# 10

# Research Methods in Occupational Health Psychology

Toon W. Taris, Annet H. de Lange, and Michiel A. J. Kompier

CHAPTER OUTLINE

Occupational health psychologists must possess a working knowledge of the most important research methods used in their field. This chapter addresses the basics of the scientific method: that is, what separates scientific from common-sense approaches. We distinguish between qualitative and quantitative methods and discuss when these are best suited for one's goals. We provide typical research designs and address a number of pitfalls that may apply. Finally, we deal with the issues of the reliability and validity of the measures that are used in occupational health psychology.

## Introduction

Suppose that the management of a large information technology company asks you to conduct a study on the health and well-being of their 10,000 employees. These workers hold different jobs, work at various locations, and differ in their background characteristics (age, family status and gender). Where would you start? You might enter the coffee room and chat with several employees over their lunch boxes, but you could also conduct interviews with a specific set of workers (e.g., union members or department leaders), send out questionnaires to all employees (e.g., using a web-based questionnaire), or, alternatively, confine yourself to examining absence figures across various groups of employees. Many such choices have to be made before you can report to the top management on their personnel's well-being, and the quality of these decisions determines the degree to which the interventions that may be conducted on the basis of your research will improve the quality of the employees' work lives.

Clearly this is not an easy task, and students in occupational health psychology (OHP) should be prepared to deal with complex decisions on how to go about things. This implies that they should possess a working knowledge of the basics of scientific research, that is, the systematic, replicable, and valid methods used to expand the knowledge of the phenomena relevant to those working in OHP. Scientific research starts with a *research question* that defines the purpose of the study. Based on this question, the researcher's knowledge and skills and all sorts of practical constraints (e.g., the budget available for the study and the degree to which the organization is willing to cooperate with the researcher), a *research design* is chosen (that is, the approach to collecting the data that will be used to address the research question). Finally, after the data have been collected and analysed, conclusions would be drawn on the phenomenon of interest, leading to interventions to ameliorate possible problems.

In this chapter we address the important steps you must take in designing and conducting research in OHP. As research in this area is usually (albeit not exclusively) conducted within organizations, in this chapter we mainly focus on research that is conducted in organizational contexts. However, many of the issues discussed here can readily be generalized towards other situations. Below we first discuss the nature of the research process, followed by a discussion of several qualitative and quantitative research designs. Finally, we address the measurement of concepts in OHP.

## Research and the Research Process

*What is science?*
*What are the goals of science?*
*What is a research question, and why is it important?*
*What is a scientific theory?*

### The scientific method

The scientific approach refers to the process or method that is used to generate a body of knowledge concerning a phenomenon of interest. In contributing to this knowledge, researchers test hypothetical propositions about the presumed relations among natural phenomena in a systematic, controlled, empirical, and critical way (Kerlinger & Lee, 2000). This approach could lead to the achievement of various goals (Christensen, 1994). First, researchers may want to provide an accurate *description* of the phenomenon of interest. For example, one could examine levels of burnout (a state of excessive work-related fatigue, cynicism towards work, and a lack of work-related efficacy) across different departments of an organization. However, it would be interesting to relate differences in levels of burnout to other phenomena as well, such as differences in work load, personality, or factors such as age and gender. In that case a second goal emerges, namely *explanation*: e.g., does having a high work load contribute to the occurrence of burnout? And are particular persons (e.g., those with high scores on neuroticism) more prone to experiencing high levels of burnout than others? If so, we may be able to anticipate which workers may come to experience burnout; *prediction* is the third goal of science. Without the ability to predict, OHP professionals have little to offer to the organizations asking for their advice. Finally, the fourth goal of science is to *obtain control* of the phenomenon of interest. That is, ideally we are able to manipulate the causes of the occurrence of this phenomenon. For example, in the case of burnout we may want to know how we can prevent burnout from occurring, e.g., by designing jobs in such a way that work stress is minimized or by developing selection methods that optimize the fit between the job and the worker hired to do the job.

Summarizing, we have attained some degree of scientific understanding only if we are able to describe, explain, predict, and control a particular phenomenon reasonably well. One way of summarizing the four goals of science is to say that we aim to develop a *theory* for the phenomenon of interest. Kerlinger and Lee (2000) defined a theory as a set of interrelated constructs (concepts), definitions, and propositions that presents a systematic view of phenomena by specifying relations among variables, with the purpose of explaining, predicting, and controlling the phenomena. Indeed, Kerlinger and Lee feel that the four goals of science discussed above are actually subsumed under a single major objective of science, namely theory development. Without an emphasis on theory development, science would be nothing more than a search for answers, lacking any framework or carefully controlled process. It is theory that allows us to explain, interpret and control the world around us, and it is as important to fundamental as to applied researchers in OHP.

Note that this does not necessarily mean that science must always start from theory. The deductive approach (in which researchers begin their studies with deriving a set of theory-based expectations or hypotheses regarding the phenomenon of interest before they collect their data) is not better or 'more scientific' than the inductive approach (where hypotheses can be construed as generalizations that are based on individual instances and observations). Indeed, the inductive approach has a long history in science. For example, the seventeenth-century philosopher Sir Francis Bacon argued that science should start with empirical observation and induction by enumeration (the generalizing from instances of phenomena to experimentally testable laws). The inductive approach is still popular today, and can fruitfully be integrated with a deductive approach: based on casual observation, researchers may develop a set of ideas (induction) that may subsequently be tried and tested (deduction) using data collected to test and refine these ideas.

## Questions and hypotheses

Every study begins with a research question. The basic aim of such a question is to guide the study: what is the phenomenon that interests us, and what do we want to do with it – describe it, explain it, predict it or gain control over it? For example, a descriptive research question might ask 'how many people in this organization suffer

**Pioneer** Frederick N. Kerlinger (1910–1991)

Frederick N. Kerlinger was born in New York City. After graduating in 1942 from New York University (majors in education and philosophy), he joined the army to become second lieutenant (1946). He returned to the United States in 1950, where Kerlinger earned degrees in educational psychology from the University of Michigan. In 1955 he was offered the chair of the doctoral programme in research and measurement of New York, where he stayed until his retirement in 1974. Subsequently he moved to the Netherlands where he chaired the Methods Department of the University of Amsterdam. After returning to the United States in 1981 he joined the School of Education at the University of Oregon. Kerlinger is best known for his textbook on *Foundations of behavioral research* (1964), in which he provided an extremely influential account of how science 'works' in the behavioural sciences.

from high levels of burnout, as compared to the level of burnout in the general population?' An explanatory research question could be 'does work load affect levels of burnout?' The latter question specifies a particular influence on burnout, suggesting that work load may well be a causal agent that leads to burnout. In order to obtain an answer to this question, researchers should collect data on at least employees' levels of burnout and work load. It is preferable to translate a research question into a *hypothesis* on the outcomes of a study; hypotheses can either be confirmed or be rejected, meaning that it is possible to test one's theory much more precisely than would be possible on the basis of a research question. It also means that researchers are forced to think about the phenomenon they are examining: what, exactly, do they expect to occur? And which mechanisms may account for these outcomes? Thus, drawing on existing theory and research, researchers will articulate their best guess concerning the study's outcome, for example, 'high work load will lead to elevated levels of burnout'. In practice most applied work takes such a deductive, hypothesis-testing approach. As Spector (2008) argues, the research question and the corresponding hypotheses form the basis of many studies in OHP: without a specific and well-formulated question it is difficult to design a study that will provide an adequate answer to it. Conversely, when the research question clearly states the study goal and the phenomenon of interest, it is relatively straightforward to choose an appropriate design for the study addressing it.

## Summary

The scientific approach refers to the process or method that is used to generate a body of knowledge concerning the phenomenon we want to study. It has various goals, including the description, explanation, prediction and control of the phenomenon, and these four goals may be construed as contributing to one overarching goal, namely theory development. This does not mean that science must always start from theory (deduction); indeed, often observation forms the starting point for a scientific study (induction). Every study begins with a research question that aims to guide the study. When this question clearly states the study goal and the phenomenon of interest, it is relatively straightforward to specify study hypotheses, and to choose an appropriate design for the study that will test these.

## Designing a Study

*What is a research design?*
*What decisions are involved in choosing a research design?*
*What is the difference between experimental and non-experimental designs?*
*How can alternative explanations for a particular finding be ruled out?*

A number of decisions must be made before we can actually conduct a study. Basically, these decisions refer to:

- the *setting* in which the study will be conducted: is it a laboratory or a field study?
- the type of participants that will be in the study;
- the constructs of interest: how can we operationalize constructs like burnout or job satisfaction into measurable variables?
- if there are different conditions in the study (e.g., a condition in which the participants are 'manipulated' in some way, versus a control condition in which nothing special happens), the assignment of participants to these conditions; and
- the data collection mode, that is, the way in which information on the variables of interest will be collected.

In conjunction, the answers to these issues determine the *research design,* that is, the basic plan for the study to be conducted.

### Setting and participants

A study can be conducted in the laboratory or the field. In a *field setting* the phenomena of interest are not manipulated by the researcher. For example, organizations are field settings in which employee behaviour can be studied 'as it occurs', without any intervention from the researcher. Conversely, the *laboratory* is an artificial setting in which the phenomena that interest us do not normally occur; these have to be created by the researcher. This distinction does not entirely coincide with that between *experimental* and *non-experimental* research. In experimental research the participants are assigned to one of at least two conditions.

In the experimental (or treatment) condition some aspect of the situation is manipulated by the researcher, whereas in the control condition nothing is changed. The two groups are then compared on the concepts that interest us: if the groups did not differ before the treatment but do thereafter *and* if other explanations for such a difference can be ruled out, the treatment has been shown to causally affect the study variables. This is a common way of conducting research in a laboratory setting, in which other factors that may affect the participants' responses can be controlled. However, experimental designs also occur in OHP. For instance, the effectiveness of intervention programmes is best studied using a *quasi-experimental* design. For example, an organization may institute a new health promotion programme at one plant location but not at another, taking measures of employee health and well-being before and after the institution of the programme at both locations. Superficially this looks like an experiment that is conducted in a field setting, the crucial difference being that participants are not randomly assigned to the experimental (new health programme) and control (business as usual) conditions. This means that it is harder to exclude alternative explanations for differences in the study outcomes between the two groups, as it is quite possible that the experimental and control group differ in more respects than only the health promotion programme they are subjected to.

Pure laboratory studies are rare in OHP, presumably because the work setting is not easily transferred to the laboratory, random assignment of participants to conditions is difficult (see Kompier & Kristensen, 2001), and because the participants in such studies are unlikely to be workers – but it is not impossible. For example, Deelstra and her colleagues (2003) examined the effects of providing instrumental social support on health experimentally in a simulated office environment (i.e., a laboratory that was equipped like any normal office, with computers, telephones, desks, chairs, and so on), among 48 temporary office workers, finding that unwanted social support may have negative effects on worker well-being. Conversely, non-experimental and quasi-experimental designs are very common in OHP. For example, many studies that examine the effectiveness of work stress prevention programmes employ a quasi-experimental design, in that groups of workers participating in the prevention programme are contrasted with other workers who do not. Common non-experimental designs in OHP include observational designs (in which the researcher observes the participants' behaviour and takes notes of what is observed: e.g., how often the participant takes a rest break or communicates with their co-workers) and survey designs (studies in which the participants complete questionnaires). These and other research designs are discussed in more detail below.

## Design: Concepts and variables

A *concept* is a property of people or things that can vary across the people or things that interest us. For instance, size or weight are properties of people. In OHP we are usually interested in concepts such as physical complaints, absence days, or work load. These concepts must be operationalized, that is, the researcher should decide upon how these concepts are measured. After the concepts have been operationalized they are referred to as *variables*.

When examining questions like: 'What is the influence of social support in the development of burnout', we can distinguish between the *independent* variable social support (i.e., the presumed cause) from the *dependent* variable burnout (the effect). Independent variables are the concepts that are either varied by the researcher (in quasi- and experimental designs) or measured (in non-experimental research) as an antecedent to other variables. The dependent variable is the variable of interest, that is, what we design our study to assess. Thus, we are usually interested in the effect of the independent variable on the dependent variable.

Note that the distinction between dependent and independent variables is a conceptual one. Although we are usually interested in drawing causal inferences (the independent variable is assumed to influence the dependent variable causally), the strength of our research design determines the degree to which such inferences are warranted. For example, the assumed dependent and independent variables are often measured at the same point in time. This implies that the temporal order of these concepts cannot be established, meaning that a positive association between work load and burnout can be interpreted theoretically in terms of 'having a high work load leads causally to high levels of burnout', but also as 'workers suffering from burnout tend to experience their work load as heavier'. Whether we can distinguish between these two accounts depends on the strength of our design, i.e., whether alternative explanations for a particular association between a pair of variables can be ruled out.

## Control

It occurs only rarely that a particular result obtained in a study can be explained in just a single way: usually several explanations for this result can be thought of. *Control* refers to the set of procedures that allow researchers to rule out alternative explanations for their findings, other than the hypotheses they intend to test. To ensure that we can draw a causal inference about the effect of our independent variable on the dependent variable in our study, we need to be able to exercise control over the study. For instance, we may want to know whether experiencing high levels of work stress increases the likelihood that workers will be absent from their jobs. We might conduct a survey in which we ask employees (1) whether they experience work stress, and (2) how many days they were absent from their job in the preceding year. We might find that the higher the work stress, the greater the number of days they were absent. However, with this sort of design there may be many uncontrolled variables that could be the real cause of absence. For instance, it would seem possible that the employees with higher absence rates are in different types of jobs than those with lower absence rates. If the high-absence workers are all in high-risk jobs and the low-absence workers are all in white collar jobs, it is difficult to conclude that work stress is the prime cause of absence – rather, it is possible that the high levels of absence in the high-risk group are due to a higher chance of experiencing accidents at work. Thus, type of job is an alternative explanation for the association between work stress and absence rates, meaning that job type should in some way be controlled for in order to rule out the possibility that job level and not job stress is the prime antecedent of absence.

There are several procedures that can be used to control for the effects of variables that do not interest us. For the most part these procedures involve either holding constant or systematically varying the levels of one or more variables. For example, in the stress-absence example above, we might hold type of job constant by collecting data for just a single occupation. Thus, all participants would have the same job, meaning that all participants have the same risk of experiencing accidents. Alternatively, we may include job type as an additional independent variable in our study design. In that way we could examine whether job stress contributed to sickness absence, net of the effects of job type.

In experimental studies, a standard way to control for factors other than the variables of interest is to assign the participants in a non-systematic way ('at random') to the experimental and control conditions. Although all participants will have their own idiosyncrasies and differ from each other in many respects, on average the participants in both groups will be more or less equivalent in their characteristics: if the samples used are large enough, the two groups should be approximately the same in terms of ability, age, gender, motivation, and so forth. Unfortunately, random assignment of participants is impossible in field studies (as here the researcher does not manipulate the independent variable, there are no experimental or control groups) and difficult in quasi-experiments. For example, although we can implement a health promotion programme in one plant location (the experimental group) but not in another (the control group), it is possible that workers of these plants differ in other respects than the location of their work as well, including the products manufactured (which could lead to differences in required skill level, pay and expected continuity of labour), quality of the management, and the neighbourhood in which the plants' workers live. In such cases it is important to consider which additional variables could affect the study's findings, and these should preferably be included in the study design in order to control for their effects. In this vein, the *internal validity* of the study (i.e., the degree to which we can draw causal inferences about our variables) can be increased.

## Summary

In order to conduct a study, a number of decisions must be made that refer to the research design, that is the setting (laboratory or the field), type of participants, variables to be assessed or varied, assignment of participants to the study conditions (i.e., random or non-random) and data collection mode (to be discussed below) that will be used. Basically, a good research design allows researchers to rule out alternative explanations for their findings, other than the hypotheses they intend to test. To ensure that causal inferences about the effect of our independent variable on the dependent variable in a study are warranted, one must be able to control for the effects of variables that may otherwise affect the study outcomes. For the most part, these procedures involve either holding constant or systematically varying the levels of one or more independent variables.

## Data Collection in Occupational Health Psychology

*What are qualitative and quantitative designs?*
*When should we prefer a qualitative design to a quantitative design, and vice versa?*
*What are typical qualitative and quantitative designs?*

After the research question and study hypotheses have been established and the focal variables have been chosen, it is time to focus on the mode of data collection. This section focuses on the most common methods of data collection in OHP. Some considerations relevant to the choice of data collection have already been mentioned in passing, but below we provide a more systematic overview of various modes of data collection.

Organizational phenomena may be studied using a wide variety of designs. It is common to juxtapose qualitative against quantitative designs. Basically, the difference between these two approaches is that quantitative approaches to examining issues in OHP emphasize its quantifiable nature and are concerned with prediction (e.g., of absence figures, levels of burnout, and so forth), classification of workers in high versus low-risk groups, or otherwise measuring distinct elements or dimensions of health and its antecedents in as objective a way as possible. Quantitative approaches tend to emphasize the testing of theory-driven hypotheses, measurement using standardized instruments, and data analysis using advanced statistical techniques. Conversely, qualitative approaches seek to characterize its rich, emergent, constructed and multi-dimensional nature using ethnographic approaches, often requiring 'psychological immersion' in an organization (such as observation of workers and in-depth interviewing) or employing a case-study approach (see Brewerton & Millward, 2001). Qualitative approaches are often used in a hypothesis-generating fashion, emphasising data exploration and the context-boundness of the phenomenon of interest, and results are often reported in narrative form.

These two classes of approaches differ in a number of respects, including philosophical orientation, question development, involvement of the researcher, tools, flexibility, and contextual influences (Bachiochi & Weiner, 2002). Clearly, it is beyond the scope of this chapter to provide an exhaustive discussion of the differences between these two. Indeed, rather than discuss these differences in detail, it is perhaps more useful to indicate when qualitative designs should be preferred to quantitative designs. According to Bachiochi and Weiner, four questions can help researchers decide whether a quantitative or a qualitative approach is best suited for their needs:

- *Is the context central to the research question?* If the specific context of the study is a key driver in the research process and/or in the interpretation of the findings, then a qualitative approach may be appropriate. For instance, we may be interested in examining the antecedents of sickness absence among certified nurses. This may require an in-depth examination of their work situation, hopefully yielding insight into the factors that could lead to elevated levels of absence. Although the findings of such a study may not be generalizable to other professions, it can provide useful information on how to reduce absence rates among nurses.
- *Is the participant's interpretation central to the research question?* Closed-ended questions, e.g., Likert-type statements that require participants to indicate how much they agree or disagree with these, may not provide the participants with sufficient opportunity to express fully why they feel the way they do. Thus, researchers may allow participants to provide answers and interpretations that were initially not anticipated. For instance, participants in a focus group may indicate that the main reasons for a high absence rate are not the high work load, the lack of opportunity for skill use or having to deal with aggressive clients, but rather lack of flexibility in their working times. This factor may not have been expected by the researcher or identified in previous quantitative research where these causes of absence were not included in the set of questions asked.
- *Is depth or richness of the data essential?* The answer to this question may seem obvious ('yes, of course rich data are essential'), but in practice it may be difficult to achieve the desired level of detail. Fortunately, many hypotheses may be tested using relatively restricted data sets. For example, in order to obtain a working understanding of the health risks in the work situation of office technology users, a relatively simple questionnaire tapping standard psychological stressors (such as work load, social support, and job autonomy) and measures of well-being may already provide a good understanding of these associations. Conversely, for research questions requiring greater depth, case studies, interviews or other qualitative approaches may provide the opportunity to collect the necessary detailed information. For example, whereas it is well-known that supervisor support is an important determinant of subordinate burnout, we may be interested in mapping the precise behaviors that contribute to the occurrence of burnout among subordinates. A combination of approaches (e.g., a qualitative pilot study combined with a quantitative survey) may provide the depth needed while still offering the sample size needed for drawing broader conclusions.
- *Is the research exploratory?* The nature of the research question may not readily lend itself to clear operational definitions, especially when a new area of research is explored. In such cases it is often better to start off with a qualitative approach, charting the domain and determining the relevant dimensions of the phenomenon of interest.

Based on such findings a quantitative follow-up may be conducted. Further, the topic may cause some degree of discomfort among the participants (such as bullying or sexual harassment at work). In such cases, a qualitative approach may lead to findings that may otherwise be missed: e.g., in survey research participants may well skip uncomfortable questions or provide a socially desirable answer to these.

## Qualitative approaches

*Observational methods*

An obvious way to collect data is to observe employees in their organizational setting. Observation may occur both in laboratory and field settings when we are interested in some behaviour that can be observed, counted or measured. Observations can be either with (*obtrusive* observation) or without (*unobtrusive* observation) the employees' knowledge. In the first case researchers may observe employees doing their jobs for a particular period of time, with the employees being aware of the fact that the researcher was collecting data on some aspect of their functioning on the job. With unobtrusive methods, the workers under study may be aware that the researcher is present, but they would not know that they were being studied. For example, a consultant may attend a meeting to observe the leadership style of the CEO and his or her interaction with the subordinates, which could be the basis of a measure of the quality of the interaction with the supervisor (a possible determinant of job stress).

One important drawback of the obtrusive observational design is that researchers may affect the phenomenon being studied: the mere fact that employees know that the researcher is monitoring their performance may be sufficient to increase or decrease their motivation to work hard. This is one reason why unobtrusive designs are to be preferred to obtrusive designs, although it should be noted that unobtrusive designs may be problematic because of ethical and legal requirements to respect people's privacy. A further problem with these designs is that they are usually time-consuming, both in terms of collecting and coding the data. Thus, this design is especially useful if data are collected on relatively small samples, examining issues that cannot be addressed using a more efficient design.

*Interviews and focus groups*

Focus groups are a method for qualitative data collection in which pre-selected groups of people (such as employees, clients, managers) have a facilitated discussion with the purpose of answering specific research questions. Interviews are typically conducted for the same purpose, but involve one-on-one sessions between an interviewer and the interviewee (Bachiochi & Weiner, 2002). Focus groups and interviews are often used prior to a quantitative study to identify the topics that should be included in this research. They are generally not used to provide representative data for a large population (quantitative approaches are often much better suited for that purpose). This also means that the conclusions of this type of research cannot be generalized too broadly; the findings are based on a limited and not necessarily representative group of participants.

Researchers intending to conduct an interview or focus group study must deal with a number of issues. First, it is important that *key groups* to be included in the study are selected, and one should decide which and how many members of these groups will be contacted for participation. The number of sessions (or interviews) to be held is determined by the point at which *theoretical saturation* is achieved, that is, when new sessions (interviews) do not lead to new information. Furthermore, it is important that a *script* is available, i.e., a structured guide or protocol that describes the course of the session and that may contain the introductory statements to be used (e.g., the goal of the session, how results will be used, and instructions for participants), the questions to be asked (usually in an open-ended format), and closing information. Of course, there are also practical issues to consider, for instance, how long will the sessions take, and will sessions be recorded or will notes be taken. Finally, researchers should consider the way the data will be analysed. As the data collected using focus groups and interviews will usually be of a textual nature (e.g., transcripts, notes, etc.), this will involve some sort of *content analysis*. Basically, the aim of content analysis is to give insight in the type of answers given by the participants, as well as the frequency of these answers. For example, in a focus group study on the sources of high job stress among firefighters, researchers might hope to obtain an idea of the type of events that elicit high levels of stress in that profession (type of stressors), as well as whether these events affect all participants equally strongly (frequency of mentioning these stressors).

*Case studies*

A case study can be defined as an examination of a single individual, group, company or society (Babbie, 1998). Researchers who develop case studies do not necessarily favour one data collection method over another. For example, data can be collected via observation, structured or semi-structured interviews, or archival research. Typically, the data set comprises a mix of verbal and numeric information, possibly including quantitative approaches (such as a personnel survey) as well. For example, Liu, Spector, and Shi (2008) combined qualitative information (stressful incidents) and quantitative information (an employee survey) in their study of gender differences in experienced job stress among university employees.

As case studies are defined in terms of the object to be studied (that is, the specific group under investigation), the issue of sampling is a central concern in the research design. Case study researchers usually follow a purposeful sampling strategy, in that objects are chosen specifically for the potential they offer for understanding the issue being researched. For example, researchers may not strive towards obtaining a random sample of all employees of a particular organization, but may contact specific employees instead. For instance, in a case study on sexual harassment in an organization, researchers may interview several current and former female workers or the organization (have they had experiences with harassment? how do they experience the organization's policy regarding incidents of harassment?), as well as the head of the HR department (what is the organization's policy in this respect? have any complaints been filed over the past couple of years?). Keyton et al. (2006) examined how Mitsubishi dealt with the claims of sexual harassment made by female employees, examining both the company and union responses to these, finding that organizations that presumptively support values such as honesty, fairness, and respect react and renegotiate these values when organizational incidents reveal value lapses.

Case studies are not typically used to test hypotheses, but they can be useful for description, for generating hypotheses, as well as for illustrating and understanding specific processes and associations (see Yin, 1994). For example, Israel and her colleagues presented four case studies to illustrate their model of occupational stress, safety and health, and discuss the implications of this model for the development of prevention interventions (Israel, Baker, Goldenhar, & Heaney, 1996).

*Action research*

Action research may be considered as a special instance of the case study design. Basically, the idea in this approach is that knowledge creation (by conducting research within a particular organization) is combined with intervention. According to Lewin (1951), the best way to understand a social system is to first introduce change into it and then observe its effects (Locke & Golden-Biddle, 2003). Action researchers thus distinguish themselves from other researchers by their dual purposes of advancing knowledge on the one hand and providing practical advice on

## Box 10.1 Action Research

An interesting combination of quasi-experimental and action research was described by Le Blanc, Hox, Schaufeli, Taris, and Peeters (2007). The aim of their study was to evaluate the impact of a team-based burnout intervention programme, which was conducted among the staff of 29 oncology wards. Before the programme started, immediately after it was completed and 6 months later quantitative surveys were conducted, in which information was collected on the participants' work situation and well-being. Nine wards were assigned to the experimental (intervention) group; the other wards served as controls.

The intervention programme for the experimental wards was designed and implemented using action-theoretical principles. As a first step, at each ward interviews with the key stakeholders (head nurses, physicians, coordinators, and team leaders) were conducted, during which the programme was clarified and information about key characteristics of the context in which the ward operated was collected. A kick-off meeting for the staff of each of the experimental wards served the same purpose. In this way the local context in which the interventions would be conducted could be taken into account. The training programme itself consisted of 6 monthly sessions of 3 hours

### Box 10.1 (*Cont.*)

each. In the first session the findings of the first questionnaire on the participants' work situation was fed back to the participants. The remaining sessions all consisted of an educational and an action part. During the educational part, general issues relating to work stress were discussed, and during the action part, participants formed problem-solving teams that collectively designed, implemented, evaluated and reformulated action plans to cope with the most important stressors in their work situation. In this way each experimental ward created its own unique intervention. Comparison of the post-intervention scores to the pre-intervention scores revealed that levels of burnout had decreased significantly for the experimental wards, whereas such was not the case for the control wards. Thus, it was concluded that the intervention program had been effective in increasing worker well-being.

the other. As a research process, action research is often conceived as an interactive and multi-phased process. The process starts with problem diagnosis and data gathering, after which an intervention is planned and implemented. This is followed by an evaluation of the effects of the intervention (leading to knowledge creation), which may lead to another cycle of intervention, and so forth. In comparison to other modes of research, the roles of the researcher and the participants are quite different in action research. Participants are considered as active shapers of their environment, not just passive recipients of environmental influences; researchers are change agents that participate in the change process as well as creators of knowledge about the change process. Box 10.1 presents an example of a diary study.

## Summary

Many phenomena in OHP may be studied using a variety of designs. A main distinction is that between quantitative versus qualitative approaches. Quantitative approaches tend to emphasize the testing of theory-driven hypotheses, measurement using standardized instruments, and data analysis using advanced statistical techniques. Conversely, qualitative approaches are often used in a hypothesis-generating mode, seeking to characterize the rich, emergent, constructed and multidimensional nature of the phenomenon of interest using an observational or case-study approach. A qualitative approach becomes more appropriate when the context or participant's interpretations are central to the research question, when answering this question requires very detailed data, or when the research has an exploratory, hypothesis-generating focus.

## Quantitative approaches

*What is the difference between quasi- and true experiments?*
*What competing explanations can be controlled by using a pre-test-post-test control group design?*
*What is the difference between cross-sectional and longitudinal designs?*
*Why is it important to have a high response rate?*

Although qualitative approaches may well be useful in OHP, it cannot be denied that most published research in this area has a strong quantitative focus. Below several designs are discussed, including experimental and quasi-experimental designs, and cross-sectional and longitudinal survey studies.

*Quasi- and true experiments*

In a *true experiment* there are two or more independent variables and one or more dependent variables, as well as random assignment of the participants (Stone-Romero, 2002). Typically, at least one of the independent variables is manipulated by the researcher, i.e., this variable has at least two levels (the experimental and the control level or condition). Due to the practical difficulty of assigning participants randomly to conditions, such true designs are relatively rare in OHP. However, *quasi-experimental designs* are much more common. In such designs one or more of the features of a true experiment have been compromised; most often the requirement of random assignment to the conditions has been dropped.

For example, Bond, Flaxman, and Bunce (2008) examined the effects of a control-enhancing intervention programme on mental health and absence rates. Participants were employees of two customer service centres of a large financial services organization. The programme was implemented in one of these centres, and before and after implementation health and absence measures were taken from both, showing that the intervention programme was effective. This type of design is commonly known as a quasi-experimental (or non-equivalent) pre-test-post-test control group design. It is quasi-experimental as participants were not randomly assigned to the conditions, and the groups may not be equivalent in terms of the properties of the members of these groups: as the levels of the main independent variable (i.e., receipt of the intervention programme) are manipulated across conditions, there is an experimental (intervention) versus a control (no intervention) group; as measures were taken before and after the intervention, this is a pre-post-test design.

In spite of the absence of random allocation of participants, the non-equivalent pre-test-post-test control group design is a strong design, in that it allows researchers to rule out a number of important competing explanations for the study findings (see Shadish, Cook, & Campbell, 2002). The most important of these are listed below:

- *History.* It is possible that events or developments, other than the researcher's intervention programme, are responsible for an improvement of health and well-being of the employees of the two customer service centres of Bond et al. (2008); for example, the introduction of new technology may affect participants' working conditions. Insofar as such events apply equally strongly to both centres, their effects will not bias the effects of the intervention. However, if the poor-performing director of the centre where the intervention programme was implemented was replaced during the programme, this would be a plausible rival hypothesis for the idea that the intervention programme led to an improvement on the outcome variables.

- *Maturation.* A change in the outcome variables may be due to development of the participants, and not the treatment: participants may become older and wiser, more experienced, and so forth. In the Bond et al. case, such processes would apply equally well to the participants in both the control and the intervention group, meaning that this alternative explanation of the findings can be ruled out.
- *Testing.* In principle, being subjected to a test (e.g., completing a questionnaire) may affect participants' answers to subsequent measurements. For example, participants may 'learn' from taking a test (e.g., they may become more familiar with the type of problems posed in intelligence tests), or being asked about their absence rates may lead participants to realize that they have been absent excessively often during the preceding year, and that they should attempt to adopt a more healthy way of living (by exercising more often, quit smoking, or get more sleep). Although such a change in lifestyle could result from completing a questionnaire at the pre-test, this would apply equally strongly to both the experimental and the control group, meaning that this rival explanation would not account for the effect of the intervention.
- *Instrumentation.* Comparison of pre-test with post-test scores implies that the measurement instruments are identical, otherwise across-time changes may be due to a change in the instruments that have been used. For example, if at the pre-test measures were taken using paper-and-pencil questionnaires and at the post-test using an internet-based questionnaire, differences between the pre and post-test could result from the difference in the measurement devices used. However, even if different types of measures were used at the pre and the post-test, this would probably affect the intervention and the control group equally strongly, again meaning that the effect of the intervention programme will not be biased.
- *Selection and regression to the mean.* Finally, the treatment difference could be due to a pre-existing difference on some other variable. It is conceivable that the intervention group was selected on the basis of a high need for the intervention (e.g., the participants in this group reported high levels of stress and health complaints). It is likely that these excessively high scores are at least partly due to random influences and unreliability of the measures. It is possible that these factors will diminish or disappear after some time, meaning that the intervention group will more or less automatically report lower levels of stress and health complaints on the post-test measure (a phenomenon known as regression to the mean: extreme scores tend to become more normal across time, see

Campbell & Kenny, 2003). However, due to the presence of a pre-test and a control group it is possible to examine the credibility of this explanation: if at baseline the control and intervention groups obtained equal scores on the pre-test, it is unlikely that selection and regression to the mean account for the effects of Bond et al.'s intervention programme.

### **Pioneer** Donald T. Campbell (1916–1996)

Donald T. Campbell was born in Grass Lake, Michigan, the son of a farmer. After high school Campbell worked on a turkey farm before completing his undergraduate education at Berkeley, graduating first in the class of 1939. He served in the Naval Reserve in World War II, after which he returned to Berkeley to receive his doctorate in social psychology in 1947. Campbell did his main research at Northwestern University in Evanston, Illinois. His long-lasting interest was the study of knowledge: how it is acquired, recognized, evaluated, refined and passed on. His paper with Donald Fiske entitled 'Convergent and discriminant validation by the multitrait-multimethod matrix' is the most widely cited paper in the *Psychological Bulletin* in the past 50 years, whereas his volume with Thomas D. Cook on quasi-experimentation has become the research bible in intervention research. When he took up his last academic post in 1982 at Lehigh University, he was designated 'university professor' with faculty listings in the departments of psychology, education, sociology and anthropology.

*Non-experimental designs*

If you want to know how people feel about something, the first thing that comes to mind is to ask them. This is the basis of one of the simplest research designs, namely the survey design. This design involves the selection of a sample of participants (such as the employees of a particular organization), who are then asked to complete a questionnaire consisting of a series of questions that are designed to tap the variables of interest. Surveys are very often used in OHP, for various reasons. First, if researchers are interested in how workers perceive their work situation or how they feel about their health (i.e., subjective perceptions of matters), there is no better source of information about these issues than the workers themselves. Second, surveys are usually easy to administer, e.g., by handing out paper-and-pencil questionnaires, or through the intra- or internet. Third, they can be administered to large groups of people at the same time. Finally, they can provide respondents with a feeling of anonymity. This could lead to more honest answers and a higher response rate, especially when the questionnaire addresses sensitive issues.

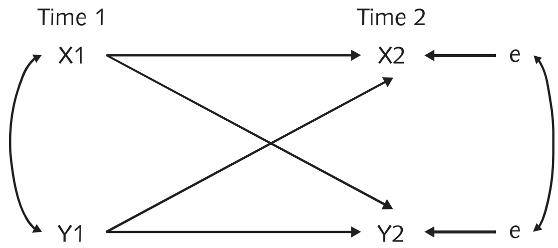
On the negative side, surveys are not well-suited for all concepts that could interest an OHP researcher. For example, whereas it is possible to ask people about the frequency and duration of their spells of sickness absence in the previous year, the answers to such questions tend to be unreliable (e.g., Van der Vaart & Glasner, 2007). Similarly, reports on their own functioning tend to be biased: for instance, Taris (2006) found that worker's self-reported performance (as measured in terms of a global evaluation of one's own functioning at work) was not significantly correlated with objective performance measures (e.g., evaluations of peers, customers, and supervisors). Thus, for some concepts other modes of data collection may be preferable; e.g., information about sickness absence may be obtained from company files.

*Cross-sectional and longitudinal designs*

Most survey studies are cross-sectional, meaning that all data are collected at (and for) a single point in time. The major drawback of such a design is that the temporal order among the variables is not known, meaning that we cannot unambiguously establish whether – say – high work load leads causally to high levels of exhaustion, or that high levels of exhaustion bring about a change in the perception of one's work load. Therefore, cross-sectional designs are of limited use in furthering the goals that are of most interest to OHP researchers, i.e., explaining, predicting and controlling the phenomena of interest. In these respects, *longitudinal designs* are often much more useful. In such designs data are collected for the same set of participants for (but not necessarily at) two or more occasions, allowing for the examination of intra-individual change across time. For example, using a longitudinal design, researchers could relate turnover (i.e., a change of employer during the observed time interval) to specific circumstances at work at the first study wave. In this way one could see whether being bullied at work predicts turnover. Similarly, a high work load at the first study wave could be related to health at a later point in time, controlling for health status at the first wave of the study.

Clearly, longitudinal designs are extremely useful in examining the temporal relationships among variables. However, in order to be able to realize this potential even longitudinal designs must meet a number of requirements. Based on common insights from general and longitudinal research methodology (e.g., Menard, 2008; Taris, 2000), De Lange, Taris, Kompier, Houtman, and Bongers (2003) formulated various design criteria to evaluate the quality of longitudinal research, including the basic design of the study (e.g., the number of measures of the study variables and the length of the interval between the study waves) and the non-response analysis.

*Study design.* Figure 10.1 presents a complete panel design for two variables X and Y. Using this design it is possible to examine the lagged effects of these two variables on each other, distinguishing between the temporal effects of variable X on Y ($X_1 \rightarrow Y_2$), and vice versa ($Y_1 \rightarrow X_2$). The advantage of this design as compared to a cross-sectional design is that they allow for a fuller understanding of




*Note* : X1(X2) = variable X as measured on time 1 (time 2),
Y1(Y2) = measure of variable Y as measured on time 1 (time 2),
e = measurement error

**Figure 10.1** The complete panel design.

the relationships between variable X and Y. For example, it is often assumed that high job demands lead causally to high levels of fatigue, and this reasoning could be tested by relating fatigue (measured at T2) to Time 1 job demands, controlling for Time 1 fatigue (this is referred to as an *incomplete* panel design, as only fatigue has been measured twice). In this way it is possible to see whether Time 1 job demands predict part of the variance of Time 2 fatigue, partialling out the Time 1-Time 2 stability of fatigue. However, whereas it is likely that demands affect fatigue across time, it is also possible that fatigue leads causally to higher job demands. For example, fatigued workers may possess less energy than non-fatigued workers, and this could lead them to evaluate their work load as heavier. Thus, in order to gain a sound understanding of the possibly reciprocal relationships between pairs of variables, a longitudinal design – preferably a complete panel design – is indispensable.

Unfortunately, even a complete panel design may be unable to demonstrate causal effects of variables over time if the interval between the study waves does not correspond with the 'true' causal lag between the variables, that is, the period that it takes for the causal variable to take effect on the outcome variable. If the interval between the study waves is too small, effects will not be significant because the causal process has not yet had enough time to unfold itself. A too-large interval will lead to bias as well, for example, because changes are often repeatable: participants may change their attitudes several times during a particular interval, or changes may only be temporary. Hence, it is imperative that the time lag between the study waves closely approximates the real causal lag. However, as the length of this lag is usually unknown, it may be difficult to satisfy this assumption (Taris, 2000).

*Diary studies.* One special instance of a longitudinal design is the so-called diary study. This design is especially suitable when one wants to learn about the fine-grained temporal relations between the variables of interest. The label 'diary' may be somewhat misleading. Ordinary diaries (the type written by Anne Frank, Samuel Pepys, and Bridget Jones) usually present a somewhat ad-hoc, impressionistic and unstructured record of what happened during the day. However, the aim of a diary *study* is the standardized and structured collection of activities and psychological reactions on a day-to-day basis and at particular well-defined times (Bolger, Davis & Rafaeli, 2003; Semmer, Grebner, & Elfering, 2004; Sonnentag, 2001). The term 'diary' thus implies a day-to-day perspective. Through this method, it is possible to assess mood states and activities of participants shortly after they actually conducted their (work) activities. This minimizes the need for retrospection and therefore the effects of recall bias (Bolger et al., 2003). Diary data may be collected by means of daily surveys, but it is also possible to employ modern means of communication. For example, many current diary studies use palmtop computers or PDAs (Personal Digital Assistants) to collect the data. The participants are expected to carry these with them, completing short diaries at frequent intervals during the day. Box 10.2 presents an example of a diary study.

## Box 10.2 A Diary Study

A recent example of a diary study is a study into daily overtime work by Beckers et al. (2008). Beckers and her co-workers collected data among 120 Dutch faculty members in a nine-day diary study: they collected precise data on the participants' daily (24 hours) overtime, work and non-work activities and work and non-work experiences. The study aim was to find out *when* faculty workers work overtime, *what activities* (e.g., research, teaching, other) were undertaken during overtime, and *how overtime was experienced* (e.g., pleasurable or stressful). Analysis of variance was used to analyse the data. The study's findings showed that overtime was very prevalent among faculty members but unevenly distributed over the week. Working overtime was common on Sunday and Monday and uncommon on Friday and Saturday. The type of overtime activities during the weekend differed from those during the workweek. Specifically, during weekend-overtime work, relatively much time was spent on research as compared to teaching-related activities. Further, overtime activities were experienced differently than activities during regular work hours: overtime work was experienced as less effortful and stressful than regular work hours, and weekend-overtime as less pleasurable than regular hours and evening-overtime.

This detailed day-to-day mapping and evaluation of overtime work thus contributed to a better understanding of overtime work by demonstrating meaningful patterns of overtime over the (work)week, and meaningful associations between overtime activities and time-contingent experiences.

*Response rates.* According to Spector (2008), perhaps the biggest problem in conducting a survey is assuring that a sufficiently high response rate is obtained.

Basically, the issue is how many of those who were intended to be part of the sample (the study population) do actually contribute data to the study (Taris, Scheurs, & Sepmeijer, 2004, for a discussion). It is important that this response rate (which is usually expressed as a percentage) is as high as possible, for a variety of reasons. For example, a high response rate means that many people are represented in the findings. This increases the credibility of the conclusions and recommendations of the study: CEOs cannot easily discard a study's findings if 90 per cent of the workers in a company indicate that they suffer from burnout, something that would be easy were a study to be based on 5 per cent of the workers. There are also more mundane reasons for striving towards a high response rate. For example, statistical tests are *more powerful* if they are applied on a large data set: obviously, in a given organization a high response rate will lead to a larger data set than a low response rate, thus increasing the usefulness of the data collected. *Outliers* (that is, participants with extreme values on the study variables) will have stronger effects on the study findings in small samples than in large samples. Finally, it is possible that those who do not respond to an invitation to participate in a survey differ systematically from those who do, that is, *non-response* may be *selective*. This may well lead to findings that do not generalize beyond the sample. Previous research revealed that survey non-response is associated with a wide array of factors, including high levels of depression, low education, and low decision latitude, low organizational commitment and job satisfaction, low organizational support, and personality variables such as low conscientiousness (Taris & Schreurs, 2007). Thus, non-response might be a problem in many organizational surveys, leading to restriction-of-range effects (that is, selective study attrition increases the homogeneity of the sample, making it more difficult to detect associations between variables) and biased descriptive statistics (non-respondents differ systematically from respondents).

The best solution to minimize the severe consequences of possible selectivity is to maximize response rates. Obviously, if there is no non-response, there will be no selective non-response either. Unfortunately, 100% response rates are extremely difficult to achieve. What, then, is an adequate response rate? Unfortunately, it is difficult to say when a particular response rate is acceptable. The main issue is the degree to which the sample is representative for the study population, and without information on the population it is difficult to say whether the sample is representative or not. Of course, one may compare figures on the composition of the population to that of the sample (e.g., percentage females, distribution of participants across age groups, jobs and departments), but such an approach does not directly focus on the study variables, meaning that bias may be present, even if the sample and study population resemble each other strongly in terms of such background variables. However, researchers have proposed some guidelines as to what response rates are acceptable. For example, based on an analysis of 490 survey studies that were published in 2000 and 2005 in seventeen refereed academic journals, Baruch and Holtom (2008) found that the average response rate for studies that utilized data collected from individuals was 52.7%

with a standard deviation of 20.4. They argued that studies with a response rate that is more than one standard deviation below the mean response rate (i.e., 32.4%) should be critically judged. These figures apply to cross-sectional research: no corresponding research has been conducted for longitudinal research, but it may be expected that the average response rates in such studies will be lower than in cross-sectional studies (Taris, 2000). Clearly, it makes sense to invest quite some effort into optimizing response rates, e.g., by keeping questionnaires as short as possible, or by rewarding participants in some way for their cooperation.

## Summary

Quantitative research designs may be true or quasi-experimental studies or field studies. Participants in true experiments are randomly assigned to the conditions of interest, but this is not always possible in OHP. In quasi-experiments (such as many intervention studies) there are at least two different conditions, but participants are not usually randomly assigned to these. Field studies can be classified as cross-sectional (single-shot) and longitudinal approaches (where measures are taken for at least two occasions). A major issue in field studies is assuring that a sufficiently high response rate is obtained: this will prevent problems with the statistical power of tests (i.e., their ability to detect effects) as well as with the degree to which sample-based findings can be generalized to the population of interest.

## Measurement and Psychometrics

*What is meant by the reliability of a variable?*
*What is meant by the validity of a variable, and how can it be assessed?*

Basically, measurement is the process of assigning numbers to characteristics of people, events or things, using rules in such a way as to represent specified attributes of these people, events or things (Stevens, 1968). For OHP, this often means that we are quantifying individuals along some attribute. An attribute is a dimension along which individuals can be measured and along which they vary, for example, number of health complaints, work load, or number of absence days. In order to do so we need an instrument to measure the location of individuals on this dimension, i.e., we must assign numbers to individuals that indicate how much or how little of an attribute they possess (Levy, 2003). One of the critical steps in planning a study is deciding how the study variables will be measured: ultimately, the quality of our measures will determine how strong our conclusions can be. Two aspects of our measures are of central importance: their reliability and their validity.

## Reliability

The reliability of a measure is the consistency or stability of a measure across repeated observations of a characteristic on the same subject, assuming that the underlying or 'true' score of this subject on this characteristic is stable across time. In order to be reliable, a measure should show little variation from observed score to observed score; if observations differ each time the subject is assessed, the measure is said to be unreliable. Reliability is an important issue, as we cannot adequately describe, explain, predict, or control attitudes, behaviour, performance or health with a variable that is not measured very well. This implies that the reliability of our variables provides an upper bound on the accuracy of our conclusions: our findings cannot be accurate or useful if our variables have been measured badly.

What makes our measures reliable or unreliable? According to *classical measurement theory,* scores on a variable consist of two components: true score and error. The true score represents the characteristic that interests us, the error component is just that – error, comprising random influences on the observed score. Normally an individual's score on a measure will not perfectly represent their true score on the characteristic of interest. For example, people may misunderstand a question or the answering categories provided for a question or assign a different meaning to these: how often should one experience a particular health complaint (say, heart palpitations) in order to check the 'sometimes' alternative for an item tapping that complaint? Note that such misunderstandings could arise even if researchers were to provide objective frequencies as answering categories for such a question, as people would have to recall the average frequency with which they experienced such complaints. Thus, there will virtually always be some potential for errors to occur.

However, as these errors are assumed to be independent of the true score, they are as likely to increase as to decrease the observed scores on our variables. This implies that if multiple observations of a variable are taken, the errors of these will cancel each other out: after averaging these observations, the errors will disappear and the resulting mean will approximate the true score. This reasoning suggests a simple way to increase the reliability of our measures, that is, to take multiple measures of the concepts of interest: when the number of items in a test increases, the reliability of that test will increase correspondingly. This is one of the reasons why the instruments used in OHP often consist of multiple items.

The reliability of a measure can be expressed in various ways. One common way to assess reliability, Cronbach's alpha, is a measure of the *internal consistency* of a measure. This refers to the interrelatedness of the test items. As these items are assumed to tap the same thing, they should be highly related: items that do not correlate highly with the other items of a test may be measuring a different thing and are usually omitted. As a rule of thumb, alpha should be at least .60 when measuring concepts at the group level; many researchers prefer a reliability of

.70 (Nunnally & Bernstein, 1994: note that for individual diagnosis alpha should be much higher, preferably 0.90 or better). Another often-used reliability index is the *test-retest reliability*, referring to the stability of test scores across time, assuming that the underlying true scores did not change. For instance, we may administer a test tapping work load to the members of an organization and then administer the same test to the same group four weeks later. Assuming that the work load did not change across that time interval, the scores of the participants should be about the same at both occasions (as expressed in a high correlation coefficient).

**Pioneer** Jum C. Nunnally, Jr. (1924–1982)

Jum C. Nunnally, Jr. was born in Binghamton, Alabama. He graduated from high school in 1942 after which he flew transport aircraft in World War II. After the war he returned to Binghamton, where he graduated in 1949 from the local college. Successively he obtained a PhD in Psychology from the University of Chicago (1952), graduating on an analysis of the therapy-induced changes in the factor structure of a client's self-description. His career then took him to the University of Illinois (1954) and finally to Vanderbilt University, where he accepted a full professorship in 1960. Nunnally is best known for his influential textbook entitled *Psychometric theory*, in which he discussed important issues in measurement avoiding unnecessarily complex mathematics.

## Validity

It is one thing for a measure to be reliable, but that does not imply that the scores on this variable represent a valid indication of a person's standing on the dimension of interest. The validity of a test is most appropriately considered in terms of its *construct validity*, that is, the degree to which a test measures the construct it was intended to measure. A construct is an abstract quality, meaning that it cannot directly be observed (unlike concepts such as level of education and gender), and that is difficult to measure. Construct validity may be demonstrated by several types of evidence (see Table 10.1). The *face validity* of a measure is high if it appears to assess what it was designed to measure. E.g., job-related fatigue may be tapped with items such as 'I feel exhausted at the end of the work day'. One problem with this type of validity is that it depends on the perspective of the person judging it: whereas I may feel that my measures have high face validity (or else I would be using different measures), others could well disagree. Therefore, this type of evidence is insufficient to demonstrate the validity of our measures.

**Table 10.1** Types of validity

| *Type* | | | *Key question* |
|---|---|---|---|
| | • Face validity | | Does the measure look like what it should assess? |
| | • Content validity | | Does the measure cover the entire underlying concept, as it is theoretically defined? |
| Construct validity | | Predictive validity<br>• convergent<br>• divergent | Does the measure predict the future criterion that it should predict (convergent validity) and not the future criterion it should not predict (divergent validity)? |
| | • Criterion validity | Concurrent validity<br>• divergent<br>• convergent | Does the measure predict the current criterion that it should predict (convergent validity) and not the current criterion it should not predict (divergent validity)? |

*Content validity* refers to the degree to which a test covers a representative sample of the characteristic being assessed. For instance, burnout is theoretically defined as a syndrome of high levels of job-related exhaustion, high levels of cynicism towards one's job and low levels of job-related self-efficacy (Maslach & Jackson, 1981). Thus, a valid measure of burnout should cover these three domains, and a measure that only covers one of these dimensions does a poor job of adequately covering the entire concept. Like face validity, content validity is often assessed by asking experts (such as fellow occupational health psychologists) to evaluate a measure.

The *criterion-related validity* of a measure refers to the degree to which this measure relates to other concepts that are theoretically assumed to be associated with this measure. One type of criterion-related validity is *predictive validity,* referring to the extent to which test scores obtained at one point in time predict criteria observed in the future. For example, theoretically it may be assumed that a measure of burnout predicts future sickness absence and turnover. Another type of criterion-related validity is the *concurrent validity,* referring to how well a test predicts a criterion that is measured at the same time as the test is conducted. For instance, a measure of burnout should correlate highly with work load and stress. Further, *convergent validity* can be demonstrated by showing that one measure of a particular concept correlates highly with other measures of the same concept. For example, a new measure of burnout has high convergent validity if it is strongly associated with existing measures of burnout. In contrast, its associations with theoretically dissimilar measures such as gender or pay level should be low (*divergent validity*).

## Summary

To be useful, the measures used in OHP research should be both reliable and valid. The reliability of a measure refers to the degree that this measure is stable across repeated observations of a characteristic on the same participant, assuming that the true score of this participant on this characteristic does not change. The validity of a measure can be established by demonstrating its face, content and criterion validity. In conjunction these three forms of validity provide evidence for the fourth and most important criterion, namely the construct validity of the measure. To the criteria of reliability and validity we might add a third, namely the *utility* of a measure. This refers to the practical applicability and the balance between the costs and benefits of a measure. Some ways of measuring stress may be reliable and valid, but – in the context of a particular study – quite impractical or overly expensive. For example, measuring cortisol levels may be a good way to obtain an objective impression of worker stress, but workers may be unwilling to participate in such a study whereas the organization they work for may be unwilling to pay for the extra costs. Thus, in practice one must frequently compromise between the validity and reliability of a measure on the one hand, and its utility on the other.

## SUMMARY AND CONCLUSIONS

Let us now return to the example we began with, that is, the occupational health psychologist who was asked by the management of a large IT firm to conduct a study on the health and well-being of their employees. Which issues should be considered before this study could actually be conducted?

- The scientific approach refers to the process or method that is used to generate a body of knowledge concerning a phenomenon deemed to be of interest. In generating (or contributing to) this body of knowledge researchers rely on formal, systematic observation to help them find answers to the questions that interest them.
- Before starting a study, it is vitally important to decide upon the goal of and basic research question for the study: will the study have a hypothesis-testing or exploratory goal? And does the researcher intend to describe, explain, predict or control the phenomenon of interest? In an applied context, it is vitally important that the study goal and research questions match with the interests of the organization, imposing practical constraints on the study's aims.

- Based on the goals and research question(s), researchers should decide upon the research design for the study: its setting (laboratory or the field), the type of participants involved, the variables to be assessed or varied, the way of assigning participants to the study conditions (i.e., random or non-random) and the data collection mode (e.g., a qualitative or a quantitative approach) that will be used.
- A good research design allows researchers to rule out alternative explanations for their findings, other than the hypotheses they intend to test. This means that researchers must be able to control for the effects of variables that are of little substantive interest.
- A main distinction can be made between quantitative versus qualitative approaches. Quantitative approaches emphasize the testing of theory-driven hypotheses and data analysis using advanced statistical techniques. Qualitative approaches are often used in an hypothesis-generating mode. Such an approach becomes more appropriate when the context or participant's interpretations are central to the research question, when answering this question requires very detailed data, or when the research has an exploratory, hypothesis-generating focus.
- Quantitative research designs may be true or quasi-experimental studies or field studies. Participants in true experiments are randomly assigned to the conditions of interest. In quasi-experiments participants are not usually randomly assigned to these. Field studies can be either cross-sectional (data are collected for one occasion) or longitudinal (measures are taken for at least two occasions).
- A major issue in field studies is assuring that a sufficiently high response rate is obtained: this will prevent problems with the statistical power of tests (i.e., their ability to detect effects) as well as with the degree to which sample-based findings can be generalized to the population of interest.
- The measures used should be reliable (stable across repeated observations of an unchanging characteristic on the same subject), valid (measure what it intends to measure), and useful in the context under study.
- The validity of a measure can be established by demonstrating its face, content and criterion validity. In conjunction these three forms of validity provide evidence for the fourth and most important criterion, namely the construct validity of the measure.

## Suggestions for Further Reading

Menard, S. (2008). *Handbook of longitudinal research: Design, measurement, and analysis.* London: Academic Press.

Nesselroade, J. R., & Baltes, P. B. (Eds.) (1979). *Longitudinal research in the study of behavior and development.* New York: Academic Press.

Robson, C. (2002). *Real world research: a resource for social scientists and practitioner-researchers* (2nd ed.). Malden, MA: Blackwell.

Rogelberg, S. G. (Ed.) (2002). *Handbook of research methods in industrial and organizational psychology.* Malden, MA: Blackwell.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Taris, T. W. (2000). *A primer in longitudinal data analysis.* London: Sage.

## References

Babbie, E. (1998). *The practice of social research.* Belmont (CA): Wadsworth.

Bachiochi, P. D., & Weiner, S. P. (2002). Qualitative data collection and analysis. In S. G. Rogelberg (Ed.), *Handbook of organizational research methods* (pp. 161–183). Malden, MA: Blackwell.

Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations, 61,* 1139–1160.

Beckers, D. G. J., Van Hooff, H. L. M., Van der Linden, D., Kompier, M. A. J., Taris, T. W., & Geurts, S. A. E. (2008). A diary study to open up the black box of overtime work among university faculty members. *Scandinavian Journal of Work, Environment and Health, 34,* 213–223.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54,* 579–616.

Bond, F. W., Flaxman, P. E., & Bunce, D. (2008). The influence of psychological flexibility on work redesign: Mediated moderation of a work reorganization intervention. *Journal of Applied Psychology, 93,* 645–654.

Brewerton, P., & Millward, L. (2001). *Organizational research methods.* London: Sage.

Campbell, D. T., & Kenny, D. A. (2003). *A primer on regression artifacts.* New York: The Guilford Press.

Christensen, L. B. (1994). *Experimental methodology.* Boston: Allyn & Bacon.

De Lange, A. H., Taris, T. W., Kompier, M. A. J., Houtman, I. L. D., & Bongers, P. M. (2003). The *very* best of the millennium: Longitudinal research on the Job demands-Control model. *Journal of Occupational Health Psychology, 8,* 282–305.

Deelstra, J. T., Peeters, M. C. W., Schaufeli, W. B., Stroebe, W., Zijlstra, F. R. H., & Van Doornen, L. P. (2003). Receiving instrumental support at work: when help is not welcome. *Journal of Applied Psychology, 88,* 324–331.

Israel, B. A., Baker, E. A., Goldenhar, L. M., & Heaney, C. A. (1996). Occupational stress, safety, and health: Conceptual framework and principles for effective prevention interventions. *Journal of Occupational Health Psychology, 1,* 261–286.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Orlando: Harcourt College Publishers.

Keyton, J., Cano, P., Clounch, T. L., Fischer, C. E., Howard, C., et al. (2006). Ethical storm or model workplace? In S. May (Ed.), *Case studies in organizational communication: Ethical perspectives and practices* (pp. 153–168). Thousand Oaks: Sage Publications.

Kompier, M. A. J., & Kristensen, T. S. (2001). Organizational work stress interventions in a theoretical, methodological and practical context. In J. Dunham (Ed.), *Stress in the workplace: Past, present and future* (pp. 164–190). London: Whurr Publishers.

Le Blanc, P. M., Hox, J. J., Schaufeli, W. B., Taris, T. W., & Peeters, M. C. W. (2007). Take care! Evaluation of a team-based burnout intervention program for oncology care providers. *Journal of Applied Psychology, 92,* 213–227.

Lewin, K. (1951). *Field theory in social science.* New York: Harper & Row.

Liu, C., Spector, P. E., & Lin, S. (2008). Use of both qualitative and quantitative approaches to study job stress in different gender and occupational groups. *Journal of Occupational Health Psychology, 13,* 357–370.

Locke, K., & Golden-Biddle, K. (2002). An introduction to qualitative research: Its potential for industrial and organizational psychology. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 99–118). Malden, MA: Blackwell.

Maslach, C., & Jackson, S. (1982). *The Maslach Burnout Inventory.* Palo Alto: Consulting Psychologists Press.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Semmer, N. K., Grebner, S. & Elfering, A. (2004). Beyond self-report: using observational, physiological, and situation-based measures in research on occupational stress. In P. L. Perrewé and D. C. Ganster (Eds.), *Research in occupational stress and well-being: Emotional and physiological processes and positive intervention* strategies (pp. 205–263). Amsterdam: Elsevier.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Sonnentag, S. (2001). Work, recovery activities, and individual well-being: A diary study. *Journal of Occupational Health Psychology, 6,* 196–210.

Spector, P. E. (2008). *Industrial and organizational behaviour* (5th ed.). Malden, MA: Wiley.

Stevens, S. (1968). Measurement, statistics, and the schemapiric view. *Science, 161,* 849–856.

Stone-Romero, E. F. (2002). The relative validity and usefulness of various empirical research designs. In S.G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 77–98). Malden, MA: Blackwell.

Taris, T. W. (2000). *A primer in longitudinal data analysis.* London: Sage.

Taris, T. W. (2006). Is there a relationship between burnout and objective performance? A critical review of 16 studies. *Work & Stress, 20,* 316–334.

Taris, T. W., & Schreurs, P. J. G. (2007). How may nonresponse affect findings in organizational studies? The tendency-to-the-positive effect. *International Journal of Stress Management, 14,* 249–259.

Taris, T. W., Schreurs, P. J. G., & Sepmeijer, K. J. (2004). Web-based data collection in occupational health psychology. In J. Houdmont and S. McIntyre (Eds.), *Occupational health psychology: Key papers of the European Academy of Occupational Health Psychology* (pp. 398–406). Maia (POR): ISMAI.

Van der Vaart, W., & Glasner, T. (2007). Applying a timeline as a recall aid in a telephone study: A record check study. *Applied Cognitive Psychology, 21,* 227–238.

Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks: Sage.