5 How to Design and Analyze Surveys in Second Language Acquisition Research

Zoltán Dörnyei and Kata Csizér

The aim of this chapter is to describe how survey studies are carried out in the context of SLA research. After a general introduction, we explain the required steps to design a survey that can provide valid and reliable data. Then we look at the main aspects of quantitative data analysis to be applied to questionnaire data, followed by a discussion of reporting survey results. Finally, we present an illustration of how the various principles of questionnaire design have been put into actual practice, with information on further resources that might be helpful in planning to conduct a questionnaire survey.

Background to the Questionnaire Surveys

Survey research is a quantitative research method which aims to collect self-report data from individuals, and the typical instrument used for this purpose is the written questionnaire (although market researchers, for example, often use structured interviews as well). Both survey methodology and questionnaire design have their origins in the social sciences. The basic idea behind survey research is the recognition that the characteristics, opinions, attitudes, and intended behaviors of a large population (e.g., second language (L2) learners in a country) can be described and analyzed on the basis of questioning only a fraction of the particular population. The development of survey methodology for research purposes has gone hand in hand with political public opinion research, as survey results about people's political preferences have an obvious link to actual election results (Babbie, 2007). However, surveys of large populations have also been employed in many other fields of study (e.g., sociology, psychology, education, and market research), and questionnaire

Research Methods in Second Language Acquisition: A Practical Guide, First Edition.
Edited by Alison Mackey and Susan M. Gass.
© 2012 Blackwell Publishing Ltd. Published 2012 by Blackwell Publishing Ltd.

UNCORRECTED PROOF

surveys have made a substantial contribution to second language acquisition (SLA) research as well (for detailed summaries, see Brown, 2001; Dörnyei, 2010). Survey studies can inform us about:

- language learners' intended language behavior, that is, how students plan to respond to certain language learning situations (e.g., how much effort they are willing to invest in L2 learning);
- people's opinions and attitudes concerning specific L2s and the language learning process in general (e.g., how much they like certain aspects of learning a foreign language);
- participants' feelings (e.g., anxiety about language use) and beliefs about certain L2-related issues (e.g., the optimal age or method of learning an L2);
- learners' knowledge of certain issues in SLA.(e.g., their awareness of different varieties of English);
- various background information and biodata from the students (e.g., facts about their language learning history).

In sum, surveys can target a wide variety of language-related issues and allow researchers to make inferences about larger L2 learning populations; this obviously facilitates decision making and policy formation in an informed and principled manner.

How to Design a Questionnaire

The backbone of any survey study is the instrument used for collecting data. The most common way of obtaining large amounts of data in a relatively short period of time in a cost-effective way is by means of standardized questionnaires. Questionnaire design requires a rigorous process if we want to produce an instrument that yields reliable and valid data and, accordingly, whole volumes have been written on how to construct instruments of good quality (for a review, see Dörnyei, 2010). In the following we discuss what we see as the six key design issues.

Issue 1: The Sampling of Questionnaire Content and the Use of "Multi-Item Scales"

The first step in preparing questionnaire items is to specify their content in explicit terms. Although this may sound obvious, it does not always happen, and vague content specifications can pose a serious threat to the validity and reliability of the instrument, particularly in two areas: (a) the *sampling of content* and (b) the preparation of "*multi-item scales*."

- *Appropriate sampling of content*: Ad hoc questionnaire design involves jotting down a few seemingly relevant questions without any rigorous procedure to ensure that the coverage is comprehensive. The problem with

UNCORRECTED PROOF

this method, as Davidson (1996, p. 10) highlights, is that "You cannot analyze what you do not measure." That is, not even the most sophisticated data analysis techniques will be able to compensate for leaving out some important questions from the data collection by accident. Certain omissions are bound to occur even in otherwise very thorough studies (as attested by the anecdotes one hears at professional conferences), but when the sampling of the content is not theory-driven, the chances of something irrecoverable happening are obviously much greater. Here are three suggestions to help to ensure appropriate content sampling. (a) Carefully clarify the research problem and identify the critical concepts that are likely to play a defining role in shaping the issue in question – these variables will need to be addressed by the questionnaire. (b) Eliminate all the questions that are only of peripheral interest but not directly related to the variables and hypotheses that the questionnaire has been designed to investigate. (c) Avoid making the questionnaire too long by covering every possible angle; focus on the key issues.

- *Using "multi-item scales"*: The notion of multi-item scales is the central component in scientific questionnaire design, yet this concept is surprisingly little known in the L2 profession. The core of the issue is that when it comes to assessing abstract, mental variables not readily observable by direct means (e.g., attitudes, beliefs, etc.), the actual wording of the questions assumes an unexpected amount of importance: minor differences in how a question is formulated and framed can produce radically different levels of agreement or disagreement, or a completely different selection of answers (Gillham, 2008). Because of the fallibility of single items, there is a general consensus among survey specialists that more than one item is needed to address each identified content area, all aimed at the same target but drawing upon slightly different aspects of it. How many is "more than one"? Professional scales often contain as many as 10–20 items focusing on a target issue, but even if we want to shorten the scales to be able to target more issues in the questionnaire it is risky to go below four items per subdomain, because if the post hoc item analysis (see below for details) reveals that certain items did not work in the particular sample, their exclusion will result in too short (or single-item) scales.

Issue 2: Main Types of Questionnaire Items

The typical questionnaire is a highly structured data collection instrument, with most items either asking about very specific pieces of information or giving various response options for the respondent to choose from, for example by ticking a box or circling the most appropriate option. Most professional questionnaires are primarily made up of "closed-ended" items, which do not require the respondents to produce any free writing. The most famous type of closed-ended item is undoubtedly the *Likert scale*, which consists of a characteristic statement accompanied by five or six response options for respondents to indicate the extent to which they "agree" or

UNCORRECTED PROOF

"disagree" with it by marking (e.g., circling) one of the responses ranging from "strongly agree" to "strongly disagree." For example:

Applied linguists are genuinely nice people.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

Another frequently applied way of eliciting a graduated response is the *semantic differential scale*, in which respondents are asked to indicate their answers by marking a continuum (with a tick or an "X") between two bipolar adjectives at the extremes. For example:

Listening comprehension tasks are:
difficult ___:___:___:___:___:_X_:___ easy
useless ___:_X_:___:___:___:___:___useful

Finally, *numerical rating scales* involve giving "so many marks out of so many" (e.g., five points to applied linguists on a scale from one to five for being nice people; see the sample questionnaire near the end of the chapter for an illustration).

Issue 3: Writing Items that Work

Over the past 50 years, survey researchers have accumulated a considerable body of knowledge and experience about what makes a questionnaire item good and what the potential pitfalls are. However, most specialists also emphasize that item construction is not a 100% scientific activity, because in order to write good questions one also needs a certain amount of creativity and lots of common sense. Indeed, it is generally recommended that when we get down to writing the actual items we should let our imagination go free and should try and create as many potential items as we can think of – the resulting collection of items is referred to as the *item pool* (DeVellis, 2003). During the generation of the item pool, item designers can draw on two sources in addition to their own verbal creativity:

- *Qualitative, exploratory data* gathered from respondents, such as *notes* taken during talks and brainstorming in focus or discussion groups; recorded unstructured/semi-structured *interviews*; and *student essays* written around the subject of the inquiry (see e.g., Tseng, Dörnyei, & Schmitt, 2006).
- *Borrowed questions* from established questionnaires. Questions that have been used frequently before must have been through extensive piloting and therefore have a certain "track record." Of course, we will need to acknowledge the sources precisely, and it is important to note that even if we adopt most items from existing instruments, our questionnaire will still need to be piloted for the specific population that we intend to use it for.

UNCORRECTED PROOF

What are the main rules about writing good items? Here are five key strategies for producing items that work:

- *Aim for short and simple items*. Whenever possible, questionnaire items should be short, rarely exceeding 20 words.
- *Use simple and natural language*. As a rule, in questionnaire items we should always choose the simplest way to say something.
- *Avoid ambiguous or loaded words and sentences*. Any element that might make the language of the items unclear, ambiguous, or emotionally loaded needs to be avoided.
- *Avoid negative constructions*. Items that contain a negative construction (i.e., including "not," "doesn't," or "don't") are deceptive because, although they read OK, responding to them – especially giving a negative response – can be problematic.
- *Avoid double-barreled questions*. Double-barreled questions are those that ask two (or more) questions in one, while expecting a single answer (e.g., *Is the relationship with your parents good?*).

Issue 4: The Format of the Questionnaire

The format and layout of the questionnaire are frequently overlooked as an important aspect of the development of the instrument. This is a mistake because producing an attractive and professional design is half the battle in motivating respondents to produce reliable and valid data. Here are some points to consider:

- *Length*: The optimal length of a questionnaire depends on how important the topic is for the respondent (if we feel very strongly about something, we are usually willing to spend longer answering questions). However, most researchers agree that anything that is more than 4–6 pages long and requires over half an hour to complete is likely to be considered too much of an imposition. So a good rule of thumb is to stay within a four-page limit, which tends not to exceed the 30-minute completion limit.
- *Space economy*: We want to make the pages full because respondents are much more willing to fill in a two-page rather than a four-page questionnaire even if the two instruments have exactly the same number of items. However, we must not make the pages look crowded by economizing on the spaces separating different sections of the questionnaire. Effective ways of achieving this trade-off involve reducing the margins, using a space-economical font, and utilizing the whole width of the page, for example by printing the response options next to each question rather than below it.
- *Mixing up the scales and items*: The items from different scales need to be mixed up as much as possible to create a sense of variety and to prevent respondents from simply repeating previous answers.
- Factual (or "personal") questions at the end: Starting the questionnaire with a rather forbidding set of personal background questions (as in passport application forms) is offputting and may also ring "privacy alarm bells" in the students – such questions are best left to the end of the questionnaire.

UNCORRECTED PROOF

Issue 5: Translating the Questionnaire

The issue of how to translate questionnaires from one language to another has typically been marginalized and treated as an addendum in questionnaire design, whereas translating questionnaires as a practice is surprisingly common, due to the frequency of multinational research teams (including supervisor–research-student teams) and the widespread – and we believe correct – belief that the quality of the obtained data improves if the questionnaire is presented in the respondents' own mother tongue (for an overview, see Harkness, 2008).

The main challenge in translating a questionnaire is to reconcile two somewhat contradictory criteria: (a) the need to produce a close translation of the original text so that we can claim that the two versions are equivalent, and (b) the need to produce natural-sounding texts in the target language. For most parts of the questionnaire we are likely to find easy solutions to this challenge, but there will be a few places where a close or literal translation will not express the real meaning and the pragmatic function of the text well. This is a point where team-based brainstorming and negotiation are particularly useful, and even in small-scale projects we should make an effort to recruit some competent help to deal with these problem issues.

After the initial translation is completed, it is necessary to ensure the equivalence of the two versions. We have two basic options: to consult bilingual external reviewers or to recruit an independent translator to back-translate the target language version into the source language (Brislin, 1970).

Issue 6: Piloting the Questionnaire

Piloting the questionnaire involves administering the instrument to a sample of participants who are similar to the target group of people for whom it has been designed. The results of the pilot study are invaluable in helping the researchers to (a) fine-tune the final version of the questionnaire in order to eliminate ambiguous, too difficult/easy, or irrelevant items; (b) improve the clarity of the item wordings and the instructions; (c) finalize the layout; (d) rehearse the administration procedures; (e) dry run the analysis in order to see whether the expected findings will potentially emerge from the data; (f) time the completion of the questionnaire; and (g) generally double-check that there are no mistakes left in the instrument.

The first stage of piloting usually involves assessing the item pool by carrying out a think-aloud protocol with three or four people (usually friends, colleagues, or family) who answer the items and provide detailed feedback. Based on their responses, we can normally put together a near-final version of the questionnaire, which is then tried out with 50–100 participants who are in every way similar to the target population the instrument is designed for. The obtained data is then submitted to item analysis, which usually involves checking three aspects of the response pattern:

UNCORRECTED PROOF

- *Missing responses* and possible signs that the instructions were not understood correctly. If some items are left out by several respondents, that should serve as an indication that something is not right.
- The *range of the responses* elicited by each item. We should avoid including items that are endorsed by almost everyone or by almost no one, because they are difficult if not impossible to process statistically.
- The *internal consistency* of the multi-item scales. Multi-item scales are only effective if the items within them work in concert, that is, if they measure the same target area. To check the coherence of each scale and to identify items that do not fit in with the others, researchers usually conduct reliability analyses (see below).

Sampling and Data Collection

The most frequent question asked by novice researchers who are planning to use questionnaires in their investigation is "How many people do I need to survey?" In measurement terms this question can be formulated as "How large should my sample be?" And a second question to follow is "What sort of people shall I select?" Or, in other words, "Whom shall my sample consist of?" Let us start answering these key questions by first looking at the second issue, the principles of quantitative sampling.

Issue 1: Sampling Procedures

The *sample* is the group of people whom the researcher actually examines and the *population* is the larger group of people whom the survey is about. That is, the target population of a study consists of all the people to whom the survey's findings are to be applied or generalized. For example, the population in a study might be English as a foreign language (EFL) learners in Taiwanese secondary schools and the actual sample might involve three Taiwanese secondary classes.

The main point of sampling is to save resources. We could, in principle, survey the whole population – as the census does – but by adopting appropriate *sampling procedures* to select a smaller number of people to be questioned we can still come up with accurate results at a much lower cost (as demonstrated by opinion polls). To achieve this, we need to choose a sample that is similar to the target population in its most important general characteristics (e.g., age, gender, ethnicity, educational background, academic capability, social class, socioeconomic status, etc.) as well as in all the more specific features that are known to be significantly related to the items included on the questionnaire (e.g., L2 learning background or the amount and type of L2 instruction received). That is, the sample needs to be *representative* of the whole population, and various sampling procedures have been developed to ensure this representativeness.

Broadly speaking, sampling strategies can be divided into two groups: (a) scientifically sound "probability sampling," which involves complex and

UNCORRECTED PROOF

expensive procedures that provide a truly representative sample; and (b) "non-probability sampling," which involves a number of strategies that try to achieve a trade-off, that is, a reasonably representative sample while using resources that are within the means of the ordinary researcher. Because probability sampling is typically beyond the means of most applied linguists, we will introduce "random sampling", the crucial element of this approach, only briefly before we describe several non-probability sampling procedures.

- *Random sampling* involves the selection of members of the population to be included in the sample on a completely random basis, a little bit like drawing numbers from a hat. In this way the selection is based entirely on chance rather than on any extraneous or subjective factors. As a result, a sufficiently large sample is generally believed to contain subjects whose characteristics are similar to the population as a whole. Combining random sampling with some form of rational/purposeful grouping is a particularly effective method for research with a specific focus: in "stratified random sampling" the population is divided into groups, or "strata," and a random sample of a proportionate size is selected from each group.
- *Convenience or opportunity sampling* is the most common non-probability sampling type in L2 research, where an important criterion of sample selection is the convenience to and resources of the researcher. Members of the target population are selected only if they meet certain practical criteria, such as geographical proximity, availability at a certain time, or easy accessibility. Captive audiences such as students in the researcher's own institution are prime examples of convenience samples. To be fair, convenience samples are rarely completely convenience-based but are usually partially purposeful, which means that besides the relative ease of accessibility, participants also have to possess certain key characteristics that are related to the purpose of the investigation.
- *Snowball sampling* involves a "chain reaction," whereby the researcher identifies a few people who meet the criteria of a particular study and then asks these participants to identify appropriate further members of the population. This technique is useful when studying groups whose membership is not readily identifiable (e.g., teenage gang members or particularly test-anxious learners).
- In *quota sampling* the researcher defines certain distinct subgroups (e.g., boys/girls or age cohorts) and determines the proportion of the population that belongs to each of these subgroups (e.g., when targeting language teachers, determining that the female–male ratio among them is 70:30 in a particular setting). The actual sample, then, is selected in a way as to reflect these proportions (i.e., 70% of the teacher sample will be women). Thus, quota sampling is similar to stratified random sampling without the "random" element.

We must not forget, however, that no matter how principled a non-probability sample strives to be, the extent of generalizability in this type of sample is often negligible. Therefore, we need to describe in sufficient detail the limitations of such

UNCORRECTED PROOF

samples when we report the results, while also highlighting the characteristics that the particular sample shares with the defined target population.

Issue 2: How Large Should the Sample Be?

When researchers ask the question, "How large should the sample be?" what they usually mean is, "How small a sample can I get away with?" Therefore, the often-quoted principle "the larger, the better" is usually rather unhelpful for them. Unfortunately, there are no hard-and-fast rules in setting the optimal sample size; the final answer to the "how large/small?" question should be the outcome of the researcher considering several broad guidelines:

1. In the survey research literature a range of between 1% and 10% of the population is usually mentioned as the "magic" sampling fraction, depending on how careful the selection has been (i.e., the more scientific the sampling procedures applied, the smaller the sample size can be, which is why opinion polls can produce accurate predictions from samples as small as 0.1% of the population).
2. From a purely statistical point of view, a basic requirement is that the results obtained from the sample should have a *normal distribution*, and a rule of thumb to achieve this, offered by Hatch and Lazaraton (1991), is that the sample should include 30 or more people. However, this is not an absolute rule, because smaller sample sizes can be compensated for by using certain special *non-parametric statistical procedures* (see Dörnyei, 2007).
3. From the perspective of *statistical significance*, the principal concern is to sample enough learners for the expected results to be able to reach significance. Because in L2 studies meaningful correlations reported in journal articles have often been as low as 0.30 and 0.40, a good rule of thumb is that we need around 50 participants to make sure that these coefficients are significant and thus we do not lose potentially important results. However, certain multivariate statistical procedures require more than 50 participants; for factor analysis or structural equation modeling, for example, we need a minimum of 100 (but preferably more) subjects.
4. A further important consideration is whether there are any distinct subgroups within the sample which may be expected to behave differently from the others. If we can identify such subgroups in advance (e.g., in most L2 studies of schoolchildren, girls have been found to perform differently from boys), we should set the sample size so that the minimum size applies to the *smallest subgroup* to allow for effective statistical procedures.
5. When setting the final sample size, it is advisable to leave a decent *margin* to provide for unforeseen or unplanned circumstances. For example, some participants are likely to drop out of at least some phases of the project; some questionnaires will always have to be disqualified for one reason or another; and – in relation to point 4 above – we may also detect unexpected subgroups that need to be treated separately.

UNCORRECTED PROOF

Issue 3: Administering the Questionnaire

There is ample evidence that questionnaire administration procedures play a significant role in affecting the quality of the elicited responses. The key question is this: why would the respondents take the survey seriously when they have usually nothing to gain from participating in the research? The answer is that people in general do not mind expressing their opinions and answering questions as long as they think that the particular survey they are invited to participate in is a serious study, related to a worthy cause, and that their opinion matters. Thus, if we take sufficient care planning and executing the administration process, we can successfully build on this human characteristic and can secure the cooperation of our informants (for a range of administration strategies suitable for different questionnaire formats, see Dörnyei, 2010).

How to Analyze Survey Results

After we have designed the questionnaire and administered it to an appropriate sample, we need to process the obtained data. The main stages of this stepwise process are as follows:

Step 1: Preparing the Raw Data to Processing

The first step in processing questionnaire data involves a series of procedures to transform the respondents' markings on the actual questionnaires into a neat data file that contains figures recorded in a way that is appropriate for statistical analysis:

- *Coding questionnaire data*: The respondents' answers are converted to numbers by means of *coding procedures* in order to be able to use the vast arsenal of statistical techniques available for numerical data. With numerical variables such as test scores, the coding is simple, and with closed-ended questionnaire items, such as Likert scales, the process is similarly straightforward (with each response option assigned a consecutive number). For simple open-ended questionnaire items (e.g., some background information), the coding frame is more complex because it can have as many as the number of the different answers in all the questionnaires, and other open-ended questions require an elaborate and principled interpretive scheme.
- *Inputting the data:* First we should create a new data file within a computer program into which the data will be recorded. Next, the data needs to be keyed in – SPSS, which is the most frequently used statistical package in the social sciences, has its own Data Editor screen, which provides a convenient, spreadsheet-like method for creating and editing data files.
- *Data cleaning*: The initial data file will always contain mistakes. Some of these are the result of human error occurring during the data entry phase (e.g., typing

UNCORRECTED PROOF

the wrong number) and some are mistakes made by the respondent when filling in the questionnaire. Data cleaning involves correcting as many of these errors and inaccuracies as possible before the actual analyses are undertaken. This is an indispensable phase of preparing the data because some mistakes can completely distort our results.

- *Data manipulation*: This involves making changes in the dataset prior to the analyses in order to make it more appropriate for certain statistical procedures. One particularly important issue here is to decide how to handle missing data, and another is to recode any negatively worded items.

Step 2: Reducing the Number of Variables in the Questionnaire

The actual analysis of questionnaire data always starts with *reducing the number of variables* measured by the questionnaire to manageable proportions so that the mass of details does not prevent us from seeing the forest for the trees. Thus, data reduction involves creating fewer but broader variables by merging items. Most researchers apply one of two approaches (or a combination of these) to determine which items belong together:

- The statistical technique of *factor analysis* is particularly suited to reducing the number of variables to a few values that still contain most of the information found in the original variables, because it explores the interrelationships of the items and tries to find patterns of correspondence – that is, common underlying themes – among them. The outcome is a small set of underlying dimensions, referred to as *factors* or *components*.
- Based on the theoretical considerations guiding the construction of the questionnaire, we form clusters of items that are hypothesized to hang together (i.e., the original multi-item scales) and then conduct an internal consistency check to determine whether our assumptions are borne out in practice. The *Reliability Analysis* procedure in SPSS not only computes "Cronbach Alpha" reliability coefficients describing the homogeneity of the items in a cluster (or, as it is usually referred to, a "scale"), but also advises us whether the exclusion of one or more items would increase the scale's internal reliability. The *Cronbach Alpha coefficient* is a figure usually ranging between 0 and +1, and during item analysis we should aim at coefficients in excess of .70; if the Cronbach Alpha of a scale does not reach .60, this should sound warning bells.

Step 3: Analyzing the Data through Statistical Procedures

The standard method of analyzing quantitative questionnaire data involves submitting it to various statistical procedures. These include a wide range of different techniques, from calculating item means on a pocket calculator to running complex statistical analyses. It is beyond the scope of this chapter to provide a detailed analysis of the available procedures (for non-technical discussions of statistics, see Dörnyei,

UNCORRECTED PROOF

2007; Pallant, 2007; Salkind, 2008). Instead, we would like to emphasize one crucial aspect of statistical data analysis that is often misunderstood or ignored by novice researchers, namely the distinction between *descriptive statistics* and *inferential statistics*:

- *Descriptive statistics*, such as mean, range, and standard deviation, are used to summarize sets of numerical data in order to conserve time and space. However, these statistics are only specific to the given sample and do *not* allow the drawing of any general conclusions that would go beyond the sample.
- *Inferential statistics* are the same as descriptive statistics except that the computer also tests whether the results observed in our sample (e.g., mean differences or correlations) are powerful enough to generalize to the whole population. If they are, we can say that our results are statistically "significant," and we can then draw some more general lessons from the study.

Thus, *statistical significance* denotes whether a particular result is powerful enough to indicate a more generalizable phenomenon. If a result is non-significant, this means that we cannot be certain whether it occurred in the particular sample only because of chance (e.g., because of the unique composition of the respondents examined). Accordingly, statistically non-significant results *must be ignored* in research studies. That is, we must not say things like "Although the mean difference between boys' and girls' scores did not reach significance, girls tended to do better than boys."

Reporting Survey Results

Survey data can be used for a great variety of purposes and each of these might require somewhat different types of summaries and reports of the findings. It is obvious, for instance, that a PhD dissertation will have to meet criteria that are very different from the presentation requirements of a summary of student achievement at a school staff meeting. There are, however, certain common issues shared by many different types of research reports of survey results. Here we highlight three such issues: the question of how much to generalize, the technical information that we need to include in a survey report, and presenting results in tables.

Issue 1: How Much to Generalize

Researchers need to exercise great caution when pitching the level of generalization in their research reports; this is particularly so in light of Lazaraton's (2005) warning that using high-powered parametric procedures may easily tempt scholars to overgeneralize their results and to make grand claims regarding their findings. The other side of the coin is, however, that research in most cases is all about the need to

UNCORRECTED PROOF

produce generalizable findings, and along with the Task Force on Statistical Inference of the American Psychological Association (TSFI; see Wilkinson & TFSI, 1999, p. 602), we would encourage researchers not to be afraid "to extend your interpretations to a general class or population if you have reasons to assume that your results apply." The question, then, is when generalization becomes *over*generalization. Unfortunately, there are no hard-and-fast rules about where the threshold is, so we need to strive for a delicate balance between the following two considerations: on the one hand, we may wish to be able to say something of a broader relevance, since without this our audience would be very limited; on the other hand, big claims can usually be made only on the basis of big studies. Having said that, Dörnyei (2010) also points out that some seminal papers in the research literature have made some very big claims based on rather small studies.

Issue 2: Technical Information to Accompany Survey Results

In order for the readers to be able to interpret (and believe) the claims made in a research report, they will have to be convinced that the methodology used to produce the particular findings was appropriate. This does not mean that we can only report results if our study did not have any methodological limitations but it does mean that we must provide a concise and yet detailed summary of the main aspects of the survey, including any known limiting factors. There is no perfect study and it is up to the readers (and the journal editors) to decide on the value of the findings. Table 5.1 presents a summary of the points to be covered by the description.

Issue 3: Presenting the Results in Tables

Questionnaire studies typically produce a wealth of data, and therefore developing effective and digestible – that is, reader-friendly – ways of presenting the data is an essential skill for the survey researcher. A rule of thumb is that we should present as much of the information as possible in *tables* rather than in the running text. Having said that, we should realize that for the sake of space economy, some international journals encourage the reporting of some statistical results within the main body of text, so we must not overdo using tables. The big advantage of tables is that they can summarize large amounts of data about the respondents and their responses, and they are also ideal for presenting the results of statistical analyses. Their drawback is that these numerical results are less digestible without any textual context, particularly for the uninitiated.

There are two technical points that we would like to highlight about tables. First, if we present statistics in tables, we should *not* repeat the figures in the text as well, except when we want to underscore some particularly noteworthy results. Second, we should note that statistics tables have certain canonical forms, both in content (i.e., what information to include) and format (e.g., usually we do not use vertical lines in them). These need to be observed closely, which means that simply importing a table from SPSS into a manuscript is most likely to be inappropriate.

UNCORRECTED PROOF

Table 5.1 Checklist for the main points to be covered by the technical description part of a survey report.

Participants
- *Description of the sample*, including the participants' total number, age, gender, ethnicity, first language, level of L2 proficiency, L2 learning history, L2 teaching institution (if applicable), type of tuition received, and any relevant grouping variable (e.g., number of courses or classes they come from).
- The *sampling method* used for the selection of the participants.
- Any necessary *additional details* depending on the particular study, such as general aptitude (or academic ability), socioeconomic background, occupation, amount of time spent in an L2 host environment, etc.

Questionnaire
- *Description* of and *rationale* for the main content areas covered by the items.
- *Factual description* of the instrument (e.g., number of items, response options, language).
- Details about the *piloting* of the instrument.
- Any available data concerning the *reliability* and *validity* of the instrument.

Questionnaire administration
- *Procedures* used to administer the questionnaire.
- *Length of time* that was needed to complete the questionnaire.
- Questionnaire *return rate*.

Variables (if the study contains several complex variables)
- *Complete list* of the variables derived from the raw questionnaire data, including details of how they were operationalized.
- With *multi-item scales*: the number of constituent items and the Cronbach Alpha internal consistency reliability coefficient for each scale.

Project Ideas and Resources

Analyzing a Questionnaire

We would like to present a questionnaire adapted from a published study (Kormos & Csizér, 2008; original version in Hungarian; for a description of the study, see study box 5.1) to illustrate how the various principles of questionnaire design have been put into actual practice. Following the questionnaire and further reading suggestions, we list some study questions that help to analyze the instrument. While no instrument is perfect and some readers might find that they would have done some aspects differently, this particular questionnaire has delivered the "proof of the pudding" – it worked.

UNCORRECTED PROOF

Study Box 5.1

Kormos, J., & Csizér, K. (2008). Age-related differences in the motivation of learning English as a foreign language: Attitudes, selves and motivated learning behavior. *Language Learning, 58*, 327–355.

Background

The aim of the investigation was two-fold. First, the authors wanted to investigate any possible differences among three distinct learner groups who studied English in the same context in Budapest, the capital city of Hungary. The second objective was to test empirically the two main constructs of Dörnyei's L2 motivational self-system, namely the Ideal L2 Self and the Ought-to L2 self, and explore the relationship of these variables with more traditional motivational and attitudinal constructs such as integrativeness and instrumentality.

Research questions

- What are the main dimensions describing students' foreign language learning motivation?
- What age-related differences can be found across the three samples?
- What dimensions influence motivated learning behavior in a significant way?

Method

A cross-sectional questionnaire survey with three independent samples of learners: secondary school students (N = 202; average age = 16.5 years), university students (N = 230; average age = 21.5 years), and adult language learners (N = 191; average age = 33.7 years). After extensive piloting, the final version of the questionnaire was mailed or personally delivered to the participating secondary schools, universities, colleges, and language schools, where a person who agreed to take charge of the administration of the questionnaires distributed them among the teachers and later collected the completed questionnaires.

Statistical tools

A range of descriptive and inferential statistical procedures including ANOVA, correlation, and regression analysis.

Results

The main factors affecting students' L2 motivation were language learning attitudes and the Ideal L2 self, which provides empirical support for the main

UNCORRECTED PROOF

construct of the theory of the L2 motivational self-system. Models of motivated behavior varied across the three investigated learner groups; for the secondary school pupils, it was their interest in English-language cultural products that affected their motivated behavior most, whereas international posture emerged as an important predictive variable in the two older age groups.

College and university student questionnaire (extract; original language: Hungarian)

We would like to ask you to help us by answering the following questions concerning foreign language learning. This is not a test so there are no "right" or "wrong" answers and you don't even have to write your name on it. We are interested in your personal opinion. Please give your answers sincerely as only this will guarantee the success of the investigation. Thank you very much for your help!

I. In the following section please answer the questions by simply giving marks from 1 to 5.

5 = very much 4 = quite a lot 3 = so-so 2 = not really 1 = not at all

For example, if you like "apples" very much, "bean soup" not very much, and "spinach" not at all, encircle the following numbers:

How much do you like apples?	(5)	4	3	2	1
How much do you like bean soup?	5	4	3	(2)	1
How much do you like spinach?	5	4	3	2	(1)

Please encircle one (and only one) number for each item, and please don't leave out any of them. Thanks.

5 = very much 4 = quite a lot 3 = so-so 2 = not really 1 = not at all

1. How much do you like the TV programs made in the United States?	5	4	3	2	1
2. How much do you think knowing English would help your future career?	5	4	3	2	1
3. How much do you like English?	5	4	3	2	1
4. How much do you like the films made in the United States?	5	4	3	2	1
5. How much do you like the pop music of the USA?	5	4	3	2	1
6. How much would you like to become similar to the people who speak English?	5	4	3	2	1

UNCORRECTED PROOF

7. How much do you like the magazines made in the United States?	**5**	**4**	**3**	**2**	**1**
8. How much do you like meeting foreigners from English-speaking countries?	5	4	3	2	10

II. Now there are going to be statements some people agree with and some people don't. We would like to know to what extent they describe your own feelings or situation. After each statement you'll find five boxes. Please put an 'X' in the box which best expresses how true the statement is about your feelings or situation. For example, if you like skiing very much, put an 'X' in the first box:

	Absolutely true	Mostly true	Partly true partly untrue	Not really true	Not true at all
I like skiing very much.	X				

***There are no right or wrong answers* – we are interested in *your* personal opinion.**

	Absolutely true	Mostly true	Partly true partly untrue	Not really true	Not true at all
9. People around me tend to think that it is a good thing to know foreign languages.					
10. My parents really encourage me to study English.					
11. Learning English is really great.					
12. The things I want to do in the future require me to speak English.					
13. I am willing to work hard at learning English.					
14. My parents encourage me to practice my English as much as possible.					
15. I really enjoy learning English.					
16. Whenever I think of my future career, I imagine myself being able to use English.					
17. Nobody really cares whether I learn English or not.					

UNCORRECTED PROOF

	Absolutely true	Mostly true	Partly true partly untrue	Not really true	Not true at all
18. It is very important for me to learn English.					
19. My parents consider foreign languages important school subjects.					
20. My parents have stressed the importance English will have for me in my future.					
21. I find learning English really interesting.					
22. I like to think of myself as someone who will be able to speak English.					
23. My parents feel that I should really try to learn English.					
24. I can honestly say that I am really doing my best to learn English.					
25. When I think about my future, it is important that I use English.					
26. I am determined to push myself to learn English.					
27. Learning English is one of the most important aspects in my life.					

III. Finally, would you please answer a few personal questions – we need this information to be able to interpret your answers properly.

28. If you had a choice, which foreign languages would you choose to learn next year at school (or work)? Please mark three languages in order of importance.

1)……..
2)……..
3)……..

29. Your gender? (Please underline): male female
30. How old are you (in years)? ……
31. What foreign language(s) are you currently learning besides English? ……..……..
32. What college/university do you attend? ……
……
33. What do you study? ……
……
34. How old were you when you started learning English? ……...

THANK YOU VERY MUCH – WE REALLY APPRECIATE YOUR HELP!

UNCORRECTED PROOF

Further Reading

- The most comprehensive text on questionnaires in L2 research is Dörnyei (2010). Brown (2001) also offers a valuable discussion of survey research, including information on interview surveys.
- For further information on various aspects of research methodology in SLA, please refer to Dörnyei (2007) and Mackey and Gass (2005).
- With regard to statistics, a good starting point is Salkind's (2008) book entitled *Statistics for People Who (Think They) Hate Statistics* – the title says it all.
- With regard to the use of SPSS, one of the most informative and user-friendly texts we are aware of is Pallant (2007) – and it also contains statistical advice.
- Finally, the largest-ever attitude/motivation survey in SLA has been Dörnyei, Csizér, and Németh (2006), which also contains extensive appendices that include all the instruments and other materials used in the study.

Study Questions

1. Look again at the questionnaire above, adapted from Kormos and Csizér (2008):

   (a) Identify the main parts of the questionnaire.
   (b) Provide numerical codes for the different types of questions in the questionnaire.
   (c) The main aim of Kormos and Csizér's (2008) study was to investigate what factors influenced students' motivated learning behavior in a significant way. Motivated learning behavior was defined as the amount of effort students were willing to invest into foreign language learning. The final multi-item scale contained *five* items, which are scattered around in the adapted version of the questionnaire presented in this chapter. Try to identify the questions that successfully measured motivated learning behavior.

2. The following set of items was used by Clément, Dörnyei, and Noels (1994) to measure the extent of group cohesiveness in learner groups. If the first item is excluded from the scale, the overall internal consistency reliability coefficient of the scale (i.e., Cronbach Alpha) goes up from .77 to .80. Discuss what could be wrong with this item.

   (a) Sometimes there are tensions among the members of my group and these make learning more difficult.
   (b) Compared to other groups like mine, I feel that my group is better than most.
   (c) There are some cliques in this group.
   (d) If I were to participate in another group like this one, I would want it to include people who are very similar to the ones in this group.

UNCORRECTED PROOF

(e) This group is composed of people who fit together.
(f) There are some people in this group who do not really like each other.
(g) I am dissatisfied with my group.

3. There are two mistakes in the following sentence. Can you spot them?

As can be seen in table below, the correlation between motivation and learning achievement is highly significant (r = .64, p < .001).

References

Babbie, E. R. (2007). *The basics of social research* (12th ed.). Belmont, CA: Thomson & Wadsworth.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*, 185–216.

Brown, J. D. (2001). *Using surveys in language programs*. Cambridge, England: Cambridge University Press.

Clément, R., Dörnyei, Z., & Noels, K. (1994). Motivation, self-confidence and group cohesion in the foreign language classroom. *Language Learning, 44*, 417–448.

Davidson, F. (1996). *Principles of statistical data handling*. Thousand Oaks, CA: Sage.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. Oxford, England: Oxford University Press.

Dörnyei, Z. (2010). *Questionnaires in second language research: Construction, administration, and processing*. New York, NY: Routledge.

Dörnyei, Z., Csizér, K., & Németh, N. (2006). *Motivation, language attitudes and globalisation: A Hungarian perspective*. Clevedon, England: Multilingual Matters.

Gillham, B. (2008). *Developing a questionnaire* (2nd ed.). London, England: Continuum.

Harkness, J. A. (2008). Comparative survey research: Goals and challenges. In E. D. De Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 56–77). New York, NY: Lawrence Erlbaum.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York, NY: Newbury House.

Kormos, J., & Csizér, K. (2008). Age-related differences in the motivation of learning English as a foreign language: Attitudes, selves and motivated learning behavior. *Language Learning, 58*, 327–355.

Lazaraton. A. (2005). Quantitative research methods. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 209–224). Mahwah, NJ: Lawrence Erlbaum.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum.

Pallant, J. (2007). SPSS survival manual: A step by step guide to data analysis using SPSS for Windows (Version 15) (3rd ed.). Maidenhead, England: Open University Press and McGraw-Hill Education.

Salkind, N. J. (2008). *Statistics for people who (think they) hate statistics*. Los Angeles, CA: Sage.

UNCORRECTED PROOF

Tseng, W.-T., Dörnyei, Z., & Schmitt, N. (2006). A new approach too assessing the strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics, 27*, 78–102.

Wilkinson, L., & TFSI. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604.

UNCORRECTED PROOF