Getting Data

Electronic supplementary material
The online version of this chapter (https://doi.org/10.1007/978-3-662-56707-4_4) contains additional material that is available to authorized users. You can also download the "Springer Nature More Media App" from the iOS or Android App Store to stream the videos and scan the image containing the "Play button".

M. Sarstedt, E. Mooi, *A Concise Guide to Market Research*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-662-56707-4_4

Learning Objectives
After reading this chapter, you should understand:
- How to find secondary data and decide on their suitability
- How to collect primary data
- How to design a basic questionnaire
- How to design basic experiments
- How to design basic qualitative research

Keywords
Back-translation • Balanced scale • Big data • Closed-ended questions • Constant sum scale • Customer relationship management • Double-barreled questions • Equidistant scale • Ethnography • Experiments • Experimental design • External secondary data • External validity • Face-to-face interviews • Focus groups • Forced-choice scale • Free-choice scale • In-depth interviews • Internal secondary data • Internal validity • Laddering • Likert scale • Mail surveys • Manipulation checks • Means-end approach • Mixed mode • Mystery shopping • Observational studies • Open-ended questions • Personal interviews • Primary data • Projective techniques • Qualitative research • Rank order scales • Reverse-scaled items • Secondary data • Semantic differential scales • Sentence completion • Social desirability bias • Social media analytics • Social networking data • Surveys • Syndicated data • Telephone interviews • Test markets • Treatments • Unbalanced scale • Verbatim items • Visual analogue scale • Web surveys

4.1 Introduction

In the previous chapter, we discussed some of the key theoretical concepts and choices associated with collecting data. These concepts and choices include validity, reliability, sampling, and sample sizes. We also discussed different types of data. Building on ► Chap. 3, this chapter discusses the practicalities of collecting data. First, we discuss how to work with secondary data. Before collecting primary data, market researchers should always consider secondary data, which are often available and do not depend on respondents' willingness to participate. Although secondary data have already been collected, you usually need to spend considerable effort preparing them for analysis, which we discuss first. If you find that the required secondary data are unavailable, outdated, or very costly, you may have to collect primary data. In the sections that follow, we discuss how to collect primary data through observations, surveys, and experiments. In ◘ Fig. 4.1, we provide an overview of some types of secondary and primary data.

4.2 Secondary Data

Secondary data are data that have already been gathered, often for a different research purpose and some time ago. Secondary data comprise internal secondary data, external secondary data, or a mix of both.

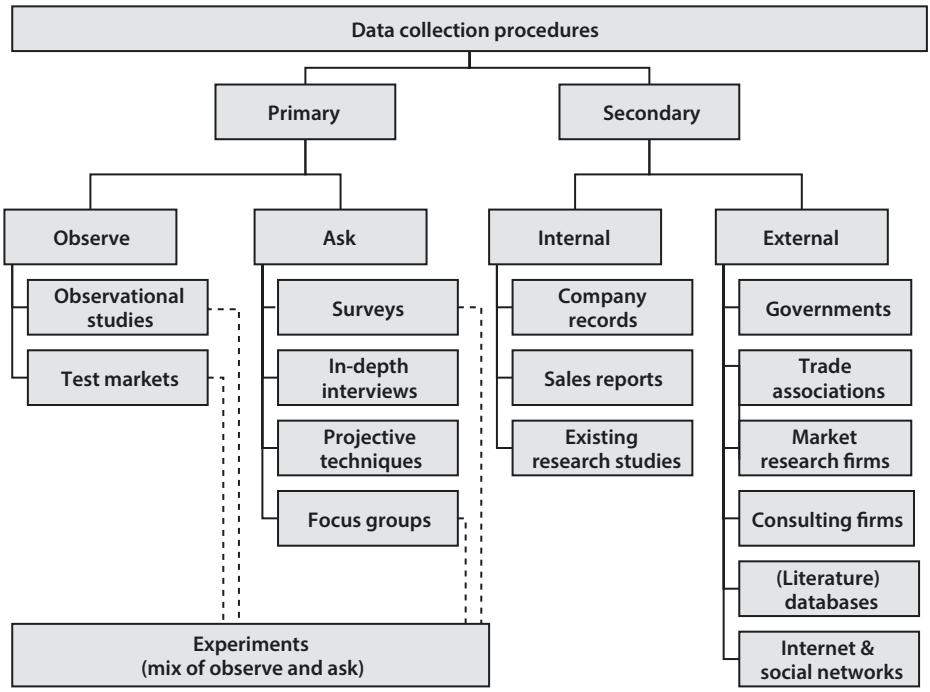


Fig. 4.1 Types of primary and secondary data sources

4.2.1 Internal Secondary Data

Internal secondary data are data that companies compiled for various reporting and analysis purposes. Much of these data have been collected and stored because "you can't manage what you don't measure."[1] Large companies have systems in place, such as *Enterprise Resource Planning systems* (usually abbreviated as ERP systems), in which vast amounts of customer, transaction, and performance data are stored. In general, internal secondary data comprise the following:

- Company records,
- sales reports, and
- existing research studies.

Company records are a firm's repository of information. They may contain data from different business functions, such as finance, or **Customer Relationship Management** (**CRM**). The finance function may provide internal reviews of an organization's financial well-being and strategic advice, as it has access to the organization's financial and operational

1 This quote has been attributed to Peter F. Drucker.

data. The term CRM refers to a system of databases and analysis software that tracks and predicts customer behavior. Firms such as IBM, Microsoft, and Oracle market the database systems that the analysis software utilizes. These database management systems often include information on, for example, purchasing behavior, (geo-)demographic customer data, and the after-sales service. This information is compiled to allow marketers to track individual customers over different sales channels and types of products in order to tailor their offerings to these customers' needs. Several information technology companies, such as SAP, Oracle, and Salesforce.com sell the analysis software that utilizes these databases. Companies use this software to, for example, identify customer trends, calculate their profitability per customer, or identify opportunities to sell new or different products. The CRM market is substantial, generating about $37 billion in 2017.

Sales reports are created when products and services are sold to business-to-business clients. Many of these reports detail discussions held with clients, as well as the products and services sold. The reports therefore provide insights into customers' needs. Sales reports are also a means of retaining customers' suggestions regarding products and services, and can be a productive source of information. For example, DeMonaco et al. (2005) found that 59 % of existing drugs had uses other than those that the producing company described. Because it is important to be aware of a drug's uses, sales discussions with doctors, hospitals, and research institutes can help this company market these drugs. When sales reports are available, they are often part of a CRM system.

Existing research studies are a good source of secondary data. You should, however, carefully consider whether existing research studies are still useful and what you can learn from them. Even if you believe their findings are outdated, their measures may be very useful. Consequently, if you wish to use existing research studies, it is important that you ascertain that enough of their details are available to make them useful.

4.2.2 External Secondary Data

External secondary data have been compiled outside a company for a variety of purposes. Important sources of secondary data, which we discuss next, include:

- Governments,
- trade associations,
- market research firms,
- consulting firms,
- (literature) databases, and
- Internet & social networks.

Governments often provide data that can be used for market research purposes. For example, The CIA World Fact Book provides information on the economy, politics, and other issues of nearly every country in the world. Eurostat (the statistics office of the European Union) provides detailed information on the economy and different market sectors of the European Union. Much of this information is free of charge and is an easy starting point for market research studies.

Trade associations are organizations representing different companies whose purpose is to promote their common interests. For example, the Auto Alliance—which consists

of US automakers—provides information on the sector and lists the key issues it faces. The European Federation of Pharmaceutical Industries and Associations represents 1900 pharmaceutical companies operating in Europe. The federation provides a detailed list of key figures and facts, and regularly offers statistics on the industry. Most of the other trade associations also provide lists of their members' names and addresses. These can be used, for example, as a sampling frame (see ► Chap. 3). Most trade associations regard ascertaining their members' opinions a key task and therefore collect data regularly. These data are often included in reports that researchers can download from the organization's website. Such reports can be a short-cut to identifying key issues and challenges in specific industries. Sometimes, these reports are free of charge, but non-members usually need to pay a (mostly substantial) fee.

Market research firms are another source of secondary data. Especially large market research firms provide syndicated data that different clients can use (see ► Chap. 1). **Syndicated data** are standardized, processed information made available to multiple (potential) clients, usually for a substantial fee. Syndicated data often allow the client's key measures (such as satisfaction or market share) to be compared against the rest of the market. Examples of syndicated data include the J.D. Power Initial Quality Study, which provides insights into the initial quality of cars in the US, and the J.D. Power Vehicle Ownership Satisfaction Study, which contains similar data on other markets, such as New Zealand and Germany. GfK's Spex, which we introduce in Box 4.1, is another important example of syndicated data.

Consulting firms are a rich source of secondary data. Most firms publish full reports or summaries of reports on their website. For example, McKinsey & Company publish the *McKinsey Quarterly*, a business journal that includes articles on current business trends, issues, problems, and solutions. Oliver Wyman publishes regular reports on trends and issues across many different industries. Other consulting firms, such as Gartner and Forrester, provide data on various topics. These data can be purchased and used for secondary analysis. For example, Forrester maintains databases on market segmentation, the allocation of budgets across firms, and the degree to which consumers adopt various innovations. Consulting firms provide general advice, information, and knowledge, while market research firms only focus on marketing-related applications. In practice, there is some overlap in the activities that consulting and market research firms undertake.

(Literature) databases comprise professional and academic journals, newspapers, and books. Two important literature databases are ProQuest (http://www.proquest.com)

Box 4.1 GfK Spex Retail

GfK is a large market research company. One of its databases, Spex Retail, provides resellers, distributors, manufacturers, and website portals with product data. In 20 languages, it offers aggregated data on more than 7 million products from 20,000 manufacturers in 30 countries. This database provides details on IT, consumer electronics, household appliances, and other products. Spex Retail also provides insight into new products being launched by providing 70,000 new information sheets that describe new products or product changes every month. The data can also be used to map product categories. Such information helps its clients understand the market structure, or identify cross-selling or up-selling possibilities. See http://www.etilize.com/retailers/#product-content for more details, including a demo of its products.

and JSTOR (https://www.jstor.org). ProQuest contains over 9000 trade journals, business publications, and leading academic journals, including highly regarded publications such as the *Journal of Marketing* and the *Journal of Marketing Research*. A subscription is needed to gain access, although some papers are published as open access. Academic institutions often allow their students and, sometimes, their alumni to access these journals. JSTOR is like ProQuest, but is mostly aimed at academics. Consequently, it provides access to nearly all leading academic journals. A helpful feature of JSTOR is that the first page of academic articles (which contains the abstract) can be read free of charge. In addition, JSTOR's information is searchable via Google Scholar (discussed in Box 4.2). Certain database firms provide firm-level data, such as names and addresses. For example, Bureau van Dijk (https://www.bvdinfo.com/en-gb), as well as Dun and Bradstreet (http://www.dnb.com), publish extensive lists of firm names, the industry in which they operate, their profitability, key activities, and address information. This information is often used as a sampling frame for surveys.

Internet data is a catch-all term that refers to data stored to track peoples' behavior on the Internet. Such data consist of *page requests* and *sessions*. A page request refers to people clicking on a link or entering a specific Internet address. A session is a series of these requests and is often identified by the IP number, a specific address that uniquely identifies the receiver for a period of time, or by means of a *tracking cookie*. With this information, researchers can calculate when and why people move from one page to another. The *conversion rate* is a specific type of information, namely the ratio of the number of purchases made on a website relative to the number of unique visitors, which often interests researchers. Facebook, Instagram, and LinkedIn, provide valuable information in the form of social networking profiles, which include personal details and information. These **social networking data** reflect how people would like others to perceive them and, thus, indicate consumers' intentions. Product or company-related user groups are of specific interest to market researchers. Take, for example, comments posted on a Facebook group site such as that of BMW or Heineken. An analysis of the postings helps provide an understanding of how people perceive these brands. Interpretations of such postings usually include analyzing five elements: the agent (who is posting?), the act (what happened, i.e., what aspect does the posting refer to?), the agency (what media is used to perform the action?), the scene (what is the background situation?), and the purpose (why do the agents act?). By analyzing this qualitative information, market researchers can gain insight into consumers' motives and actions. Casteleyn et al. (2009), for example, show that the Heineken Facebook posts reveal that the brand has a negative image in Belgium. The task of collecting, processing, analyzing, and storing social networking data is very challenging, due to the data's complexity and richness. To enable these tasks, researchers have combined theories and methods from a variety of disciplines (e.g., computer science, linguistics, statistics) in the emerging research field of **social media analytics** to develop new approaches to and method for analyzing social networking data. These include (1) *text mining* to derive high-quality information from text, (2) *social network analysis* to study the structure of the relationships between persons, organizations, or institutions in social networks, and (3) *trend analysis* to predict emerging topics in, for example, Twitter tweets or Facebook posts (Stieglitz et al. 2014). Several companies, such as Talkwalker (www.talkwalker.com), aggregate data from different websites including blogs and social media platforms, such as Twitter, Facebook, and Instagram, which they then analyze. These sites also provide statistics, such as the number of mentions or complaints, thereby providing insight into

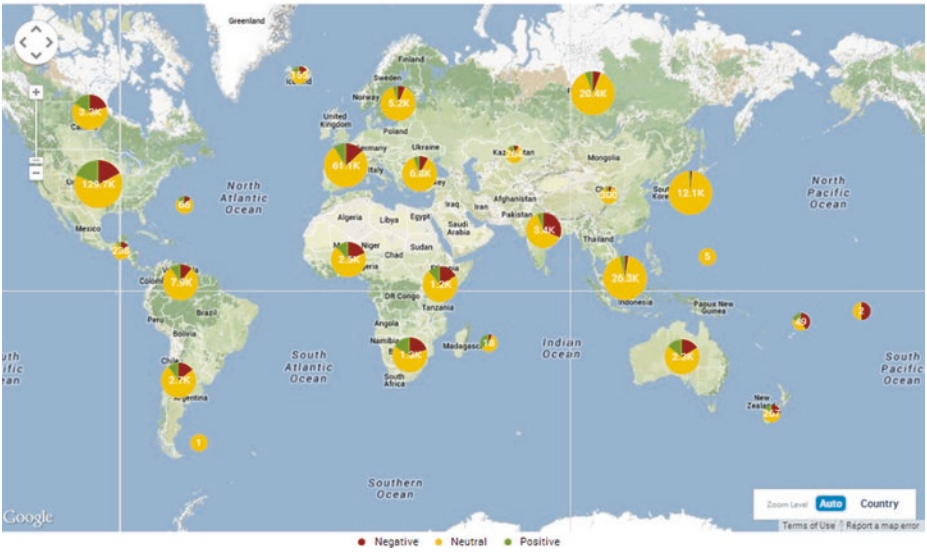


Fig. 4.2 Snapshot of sentiment analysis via www.talkwalker.com

people, brands, and products rated on various dimensions. Talkwalker, for example, uses AI to power it's sentiment detection and conducted a sentiment analysis of a major product recall by Toyota. Toyota found that its social media mentions increased sharply and were far more negative in the US and Europe than in Indonesia and Japan (see https://www.talkwalker.com/blog/navigate-the-automotive-recall-storm-with-social-media-monitoring). Fig. 4.2 shows a snapshop of a sentiment analysis via Talkwater.

Social websites also provide quantitative information. For example, Facebook's Ad Manager provides information on the effectiveness of advertising on Facebook, including on measures, such as the click-through-rate, and on demographics, such as gender or location.

Big data is an important term in the context of Internet and social networking data. The term big data describes very large datasets, generally a mix of quantitative and qualitative data in very large volumes, which are automatically analyzed, often with the aim of making predictions. There is no commonly accepted definition of the term, but the use of big data has become very important very quickly. Big data's use is not unique to market research, but spans boundaries and often includes IT, operations, and other parts of organizations. Netflix, a provider of videos and movies, relies on big data. Netflix faces the challenge that it pays upfront for the videos and the movies it purchases, and therefore needs to understand which, and how many, of its customers will watch them. Netflix analyzes 2 billion hours of video each month in an endeavor to understand its customers' viewing behavior and to determine which videos and movies will become hits. Walmart, the largest retailer in the world, also uses big data. One of Walmart's challenges is to proactively suggest products and services to its customers. Using big data, Walmart connects information from many sources, including their location, and uses a product database to find related products. This helps Walmart make online recommendations.

4.3 Conducting Secondary Data Research

In ▶ Chap. 2, we discussed the market research process, starting with identifying and formulating the research question, followed by determining the research design. Once these two have been done, your attention should turn to designing the sample and the method of data collection. In respect of secondary data, this task involves the steps shown in ◘ Fig. 4.3.

4.3.1 Assess Availability of Secondary Data

Search engines (such as Google or Bing) provide easy access to many sources of the secondary data we have just discussed. Furthermore, many (specialist) databases also provide access to secondary data.

Search engines crawl through the Internet, regularly updating their contents. Algorithms, which include how websites are linked, and other peoples' searches evaluate this content and present a set of results. Undertaking searches by means of search engines requires careful thought. For example, the word order is important (put keywords first) and operators (such as +, −, and ~) may have to be added to restrict searches. In Box 4.2, we discuss the basics of using Google to search the Internet.

Box 4.2 Using Google to searching for secondary data

Most people use Google daily, so why not use it for market research purposes too? Google has an easy interface, but if you use the standard search box, you may not find the data you are looking for. What must you then do to find useful data?

- You could use Google Scholar (https://scholar.google.com) if you are looking for scholarly information such as that found in academic journals. While you can search for any information, specific search items can usually only be accessed if you have an organizational, university, or library password.
- By using Google Books (https://books.google.com/books), you can enter several keywords to easily search through a very large catalogue of books. Google clearly indicates the books in which the search results are found and the pages on which the keywords occur. The Ngram Viewer (https://books.google.com/ngrams/) shows the relative frequency with which words are used and is a cool Google books tool.
- If you cannot find what you are looking for, Google allows you to refine your search. After you have searched you can typically select *All, Images, Videos, News, Maps*, and *More* to select the type of result you are interested in. Under *Tools*, you can select the country and time range from which results are to be shown.
- Google Public Data Explorer (https://www.google.com/publicdata/directory) facilitates exploration of a variety of public-interest datasets. These include the US Census Bureau, Eurostat, and the OECD datasets. This site is particularly useful if you want to obtain visualized data on economic indicators.
- Try using operators, which are signs that you use to restrict your research. For example, putting a minus symbol (–) (without a space) before a search word *excludes* this word from your findings. Putting a sequence of words, or an entire sentence in quotation marks (e.g., "a concise guide to market research") indicates that Google should only search for exact matches.

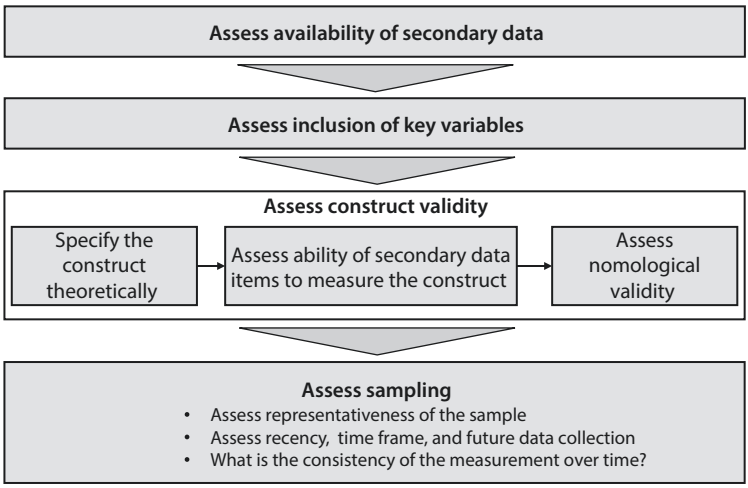


Fig. 4.3 Assessing secondary data

Databases contain existing data that can be used for market research purposes, which we discussed in the ► Sect. 4.2.2. Lightspeed Research (http://www.lightspeedresearch.com), for example, maintains several databases, such as a B2B and Millenial Panel, providing details on selected market segments. GfK provides several databases that track retail sales. Nielsen maintains a large consumer panel of some 250,000 households in 27 countries. It is clearly not possible to provide an exhaustive list of the databases available, but an online search, a market research agency, or an expert should help you identify the options quickly.

Once a potential secondary data source has been located, the next task is to evaluate the available data. It is important to critically assess the (potential) data's fit with your needs. Fig. 4.3 provides a set of criteria to help evaluate this fit.

4.3.2 Assess Inclusion of Key Variables

Measurement is the first element to assess. It consists of a set of criteria. You should first check whether the desired variables are included in the source. The key variables in which you are interested, or could use, should obviously be part of the data. Also check if these variables are included at the required level of analysis, which is called the aggregation level (see ► Chap. 3). For example, the *American Customer Satisfaction Index* (*ACSI*) satisfaction dataset reports on satisfaction at the company level;[2] therefore, if researchers need the measurement of satisfaction at a product, service, or store level, these data are inappropriate.

2 See https://www.theacsi.org for a detailed description of how ACSI data are collected.

4.3.3 Assess Construct Validity

After checking that the desired variables are included in the source, the construct validity should be assessed (see ► Chap. 3 for a discussion of validity). Validity relates to whether variables measure what they should measure (see ► Chap. 3). *Construct validity* is a general term relating to how a variable is defined conceptually and its suggested (empirical) measure. Houston (2002) establishes a three-step method to assess the construct validity of secondary data measures.

- First, specify the theoretical definition of the construct in which you are interested. Satisfaction is, for example, often defined as the degree to which an experience conforms to expectations and the ideal.
- Second, compare your intended measure against this theoretical definition (or another acceptable definition of satisfaction). Conceptually, the items should fit closely.
- Third, assess if these items have *nomological validity* (see ► Chap. 3). For example, customer expectations, perceived quality, and value have a significant influence on customer satisfaction. Similarly, satisfaction generally relates positively to customer loyalty. As such, you would expect the measure of satisfaction that you are evaluating to correlate with these measures (if included in the database).

Taking these three steps is important, as the construct validity is often poor when secondary data are used.[3] See Raithel et al. (2012) for an application of this three-step process.

If there are multiple sources of secondary data, identical measures can be correlated to assess the construct validity. For example, the Thomson Reuter SDC Platinum and the Bioscan database of the American Health Consultants both include key descriptors of firms and financial information; they therefore have a considerable overlap regarding the measures included. This may raise questions regarding which databases are the most suitable, particularly if the measures do not correlate highly. Fortunately, databases have been compared; for example, Schilling (2009) compares several databases, including SDC Platinum.

4.3.4 Assess Sampling

Next, the sampling process of the collected secondary data should be assessed. First assess the population and the representativeness of the sample drawn from it. For example, the sampling process of Nielsen Homescan's data collection effort is based on probability sampling, which can lead to representative samples (see ► Chap. 3). Sellers of secondary data often emphasize the size of the data collected, but good sampling is more important than sample size! When sampling issues arise, these can be difficult to detect in secondary data, because the documents explaining the methodology behind secondary data (bases) rarely discuss the data collection's weaknesses. For example, in many commercial mailing lists 25% (or more) of the firms included routinely have outdated contact information, are

3 Issues related to construct validity in business marketing are discussed by, for example, Rindfleisch and Heide (1997). A more general discussion follows in Houston (2002).

bankrupt, or otherwise not accurately recorded. This means that the number of contactable firms is much lower than the number of firms listed in the database. In addition, many firms may not be listed in the database. Whether these issues are problematic depends on the research purpose. For example, descriptive statistics may be inaccurate if data are missing.

The recency of the data, the time period over which the data were collected, and future intentions to collect data should be assessed next. The data should be recent enough to allow decisions to be based on them. The data collection's timespan and the intervals at which the data were collected (in years, months, weeks, or in even more detail) should match the research question. For example, when introducing new products, market competitors' market share is an important variable, therefore such data must be recent. Also consider whether the data will be updated in the future and the frequency with which such updates will done. Spending considerable time on getting to know data that will not be refreshed can be frustrating! Other measurement issues include definitions that change over time. For example, many firms changed their definitions of loyalty from behavioral (actual purchases) to attitudinal (commitment or intentions). Such changes make comparisons over time difficult, as measures can only be compared if the definitions are consistent.

4.4 Conducting Primary Data Research

Primary data are gathered for a specific research project or task. There are two ways of gathering primary data. You can observe consumers' behavior, for example, by means of observational studies, or test markets. Alternatively, you can ask consumers directly by means of surveys, in-depth interviews, predictive techniques, or focus groups. Experiments are a special type of research, which is normally a combination of observing and asking. We provide an overview of the various types of primary data collection methods in Fig. 4.1.

Next, we briefly introduce observing as a method of collecting primary data. We proceed by discussing how to conduct surveys. Since surveys are the main means of collecting primary data by asking, we discuss the process of undertaking survey research in enough detail to allow you to set up your own survey-based research project. We then discuss in-depth interviews, including a special type of test used in these interviews (projective techniques). Last, we discuss combinations of observing and asking—the basics of conducting experimental research.

4.4.1 Collecting Primary Data Through Observations

Observational studies can provide important insights that other market research techniques do not. Observational techniques shed light on consumers' and employees' behavior and can help answer questions such as: How do consumers walk through supermarkets?; how do they consume and dispose of products?; and how do employees spend their working day? Observational techniques are normally used to understand what people are doing rather than *why* they are doing it. They work well when people find it difficult to put what they are doing into words, such as shoppers from different ethnic backgrounds.

Most observational studies use video recording equipment, or trained researchers, who unobtrusively observe what people do (e.g., through one-way mirrors or by using

recording equipment). Another example is the use of eyetracking equipment that measure what catches customers' eyes and how long their gaze remains (see Box 4.3 for a video on how market researchers use eyetracking for studies in supermarkets).

Recently, researchers started using computer chips (called RFIDs) as observational equipment to trace consumers' shopping paths within a supermarket ► Chap. 2. Almax, an Italian company, has developed a special type of observational equipment. Their EyeSee product is an in-store mannequin equipped with a camera and audio recording equipment. The product also comprises software that analyze the camera recordings and provides statistical and contextual information, such as the shoppers' demographics. Such information is useful for developing targeted marketing strategies. For example, a retail company found that Chinese visitors prefer to shop in Spanish stores after 4 p.m., prompting these stores to increase their Chinese-speaking staff at these hours. ◘ Fig. 4.4 shows what the EyeSee Mannequin looks like.

Mystery shopping, when a trained researcher is asked to visit a store or restaurant and consume the products or services, is a specific type of observational study. For example, McDonalds and Selfridges, a UK retail chain, both use mystery shoppers to ensure the quality of their services and products (see Box 4.4 for an MSNBC video on mystery shopping).

Sometimes observational studies are conducted in households, with researchers participating in them to see how the inhabitants buy, consume, and dispose of products or services. The type of study in which the researcher is a participant is called an **ethnography**. An example of an ethnography is Volkswagen's Moonraker project, in which several Volkswagen employees followed American drivers to gain an understanding of how their usage of and preferences for automobiles differ from those of European drivers (Kurylko 2005).

Box 4.3 Eyetracking study in a supermarket

© gopixa/Getty Images/iStock
https://www.youtube.com/watch?v=Hatmm84sqm0&index=3&list=PLMM7ZjFcpspWOBnq-rdg7P1fBXcJY7tWL

Fig. 4.4 The EyeSee mannequin. (© almax) https://www.youtube.com/watch?v=NA-fFv1uB-c

Test markets are a useful, but costly, type of market research in which a company introduces a new product or service to a specific geographic market. Test markets are sometimes also used to understand how consumers react to different marketing mix instruments, such as changes in pricing, distribution, or advertising and communication. Test marketing is therefore about changing a product or service offering in a real market and gauging consumers' reactions. While the results from such studies provide important insights into consumer behavior in a real-world setting, they are expensive and difficult to conduct. Some frequently used test markets include Hassloch in Germany, as well as Indianapolis and Nashville in the US.

Box 4.4 Using mystery shopping to improve customer service

© andresr/Getty Images/iStock
https://www.youtube.com/watch?v=ErfUVt0fyK0

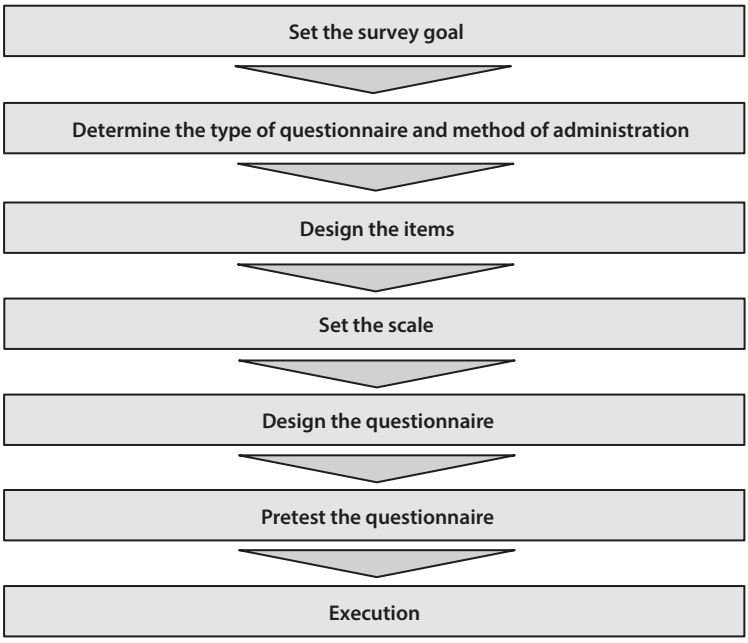
4.4.2 Collecting Quantitative Data: Designing Surveys

There is little doubt that **surveys** are the mainstay of primary market research. While it may seem easy to conduct a survey (just ask what you want to know, right?), there are many issues that could turn good intentions into bad results. In this section, we discuss the key design choices for good surveys. Designing a good survey requires several steps. First, determine the survey goal. Next, determine the type of questionnaire required and the administration method. Thereafter, decide on the questions and the scale, as well as the design of the questionnaire. Conclude by pretesting and administering the questionnaire. We show these steps in ◘ Fig. 4.5.

4.4.2.1 Set the Survey Goal

Before you start designing the questionnaire, it is vital to consider the survey goal. Is it to collect quantitative data on customers' background, to assess customer satisfaction, or do you want to understand why and how customers complain? These different goals influence the type of questions asked (such as open-ended or closed-ended questions), the method of administration (e.g., by mail or on the Web), and other design issues discussed below. Two aspects are particularly relevant when designing surveys:

First, consider the information or advice you want to emerge from the study for which the survey is required. Say you are asked to help understand check-in waiting times at an airport. If the specific study question is to gain an understanding of how many minutes travelers are willing to wait before becoming dissatisfied, you should be able to provide an answer to the question: How much does travelers' satisfaction decrease with increased waiting time? If, on the other hand, the specific question is to understand how people



◘ **Fig. 4.5** Steps in designing surveys

perceive waiting time (short or long), your questions should focus on how travelers perceive this time and, perhaps, what influences their perception. Thus, the information or advice you want to provide influences the questions that you should pose in a survey.

Second, consider the method required for the study early in the design process. For example, if a study's goal is to determine market segments, you should probably use cluster analysis (see ► Chap. 9). Similarly, if the study's goal is to develop a way to systematically measure customer satisfaction, you are likely to use factor analysis (see ► Chap. 8). This approach is crucial, as each method requires different types of data. Cluster analysis, for example, generally requires variables that are not too highly correlated, meaning that researchers need to use a type of questionnaire that can produce these data. On the other hand, factor analysis requires data that include different, but highly correlated, variables. If you use factor analysis to distinguish between the different aspects of consumer satisfaction, you need to design a survey that will produce data allowing you to conduct factor analysis.

4.4.2.2 Determine the Type of Questionnaire and Method of Administration

After determining the survey goal, you need to decide on the type of questionnaire you should use and how it should be administered. There are four key ways of administering a survey:

- Personal interviews,
- telephone interviews,
- web surveys, and
- mail surveys.

In some cases, researchers combine different ways of administering surveys. This is called a mixed mode.

- **Personal Interviews**

Personal interviews (or **face-to-face interviews**) can obtain high response rates, since engagement with the respondents are maximized, allowing rich information (visual expressions, etc.) to be collected. Moreover, since people find it hard to walk away from interviews, it is possible to collect answers to a reasonably lengthy set of questions. Consequently, personal interviews can support long surveys. It is also the best type of data collection for open-ended responses. In situations where the respondent is initially unknown, this may be the only feasible data collection type. Consequently, in-depth interviews may be highly preferable, but they are also the costliest per respondent. This is less of a concern if only small samples are required (in which case personal interviewing could be the most efficient). Other issues with personal interviews include the possibility of *interviewer bias* (i.e., a bias resulting from the interviewer's behavior, e.g., in terms of his/her reactions or presentation of the questions), *respondent bias* to sensitive items, and the data collection usually takes more time. Researchers normally use personal interviewing when they require an in-depth exploration of opinions. Such interviewing may also help if drop out is a key concern. For example, if researchers collect data from executives around the globe, using methods other than face-to-face interviewing may lead to excessive non-response in countries such as Russia or China where face-to-face interviews are a sign of respect for and appreciation of the time taken. *CAPI*, which is the abbreviation of *computer-assisted personal interviews*, is a frequently used term in

the context of in-depth interviewing. CAPI involve using computers during the interviewing process to, for example, route the interviewer through a series of questions, or to enter responses directly. Similarly, in *CASI* (*computer-assisted self-interviews*), the respondent uses a computer to complete the survey questionnaire without an interviewer administering it.

■ Telephone Interviews

Telephone interviews allow researchers to collect data quickly. These interviews also support open-ended responses, although not as well as personal interviews. Moreover, interviewer bias can only be controlled moderately, since the interviewers follow predetermined protocols, and the respondent's interactions with others during the interview is strongly controlled. Telephone interviewing can be a good compromise between mailed interviews' low cost and the richness of in-depth interviews. *CATI* refers to *computer-assisted telephone interviews*, which are an important method of administering surveys. Until the 1990s, telephone interviews were generally conducted via landlines, but mobile phone usage has soared in the meantime. In many countries, mobile phone adoption rates are higher than landline adoption. This holds especially for African countries, India, and many European countries if younger consumers are the targeted interviewees (Vincente and Reis 2010). Consequently, *mobile phone surveys* have become dominant in market research.

A decade ago, the differences between landline and mobile phone surveys could be large, with younger and richer individuals being overrepresented in mobile phone surveys (Vincente et al. 2008). As the adoption of mobile phones increased, the use of landlines decreased, which introduced new problems in terms of sampling errors (Stern et al. 2014). For example, younger people are now less likely to have landlines. An additional issue is that landlines are fixed to a geographical area whereas mobile phones are not. As people move or travel, mobile phones are far less likely to give useful information about geographic areas. While recent research has shown that differences in response accuracy between mobile phone and landline surveys are small (e.g., with regard to social desirability bias), especially for questions that are not cognitively demanding (e.g., Kennedy and Everett 2011; Lynn and Kaminska 2013), the samples from which they are drawn can be very different in practice. There are some other differences as well. For example, the likelihood of full completion of surveys is higher for mobile calling, even though completion takes around 10–15 % longer (Lynn and Kaminska 2013).

■ Web Surveys

Web surveys (sometimes referred to as *CAWI*, or *computer-assisted web interviews*) are often the least expensive to administer and can be set up very quickly. Researchers can administer web surveys to very large populations, even internationally, because, besides the fixed costs of setting up a survey, the marginal costs of administering additional web surveys are relatively low.

Many firms specializing in web surveys will ask $0.30 (or more) for each respondent, which is substantially lower than the costs of personal interviews, telephone interviews, and mail surveys. It is also easy to obtain precise quotes quickly. Qualtrics (http://www.qualtrics.com) is a leading web service provider that allows a specific type of respondent and a desired sample size to be chosen. For example, using Qualtrics's sample to survey 500 current owners of cars to measure their satisfaction costs $2500 for a 10-minute survey. This cost increases sharply if samples are hard to access and/or require compensation for their time. For example, surveying 500 purchasing managers by means of a 10-minute survey costs approximately $19,500.

Web surveys also support complex survey designs with elaborate branching and skip patterns that depend on the response. For example, web surveys allow different surveys to be created for different types of products. Further, since web surveys reveal the questions progressively to the respondents, there is an option to channel them to the next question based on their earlier responses. This procedure is called *adaptive questioning*. In addition, web surveys can be created that allow respondents to automatically skip questions if they do not apply. For example, if a respondent has no experience with an iPad, researchers can create surveys that do not ask questions about this product. However, web surveys impose similar burdens on the respondents as mail surveys do (see below), which they may experience as an issue. This makes administering long web surveys difficult. Moreover, open-ended questions tend to be problematic, because few respondents are likely to provide answers, leading to a low item response. There is evidence that properly conducted web surveys lead to data as good as those obtained from mail surveys; in addition, the lack an interviewer, which can result in *interviewer bias,* means they can provide better results than personal interviews (Bronner and Ton 2007; Deutskens et al. 2006). In web surveys, the respondents are also less exposed to evaluation apprehension and less inclined to respond with socially desirable behavior.[4]

To complete web surveys, most respondents still use their personal computers, but the popularity of mobile devices (i.e., smartphones and tablets) has risen significantly in recent years. Brosnan et al. (2017) have found that survey completion rates vary across devices with personal computers yielding the highest rate (83 %), followed by tablets (66 %) and smartphones (63 %). In addition, since device use is significantly associated with specific socio-demographic characteristics, you should not limit respondents to use a specific device for completing a survey as this may negatively affect representativity.

It is important to distinguish between true web-based surveys used for collecting information on which marketing decisions will be based and polls, or very short surveys, on websites used to increase interactivity. These polls/short surveys are used to attract and keep people interested in websites and are not part of market research. For example, USA Today (http://www.usatoday.com), an American newspaper, regularly publishes short polls on their main website.

4 For a comparison of response behavior between CASI, CAPI, and CATI, see Bronner and Ton (2007).

Mail Surveys

Mail surveys are paper-based surveys sent out to respondents. They are a more expensive type of survey research and are best used for sensitive items. Since no interviewer is present, there is no interviewer bias. However, mail surveys are a poor choice for complex survey designs, such as when respondents need to skip a large number of questions depending on previously asked questions, as this means that the respondent needs to correctly interpret the survey structure. Open-ended items are also problematic, because few people are likely to provide answers to such questions if the survey is administered on paper. Other problems include a lack of control over the environment in which the respondent fills out the survey and that mail surveys take longer than telephone or web surveys. However, in some situations, mail surveys are the only way to gather data. For example, while executives rarely respond to web-based surveys, they are more likely to respond to paper-based surveys. Moreover, if the participants cannot easily access the web (such as employees working in supermarkets, cashiers, etc.), handing out paper surveys is likely to be more successful.

Tip

The method of administration also has a significant bearing on a survey's maximum duration (Vesta Research 2016). As a rule of thumb, telephone interviews should be no longer than 20 minutes. When calling a on a mobile phone, however, the survey should not exceed 5 minutes. The maximum survey duration of web surveys is 20 minutes—10 minutes for social media-based surveys. Personal interviews and mail surveys can be much longer, depending on the context. For example, surveys comprising personal interviews on topics that respondents find important, could take up to 2 hours. However, when topics are less important, mail surveys and personal interviews need to be considerably shorter.

Market researchers are increasingly using **mixed mode** approaches. An example of a mixed mode survey is when potential respondents are first approached by phone, asked to participate and confirm their email addresses, after which they are given access to a web survey. Mixed mode approaches are also used when people are first sent a paper survey and then called if they fail to respond.

Mixed mode approaches may help, because they signal that the survey is important. They may also help improve response rates, as people who are more visually oriented prefer mail and web surveys, whereas those who are aurally oriented prefer telephone surveys. However, there is only evidence of increased response rates when modes are offered sequentially and of mixed modes reducing response rates when offered simultaneously (Stern et al. 2014). By providing different modes, people can use the mode they most prefer. A downside of mixed mode surveys is that they are expensive and require a detailed address list (including a telephone number and matching email address). However, systematic (non)response is the most serious mixed mode survey issue. For example, when filling out mail surveys, the respondents have more time than when providing answers

by telephone. If respondents need this time to think about their answers, the responses obtained from mail surveys may differ systematically from those obtained from telephone surveys.

4.4.2.3 Design the Items

Designing the items, whether for a personal interview, web survey, or mail survey, requires a great deal of thought. Take, for example, the survey item shown in ◘ Fig. 4.6:

It is unlikely that people can give meaningful answers to such an item. First, using a negation ("not") in sentences makes questions hard to understand. Second, the reader may not have an iPhone, or may not have experience using it. Third, the answer categories are unequally distributed. That is, there is one category above neutral while there are two below. These issues are likely to create difficulties with understanding and answering questions, which may, in turn, cause validity and reliability issues. While the last aspect relates to the properties of the scale (see ► Sect. 4.4.2.4), the first two issues refer to the item content and wording, which we will discuss next (see ◘ Table 4.1 for a summary).

- **Item content**

When deciding on the *item content*, there are at least three essential rules you should keep in mind.

As a *first rule*, ask yourself whether everyone will be able to answer each item. If the item is, for example, about the quality of train transport and the respondent always travels by car, his or her answers will be meaningless. However, the framing of items is important, since, for example, questions about why the specific respondent does not use the train can yield important insights.

As a *second rule*, you should check whether respondents can construct or recall an answer. If you require details that possibly occurred a long time ago (e.g., what information did the real estate agent provide when you bought/rented your current house?), the respondents may have to "make up" an answer, which leads to validity and reliability issues.

As a *third rule*, assess whether the respondents are willing to answer an item. If contents are considered sensitive (e.g., referring to sexuality, money, etc.), respondents may provide more socially desirable answers (e.g., by reporting higher or lower incomes than are actually true). Most notably, respondents might choose to select a position that they believe society favors (e.g., not to smoke or drink, or to exercise), inducing a **social desirability bias**. They may also not answer such questions at all. You should determine whether these items are necessary to attain the research objective. If they are not, omit them from the survey. The content that respondents may regard as sensitive is subjective and differs across cultures, age categories, and other variables. Use your common sense and, if necessary, use experts to decide whether the items are appropriate. In addition, make sure you pretest the survey and ask the pretest participants whether they were reluctant to provide certain answers. To reduce respondents' propensity to give socially desirable answers, use

◘ **Fig. 4.6** Example of a bad survey item

indirect questioning (e.g., "What do you believe other people think about … ?"), frame questions as neutrally as possible, and suggest that many people exhibit behavior different from the norm (e.g., Brace 2004). Adjusting the response categories also helps when probing sensitive topics. For example, instead of directly asking about respondents' disposable income, you can provide various answering categories, which will probably increase their willingness to answer this question.

Designing the items also requires deciding whether to use open-ended or closed-ended questions. **Open-ended questions** (also called **verbatim items**) provide little or no structure for respondents' answers. Generally, the researcher asks a question and the respondent writes down his or her answer in a box. Open-ended questions are flexible and allow for explanation, making them particularly suitable for exploratory research. However, the drawback is that the respondents may be reluctant to provide detailed information. Furthermore, answering open-ended questions is more difficult and time consuming. In addition, their interpretation requires substantial coding. Coding issues arise when respondents provide many different answers (such as "sometimes," "maybe," "occasionally," or "once in a while") and the researcher has to divide these into categories (such as very infrequently, infrequently, frequently, and very frequently) for further statistical analysis. This coding is very time-consuming, subjective, and difficult. **Closed-ended questions** provide respondents with only a few categories from which to choose, which drastically reduces their burden, thereby inducing much higher response rates compared to open-ended questions. Closed-ended questions also facilitate immediate statistical analysis of the responses as no ex post coding is necessary. However, researchers need to identify answer categories in advance and limit this number to a manageable amount. Closed-ended questions are dominant in research and practice.

▪ Item wording

The golden rule of *item wording* is to keep it short and simple. That is, use simple words and avoid using jargon or slang as some respondents may misunderstand these. Similarly, keep grammatical complexities to a minimum by using active rather than passive voice, repeat the nouns instead of using pronouns, and avoiding possessive forms (Lietz 2010). Use short sentences (20 words max; Oppenheim 1992) to minimize the cognitive demands required from the respondents (Holbrook et al. 2006) and to avoid the risk of **double-barreled questions**, causing respondents to agree with one part of the question but not the other, or which they cannot answer without accepting a particular assumption. An example of such a double-barreled question is: "In general, are you satisfied with the products and services of the company?" A more subtle example is: "Do you have the time to read the newspaper every day?" This question also contains two aspects, namely "having time" and "reading the paper every day." The question "Do you read the newspaper every day?" followed by another about the reasons for a negative or a positive answer would be clearer (Lietz 2010).

Moreover, avoid using the word *not* or *no* where possible. This is particularly important when other words in the same sentence are negative, such as "unable," or "unhelpful," because sentences with two negatives (called a double negative) are hard to understand. For example, a question such as "I do not use the email function on my iPhone because it is unintuitive" is quite hard to follow and takes longer to process (Peng and Finn 2016). Also, avoid using *vague quantifiers*, such as "frequent" or "occasionally" (Dillman et al. 2014), which make it difficult for respondents to answer questions (what exactly is meant by

"occasionally"?). They also make comparing responses difficult. After all, what one person considers "occasionally," may be "frequent" for another.[5] Instead, it is better to use frames that are precise ("once a week").

Never suggest an answer. For example, a question like "Company X has done very well, how do you rate it?" is highly suggestive and would shift the mean response to the positive end of the scale.

Many researchers recommend including **reverse-scaled items** in surveys. Reverse-scaled means the question, statement (if a Likert scale is used), or word pair (if a semantic differential scale is used) are reversed compared to the other items in the set. Box 4.5 shows an example of a five-item scale for measuring consumers' attitude toward the brand (e.g., Sarstedt et al. 2016) with one reverse-scaled item printed in bold. Reverse-scaled items act as cognitive "speed bumps" that alert inattentive respondents that the content varies (Weijters and Baumgartner 2012) and help reduce the effect of *acquiescence*, which relates to a respondent's tendency to agree with items regardless of the item content (Baumgartner and Steenkamp 2001; see also ► Chap. 5). Furthermore, reverse-scaled items help identify straight-lining, which occurs when a respondent marks the same response in almost all the items (► Chap. 5). However, numerous researchers have shown that reverse-scaled items create problems, for example, by generating artificial factors, as respondents have greater difficulty verifying the item. This creates *misresponse rates* of 20 % and higher (e.g., Swain et al. 2008; Weijters et al. 2013). Given these results, we suggest employing reverse-scaled items sparingly (Weijters and Baumgartner 2012), for example, only to identify straight-lining. If used, reverse-scaled items should not contain a particle negation such as "not" or "no."

Finally, when you undertake a survey in different countries (or adapt a scale from a different language), use professional translators, because translation is a complex process. Functionally, translating one language into another seems quite easy, as there are many websites, such as Google translate (translate.google.com), that do this. However, translating surveys requires preserving the conceptual equivalence of whole sentences and paragraphs; current software applications and websites cannot ensure this. In addition, cultural differences normally require changes to the instrument format or procedure. Back-translation is a technique to establish conceptual equivalence across languages. **Back-translation** requires translating a survey instrument into another language, after which another translator takes

Box 4.5 An example of a scale with reverse-scaled items (in bold).

Please rate the *brand* in the advertisement on the following dimensions:

Dislike	1	2	3	4	5	6	7	Like
Unpleasant	1	2	3	4	5	6	7	Pleasant
Good	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Bad**
Expensive	1	2	3	4	5	6	7	Inexpensive
Useless	1	2	3	4	5	6	7	Useful

5 Foddy (1993) reported 445 interpretations of the word "usually," with the meaning assigned to the word varying, depending on, for example, the type of activity or who was asked about the activity.

the translated survey instrument and translates it back into the original language (Brislin 1970). After the back-translation, the original and back-translated instruments are compared and points of divergence are noted. The translation is then corrected to more accurately reflect the intent of the wording in the original language.

4.4.2.4 Set the Scale

When deciding on scales, two separate decisions need to be made. First, you need to decide on the type of scale. Second, you need to set the properties of the scale you choose.

- **Type of Scale**

Marketing research and practice have provided a variety of *scale types*. In the following, we discuss the most important (and useful) ones:

- Likert scales,
- semantic differential scales, and
- rank order scales.

The most popular scale type used in questionnaires is the **Likert scale** (Liu et al. 2016). Likert scales are used to establish the degree of agreement with a specific statement. Such a statement could be "I am very committed to Oddjob Airways." The degree of agreement is usually set by scale endpoints ranging from strongly disagree to strongly agree. Likert scales are used very frequently and are relatively easy to administer. Bear in mind that if the statement is too positive or negative, it is unlikely that the endings of the scale will be used, thereby reducing the number of answer categories actually used. We show an example of three Likert-scale-type items in ◘ Fig. 4.7, in which respondents assess the personality of the Oddjob Airways brand.

The semantic differential scale is another scale type that features prominently in market research. **Semantic differential scales** use an opposing pair of words, normally adjectives (e.g., young/old, masculine/feminine) constituting the endpoint of the scale. Respondents then indicate how well one of the word in each pair describes how he or she feels about the object to be rated (e.g., a company or brand). These scales are widely used in market research. As with Likert scales, 5 or 7 answer categories are commonly used (see the next section on the number of answer categories you should use). We provide an example of the semantic differential scale in ◘ Fig. 4.8, in which respondents provide their view of Oddjob Airways.

Rank order scales are a unique type of scale, as they force respondents to compare alternatives. In its basic form, a rank order scale allows respondents to indicate which alternative they rank highest, which alternative they rank second highest, etc. ◘ Fig. 4.9 shows an example. The respondents therefore need to balance their answers instead of merely

◘ **Fig. 4.7** Example of a 7-point Likert scale

Fig. 4.8 Example of a 7-point semantic differential scale

Fig. 4.9 Example of a rank order scale

stating that everything is important. In a more complicated form, rank order scales ask respondents to allocate a certain total number of points (often 100) to a number of alternatives. This is called the **constant sum scale**. Constant sum scales work well when a small number of answer categories are used (normally up to 5). Generally, respondents find constant scales that have 6 or 7 answer categories somewhat challenging, while constant scales that have 8 or more categories are very difficult to answer. The latter are thus best avoided.

In addition to these types of scaling, there are other types, such as graphic rating scales, which use pictures to indicate categories, and the *MaxDiff scale*, in which respondents indicate the most and least important items. We introduce the MaxDiff scale in the ↓ Web Appendix (→ Downloads). The MaxDiff scale is a form of *Best-Worst scaling*.

© EtiAmmos/Getty Images/iStock
https://www.guide-market-research.com/app/download/13488664327/SPSS+3rd_Chapter+4_MaxDiff+scales.pdf?t=1516713250

Properties of the Scale

After selecting the type of scale, we need to set the *scale properties*, which involves making several decisions:

- Decide on the number of response categories,
- choose between forced-choice scale and free-choice scales,
- design the response categories,
- label the response categories,
- decide whether to use a "don't know" option, and
- choose between a balanced and unbalanced scale.

Decide on the number of response categories: When using closed-ended questions, the number of answer categories needs to be determined. In its simplest form, a survey could use just two answer categories (yes/no). Multiple categories (such as, "completely disagree," "disagree," "neutral," "agree," "completely agree") are used more frequently to allow for more nuances. When determining how many scale categories to use, we face a trade-off between having more variation and differentiation in the responses versus burdening the respondents too much, which can trigger different types of response biases (e.g., Weijters et al. 2010). Because 5-point scales and 7-point scales are assumed to achieve this trade-off well, their use has become common in marketing research (e.g., Fink 2003; Peterson 1997). Research on these two scale types in terms of their reliability and validity is inconclusive, but the differences between them are generally not very pronounced (e.g., Lietz 2010; Peng and Finn 2016; Weijters et al. 2010). Ten-point scales are often used in market research practice. However, scales with many answer categories often confuse respondents, because the wording differences between the scale points become trivial. For example, the difference between "tend to agree" and "somewhat agree" are subtle and respondents may not be able to differentiate between them. In such a case, respondents tend to choose categories in the middle of the scale (Rammstedt and Krebs 2007). Given this background, we recommend using 5-point or 7-point scales.

Finally, many web surveys now use levers that allow scaling on a continuum without providing response categories. This scale type is called a **visual analogue scale** and is especially recommended if small differences in response behavior need to be detected. Visual analogue scales are well known in paper-and-pencil-based research (especially in the medical sector) and have become increasingly popular in web surveys. A visual analogue scale consists of a line and two anchors, one at each end. The anchors often consist of verbal materials that mark opposite ends of a semantic dimension (e.g., good and bad). However, the anchors may also be pictures, or even sound files. Visual anchors, such as smileys, are often used with participants who may not fully grasp the meaning of verbal materials—for example, preschool children (Reips and Funke 2008). Compared to ordinal scales, measurement by means of visual analogue scales is more exact, leading to more narrow confidence intervals, and higher power statistical tests. This exactness helps detect smaller effects that may be unobservable with ordinal scales (Reips and Funke 2008). However, although not all web survey platforms offer visual analogue scales, you should use them if possible. ◘ Fig. 4.10 shows an example of a visual analogue scale that measures customers' expectations of Oddjob Airways.

Choose between a forced-choice scale and a free-choice scale: Sometimes, researchers and practitioners use 4-point or 6-point scales that omit the neutral category, thereby forcing the respondents to be positive or negative. Using such a **forced-choice scale** could bias the

Fig. 4.10 Example of a visual analogue scale

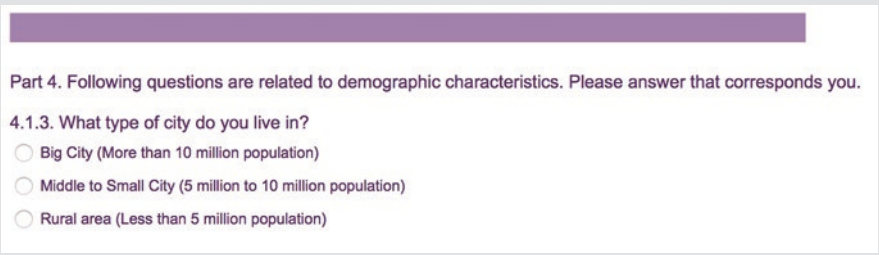
answers, leading to validity issues.[6] By providing a neutral category choice (i.e., a **free-choice scale**), the respondents are not forced to give a positive or negative answer. Many respondents feel more comfortable about participating in a survey using a free-choice scale (Nowlis et al. 2002). Furthermore, research has shown that including a neutral category minimizes response bias (Weijters et al. 2010). Therefore, we strongly suggest using free-choice scales and including a neutral category.

Design the response categories: When designing response categories for categorical variables, use response categories that are exclusive, so that answers do not overlap (e.g., age categories 0–4, 5–10, etc.). The question here is: How do we decide on the spacing between the categories? For example, should we divide US household income in the categories 0–$9999, $10,000–$19,999, $20,000-higher, or use another way to set the categories? One suggestion is to use narrower categories if the respondent can recall the variable easily. A second suggestion is to space the categories so that we, as researchers, expect an approximately equal number of observations per category. In the example above, we may find that most households have an income of $20,000 or higher and that categories 0–$9999 and $10,000–$19,999 are infrequently used. It is best to choose categories where equal percentages are expected such as 0–$24,999, $25,000–$44,999, $45,000–$69,999, $70,000–$109,999, $110,000–and higher. Although the range in each category differs, we can reasonably expect each category to hold about 20 % of the responses if we sample US households randomly (https://en.wikipedia.org/wiki/Household_income_in_the_United_States#Household_income_over_time). Box 4.6 shows a real-life example of oddly chosen response categories (and there are a few other issues too!).

Note that the response categories also give the respondents the range of acceptable answers. Respondents generally view the middle of the scale as normal, or most common, and position themselves in relation to this (e.g., Revilla 2015). For example, Tourangeau and Smith (1996) found that the reported number of sexual partners in an open-ended question is more than twice as high when the answer category labels used in a previously asked closed-ended question on the same topic shifted to indicate higher numbers (from 0, 1, 2, 3, 4, 5 or more to 0, 1–4, 5–9, 10–49, 50–99, 100 or more).

6 Forced-choice scales can, of course, also be used for uneven response categories such as 5-point or 7-point Likert scales.

Box 4.6 Oddly chosen response categories



Label the response categories: A common way of labeling response categories is to use endpoint labels only, omitting intermediary labels. For example, instead of labeling all five points of a Likert scale (e.g., "completely disagree," "disagree," "neutral," "agree," and "completely agree"), we only label the endpoints (e.g., "completely disagree" and "completely agree"). While this approach makes response category labeling easy, it also amplifies acquiescence ► Chap. 5, because the endpoints reinforce the respondents' agreement (Weijters et al. 2010). Conversely, if all the categories are labeled, this helps the respondents interpret and differentiate, making the midpoint more salient and accessible. Labeling the response categories fully also increases the scale reliability (Weng 2004), but reduces criterion validity. Drawing on Weijters et al. (2010), we generally recommend only labeling the endpoints. However, when using the items for prediction-type analyses, such as in correlation and regression analyses, it is beneficial to label all the categories.[7]

Decide whether or not to use a "don't know" option: Another important choice to make is whether or not to include a "don't know" option in the scaling (Lietz 2010). Using a "don't know" option allows the researcher to distinguish between those respondents who have a clear opinion and those who do not. Moreover, the respondents may find that answering the survey is slightly easier. While these are good reasons for including this category, the drawback is that there will then be missing observations. If many respondents choose not to answer, this will substantially reduce the number of surveys that can be used for analysis. Generally, when designing surveys, you should only include a "don't know" (or "undecided") option as an answer to questions that the respondents might genuinely not know, for example, when requiring answers to factual questions. If included, the option should appear at the end of the scale. The "don't know" option should not be included in other types of questions (such as on attitudes or preferences), as researchers are interested in the respondents' perceptions regardless of their knowledge of the subject matter.

7 Sometimes, researchers number the response options. For example, when numbering the response options of a 7-point scale you can use only positive numbers (1 to 7), or positive and negative numbers (-3 to +3). For recommendations, see Cabooter et al. (2016).

Choose between a balanced and unbalanced scale: A **balanced scale** has an equal number of positive and negative scale categories. For example, in a 5-point Likert scale, we may have two negative categories (completely disagree and disagree), a neutral option, and two positive categories (agree and completely agree). Besides this, the wording in a balanced scale should reflect equal distances between the scale items. This is called an **equidistant scale**, which is a requirement for analysis techniques such as factor analysis (► Chap. 8). Consequently, we strongly recommend using a balanced scale instead of an **unbalanced scale**. A caveat of balanced scales is that many constructs cannot have negative values. For example, one can have some trust in a company or very little trust, but negative trust is highly unlikely. If a scale item cannot be negative, you will have to resort to an unbalanced scale in which the endpoints of the scales are unlikely to be exact opposites. ◘ Table 4.1 summarizes the key choices we have to make when designing survey items.

4.4.2.5 Design the Questionnaire

After determining the individual questions, you have to integrate these, together with other elements, to create the questionnaire. *Questionnaire design* involves the following elements:

- Designing the starting pages of the questionnaire,
- choosing the order of the questions, and
- designing the layout and format.

Starting pages of the questionnaire: At the beginning of each questionnaire, the importance and goal are usually described to stress that the results will be treated confidentially, and to mention what they will be used for. This is usually followed by an example question (and answer), to demonstrate how the survey should be filled out. Keep this page very short when using mobile surveys.

If questions relate to a specific issue, moment, or transaction, you should indicate this clearly at the very beginning. For example, "Please provide answers to the following questions, keeping the purchase of product X in mind." If applicable, you should also point out that your survey is conducted in collaboration with a university, a recognized research institute, or a known charity, as this generally increases respondents' willingness to participate. Moreover, do not forget to provide a name and contact details for those participants who have questions, or if technical problems should arise. Consider including a photo of the research team, as this increases response rates. Lastly, you should thank the respondents for their time and describe how the questionnaire should be returned (for mail surveys).

Order of the questions: Choosing an appropriate *question order* is crucial, because it determines the questionnaire's logical flow and therefore contributes to high response rates. The order of questions is usually as follows:

1. *Screening questions* (typically simply referred to as *screeners*) come first. These questions determine what parts of the survey a respondent should fill out.
2. Next, ask questions relating to the study's key variables. This includes the dependent variables, followed by the independent variables.
3. Use a *funnel approach*. That is, ask questions that are more general first and then move on to details. This makes answering the questions easier as the order helps the respondents recall. Make sure that sensitive questions are put at the very end of this section.
4. Questions related to demographics are placed last if they are not part of the screening questions. If you ask demographic questions, always check whether

Table 4.1 A summary of some of the key choices when designing survey items

Aspect	Recommendation
Item content	
Can all the respondents answer the question asked?	Ensure that all potential respondents can answer all items. If they cannot, ask screener questions to direct them. If the respondents cannot answer questions, they should be able to skip them.
Can the respondents construct or recall the answers?	If the answer is no, you should use other methods to obtain information (e.g., secondary data or observations). Moreover, you should ask the respondents about major aspects before zooming in on details to help them recall answers.
Do the respondents want to answer each question?	If the questions concern "sensitive" subjects, check whether they can be omitted. If not, stress the confidentiality of the answers and mention why these answers are useful for the researcher, the respondent, or society before introducing them.
Open-ended or closed-ended questions?	Keep the subsequent coding in mind. If easy coding is possible beforehand, design a set of exhaustive answer categories. Further, remember that open-ended scale items have a much lower response rate than closed-ended items.
Item wording	
Grammar and sentences	Use simple wording and grammar. Keep sentences short. Avoid negations, vague quantifiers, and double-barreled questions. Employ reverse-scaled items sparingly. Use back-translation if necessary.
Type of scale	
Number of response categories	Use visual analogue scales. If not available, use 5-point or 7-point scales.
Forced-choice or free-choice scale?	Use a free-choice scale by including a neutral category.
Design of the response categories	Ensure that the response categories are exclusive and that each category has approximately the same percentage of responses.
What scaling categories should you use (closed-ended questions only)?	Use visual analogue scales if possible, otherwise Likert scales. If the question requires this, use semantic differential scales, or rank order scales.
Labeling of response categories	Label endpoints only. When using the items for prediction-type analyses, label all categories.
Inclusion of a "Don't know" option	Include a "Don't know" option only for items that the respondent might genuinely not know. If included, place this at the end of the scale.
Balanced or unbalanced scale?	Always use a balanced scale. There should be an exact number of positive and negative wordings in the scale items. The words at the ends of the scale should be exact opposites.

they are relevant to the research goal and if they are not already known.[8] In addition, check if these demographics are likely to lead to non-response. Asking about demographics, like income, educational attainment, or health, may result in a substantial number of respondents refusing to answer. If such sensitive demographics are not necessary, omit them from the survey. Note that in certain countries asking about a respondent's demographic characteristics means you must adhere to specific laws, such as the Data Protection Act 1998 in the UK.

If your questionnaire comprises several sections (e.g., in the first section, you ask about the respondents' buying attitudes and, in the following section, about their satisfaction with the company's services), you should make the changing context clear to the respondents.

Layout and format of the survey: The layout of both mail and web-based surveys should be concise and should conserve space where possible, particularly in respect of mobile surveys. Avoid using small and colored fonts, which reduce readability. Booklets work well for mail-based surveys, since postage is often cheaper if surveys fit in standard envelopes. If this is not possible, single-sided stapled paper can also work. When using web-based surveys, it is good to have a counter that tells the respondents what percentage of the questions they have already filled out. This gives them some indication of how much time they are likely to need to complete the survey. Make sure the layout is simple and compatible with mobile devices and tablets. Qualtrics and other survey tools offer mobile-friendly display options, which should always be ticked.

4.4.2.6 Pretest the Questionnaire and Execution

We have already mentioned the importance of *pretesting* the survey several times. Before any survey is sent out, you should pretest the questionnaire to enhance its clarity and to ensure the client's acceptance of the survey. Once the questionnaire is in the field, there is no way back! You can pretest questionnaires in two ways. In its simplest form, you can use a few experts (say 3–6) to read the survey, fill it out, and comment on it. Many web-based survey tools allow researchers to create a pretested version of their survey, in which there is a text box for comments behind every question. Experienced market researchers can spot most issues right away and should be employed to pretest surveys. If you aim for a very high quality survey, you should also send out a set of preliminary (but proofread) questionnaires to a small sample of 50–100 respondents. The responses (or lack thereof) usually indicate possible problems and the preliminary data can be analyzed to determine the potential results. Never skip pretesting due to time issues, since you are likely to run into problems later!

Motivating potential respondents to participate is an increasingly important aspect of survey research. In addition to Dillman et al.'s (2014) recommendations on how to increase *response rates* (Box 4.7), incentives are used. A simple example of such an incentive is to provide potential respondents with a cash reward. In the US, one-dollar bills are often used for this purpose. Respondents who participate in (online) research panels often receive points that can be exchanged for products and services. For example, Research

8 The demographics of panels (see ► Sect. 4.4.2.2) are usually known.

Box 4.7 Dillman et al.'s (2014) recommendations on how to increase response rates

It is becoming increasingly difficult to get people to fill out surveys. This may be due to over-surveying, dishonest firms that disguise sales as research, and a lack of time. Dillman et al. (2014) discuss four steps to increase response rates:

1. Send out a pre-notice letter indicating the importance of the study and announcing that a survey will be sent out shortly.
2. Send out the survey with a sponsor letter, again indicating the importance of the study.
3. Follow up after 3–4 weeks with a thank you note (for those who responded) and the same survey, plus a reminder (for those who did not respond).
4. Call or email those who have still not responded and send out a thank you note to those who replied during the second round.

Now, a market research company, gives its Canadian panel members AirMiles that can be exchanged for, amongst others, free flights. A special type of incentive is to indicate that some money will be donated to a charity for every returned survey. ESOMAR, the world organization for market and social research (see ► Chap. 10), suggests that incentives for interviews or surveys should "be kept to a minimum level proportionate to the amount of their time involved, and should not be more than the normal hourly fee charged by that person for their professional consultancy or advice."

Another incentive is to give the participants a chance to win a product or service. For example, you could randomly give a number of participants gifts. The participants then need to disclose their name and address in exchange for a chance to win a gift so that they can be reached. While this is not part of the research itself, some respondents may feel uncomfortable providing their contact details, which could potentially reduce the response rate.

Finally, reporting the findings to the participants is an incentive that may help participation (particularly in professional settings). This can be done by providing a general report of the study and its findings, or by providing a customized report detailing the participant's responses and comparing them with all the other responses (Winkler et al. 2015). Obviously, anonymity needs to be assured so that the participants cannot compare their answers to those of other individual responses.

4.5 Basic Qualitative Research

Qualitative research is mostly used to gain an understanding of *why* certain things happen. It can be used in an exploratory context by defining problems in more detail, or by developing hypotheses to be tested in subsequent research. Qualitative research also allows researchers to learn about consumers' perspectives and vocabulary, especially if they are not familiar with the context (e.g., the industry). As such, qualitative research offers importance guidance when little is known about consumers' attitudes and perceptions, or the market (Barnham 2015).

As discussed in ► Chap. 3, qualitative research leads to the collection of qualitative data. One can collect qualitative data by explicitly informing the participants that you are

doing research (directly observed qualitative data), or you can simple observe the participants' behavior without them being explicitly aware of the research goals (indirectly observed qualitative data). There are ethical issues associated with conducting research if the participants are not aware of the research purpose. Always check the regulations regarding what is allowed in your context and what not. It is, in any case, good practice to brief the participants on their role and the goal of the research once the data have been collected.

The two key forms of directly observed qualitative data are in-depth interviews and focus groups. Together, they comprise most of the conducted qualitative market research. First, we will discuss in-depth interviews, which are—as the terms suggests—interviews conducted with one participant at a time, allowing for high levels of personal interaction between the interviewer and respondent. Next, we will discuss projective techniques, a frequently used type of testing procedure in in-depth interviews. Lastly, we will introduce focus group discussions, which are conducted with multiple participants.

4.5.1 In-depth Interviews

In-depth interviews are qualitative conversations with participants on a specific topic. These participants are often consumers, but, in a market research study, they may also be the decision-makers, who are interviewed to gain an understanding of their clients' needs. They may also be government or company representatives. The structure levels of interviews vary. In their simplest form, interviews are unstructured and the participants talk about a topic in general. This works well if you want to obtain insight into a topic, or as an initial step in a research process. Interviews can also be fully structured, meaning all the questions and possible answer categories are decided beforehand. This way allows you to collect quantitative data. However, most in-depth interviews for gathering qualitative data are semi-structured and contain a series of questions that need to be addressed, but have no specific format regarding what the answers should look like. The person interviewed can make additional remarks, or discuss somewhat related issues, but is not allowed to wander off too far. In these types of interviews, the interviewer often asks questions like "that's interesting, could you explain?," or "how come … ?" to gain more insight into the issue. In highly structured interviews, the interviewer has a fixed set of questions and often a fixed amount of time for each person's response. The goal of structured interviews is to maximize the comparability of the answers. Consequently, the set-up of the questions and the structure of the answers need to be similar.

In-depth interviews are unique in that they allow probing on a one-to-one basis, thus fostering an interaction between the interviewer and interviewee. In-depth interviews also work well when those being interviewed have very little time and when they do not want the information to be shared with the other study participants. This is, for example, probably the case when you discuss marketing strategy decisions with CEOs. The drawbacks of in-depth interviews include the amount of time the researcher needs to spend on the interview itself and on traveling (if the interview is conducted face-to-face and not via the telephone), as well as on transcribing the interview.

When conducting in-depth interviews, a set format is usually followed. First, the interview details are discussed, such as confidentiality issues, the interview topic, the structure, and the duration. Moreover, the interviewer should disclose whether the interview is being recorded and inform the interviewee that there is no right or wrong answer, just opinions on the subject. The interviewer should also try to be open and maintain eye contact with the interviewee. Interviewers can end an interview by informing their respondents that they have reached the last question and thanking them for their time.

Interviews are often used to investigate means-end issues, in which researchers try to understand what ends consumers aim to satisfy and which means (consumption) they use to do so. A **means-end approach** involves first determining a product's attributes. These are the functional product features, such as the speed a car can reach or its acceleration. Subsequently, researchers look at the functional consequences that follow from the product benefits. This could be driving fast. The psychosocial consequences, or personal benefits, are derived from the functional benefits and, in this example, could include an enhanced status, or being regarded as successful. Finally, the psychosocial benefits are linked to people's personal values or life goals, such as a desire for success or acceptance. Analyzing and identifying the relationships between these steps is called **laddering**.

4.5.2 Projective Techniques

Projective techniques describe a special type of testing procedure, usually used as part of in-depth interviews. These techniques provide the participants with a stimulus and then gauge their responses. Although participants in projective techniques know that they are participating in a market research study, they may not be aware of the research's specific purpose. The stimuli that the projective techniques provide are ambiguous and require a response from the participants. *Sentence completion* is a key form of projective techniques.

An iPhone user is someone who: .....
The Apple brand makes me think of: .....
iPhones are most liked by: .....

In this example, the respondents are asked to express their feelings, ideas, and opinions in a free format.

Projective techniques' advantage is that they allow for responses when people are unlikely to respond if they were to know the study's exact purpose. Thus, projective techniques can overcome self-censoring and allow expression and fantasy. In addition, they can change a participant's perspective. Think of the previous example. If the participants are iPhone users, the sentence completion example asks them how they think other people regard them, not what they think of the iPhone. A drawback is that projective techniques require the responses to be interpreted and coded, which can be difficult.

4.5.3 Focus Groups

Focus groups are interviews conducted with a number of respondents at the same time and led by a moderator. This moderator leads the interview, structures it, and often plays a central role in transcribing the interview later. Focus groups are usually semi or highly structured. The group usually comprises between 6 and 10 people to allow for interaction between the participants and to ensure that all the participants have a say. The duration of a focus group interview varies, but is often between 30 and 90 minutes for company employee focus groups and between 60 to 120 minutes for consumers. When focus groups are held with company employees, moderators usually travel to the company and conduct their focus group in a room. When consumers are involved, moderators often travel to a market research company, or hotel, where the focus group meets in a conference room. Market research companies often have specially equipped conference rooms with, for example, one-way mirrors, built-in microphones, and video recording devices.

Learn more about focus groups in this TED-Ed video.

© ALotOfPeople/Getty Images/iStock
https://www.youtube.com/watch?v=3TwgVQIZPsw

How are focus groups structured? They usually start with the moderator introducing the topic and discussing the background. Everyone in the group is introduced to all the others to establish rapport. Subsequently, the moderator encourages the members of the focus group to speak to one another, instead of asking the moderator for confirmation. Once the focus group members start discussing topics with one another, the moderator tries to stay in the background, merely ensuring that the discussions stay on-topic. Afterwards, the participants are briefed and the discussions are transcribed for further analysis.

Focus groups have distinct advantages: they are relatively cheap compared to in-depth interviews, work well with issues that are socially important, or which require spontaneity. They are also useful for developing new ideas. On the downside, focus groups do not offer

Table 4.2 Comparing focus groups and in-depth interviews

	Focus groups	In-depth interviews
Group interactions	Group interaction, which may stimulate the respondents to produce new thoughts.	There is no group interaction. The interviewer is responsible for stimulating the respondents to produce new ideas.
Group/peer pressure	Group pressure and stimulation may clarify and challenge thinking.	In the absence of group pressure, the respondents' thinking is not challenged.
	Peer pressure and role playing.	With just one respondent, role playing is minimized and there is no peer pressure.
Respondent competition	Respondents compete with one another for time to talk. There is less time to obtain in-depth details from each participant.	Individuals are alone with the interviewer and can express their thoughts in a non-competitive environment. There is more time to obtain detailed information.
Peer influence	Responses in a group may be biased by other group members' opinions.	With one respondent, there is no potential of other respondents influencing this person.
Subject sensitivity	If the subject is sensitive, respondents may be hesitant to talk freely in the presence of other people.	If the subject is sensitive, respondents may be more likely to talk.
Stimuli	The volume of stimulus materials that can be used is somewhat limited.	A fairly large amount of stimulus material can be used.
Interviewer schedule	It may be difficult to assemble between 6 and 10 respondents if they are a difficult type to recruit (such as busy executives).	Individual interviews are easier to schedule.

the same opportunity for probing as interviews do, and also run a greater risk of going off-topic. Moreover, a few focus group members may dominate the discussion and, especially in larger focus groups, "voting" behavior may occur, hindering real discussions and the development of new ideas. Table 4.2 summarizes the key differences between focus groups and in-depth interviews.

4.6 Collecting Primary Data through Experimental Research

In ► Chap. 2, we discussed *causal research* and briefly introduced experiments as a means of conducting research. The goal of experiments is to avoid unintended influences through the use of randomization. Experiments are typically conducted by manipulating one variable, or a few, at a time. For example, we can change the price of a product, the type of product, or the package size to determine whether these changes affect important outcomes such as attitudes, satisfaction, or intentions. Simple field observations are often

unable to establish these relationships, as inferring causality is problematic. Imagine a company that wants to introduce a new type of soft drink aimed at health-conscious consumers. If the product were to fail, the managers would probably conclude that the consumers did not like the product. However, many (usually unobserved) variables, such as competitors' price cuts, changing health concerns, and a lack of availability, can also influence new products' success.

4

4.6.1 Principles of Experimental Research

Experiments deliberately impose one or more **treatments** and then observe the outcome of a specific treatment (Mitchell and Jolley 2013). This way, experiments attempt to isolate how one change affects an outcome. The outcome(s) is (are) the dependent variable(s) and the independent variable(s) (also referred to as factors) are used to explain the outcomes. To examine the influence of the independent variable(s) on the dependent variable(s), the participants are subjected to treatments. These are supposed to manipulate the participants by putting them in different situations. A simple form of treatment could be an advertisement with and without humor. In this case, the humor is the independent variable, which can take two levels (i.e., with or without humor). If we manipulate, for example, the price between low, medium, and high, we have three levels. When selecting independent variables, we normally include those that marketers care about and which are related to the marketing and design of the products and services. To assure that the participants do indeed perceive differences, **manipulation checks** are conducted. For example, if two different messages with and without humor are used as stimuli, we could ask the respondents which of the two is more humorous. Such manipulation checks help establish an experiment's validity.

Experiments can be incredibly useful but conducting experiments without user consent can result in a consumer backlash, ethics concerns, or may even be illegal as Facebook found out.

© bombuscreative/Getty Images/iStock
https://www.youtube.com/watch?v=VmQAdsCSonA

Care should be taken not to include too many of these variables in order to keep the experiment manageable. An experiment that includes four independent variables, each of which has three levels and includes every possible combination (called a *full factorial design*) requires $4^3 = 64$ treatments. Large numbers of levels (5 or more) will dramatically increase the complexity and cost of the research. Finally, *extraneous variables*, such as the age or income of the experiment participant, are not changed as part of the experiment. However, it might be important to control for their influence when setting up the experimental design.

Experiments can run in either a lab or field environment. *Lab experiments* are performed in controlled environments (usually in a company or academic lab), thereby allowing for isolating the effects of one or more variables on a certain outcome. Extraneous variables' influence can also be controlled for. Lab experiments stand out in terms of **internal validity**, which is the extent to which we can make causal claims based on the study. However, as they take place in controlled experimental settings, which typically ignore real-world business conditions, they therefore usually lack **external validity**, which is the extent to which the study results can be generalized to similar settings. *Field experiments*, on the other hand, are performed in natural environments (e.g., in a supermarket or in home) and, as such, generally exhibit high degrees of external validity. However, since controlling for extraneous variables' impact is difficult in natural environments, field experiments usually lack internal validity (Gneezy 2017).

4.6.2 Experimental Designs

Experimental design refers to an experiment's structure. There are various types of experimental designs. To clearly separate the different experimental designs, researchers have developed the following notation:

O:	A formal observation or measurement of the dependent variable. Subscripts below an observation O such as 1 or 2, indicate measurements at different points in time.
X:	The test participants' exposure to an experimental treatment.
R:	The random assignment of participants. Randomization ensures control over extraneous variables and increases the experiment's reliability and validity.

In the following, we will discuss the most prominent experimental designs:

- One-shot case study,
- before-after design,
- before-after design with a control group, and
- Solomon four-group design.

The simplest form of experiment is the *one-shot case study*. This type of experiment is structured as follows:[9]

9 If one symbol precedes another, it means that the first symbol precedes the next one in time.

X	O_1

This means we have one treatment (indicated by X), such as a new advertising campaign. After the treatment, we await reactions to it and then measure the outcome of the manipulation (indicated by O_1), such as the participants' willingness to purchase the advertised product. This type of experimental set-up is common, but does not tell us if the effect is causal. One reason for this is that we did not measure anything before the treatment and therefore cannot assess what the relationship between the treatment and outcome is. The participants' willingness to purchase the product was perhaps higher before they were shown the advertisement, but since we did not measure their willingness to purchase before the treatment, this cannot be ruled out. Consequently, this design does not allow us to establish causality.

The simplest type of experiment that allows us to make causal inferences—within certain limits—is the *before-after design* used for one group. The notation for this design is:

O_1	X	O_2

We have one measurement before (O_1) and one after a treatment (O_2). Thus, this type of design can be used to determine whether an advertisement has a positive, negative, or no effect on the participants' willingness to purchase a product.

A problem with this type of design is that we do not have a standard of comparison with which to contrast the change between O_1 and O_2. While the advertisement may have increased the participants' willingness to purchase, this might have been even higher if they had not seen the advertisement. The reason for this is that the before-after-design does not control for influences occurring between the two measurements. For example, negative publicity after the initial measurement could influence the subsequent measurement. These issues make the "real" effect of the advertisement hard to assess.

If we want to have a much better chance of identifying the "real" effect of a treatment, we need a more complex setup, called the *before-after design with a control group*. In this design, we add a control group who is not subjected to the treatment X. The notation of this type of experiment is:

Experimental group (R)	O_1	X	O_2
Control group (R)	O_3		O_4

The effect attributed to the experiment is the difference between O_1 and O_2 minus the difference between O_3 and O_4. For example, if the participants' willingness to purchase increases much stronger in the experimental group (i.e., O_2 is much higher than O_1) than in the control group (i.e., O_4 is slightly higher than or equal to O_3), the advertisement has had an impact on the participants.

The random assignment of the participants to the experimental and the control groups (indicated by R), is an important element of this experimental design. This means that, for any given treatment, every participant has an equal probability of being chosen for one of the two groups. This ensures that participants with different characteristics

are spread randomly (and, it is hoped, equally) between the treatment(s), which will neutralize *self-selection*. Self-selection occurs when participants can select themselves into either the experimental or the control group. For example, if participants who like sweets participate in a cookie tasting test, they will certainly try to give the treatment (cookie) a try! See Mooi and Gilliland (2013) for an example of self-selection and an analysis method.

However, the before-after experiment with a control group does have limitations. The initial measurement O_1 may alert the participants that they are being studied, which may bias the post measurement O_2. This effect is also referred to as the *before measurement effect*, or the *testing effect* (Campbell and Stanley 1966). Likewise, the initial measurement O_1 may incite the participants to drop out of the experiment, leading to no recorded response for O_2. If there is a systematic reason for the participants to drop out, this will threaten the experiment's validity.

The *Solomon four-group design* is an experimental design accounting for before measurement effects and is structured as follows (Campbell and Stanley 1966):

Experimental group 1 (R)	O_1	X	O_2
Control group 1 (R)	O_3		O_4
Experimental group 2 (R)		X	O_5
Control group 2 (R)			O_6

The design is much more complex, because we need to measure the effects six times and administer two treatments. This method provides an opportunity to control for the before measurement effect of O_1 on O_2. The design also provides several measures of the treatment's effect (i.e., $(O_2–O_4)$, $(O_2–O_1)–(O_4–O_3)$, and $(O_6–O_5)$). If these measures agree, the inferences about the treatment's effect are much stronger.

In ► Chap. 6, we discuss how to analyze experimental data using ANOVA and various other tests.

4.7 Oddjob Airways (Case Study)

Case Study

Oddjob Airways

Founded in 1962 by the Korean businessman Toshiyuki Sakata, Oddjob Airways is a small premium airline, mainly operating in Europe, but also offering flights to the US.[10] The company regularly conducts market research in order to monitor its service quality and to improve customer satisfaction. The market research department has uploaded a series of videos on https://www.oddjobairways.com/market-research/, which document some of the recent projects. Please help the department by answering the following questions:

1. Go to https://www.oddjobairways.com/market-research/interviews/ and watch the three videos. Try to identify all errors that the interviewer made. You can find the solutions under „Corrections."
2. Under https://www.oddjobairways.com/market-research/experiments/, you will find two videos, which show safety videos with male and female voice-overs. Carry out an experiment, which contrasts viewers' perceptions of the safety video (e.g., in terms of satisfaction) when the voice-over is male vs. female. Compute the mean values (► Chap. 5) for each group. In a further step, carry out independent samples *t*-tests (► Chap. 6) to check for significant differences between the two groups.
3. Go to https://www.oddjobairways.com/market-research/focus-group/ and watch the video on focus groups. Would you have spotted all errors?

4.8 Review Questions

1. What is the difference between primary and secondary data? Can primary data become secondary data?
2. Please search for and download two examples of secondary data sources found on the Internet. Discuss two potential market research questions that each dataset can answer.
3. Imagine you are asked to understand what consumer characteristics make these consumers likely to buy a BMW i3 (http://www.bmw.com). How would you collect the data? Would you start with secondary data, or would you start collecting primary data directly? Do you think it is appropriate to collect qualitative data? If so, at what stage of the process should you do so?
4. What are the different reasons for choosing interviews rather than focus groups? What choice would you make if you want to understand CEOs' perceptions of the economy, and what would seem appropriate when you want to understand how consumers feel about a newly introduced TV program?
5. Consider the following examples of survey items relating to how satisfied iPhone users are with their performance, reliability, and after-sales service. Please assess their adequacy and make suggestions on how to revise the items, if necessary.

10 All chapters dealing with research methods (i.e., Chapters 5-9) also draw on the Oddjob Airways case study. We introduce the case study in greater detail in section 5.8.

	Strongly disagree	Somewhat disagree	Some-what agree	Agree	Completely agree
I am satisfied with the performance and reliability of my iPhone	☐	☐	☐	☐	☐

	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Completely agree
I am satisfied with the after-sales service of my iPhone	☐	☐	☐	☐	☐

Which of the following iPhone features do you find most important?	Not at all important	Not important	Neutral	Important	Very important
Camera	☐	☐	☐	☐	☐
Music player	☐	☐	☐	☐	☐
App store	☐	☐	☐	☐	☐
Web browser	☐	☐	☐	☐	☐
Mail	☐	☐	☐	☐	☐

6. Make recommendations regarding the following aspects of scale design: the number of response categories, the design and labeling of response categories, and the inclusion of a "don't know" option.
7. Describe the Solomon four-group design and explain which of the simpler experimental design problems it controls for.
8. If you were to set up an experiment to ascertain what type of product package (new or existing) customers prefer, what type of experiment would you choose? Please discuss.

References

Barnham, C. (2015). Quantitative and qualitative research: Perceptual foundations. *International Journal of Market Research*, *57*(6), 837–854.

Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156.

Brace, I. (2004). *Questionnaire design. How to plan, structure and write survey material for effective market research*. London: Kogan Page.
Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*(3), 185–216.
Bronner, F., & Ton, K. (2007). The live or digital interviewer. A comparison between CASI, CAPI and CATI with respect to differences in response behaviour. *International Journal of Market Research*, *49*(2), 167–190.
Brosnan, K., Grpn, B., & Dolnicar, S. (2017). PC, phone or tablet? Use, preference and completion rates for web surveys. *International Journal of Market Research*, *59*(1), 35–55.
Cabooter, E., Weijters, B., Geuens, M., & Vermeir, E. (2016). Scale format effects on response option interpretation and use. *Journal of Business Research*, *69*(7), 2574–2584.
Casteleyn, J., André, M., & Kris, R. (2009). How to use facebook in your market research. *International Journal of Market Research*, *51*(4), 439–447.
DeMonaco, H. J., Ayfer, A., & Hippel, E. V. (2005). The major role of clinicians in the discovery of off-label drug therapies. *Pharmacotherapy*, *26*(3), 323–332.
Deutskens, E., De Jong, A., De Ruyter, K., & Martin, W. (2006). Comparing the generalizability of online and mail surveys in cross-national service quality research. *Marketing Letters*, *17*(2), 119–136.
Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Hoboken, NJ: John Wiley & Sons.
Fink, A. (2003) *How to ask survey questions*. Thousand Oaks, CA: Sage.
Foddy, W. (1993). *Constructing questions for interviews and questionnaires. Theory and practice in social science research*. Cambridge: Cambridge University Press.
Gneezy, A. (2017). Field experimentation in marketing research. *Journal of Marketing Research*, *54*(1), 140–143.
Holbrook, A., Cho, Y. I. K., & Johnson, T. (2006). The impact of question and respondent characteristics on comprehension and mapping difficulties. *Public Opinion Quarterly*, *70*(4), 565–595.
Houston, M. B. (2002). Assessing the validity of secondary data proxies for marketing constructs. *Journal of Business Research*, *57*(2), 154–161.
Kennedy, C., & Everett, S. E. (2011). Use of cognitive shortcuts in landline and cell phone surveys. *Public Opinion Quarterly*, *75*(2), 336–348.
Kurylko, D. T. (2005). Moonraker project seeks marketing savvy for VW. *Automotive News Europe*, *10*(17), 22.
Lietz, P. (2010). Research into questionnaire design. A summary of the literature. *International Journal of Market Research*, *52*(2), 249–272.
Liu, M., Lee, S., & Conrad, F. G. (2016). Comparing extreme response styles between agree-disagree and item-specific scales. *Public Opinion Quarterly*, *79*(4), 952–975.
Lynn, P., & Kaminska, O. (2013). The impact of mobile phones on survey measurement error. *Public Opinion Quarterly*, *77*(2), 586–605.
Mooi, E., & Gilliland, D. I. (2013). How contracts and enforcement explain transaction outcomes. *International Journal of Research in Marketing*, *30*(4), 395–405.
Mitchell, M. L., & Jolley, J. M. (2013). *Research design explained* (8th ed.). Belmont, CA: Wadsworth.
Nowlis, S. M., Kahn, B. E., & Dhar, R. (2002). Coping with ambivalence: The effect of removing a neutral option on consumer attitude and preference judgments. *Journal of Consumer Research*, *29*(3), 319–334.
Oppenheim, A. N. (1992). *Questionnaire design, Interviewing and attitude measurement*. London: Pinter.
Peng, L., & Finn, A. (2016). Assessing response format effects on the scaling of marketing stimuli. *International Journal of Market Research*, *58*(4), 595–619.
Peterson, R. A. (1997). A quantitative analysis of rating-scale response variability. *Marketing Letters*, *8*(1), 9–21.
Raithel, S., Sarstedt, M., Scharf, S., & Schwaiger, M. (2012). On the value relevance of customer satisfaction. Multiple drivers in multiple markets. *Journal of the Academy of Marketing Science*, *40*(4), 509–525.
Rammstedt, B., & Krebs, D. (2007) Does response scale format affect the answering of personality scales? *European Journal of Psychological Assessment*, *23*(1), 32–38.
Reips, U.-D., & Funke, F. (2008). Interval-level measurement with visual analogue scales in Internet-based research: VAS generator. *Behavior Research Methods*, *40*(3), 699–704.

Revilla, M. (2015). Effect of using different labels for the scales in a web survey. *International Journal of Market Research*, 57(2), 225–238.
Rindfleisch, A., & Heide, J. B. (1997). Transaction cost analysis: Past, present, and future applications. *Journal of Marketing*, *61*(4), 30–54.
Sarstedt, M., Diamantopoulos, A., Salzberger, T., & Baumgartner, P. (2016). Selecting single items to measure doubly concrete constructs: A cautionary tale. *Journal of Business Research*, *69*(8), 3159-3167.
Schilling, M. A. (2009). Understanding the alliance data. *Strategic Management Journal*, *30*(3), 233–260.
Stern, M. J., Bilgen, I., & Dillman, D. A. (2014). The state of survey methodology challenges, dilemmas, and new frontiers in the era of the tailored design, *Field Methods*, 26(3), 284–301.
Stieglitz, S., Dang-Xuan, L., Bruns, A., & Neuberger, C. (2014). Social media analytics, An interdisciplinary approach and its implications for information systems. *Business & Information Systems Engineering*, *6*(2), 89–96.
Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, *45*(1), 116–131.
Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: the impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, *60*(2), 275–304.
Vesta Research. (2016). Rules of thumb for survey length. *Vesta Research Blog*. Available at: http://www.verstaresearch.com/blog/rules-of-thumb-for-survey-length/
Vicente, P., & Reis, E. (2010). Marketing research with telephone surveys: Is it time to change? *Journal of Global Marketing*, *23*(4), 321–332.
Vincente, P., Reis, E., & Santos, M. (2008). Using mobile phones for survey research. *International Journal of Market Research*, *51*(5), 613–633.
Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research*, *49*(5), 737–747.
Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247.
Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods*, *18*(3), 320–334.
Weng, L.-J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, *64*(6), 956–972.
Winkler, T. J., Sarstedt, M., Keil, M., & Rost, P. (2015). Selfsurvey.org: A platform for prediction-based benchmarking and feedback-enabled survey research. Proceedings of the European Conference on Information Systems, Paper 204, Münster, Germany.

Further Reading

Barnham, C. (2015). Quantitative and qualitative research: Perceptual foundations. *International Journal of Market Research*, *57*(6), 837–854.
Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Wadsworth Publishing.
Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Wadsworth Publishing.
Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Hoboken, NJ: John Wiley & Sons.
FocusGroupTips.com.
Gideon, L. (2012). The art of questionnaire phrasing. In: L. Gideon (Ed.), *Handbook of survey methodology for the social sciences* (pp. 91–108). New York, NJ: Springer.
Hulland, J., Baumgartner, H., & Smith, K.M. (2018). Marketing survey research best practices: Evidence and recommendations from a review of *JAMS* articles. *Journal of the Academy of Marketing Science*, *46*(1), 92–108.
Lichters, M., Sarstedt, M., & Vogt, B. (2015). On the practical relevance of the attraction effect: a cautionary note and guidelines for context effect experiments. *AMS Review*, *5*(1-2), 1–19.
Lietz, P. (2010). Current state of the art in questionnaire design: a review of the literature. *International Journal of Market Research*, *52*(2), 249–272.

Morales, A. C., Amir, O., & Lee, L. (2017). Keeping it real in experimental research. Understanding when, where, and how to enhance realism and measure consumer behavior. *Journal of Consumer Research*, *44*(2), 465–476.

Mystery Shopping Providers Association (www.mysteryshop.org).

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632.

Smyth, J.D., Olson, K., & Burke, A. (2018). Comparing survey ranking question formats in mail surveys. *International Journal of Market Research*, forthcoming.

Veludo-de-Oliveira, T. M., Ikeda, A. A., & Campomar, M. C. (2006). Laddering in the practice of marketing research: Barriers and solutions. *Qualitative Market Research: An International Journal*, *9*(3), 297–306.