are guided by diverse research questions. For example, why do people

buy environmentally friendly products? How do we reduce littering behavior in public places? Or, what is the best way to reduce household electricity consumption? Some studies focus on demonstrating relationships between variables, such as the relationship between environmental values and preference for green products. Other studies manipulate one or more variables in an effort to identify the causes of beliefs or behavior, such as whether the provision of information about climate change causes a change in attitudes toward climate change.

In the first part of this chapter, we describe the defining features of three main research designs used in environmental psychology, namely experimental, quasi-experimental, and correlational studies (for related chapters on survey research, see Chapter 5; for field studies, see Chapters 15 and 16). We highlight some of the main advantages and disadvantages of each design. At the end of the chapter, we discuss some general issues in conducting research, including measurements of variables, mediator and moderator variables, validity, and replication.

Choosing a Research Design

An **experimental design** includes the systematic variation of one or more independent variables, and is characterized by random assignment of participants to conditions, or groups. In a **quasi-experimental design**, one or more independent variables are systematically varied, but random assignment of participants to condition does not occur. A study using a **correlational design** describes the relationship between independent and dependent variables, without systematic variation of these variables or random assignment of participants (see also Pelham & Blanton, 2007). In research on human–environment issues, researchers typically use correlational and quasi-experimental designs to examine their research questions (see, e.g., De Groot & Steg, 2008). Increasingly, however, experimental studies are being conducted to examine environmental issues (e.g., Griskevicius, Tybur, & Van den Bergh, 2010).

Experimental designs

Experimental designs are **true experiments** that address research questions that aim to find out the cause of a particular phenomenon (see also Pelham & Blanton, 2007). Experimental designs share three important characteristics: manipulation of relevant variables, random assignment to treatments; and the use of comparison groups to eliminate the effects of other variables. **Manipulation** of variables occurs when one or more variables is systematically varied, while holding the other variables constant.

For example, consider an experiment designed to determine whether exposing people to environmental concepts affects people's preferences for green products. The researcher may ask one group of participants to complete a word association task that contains environmental words (e.g., green, conservation), while another group of participants (the control group) is asked to fill out a word association task that consists of neutral words (e.g., desk, shoe), before completing the same product preference task. **Random assignment** of participants to groups occurs when the

researcher arbitrarily assigns participants to experimental and control groups. This way, each participant has an equal chance of ending up in either group (e.g., exposure to environmental words or exposure to neutral words). Any difference between groups in the outcome measure (e.g., preference for green products) is very likely attributable to the manipulation (e.g., activation of environmental concepts), not to group differences in other characteristics.

In experimental designs, the variable that is systematically varied is referred to as the **independent variable** and the outcome measure is referred to as the **dependent variable**. The group of participants that is not exposed to the manipulation is generally referred to as a **control group** or a comparison group. A control group consists of individuals who complete a similar task or activity to the manipulated variable, while a comparison group often does not engage in any activity.

Experimental designs have a number of advantages and disadvantages (see also Creswell, 2009; Pelham & Blanton, 2007). Experimental designs are generally high in **internal validity**, which refers to the ability to infer causal relationships between independent and dependent variables. In lab studies, the researcher has a great deal of control over what happens, such as which variables are varied and which variables are held constant. Because participants are randomly assigned to groups, any differences between groups on the outcome measure may be attributed to the treatment or the manipulation. It is less likely that group differences are attributable to systematic differences between groups.

Further, in experimental designs, the researcher can minimize the effect of confounding variables. A confounding variable is an additional variable that is correlated with the independent variable and affects the dependent variable. An **operational confound** occurs when a measure designed to assess a particular construct also measures something else. For example, if a researcher measures people's product preferences with a choice task, this may also gauge people's decision-making abilities. It is important to carefully select the dependent and independent measures. This can be done for example by pre-testing a measure in a pilot study. A **procedural confound** occurs when the researcher inadvertently varies more than one independent variable.

Imagine, for example, that all sessions of one experimental condition (e.g., control condition) were run in the same week, in which it happened to be raining every day and that the other sessions (e.g., exposure to environmental concepts) were run the week after, in which it happened to be sunny every day. Perhaps the good weather put people in a good mood, which made them more likely to choose environmentally friendly concepts. The researcher would not be able to draw a causal conclusion between the independent variable (exposure to environmental concepts) and the dependent variable (preference for green products) because of a confounding factor (the weather). Obviously, one cannot control the weather, but the researcher does have a great deal of control over procedures (e.g., time of day, day of the week, location, training of research assistants) to ensure that there is no systematic variation between groups as a result of how the research is set up.

A **person confound** refers to an individual difference variable among participants in the study, which may affect the outcome measure. For example, people may differ in their preference for green products, or people may have different levels of environmental concern from the outset. Because experimental designs use random

assignment of participants to conditions, they are generally protected against this type of confound, as normally these individual differences would be equally distributed across conditions and not vary systematically with the outcome measure.

An experimental design also allows for the test of **interaction** effects. Interaction effects indicate that the effect of an independent variable on the dependent variable depends on another independent variable. This type of design, where two independent variables are systematically varied to assess interaction effects, is called a factorial design. For example, a researcher could examine the effect of exposure to environmental concepts and social influence on people's preference for green products, whereby both exposure to environmental concepts and social influence are systematically varied. The researcher may find, for example, that the effect of exposure to environmental concepts on people's preference for green products depends on whether or not participants have been told that a majority of other people prefer green products. Interactions can help in providing insight into the conditions under which a hypothesized relationship does and does not apply. For example, as a general theory, exposure to environmental concepts can be said to increase people's preference for green products, but only when people are told that a majority of other people also prefer green products.

A disadvantage of experimental designs is that they generally have low **external validity**, which refers to the ability to generalize findings from a study to different situations or populations of individuals. Lab studies are conducted in artificial situations that are unlikely to occur in everyday settings. While experimental studies provide an opportunity to examine one or two variables in isolation, this is not representative of real-life situations. One possible solution is to make the lab setting as realistic as possible, or alternatively, to conduct a study in a naturally occurring setting which allows a similar level of control over variables.

Yet another possibility is to replicate the findings using other research designs (see also the section on replication at the end of the chapter). Experimental studies often make use of convenience samples (e.g., undergraduate psychology students). The use of such samples is practical (e.g., labs are usually situated on campus, ease of recruitment of students). Another advantage of convenience samples is that the sample is relatively homogeneous, because variables such as gender, age, and education are held constant. However, this limits the possibility to generalize the findings to other populations. A possible solution to this problem is to conduct follow-up studies with different samples to examine whether the findings are generalizable to other populations.

Quasi-experimental designs

Quasi-experimental designs can be described as holding the middle ground between true experiments and correlational designs. While quasi-experimental designs are generally characterized by a relatively high level of control over variables, manipulation and random assignment are often not possible.

Many different types of quasi-experimental designs are known (see Campbell & Stanley, 1963). The most commonly used design is the **person-by-treatment quasi-experiment**. In this design, one independent variable is systematically varied and the other independent variable is measured. For example, a quasi-experimental

Box 4.1 Example of a true experiment.

Griskevicius et al. (2010) examined why people purchase green products, given that green products are often more expensive than conventional products. They argued that a higher price may be associated with a higher status and that buying a more costly, green product reflects positively on one's social status.

To examine this, Griskevicius and colleagues set up an experiment in which they systematically varied status motivations. They examined whether activating a desire for social status among participants affected their choice of green versus conventional products. One hundred and eighty undergraduate students were recruited and randomly assigned to one of two conditions. In one condition, the students read a scenario that was aimed to elicit a desire for social status. As part of this scenario, participants were asked to imagine graduating from college, landing their first job, with immediate possibilities for promotion, and the potential of moving up in social status relative to their peers. In a control condition, a neutral scenario was presented. Participants then completed an (ostensibly unrelated) product preference task. For each product, participants were asked to choose between a green or a conventional (more luxurious) product.

The results provided support for the study's main hypothesis. In the control group, the non-green products were chosen more often than the green products, but when a desire for social status was activated, participants were more likely to choose a green product. This study found that activating status motives increased the likelihood that people choose a green product over a non-green (but more luxurious) product. In their second experiment, they examined how status motives affected preference for green products when people considered shopping in a public setting (a store) or a private setting (online at home). This study used the same manipulation of status motive as in study 1. Then, participants were asked to choose between green and non-green items, but one group of participants were asked to imagine they were online shopping at home alone and another group of participants were asked to imagine they were out shopping at a store. The researchers found a significant interaction effect between status manipulation and shopping in the public or private sphere. Activating status motives encouraged people to choose green products more often, in particular when shopping in public (see Figure 4.1).

design can be set up to address the question of whether the effect of information provision about the environmental benefits of eating organic food is different for people with different environmental values. In this case, one of the independent variables is manipulated (environmental information or neutral information) and one of the independent variables is measured (environmental values).

Imagine a study in which all participants fill out a survey to measure environmental values. Then, one group of participants watches a video with information

Figure 4.1 Preference for green products versus non-green products (y-axis) in the control group and the status group. The dark gray bar indicates the choice was made in the private sphere (at home, online), the light gray bar indicates the choice was made in the public sphere (in a store).

about organic food, while another group watches a video on an unrelated topic. This study may find that the provision of information about environmental benefits of organic food is effective in encouraging organic food choices, but that this effect depends on the extent to which people endorse environmental values. For example, environmental information may only be effective when people already have relatively strong environmental values, but when they have weak environmental values, the information does not affect their choices.

A drawback of quasi-experimental designs is that random assignment to conditions often does not occur, which may introduce systematic differences between the groups. However, researchers are able to include **covariates**, which allow them to examine the relationship between the dependent and independent variables of interest, while controlling for other relevant variables. In our example, the researchers might consider including covariates such as gender, income, or scores on an environmental knowledge test. By including these covariates, the researcher can examine whether the difference in preference for organic foods are indeed related to the variable of interest (information provision). By statistically eliminating the effect of other, relevant variables, one can be more certain of the existence of a relationship between the independent and dependent variables.

One of the main advantages of quasi-experimental designs is that they allow researchers to include variables that are difficult to mimic in experimental lab studies, which makes them very suitable for examining applied, real-life issues, such as environmental issues. Also, in quasi-experimental designs, researchers have a relatively large amount of control over the variables they wish to include. By including covariates, it can help make alternative explanations less plausible. Of course, one would need to know beforehand which covariates to include (based, e.g., on previous research), and sometimes not all potential covariates are known. Research in environmental psychology often makes use of quasi-experimental designs (see Box 4.2).

Box 4.2 Example of a quasi-experiment.

Verplanken and Holland (2002) examined whether the activation of environmental values would affect environmental decision making. They were interested to know whether this effect depended on how central or important environmental values were for participants. To measure the centrality of environmental values, the researchers used the Schwartz Values Inventory. To activate environmental values, participants were given a description of a person and they were asked to form an impression of this person. In one condition, the person was described in terms of values related to the environment (e.g., concerned about nature), and in the control condition, the values that described this person were unrelated to the environment (e.g., perfectionism). Then, participants were asked to complete an (ostensibly unrelated) consumer choice task. Participants were asked to choose between 20 television sets. Some television sets scored very well on environmental aspects (e.g., efficient use of electricity) while other television sets scored poorly on environmental aspects.

The results provided support for the environmental value activation hypothesis. That is, when environmental values were activated, participants were more likely to choose a television set that scored more favorably on environmental aspects, compared to when environmental values were not activated. In addition, this effect was stronger for participants for whom environmental values were more important. These results are important in that they suggest that making environmental values salient (e.g., in an information campaign) will be more effective in encouraging environmentally related choices among people for whom environmental values are important, compared to people for whom environmental values are less important.

Correlational designs

Correlational research examines the relationship between two or more variables in a naturally occurring setting (Pelham & Blanton, 2007). For example, a researcher may be interested in the relationship between people's level of physical activity and level of access to green spaces. Correlational studies do not involve a form of manipulation of variables as is the case in lab studies; rather, all variables are assessed as they occur, for example, by means of a survey (see Chapter 5) or observation (see Chapter 2). The **correlation coefficient** gives an indication of the strength and the direction of the association between two variables, ranging from -1 to $+1$. The strength of the association is represented by the value of the correlation, the higher the number the stronger the two variables are related to one another. The sign indicates the direction of the association: a negative correlation suggests that an increase in one variable is associated with a decrease in the other variable; a positive correlation indicates that an increase in one variable is associated with an increase in the other variable.

The main advantage of correlational designs is that they possess relatively high external validity. Correlational studies are often conducted in real-life settings among

representative groups of the population and they measure (rather than manipulate) variables. These features of correlational studies contribute to a greater degree of generalizability of the findings to other situations, compared to experimental studies. In addition, when the study is conducted among representative samples of participants, the study findings may be more readily generalizable to other people.

Correlational studies have a number of important drawbacks. In correlational studies researchers have much less experimental control over the variables, which means that these studies can only infer correlation and not causality. One of the main issues related to correlational studies is **reverse causality**. Reverse causality refers to the fact that there are different ways of interpreting the same correlation. For example, a significant positive correlation between people's environmental concern and their engagement in environmentally friendly behavior may signify that because people have a higher environmental concern, they are more likely to engage in environmentally friendly behavior. But it may also be that because people engage in environmentally friendly behavior, they develop a stronger environmental concern. In other words, the direction of the relationship may be reversed.

Another disadvantage of correlational studies is the potential presence of a third variable which affects both variables. The **third-variable problem** refers to the

Box 4.3 Example of a correlational study.

What requires more energy: transporting goods via airplanes, or via trucks? What requires more energy: making a glass bottle or an aluminum can? Attari, DeKay, Davidson, and de Bruin (2010) examined the accuracy of people's perceptions of the energy use of several appliances and activities. They conducted an online survey among 505 participants in seven US metropolitan areas (New York, Philadelphia, Washington DC, Houston, Dallas, Denver, and Los Angeles). Participants were asked to estimate the energy requirements of a number of devices and appliances and the energy saved by several household activities (the energy used by a 100-watt incandescent light bulb in one hour was provided as a reference point). The results indicate that participants' estimates of energy use and savings were not always accurate (see Figure 4.2). Note that in most cases the estimates fall below the accurate-estimate line, indicating that people tend to underestimate the amount of energy required by different devices or saved through behavior changes. For example, participants incorrectly reported that making a glass bottle required less energy than making an aluminum can (the reverse is true). However, participants did correctly indicate that transporting goods via airplanes required more energy compared with other modes of transportation. The study also examined individual differences in these estimates and found that people with more positive environmental attitudes (measured by the New Environmental Paradigm; see also Chapter 6) had more accurate estimates. The study reported a positive correlation between environmental concern and accuracy of estimates, $r = .38$.

Figure 4.2 Mean perceptions of energy used or saved in relation to actual energy use or saved. The diagonal line represents accurate estimates. Note that in most cases the estimates fall below the accurate-estimate line, indicating that people tend to underestimate the amount of energy required by different devices or saved through behavior changes. (Source: Attari et al., 2010.)

existence of a third variable (not included in the study), which relates to both variables involved, that better explains their relationship. Consider the finding that city dwellers living in areas with little greenery are more likely to be obese than those living in pleasant areas with lots of greenery. This may also be explained by other features of these areas (other than lack of greenery, such as lack of walking trails, or access to fast food places). So-called third variables are not included in a particular study, but they may account for the observed effects.

Measurement

Measurement is at the heart of research. In the research process, once the hypotheses are formalized, the researcher must operationalize them. This involves providing clear definitions of the constructs implicated in the hypotheses, and specifying how these constructs can be measured in the research process. While measurement typically focuses on the dependent variable, measured variables can be used in a variety of other ways, including as non-manipulated independent variables in quasi-experimental and correlational studies. We discuss a number of issues related to measurement, including the four primary levels of measurement, different types of measures, and the uses of measured variables in the research process.

Levels of measurement

Variables can be measured on a variety of levels, and in a variety of ways. Measured variables can be classified into four levels: nominal, ordinal, interval, and ratio. These four levels range from the most basic level of measurement (nominal), to the most complex (ratio).

- **Nominal:** Measures that involve the categorization of observed events. For example, coding whether or not a person recycles a beverage container, where “yes” is coded as the numeric value of 1, and “no” is coded as the numeric value of “0.” Note that the designation of 1 and 0 is arbitrary, and the observations could have just as easily been coded as the reverse. Nominal variables can be dichotomous, as in the above example, or polytomous, with many categories. For example, the different types of material recycled might include paper, plastic, glass, and organics, coded from 1 to 4 respectively.
- **Ordinal:** Measures that assign numeric values to observed events, and the order of these values is meaningful. For example, a researcher staffing a household waste collection event on a Saturday might count the sequence of residents who come to the event. The first resident to show up would be coded 1, the second 2, and so on until the last resident. These numbers indicate the time when residents arrived, and higher scores indicating later arrivals. But note that the distance between the scores is not meaningful. It could be that in the morning, only a few residents arrived, whereas the afternoon was crowded. This distance between residents would not be represented in the ordinal level measure.
- **Interval:** Measures in which the interval between the units is meaningful, but where 0 does not indicate the complete absence of the construct. For example, a researcher studying environmental concern might ask participants in a survey to rate their level of agreement with a statement from 1 = strongly disagree to 5 = strongly agree. These Likert-style measures allow researchers to quantify psychological constructs like “environmental concern,” and to create scales that allow for comparisons across individuals. But note that 0 does not indicate the absence of concern.
- **Ratio:** Measures in which the interval between units is meaningful, and where 0 indicates the complete absence of the construct. For example, a researcher studying recycling by businesses might quantify the behavior by weighing the amount of materials collected each week in kilograms. Other ratio-level measures include physical properties like distance, length, mass, speed, and volume.

A large amount of the work involved in scientific research focuses on measurement. Studies are designed to use a specific measurement protocol, using specific tools, and to obtain observations from a sample of the target population on a specific level. Where possible, researchers strive for ratio-level measures. These measures contain the maximum amount of information, and most parametric statistical techniques were developed for analyzing ratio-level data. In contrast to ratio-level measures, ordinal and nominal levels may restrict which analyses one can conduct.

Types of measures

Measures can be obtained in a variety of ways, and environmental psychologists are often quite creative in developing their measurement protocols. Observational data are obtained by coding an event that is personally witnessed. For example, measuring whether or not a walker in a park sits on a bench, which would be coded using a nominal-level measure (1 = yes, 0 = no). The researcher could have also opted for a ratio-level measure, recording the number of seconds that each walker in the park spent sitting on a bench (from 0 to a defined maximum value).

A second type of measure in environmental psychology involves asking individuals to answer questions. These questions can be asked in a structured interview, using paper and pencil, a webpage, or on a mobile electronic device like an iPad. The questions can be asked using an open-ended format (see also Chapter 7 on qualitative measures), in which respondents elaborate in their own words, or closed ended in which respondents use a specified response scale. Note that self-report data can be obtained on any of the four measurement levels. For example, asking a person to report the amount of their last utility bill (in dollars) would provide a ratio-level measure, whereas asking a participant if they paid their last utility bill (1 = yes, 0 = no) would provide nominal-level data.

A third type of measure in environmental psychology involves the coding of residual data. In this type of measure, the researcher codes the residual evidence of a past behavior. For example, a researcher might count the number of littered items in a park as a measure of littering behavior. The researcher did not directly observe the littering, so the residual evidence is used as a proxy for the behavior, and measured at a ratio level. As with the other types of measures, residual data can be quantified at any of the four levels. Other examples of residual data include, for example, utility meter readings or annual mileage.

Psychometrics

Measured variables differ in their quality, and some measures are better than others. In assessing the quality of a measure, researchers consider the reliability and validity. *Reliability* refers to the likelihood that if the same measure was applied to the same event, it would yield the same score. Every measure has some level of unreliability, but the goal in developing good measures is to maximize the reliability. The reliability of a measure is typically quantified as the strength of the relationship between two independent assessments of the same construct.

This can be done in a variety of ways, depending on the type of measure. With observational data, estimates for reliability are often obtained by having two or more coders record the same event – a procedure called inter-rater reliability. With self-report data, researchers often use multiple questions to measure the same construct, and then calculate the strength of the relationship among the items. This is referred to as the internal consistency of a measure and often quantified using Cronbach's alpha (which ranges from 0 to 1, with 1 indicating perfect reliability).

Reliable measures are critical for any area of research, and researchers often spend considerable time refining their measures. In the case of inter-rater reliability, it is important to provide coders with very specific guidelines for recording

their observations. Researchers typically develop and refine their measurement protocols over many days and weeks, coding, discussing, and assessing reliability until a satisfactory level is achieved.

With self-report data, researchers typically utilize existing measures that have been shown to provide good levels of internal consistency in previous research. Note that in utilizing existing measures, it is important to use the full set of items that have been previously tested. While it is tempting to pick and choose items that appear most relevant to the current project, any change to the items can undermine the scale's reliability. It is not always possible to use existing scales and for more information about creating your own scale, see Chapter 3.

Validity refers to a measure that accurately reflects the construct of interest. Consider the case where a researcher conducts a survey asking residents about the degree to which they support a 10-cent deposit on beverage containers, using a Likert-style response scale. Because the researcher has a basic understanding of research design, she asks a series of questions about support and opposition for bottle bills in order to establish that the measure is reliable. But she then (mistakenly) concludes that higher scores on the scale indicates a respondent's level of environmental concern. While environmental concern might be correlated with support for bottle bills, the two constructs are different and obtaining self-report data about bottle bills is not a valid measure of environmental concern.

Typically, the validity of a measure is established in prior research. (Again, note the importance of utilizing measures developed and tested in prior studies.) The validity of a measure can be assessed by examining its correlation with measures of similar constructs. This approach is called convergent validity, and in establishing the validity of a measure it is important that it correlates with other similar measures. For example, in the previous example of attitudes toward bottle deposits, a researcher might ask residents to sign a petition supporting a new law. This behavior (signing the petition) should be correlated with their expressed attitudes toward the policy. In addition, a valid measure should also possess divergent validity. That is, the measure should *not* correlate with unrelated measures. Continuing the previous example, attitudes toward bottle deposits should not be strongly related to support for increasing property taxes to provide better schools in the local area. In establishing a new measure, researchers must show both divergent and convergent validity.

Uses of measured variables

As discussed in the opening section of this chapter, measured variables can serve as independent or dependent variables in a study. While these are the primary uses of measured variables, contemporary research in environmental psychology has used measured variables in other notable ways.

Moderator variables are those that affect the relation between an independent and dependent variable. Consider an experiment in which residents are provided with real-time information about the costs of electricity consumption through an in-home display mounted on the kitchen counter. The display shows the current hourly cost at their given rate of consumption, along with the total amount accumulated over the billing cycle. In this experiment, households are randomly assigned to receive the in-home display, or not, and the average daily consumption in kilowatts

is measured. The independent variable is the presence or absence of the display (a nominal measure), and the dependent variable is the daily amount of electricity consumed (a ratio-level measure of a residual type). The results show that households receiving financial information use less electricity over the next month than do households without a display.

In the energy study described above, there are a number of potential moderators that could affect the impact of the in-home displays. One potential moderator is household income, and it is likely that households with higher incomes will be less affected by the display than low-income households. Other potential moderators might include home ownership, presence of children in the home, or hours spent working outside the home. Such moderators are typically indicated as interactions, whereby the independent variable “interacts” with the moderator variable (see also interaction effects in experimental designs section above).

Mediator variables explain why an effect occurs. Like moderators, mediators are third variables that are measured as part of the research process, but mediators are the generative mechanism by which the independent variable affects the dependent variable. Consider the case of the energy study described above. As a starting point, the research has shown that the presence of the in-home display reduces the amount of electricity consumed per day. But why? One possibility is that residents in the home engage in specific behaviors, such as turning off lights when they leave the room. To identify the mediating mechanism for the effect of the in-home displays, the researcher must measure the behavior (e.g., turning off the lights), establish that the behavior changed as a result of the independent variable (e.g., the display), and finally show that the change in behavior accounted for the reduction in electricity consumption.

A third use of a measured variable is to account for confounding. Confounding occurs when the independent variable co-varies with one or more other measured variables. In the case of an experiment where participants are randomly assigned to condition, confounding is not a concern because random assignment generally removes the possibility for systematic co-variation. However, in the case of quasi-experiments or correlational studies, confounding is an ever-present concern. Consider a research example testing whether employees at a large company who drive a personal car to work are more likely to eat organic food for lunch, compared with employees who take public transport. The independent variable (transportation mode) and the dependent variable (eating organic food) are measured variables and neither is manipulated. One potential confounding variable is income: it is likely that employees who drive to work are more affluent and therefore more likely to purchase organic food. In this case, transportation mode and income are confounded. By measuring income, the researcher can test the possibility that income is confounding the results and address it as a limitation in the study.

General Issues

In this final section, we discuss some general issues that are relevant when conducting scientific research. We first discuss the added value of combining multiple research designs as a key way to secure internal and external validity of research.

Next, we discuss the value of replication and meta-analyses. Finally, we elaborate on possible threats to scientific progress, including the file drawer problem, voodoo correlations, and selective reporting.

Internal and external validity

As described above, internal and external validity of findings are of key importance in scientific research. Internal validity implies that we understand which variables cause certain effects, while external validity implies that results of a specific study can be generalized to other situations or to the population at large. Yet, enhancing internal validity often involves a loss of external validity, and the other way around. One way of securing both internal and external validity is by combining multiple research designs in a series of studies examining the same issues.

Consider a study that is aimed at understanding whether environmental knowledge encourages pro-environmental actions. This could be examined with a questionnaire study among a representative sample of the population, including questions on environmental knowledge and pro-environmental behavior, and study the correlation between both constructs. The external validity of this study would be rather high, given that the reliability and validity of the measures are high. This implies that results can probably be generalized to different populations or different situations (e.g., you are likely to find similar correlations when you conduct the study again one year later, in a comparable sample). Yet, internal validity is likely to be low, because it is not clear whether environmental knowledge caused pro-environmental actions (or the other way around), or whether a third variable may have accounted for the observed correlation. For example, environmental concern may have increased both environmental knowledge and pro-environmental behavior.

To address these concerns, you could run an experimental study among university students in which you manipulate environmental knowledge (e.g., by presenting information on environmental problems via a documentary or a leaflet), and examine its effect on pro-environmental behavior. If you find that the experimental group (who received the information) scores better on a subsequent knowledge test (which serves as a manipulation check) and more often engages in subsequent pro-environmental behavior than a control group who received unrelated information (e.g., on university policies), you would be reasonably sure that environmental knowledge caused pro-environmental actions. In this case, internal validity would be high, while external validity would be low. Yet, by combining both studies, it is possible to compensate for the weakness of one research design with the strength of another design.

Replication and meta-analysis

Next to combining research methods, replication of research is another important way to increase confidence in and validity of research outcomes. Replication involves finding consistent support for certain findings by running similar studies in other settings. By doing so, one can rule out the possibility that the findings occurred by chance because the researcher just happened to create a particular set of conditions that yielded the effects. Typically, replication studies imply a test of basic ideas or findings in a different way. This may involve the use of a different sample, or the use

of slightly different operationalizations of key constructs included in the study. For example, the manipulation of the independent variables may be changed, or different items to measure a construct can be included.

Additionally, replication may involve running a similar study in a different lab, in order to rule out the possibility that results are due to specific conditions in a lab that have not been reported by the researcher or that occurred outside the awareness of the researcher. For this reason, it is important that researchers report their research method in sufficient detail, so that others are able to follow the same procedure and use the same (or at least similar) measures. This is also important in identifying possible confounds. For example, it may be that a researcher gave some additional instructions or delivered additional interventions that were deemed to be unrelated to the task, but were crucial to find the effects.

Another important way to study consistency and validity of research findings is by conducting a *meta-analysis* (see also Chapter 20). Meta-analysis refers to using statistical techniques to analyze the results of a group of studies to establish the relative strength of observed effects, the variability in effect sizes (reflecting whether effects are similar or different across studies), and the conditions under which effects were particularly small or large.

An important issue for researchers conducting meta-analysis is that studies yielding non-significant results tend to be underreported in the literature (and may hence not have been included in the meta-analysis). This is referred to as the *file drawer problem* (Rosenthal, 1979): a bias toward publishing positive results (supporting the hypotheses), while negative results (supporting the null hypothesis) are less likely to be published. This may be because researchers themselves decide not to publish their null results, or because journal editors (and reviewers) prefer publishing significant results rather than null results.

Publication bias is a serious problem, as it may result in concluding that some effects are significant and important, while in fact the majority of (unpublished) studies clearly suggest otherwise. Therefore, it is of great importance to also make null results available. In fact, there are a number of websites that collect non-significant findings to make these available in the public domain (see Spellman, 2012).

Researcher degrees of freedom

The overrepresentation of significant effects in the literature may not only be influenced by publication bias and file drawer problems, but can also be caused by the publication of inflated effects. That is, researchers may be motivated by finding significant effects, and adjust their research until effects become significant (e.g., McGuire, 1973). This may result in developing and selecting very specific stimuli, task settings, instructions, dependent and independent variables, treatment levels, mediators, moderators, and favorable boundary conditions that are most likely to yield the expected effects (Fiedler, 2011). By doing so, a researcher capitalizes on chance in model testing. This will likely result in so-called *voodoo correlations*, that is, inflated correlations in which the effects of phenomena are overestimated as they only occur in very specific conditions that have not been specified. This again demonstrates the need for replicating research in different conditions and trying to find convergent evidence for an effect across different methods and situations.

Another possible bias in publication lies in selective or inaccurate reporting of the research methods or results. For example, researchers may decide to only report the results of manipulations that yielded significant results. This is related to the file drawer problem, but in this case it implies that researchers only publish part of their research, and leave out unfavorable parts. Yet, the result of this practice is very similar: inflated effects.

Conclusion and Summary

In environmental psychology, research questions and theories are used to make predictions about human behavior in relation to environmental issues, and various research designs and measurement techniques are used to test these hypotheses. Each of these has advantages and disadvantages. One important difference is that internal and external validity can differ greatly across research designs. Internal validity implies that we understand which variables cause certain effects, while external validity implies that results of a specific study can be generalized to other situations or to the population at large.

Experimental studies have high experimental control and therefore high internal validity, but the findings are more difficult to generalize to other people and other situations. Quasi-experimental studies have less control over the variables, but the findings are typically better generalizable to other people. Correlational studies do not involve a form of manipulation of variables as is the case in lab studies and therefore have low internal validity. But because the variables are assessed as they occur naturally, the results can be more easily generalized to other situations. To measure any changes in the variables of interest, environmental psychologists might use observational data (personally witnessing an event, such as people littering in a park), self-report data (asking questions such as “how often did you recycle in the past month?”), or collect residual data (e.g., counting the number of littered items in a park). Researchers can assess the reliability and validity of each of these types of measures to ensure quality of measurement. Combining research methods (e.g., first a lab study, then a correlational study, or vice versa) and replication of research findings (e.g., testing the same hypothesis among a sample of students and a sample of employees) are important steps to help increase confidence in the research outcomes.

Glossary

Confound An additional variable that affects the dependent variable and varies systematically with the independent variable(s).

Control group A group of participants in a study that are not exposed to the manipulation.

Correlation coefficient A statistic used to represent the association between two variables, ranges from -1 to $+1$.

Correlational design A design that describes the relationship between independent and dependent variables, without systematic variation of these variables or random assignment of participants.

- Correlational research** A study in which the investigator measures naturally occurring variables and the relations between these variables.
- Covariate** A variable that is related to both dependent and independent variables.
- Dependent variable** A variable in an experiment that is measured.
- Experimental design** The systematic variation of one or more independent variables, characterized by random assignment of participants to conditions, or groups.
- External validity** The ability to generalize findings from a study to different situations or populations of individuals.
- Independent variable** A variable in an experiment that is systematically varied by the researcher.
- Interaction** The effect of an independent variable on the dependent variable depends on another independent variable.
- Internal validity** The ability to infer causal relationships between independent and dependent variables.
- Manipulation** Occurs when one or more variables is systematically varied, while holding the other variables constant.
- Operational confound** This occurs when a measure designed to assess a particular construct also measures something else.
- Person confound** An individual difference variable among participants in the study, which may affect the outcome measure.
- Person-by-treatment quasi-experiment** A design in which one independent variable is systematically varied and the other independent variable is measured.
- Procedural confound** This occurs when a researcher inadvertently varies more than one independent variable.
- Quasi-experimental design** A design in which one or more independent variables are systematically varied, but random assignment does not occur.
- Random assignment** Used to ensure that every study participant has the same chance of being assigned to the experimental or control conditions.
- Reverse causality** The observation that when two variables are related, it is not possible to tell whether variable 1 caused variable 2, or the other way around.
- Third-variable problem** When two variables are related, this may be because a third variable that was not included in the study is in fact related to the two variables.
- True experiment** A study characterized by random assignment and systematic variation of variables of interest.

References

- Attari, S. Z., DeKay, M. L., Davidson, C. I., & de Bruin, W. B. (2010). Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Sciences*, 107, 16054–16059.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. (3rd ed.). Thousand Oaks, CA: Sage.
- De Groot, J. I., & Steg, L. (2008). Value orientations to explain beliefs related to environmental significant behavior: How to measure egoistic, altruistic, and biospheric value orientations. *Environment and Behavior*, 40, 330–354.

- Fiedler, K. (2011). Voodoo correlations are everywhere – Not only in neuroscience. *Psychological Science*, 6, 163–171.
- Griskevicius, V., Tybur, J. M., & Van den Bergh, B. (2010). Going green to be seen: Status, reputation, and conspicuous conservation. *Journal of Personality and Social Psychology*, 98, 392–404.
- McGuire, W. J. (1973). The yin and yang of progress in social psychology: Seven koan. *Journal of Personality and Social Psychology*, 26, 446–456.
- Pelham, B. W., & Blanton, T. (2007). *Conducting research in psychology: Measuring the weight of smoke* (3rd ed.). Belmont, CA: Wadsworth/Thomson Learning.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638–641.
- Spellman, B. A. (2012). Data, data, everywhere ... especially in my file drawer. *Perspectives on Psychological Science*, 7, 58–59.
- Verplanken, B., & Holland, R. W. (2002). Motivated decision making: Effects of activation and self-centrality of values on choices and behavior. *Journal of Personality and Social Psychology*, 82, 434.

Suggested Readings

- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage.
- Pelham, B. W., & Blanton, T. (2007). *Conducting research in psychology: Measuring the weight of smoke* (3rd ed.). Belmont, CA: Wadsworth/Thomson Learning.