CHAPTER FIVE

Survey Research

Elvis Wagner

Virtually everyone is familiar with survey research. People are constantly surveyed about their buying habits, political views, personal beliefs and just about everything else, and the results of this type of research are published and consumed constantly in newspapers, magazines, on television and on the internet. Survey research is a very powerful tool that is used extensively by governments, businesses, educational institutions and by individual researchers, and it has long had an important role in the field of applied linguistics. Researchers have used survey research to investigate the characteristics, attitudes and opinions of language learners. Perhaps most prominently, applied linguistics survey research has provided valuable insights in the areas of learner beliefs (e.g. Kormos et al. 2011), learning strategies (e.g. Chamot & Küpper 1989; Purpura 1999), learner motivation (e.g. Dörnyei 2001; Gardner 1985) and language learning anxiety (Horwitz 2010), among others. As Brown (2001) describes, these notions (i.e. beliefs, strategies, motivation, anxiety) are psychological constructs, abstract notions that are assumed to exist, although they cannot be observed directly. Therefore, survey research instruments allow researchers to operationalize (and consequently, measure) these constructs.

Underlying assumptions and methodology

Survey research differs from experimental research in that the researcher does not manipulate the setting or environment in order to investigate how this affects particular variables or the relationship between variables (see Gass this volume; Nunan 1992). Instead, the goal of survey research is to get

EBSCO Publishing : eBook Collection (EBSCOhost) - printed on 6/11/2025 3:25 AM via UNIVERSITY OF QUEENSLAND (UQ)
AN: 3315834 ; Brian Paltridge, Aek Phakiti.; Research Methods in Applied Linguistics : A Practical Resource
Account: s1097571.main.ehost

information about learners' characteristics, beliefs or attitudes; information that is usually not available from production data, such as performance or observational data (Mackey & Gass 2005). Brown (2001) describes survey research as primary research distinct from both qualitative and statistical research, although survey research will often have both qualitative and quantitative components.

Steps in doing survey research

Because survey research is so prevalent in the social sciences, virtually everyone is familiar with the methodology. In many ways, this presents advantages to the beginning applied linguistics researcher who wants to conduct survey research. The methodology is familiar, and the experience of taking surveys can be useful in creating survey instruments. However, this familiarity can also present unique challenges. Because people are familiar with the methodology, they sometimes gloss over or skip entirely some of the important steps in planning and creating viable research projects. Therefore, a framework for the steps a researcher should follow might be useful. In reality, these steps are not very different from research projects utilizing other methodologies:

- Planning the project
- Designing the survey
- Administering the survey
- Analysing the data

It is very important to note, however, that while these are useful steps to follow when conducting survey research, it is vital that the researcher consider all of these steps during the 'planning the project' phase. That is, unless the researcher has considered how she will administer the survey and analyse the data *before* designing the survey, the research project is almost certainly destined for failure.

Planning the project

When doing the initial planning of the project, the researcher must consider *what* she wants to investigate (reviewing the literature and formulating the research questions), *why* she wants to investigate this topic (identifying the gap in the literature) and *who* she wants to investigate (the population and the sampling techniques that will be used). Based on this planning and research, the researcher then must decide *how* she is going to investigate this topic (the methodology that will be used; the instruments that will be used

for the data collection; whether the data will be qualitative, quantitative or both; and how the data will be analysed).

One of the fundamental underlying assumptions to consider when planning survey research is the issue of sampling. Survey research involves trying to find out information about a particular population (e.g. all ESL learners in Great Britain; all university EFL (English as foreign language) learners in Malaysia; all of the EFL learners in a particular language program in Korea). Rarely is an entire population surveyed (an obvious exception is a national census, where a government seeks to obtain information from every member of the population). Because of the huge amount of resources needed to actually complete a census, most research utilizes surveys, in which information is obtained from a sample of the population. While it might be possible for a government to attempt to survey all L2 language users in a country, it is much more feasible (and often useful) to sample from this population. The notion of sampling is a very important consideration in survey research, because the quality and representativeness of the sampling determines the extent of the generalizability of the results of the research of the sample to the larger population (Vogt 2007). Unfortunately, sampling is a consideration that is often overlooked or neglected by educational researchers (including applied linguistics researchers). Because of its importance in a sound research design, a quick overview of the two basic sampling procedures, probability and nonprobability sampling, is provided here.

Types of sampling

Probability sampling techniques (i.e. random, stratified random, systematic and cluster) are used to select a small group (the sample) from a larger group (the population), in order to create samples that are representative of the population. These techniques can be utilized according to the needs, resources, size of the population and quest for generalizability of the individual research projects. *Random sampling* is a variety of sampling in which the researcher seeks to include a truly representative sample of the population in the study. The goal is to assure that every member of the population has an equal chance of being included in the sample. Unfortunately, it is actually very difficult to obtain a truly random sample, and the representativeness of the sample will always be affected by sampling error. A *stratified random sample* is a version of random sampling in which subgroups are selected from within a particular population, and samples are then generated for each of these subgroups. *Systematic sampling*, a more widely used technique, is a sampling technique in which every *nth* person is selected. Finally, *cluster sampling* is a technique in which natural subgroups (clusters) within a population can be identified, and then random samples are generated for each of the clusters.

An example might serve to make the notion of sampling clearer. Say that a researcher is hoping to explore why students at a particular university are studying English as a foreign language. If a list of all the EFL students existed, the researcher could survey each of these learners (in essence, conduct a census) about their motivation for learning English. However, this list might include 1,000 learners, and surveying 1,000 participants would require huge amounts of resources, so the researcher instead chooses to survey a random sampling of these learners. She could have a computer randomly choose 100 participants from the list. If the researcher were interested in comparing particular groups within this population of EFL learners (e.g. to compare the motivation of English majors with non-English majors), she might choose instead to perform stratified random sampling by having a computer randomly choose fifty English majors and fifty non-English majors). She could also perform systematic sampling by generating a list of the 1,000 EFL learners in alphabetical order, randomly choosing one of the first ten names on the list and then selecting every tenth learner to survey.

Random sampling, stratified random sampling and systematic sampling all assume that a 'list' of all the learners in the population exists (in this example, all of the EFL students in this particular university). In reality, however, for many researchers, no such list of all the individuals in a chosen population exists or else the researcher does not have access to this list. If this is the case, then cluster sampling is an option. Again, using the example of our university EFL students, it is possible or even probable that a list of all the current EFL students does not exist (or is out of date or the researcher does not have access to the list). But the researcher does know that there are eight different levels of EFL classes offered at the university. The researcher might then choose to sample eight different classes of students, randomly choosing one class from each of the eight different levels. She could also randomly choose twelve students within each class to survey.

Using appropriate probability sampling helps support the generalizability of the results of survey research. Nevertheless, much of the survey research conducted in applied linguistics utilizes nonprobability sampling – convenience samples. As its name implies, *convenience sampling* involves surveying individuals who are readily available and who the researcher has access to. The use of convenience samples can be informative and can yield interesting and useful results. However, the drawback is that it is not possible to generalize the results to a larger population. Returning to our EFL university student example, if a researcher surveyed the students in her own two classes (a convenience sample) about their motivation for learning English, she might find useful information about the students in these classes, but she could not generalize her results to the rest of the population (all of the university EFL students).

The survey researcher must also be concerned with other sampling issues, including *nonresponse bias*, *sampling error*, *self-selection* and *sampling with replacement* (see Fink 1995; Fowler 1993; Perry 2011; Vogt 2007).

Designing the survey

After the initial planning stage, the researcher needs to design the survey instruments that will be used in the data collection. The design process includes selecting the instruments to be used (see the section below for the different types of survey instruments), creating the instruments (or adapting instruments used by other researchers), piloting the instruments and revising the instruments based on the piloting process.

Questionnaires and interviews

There are two different types of survey data collection instruments: questionnaires and interviews. Questionnaires and interviews are differentiated by the mode in which they are administered and the type of information that each are designed to elicit. Typically, questionnaires are given in written form and are used in order to get information from or about a large number of individuals, while interviews are administered orally and are used to get more in-depth information from a smaller sample of individuals. A questionnaire is a written instrument in which respondents read questions or statements and respond to these questions by selecting a choice offered or writing their own response. Interviews are conducted orally by a researcher in order to elicit oral responses from a participant. An interview schedule refers to the list of questions that the interviewer will ask the participant and serves to ensure that all participants will be asked all of the same questions, in the same order.

In many ways, questionnaires and interviews reflect the larger quantitative versus qualitative data analysis continuum. Questionnaires are designed for efficiency; they can be administered to a large number of participants easily, they can be objectively scored, and the data can be analysed quantitatively. The shortcomings of questionnaires, however, are also readily apparent. The data derived from questionnaires often provide only a superficial assessment of sometimes very complex constructs. Conversely, the data derived from interviews can be quite rich and in-depth. However, this richness comes at a cost of efficiency, in that the researcher (usually) has to meet individually with the participant to administer the interview. The data elicited must then be transcribed and then analysed qualitatively. Questionnaires and interviews are not mutually exclusive – many survey research projects will employ both instrument types in order to get richer interview data that complement the broader questionnaire data. The purpose of the study and the research questions investigated

will dictate the type of data collection instruments used. Because of space constraints, the focus here will be on the design of questionnaires.

Questionnaires can be either closed or open-ended. Closed-ended questionnaires have a stimulus (these can be questions or statements, although I will refer to these from this point on as questions) that the participants read and then choose the most appropriate response from a list of possible responses. These possible responses can be in a variety of formats. They can be dichotomous choices (e.g. *yes/no, true/false, agree/disagree*); they can be in multiple-choice format where all of the possible answers are listed; or they can be Likert scale items, in which the respondents have a number of possible responses to choose from. Open-ended questionnaires require the participant to write an answer in response to the stimulus question. The responses elicited with open-ended questionnaires can vary from one word (in response to 'What is your native language?') to extensive written texts (e.g. in response to 'What is your motivation for learning English?').

Questionnaires (closed and open-ended) can be designed to elicit objective data – information about the characteristics of the participants such as age, length of residence in an English speaking country, years of English study, etc., and they can elicit subjective data – information about the beliefs, attitudes and values of the participants. Often, a particular survey will elicit both types of information, such as the background characteristics of the participants as well as their attitudes about learning English.

One of the most commonly used items in survey research is the Likert scale item. This type of item usually includes a statement and then generally has four or five response options, typically including *strongly agree, agree, don't know* (or *no opinion* or *neutral*), *disagree* and *strongly disagree* or some variation of these. These response options are then assigned a number by the researcher (typically 5 for *strongly agree* and 1 for *strongly disagree*), which can be used for quantitative analysis. Some researchers prefer to use more than five categories, with response options such as *somewhat* agree, *slightly* agree, *somewhat* disagree, etc. Including a larger range of response options can serve to improve the psychometric properties of the questionnaire but can also serve to make it more difficult for participants to respond in that they might have difficulty differentiating between the different degrees of agreement or disagreement. There is no consensus among researchers about the use of the *no opinion* or *don't know* response. Some researchers suggest not giving the survey takers this option, because participants who do not have strong feelings about the material in the survey tend to select this category, and if many of the participants choose this category, the results of the overall survey often will not reach statistical significance. Thus, not giving the participants this option can lead to more interpretable results. However, Vogt (2007) argues that respondents choose this neutral response because it most accurately describes their response to the statement and that it is inappropriate to not offer this response simply because it is inconvenient for the researcher.

One of the reasons that these Likert scale items are particularly useful is that a number of items can be used to try and assess the same construct. So, to return to our motivation example, the researcher might want to investigate a particular population's motivation for learning English. Based on her personal experience and the review of the literature conducted in planning the project, she might want to investigate a number of possible external motivators (e.g. *needed for a good job, desire to communicate with English speakers, desire to live or travel in an English speaking country*). The researcher can then create a number of Likert scale items to assess the participants' attitudes for each of these different components of the motivation construct. Then, the items measuring a particular component can be summed or averaged to form a composite scale, although the idea of Likert scale items being treated as composite variables is not universally accepted by statisticians. Likert scales are not really interval scales in that the intervals between the different responses are probably not equal intervals, and thus they are more accurately seen as ordinal data. Nevertheless, many, if not most, applied linguists do treat Likert scale items as interval data.

Generally speaking, the more items that are used to measure a particular concept, the more reliable and accurate is the overall scale. Often, items that are measuring the same construct will include *reverse-coded* items. So, for example, one item in the 'job' motivation scale might state: 'Speaking English is necessary for my job', while the reverse coded item would be something like, 'Learning English is not important for my occupation.'

After the individual items have been created, the questionnaire is then compiled and formatted. The importance of the formatting cannot be underestimated. Brown (2001) and Dörnyei (2010) stress the importance of creating a questionnaire that looks professional, with no typographical errors or formatting inconsistencies. The respondents will automatically make a number of assumptions about a questionnaire based on its appearance. If the questionnaire looks professional, participants are more likely to respond to it seriously. For more information on formatting issues, the reader is urged to consult Brown (2001), Dörnyei (2010), Gillham (2007) and Phakiti (2014).

For reliability purposes, the goal for the researcher is to standardize the questionnaire so that every one of the respondents gets the exact same items, in the same order. Brown (2001) suggests that questions should be ordered from shortest and easiest to answer to longest and most difficult to answer. He also suggests keeping together all questions of a single type, of a single function, of a single response format and question form and all questions on a given topic.

After the initial questionnaire has been created, it is important to pilot the questionnaire to see how it performs. Ideally, the questionnaire can be piloted on participants that are members of the target population and then analysed statistically (see the 'analysis' section below). Based on this analysis of which items seem to be working well and which items are problematic,

the developer can revise or delete certain items and add different items if necessary. Then the researcher can do another pilot administration of the questionnaire, analyse the results and revise as needed.

Unfortunately, this extensive piloting and revision process is not always feasible. Nevertheless, it is imperative that some sort of piloting is performed. Resources spent piloting the instrument at this phase of the project are generally resources well spent. It is vital that someone other than the questionnaire developer actually take the questionnaire. At the very least, the researcher can trial the questionnaire on friends or colleagues and get their feedback about the questionnaire. Even if these respondents are not part of the population for which the questionnaire is intended, they can give the researcher valuable information that will assist in the revision process. Often the developer is too close to the questionnaire – he or she cannot see the problems or issues that might be obvious to someone who has not spent many hours researching and developing this survey instrument.

Administering the survey

Only after the instrument has been piloted and revised should it be administered to actual participants. With interviews, the researcher should make sure to schedule enough time to complete the interview. The piloting process should help make clear how much time is needed, but in my experience interviews often take longer than anticipated, and the data from a rushed interview is suspect. The researcher should always strive to be professional, and the interview space should be clean, quiet and unobtrusive. Part of being professional is to be familiar with the interview schedule and rehearsing the questions to be asked. For reliability purposes, it is important that the interviewer follow the interview question wording exactly (Brown 2001). The interviewer should strive to make the interviewee comfortable, welcoming the interviewee at the beginning of the session, thoroughly explaining the research purposes and interview procedure and answering any questions the participant may have. If audio- or video-recording equipment is used to record the interview, it should be as unobtrusive as possible and tested beforehand to make sure it is recording properly.

For questionnaire administration, which often includes larger groups of participants, if at all possible, the researcher should attend the administration in person. While it might be possible for a classroom teacher to administer the questionnaires to his or her class of students, the classroom teacher will invariably not be as invested in the research as the primary researcher nor will he or she be as knowledgeable about the research or able to answer all the questions participants might have. Administering the questionnaire in person, in a polite and professional manner, generally leads to a higher response rate and better results.

It is very important that the researcher anticipate issues that might arise during the administration. Again, the process of piloting the questionnaire should help the researcher anticipate possible problems. A useful suggestion is, as much as possible, to prepare the participants for the questionnaire *before* they appear for the large-scale administration. This can be accomplished by providing participants in advance with information about the study and its purpose (email is often useful for this) and about the questionnaire procedure itself. The researcher should check out the space where the administration will occur, making sure that there is adequate and comfortable space for the respondents. While it may seem obvious to mention that the researcher should double-check that she has enough copies of the questionnaire, and that these copies are collated and in the appropriate order, numerous questionnaire administrations have been scuttled by photocopying problems. As with interviews, when the participants arrive for the administration, the researcher should welcome them, make them feel as comfortable as possible and explain the purpose of the research, the instructions for responding to the questionnaire and answer any questions respondents might have. Again, this might seem obvious, but large-scale administrations can be quite stressful, and if not adequately prepared, the researcher can forget to make the respondents as comfortable as possible. The goal, of course, is to make participants want to respond to the interview or questionnaire as fully and truthfully as possible. Respondents who feel unwanted or unwelcome are much less likely to devote their time and attention to the survey, and the results will be of dubious value. After the administration is completed, the researcher should thank the respondents for their participation.

Sampling issues were described in the 'Planning the Project' section, but sampling is also an administration issue. Appropriate sampling procedures necessitate that the researcher try to get everyone that was targeted in the sampling procedure to actually perform the interview or complete the questionnaire. Even the most conscientious sampler can end up with an unrepresentative sample if not everyone in the targeted sample completes the survey. This is called *nonresponse bias*. For example, if a researcher is exploring students' motivation for learning English and uses a sampling procedure that randomly chooses four classes in a language program to be surveyed, it is important that every student in those four classes be surveyed. If the researcher administers the questionnaire in the four classes in which seventy-five of the eighty students are present, it is very important that the researcher follow up and try to get the remaining five students who were not present in their classes that day to complete the questionnaire. This might sound trivial, but these missing five students could represent serious nonresponse bias. Perhaps these five students often skip class because they are not motivated to learn English. By not getting these students' input, the results of the study are of questionable value.

Questionnaires do not have to be administered in person, of course. Often, questionnaires are sent through the mail to possible respondents, so that the respondent can complete the questionnaire and return it to the researcher. This presents special problems related to response rate and is beyond the purview of this chapter (see Vogt 2007 for further information about postal surveys and questionnaires).

Surveys and questionnaires can also be conducted on the internet or via email. This type of survey research has a number of advantages, the most obvious being the lower costs involved. Rather than having to create and administer physical questionnaires (either in person or via postal mail), the questionnaires can be presented on a website or emailed to potential participants, allowing the participants to respond to the questionnaires when it is convenient for them. The other major advantage is that much of the data entry can be done automatically. When using a web-based survey service, the participants usually complete the questionnaires online, the data is entered automatically, and the results are immediately available for the researcher. Web-based surveys do pose special research problems that have to be addressed, however. Like all other surveys, it is important that adequate sampling procedures be followed, and the researcher must endeavour to follow up with potential participants who did not respond to the initial solicitation. Web-based surveys might be especially prone to bias issues like self-selection bias. Another issue is the need to make online surveys password protected. When participants are solicited for participation in the survey, they should be given a special password that is only able to be used once or else participants might skew the results of the survey by responding multiple times. Web-based surveys also present security issues, and special care should be taken to ensure that secure connections be used so that anonymity can be ensured, especially if the survey is of a sensitive nature.

Analysing the data

After the interviews or questionnaires have been administered, it is necessary to analyse the data. For interviews, the data needs to be transcribed and then coded and analysed qualitatively (see Holliday this volume for more information about analysing qualitative data). Questionnaire data, however, usually entails larger sample sizes and generally necessitate inputting the data into a spreadsheet such as Excel or a statistical program such as SPSS, so that the data can be analysed statistically. One of the most important things to consider is the issue of *reverse-coded items*. Although it might seem obvious, it is imperative that the researcher remember to reverse-code these items when inputting the data. That is, if a Likert scale was used with 5 for *strongly agree* and 1 for *strongly disagree*, when inputting the values for the reverse-coded items, it is necessary to input 1 for *strongly agree*, 2 for *agree*, etc.

After the data are inputted, descriptive statistics (mean, standard deviations, skewness and kurtosis) can be computed. If statistical procedures that require a normal distribution are going to be used in the analysis, these descriptive statistics can be consulted to check that the assumptions regarding normality are met (see Phakiti this volume).

It is also important to examine (and report) the reliability of the questionnaire. Usually, the internal consistency reliability will be estimated using Cronbach's alpha. Internal consistency reliability is used to estimate the extent to which scores on the different items correlate with each other, and Cronbach's alpha is a coefficient (ranging from 0 to 1) that indicates the extent to which the items are measuring a single (unidimensional) construct. The closer this coefficient is to 1, the more consistently the items are measuring the same thing. Depending on what the questionnaire is intended to assess, the reliability can be estimated for the overall questionnaire, for each subsection of the questionnaire or for the individual scales. If the questionnaire is narrow in scope, composed of ten items that are all intended to assess the unidimensional construct 'Learners' motivation in their English classroom', for example, then the reliability should be assessed for the overall questionnaire. If the questionnaire is broader in scope, with three subsections, all designed to measure different components of motivation for learning English (e.g. 'English in the classroom', 'English on the job' and 'English for tourism'), then it would be appropriate to estimate reliability for each of the distinct subsections. If the questionnaire is very broad in scope, with numerous scales all designed to measure different components of motivation, then it would be most appropriate to estimate reliability for each scale. Internal consistency reliability is an estimate of the degree to which the different items are consistently measuring a single construct (see Phakiti 2003). If the questionnaire is designed to measure different things (multiple constructs), then one would expect that the individual items indeed are not measuring consistently.

One of the difficulties with estimating internal consistency reliability for a composite scale (that might be measured by only three or four items) is that often the reliability estimate will be quite low, because Cronbach's alpha is affected by the number of items in the scale. While it is impossible to give a pre-determined cut-off on how high the alpha should be for a questionnaire, Dörnyei (2010) suggests that even short scales composed of only three or four items should have a reliability coefficient alpha of at least 0.70 (i.e. 70 per cent of the items are reliable in terms of measuring the construct), and he suggests that an alpha that is less than 0.60 is problematic. Similarly, Perry (2011) advises that while questionnaires tend to have lower reliability coefficients than test instruments, a reliability coefficient below 0.60 is considered low. Vogt (2007) is even more stringent, stating that an alpha of 0.70 is the minimum acceptable. A reliability coefficient much below 0.70 indicates that the items are not consistently measuring the same unidimensional construct but may in fact be measuring a number of different things.

If the questionnaire has a large number of items and a large number of scales, and if the questionnaire has been given to a large group of respondents, it is more appropriate to use factor analysis in addition to internal consistency reliability to investigate the extent to which the different items designed to measure a specific component of the construct are indeed measuring the same thing. Factor analysis is a more sophisticated statistical technique than reliability analyses and can give the researcher more information about what the different items in the questionnaire are actually measuring (see Vogt 2007 for more information about factor analysis).

Factors affecting the validity of survey research

There are a number of issues that the survey researcher must be aware of and address in order to make the results of survey research as trustworthy and valid as possible. Perhaps the most important consideration involves the issue of sampling. This has been addressed above, but it is worth stressing the notion that unless adequate sampling procedures have been instituted, the research is of little use outside the immediate context of the research and generalizing to a larger population is inappropriate.

Research construct

Another difficulty inherent in much survey research is related to the nature of what is being investigated. As noted, much survey research in applied linguistics is aimed at exploring abstract constructs like motivation, strategy use, attitudes and the anxiety of language learners. Dörnyei, perhaps the pre-eminent researcher of L2 motivation and who uses questionnaires extensively in his research, describes how motivation is an abstract term for a construct that is unobservable, multidimensional and inconstant (Dörnyei 2001). Needless to say, it is no easy task to measure an unobservable, multidimensional and inconstant construct using only a number of statements or questions that a participant has to respond to. To complicate matters even further, many of the constructs being investigated in survey research include behaviours that are unconscious – automatized behaviours that the language learner may do without even being consciously aware of doing so. There is disagreement about whether unconscious behaviours can actually be accessed by conscious recollection, thus calling into doubt the validity of the results of this type of survey research. In addition, questionnaires using Likert scale items to measure these abstract constructs compound the difficulty in that they require respondents to note the strength of choice (e.g. the extent that they *strongly agree* versus just *agree* with the statement). As argued by Vogt (2007), one participant's conception of *agree* might be very different from another

participant's conception of *agree*. Even *yes* or *no* answers are problematic in that for many statements, they provide a false dichotomy. For example, a participant asked to respond *Yes* or *No* to a statement 'I listen to the radio to learn English' might become frustrated if she used to listen to the radio in English, but does not anymore, or if she has listened only once or twice. What is the correct answer? While the researcher needs to try and anticipate these types of issues, often they will only become apparent in the trialling and piloting phase.

Again, investigating these abstract constructs is difficult under the best of conditions, but the applied linguistics survey researcher also has to address language and literacy issues. By definition, language learners have an imperfect control of language. If the interviews or questionnaires are presented to the participants in the target language, it is imperative that the language used in eliciting the data be at a level comprehensible to the respondents. This is, of course, much easier said than done. One way around this issue is to elicit the data in the native language of the participants, but this is not always feasible and also introduces translation issues. Dörnyei (2010) describes the two main challenges in translating questionnaires as the need to create a very close translation so that the versions are equivalent; and to produce a text that sounds natural in the target language. Dörnyei (2010, p. 51) provides two possibilities for doing this. The first includes the use of a team of external reviewers who review the two versions of the questionnaire and assess the equivalence of the two versions and the naturalness of the language. The second possibility involves back-translation, in which the questionnaire is translated from the original language to the target language, and then a different person (or persons) translates the target language version of the questionnaire back into the original version. These two original language versions are then compared to ensure the accuracy of the translation.

Fatigue

Another issue that can affect the validity and trustworthiness of survey research is the issue of fatigue. As a general rule of thumb, the more items on a data collection instrument (e.g. a test, questionnaire or interview), the more reliable that instrument will be. That is, the Cronbach's alpha for a questionnaire that has twenty items will tend to be higher than for a questionnaire that has ten items. There is a point, however, where additional items will lead to diminishing returns. Anyone who has ever completed a questionnaire that was too long can attest to this phenomenon. Respondents can read and answer the questionnaire items for only so long before they become fatigued and lose concentration. At this point, the responses provided by the participant become suspect, the reliability of the responses suffers and the validity of the research is questionable. With respondents

who are reading and responding in a second language, it is likely that fatigue will set in even sooner. Obviously, the point at which fatigue sets in will vary by the individual respondent, but Dörnyei (2010) suggests that no survey or questionnaire require more than 30 minutes to complete.

Bias

Yet another issue that the applied linguistics survey researcher has to acknowledge and try to address is the issue of respondents who are unmotivated or unreliable. Often, a survey researcher will contact a teacher or a language program administrator and get permission to administer her survey or questionnaire to a classroom of students. These students are a captive audience for the researcher, and while they might complete the questionnaire as instructed, if they are unmotivated to concentrate on the questionnaire and answer truthfully, the reliability and validity of the results are questionable. At the same time, survey research is prone to many different forms of bias. That is, the respondents will often respond (consciously or unconsciously) to the survey prompts by giving answers that will enhance their own standing (*prestige bias*) or their responses will reflect how they would like to think of themselves as acting rather than how they really act (*self-deception bias*). Some respondents are also prone to responding to the interview or questionnaire according to how they think the researcher wants them to respond (*acquiescence bias*).

The point here is thus twofold. It is vital that the researcher pilot and validate the instruments used in survey research (see Brown 2001 for further information about validating survey instruments). In addition, while survey research can be quite informative, the inherent limitations in this type of research must always be remembered and acknowledged by the researcher.

A sample study

Purpura (1999) employed questionnaires to investigate low-ability and high-ability learners' cognitive and metacognitive strategy use and how strategy use affected second language test performance. Drawing on a review of the literature of previous strategy use studies, Purpura created cognitive strategy and metacognitive strategy use questionnaires based on a model of human information processing. The cognitive strategy use questionnaire contained forty questionnaire items designed to measure twelve different cognitive strategy types (e.g. *analysing contrastively, linking with prior knowledge, practising naturalistically*). Each cognitive strategy type was measured by at least two different questionnaire items, although some strategy types were measured by up to five items. The items that were measuring the same cognitive strategy type were averaged to create a composite scale of that

variable. Similarly, the metacognitive strategy use questionnaire contained forty questionnaire items designed to measure five different metacognitive strategy types (e.g. *setting goals, assessing the situation, formulating a plan*). Each metacognitive strategy type was measured by at least three items, and one metacognitive strategy type was measured by ten questionnaire items. Again, the items that were measuring the same metacognitive strategy type were averaged to create a composite scale of that variable.

These strategy questionnaires were then administered to 1,382 English language learners. The participants also took the First Certificate of English Anchor Test, so that the strategy use of the learners could be investigated in conjunction with second language test performance. Purpura used a number of statistical procedures to analyse the questionnaire data. He first calculated descriptive statistics (including the means, standard deviations, skewness and kurtosis). Calculating these statistics allowed him to check that the assumptions regarding normality for each variable were met and that statistical procedures requiring a normal distribution would be appropriate for these data. He then performed a series of reliability analyses using Cronbach's alpha for the strategy use questionnaires, estimating the reliability coefficient for both the cognitive questionnaire (alpha = 0.84) and metacognitive questionnaires (alpha = 0.89) as well as the reliability coefficient for each of the scale variables (the twelve different cognitive strategy types and the five different metacognitive strategy types). These reliability analyses allowed Purpura to investigate the extent to which the different items within each scale were actually measuring the same underlying trait, in the hopes of validating the questionnaire. In other words, if three items were designed to measure the strategy *analysing contrastively*, then it would be expected that these three items would have a very high internal consistency. And indeed, a number of the scale reliability coefficients were relatively high (above 0.70), suggesting that the items in these scales were measuring the same construct. However, a number of the composite scale variables were much lower, suggesting that while the items measuring these strategies were designed to measure the same trait, they might in fact have been measuring different traits.

Purpura then used exploratory factor analysis to examine the underlying factorial structure of the two questionnaires. These factor analyses suggested that the cognitive strategy questionnaire was measuring eleven different factors, rather than the twelve different factors (hypothesized cognitive strategies) that Purpura had expected and that the metacognitive strategy questionnaire appeared to be measuring four different factors, rather than the five different factors (hypothesized metacognitive strategies) that he had originally expected.

Based on the results of these exploratory factor analyses, substantive rationale and a rethinking of the original theoretical models, Purpura then used confirmatory factor analysis to test a number of competing models of cognitive strategy use and metacognitive strategy use. In other words, he

used these reliability and factor analyses in order to 'generate empirically-based, composite variables, which could then be used to posit a model of strategy use' (Purpura 1999, p. 67), which could be related to a model of second language test performance.

Resources for further reading

Brown, JD 2001, *Using Surveys in Language Programs*, Cambridge University Press, Cambridge.

This book, by a leading authority on statistics and research in applied linguistics, is aimed at teachers and researchers in language programs that are interested in conducting survey research. It provides in-depth coverage of issues related to survey research. The book is organized according to the steps involved in conducting survey research (*planning*, *designing the instrument*, *gathering data*, *analysing data statistically*, *analysing data qualitatively* and *reporting*). The chapter devoted to statistical analysis is especially thorough and useful for researchers interested in using survey methodology.

Dörnyei, Z with Taguchi, T 2010, *Questionnaires in Second Language Research: Construction, Administration, and Processing*, 2nd edn, Routledge, New York, NY.

This book is a very practical guide to the construction and administration of questionnaires in L2 research from a researcher who has used questionnaires extensively in his research, especially pertaining to the cognitive components of language learning. Perhaps the book's greatest strength is its practical nature – the author gives specific instructions and suggestions for creating reliable, valid and useful questionnaires.

Some useful websites for hosting online surveys
    www.surveymonkey.com, viewed 6 May 2014
    www.questionpro.com, viewed 6 May 2014
    www.KeySurvey.com, viewed 6 May 2014
    www.polldaddy.com, viewed 6 May 2014
    http://freeonlinesurveys.com, viewed 6 May 2014

There are numerous websites and software packages devoted to the creation and administration of surveys and questionnaires. Most of these websites allow users to try out the service for free.

http://nces.ed.gov/statprog/2002/stdtoc.asp, viewed 6 May 2014

Part of the United States Department of Education, The National Center for Educational Statistics administers this website that provides and sets standards for survey data sampling and collection. The Center provides these standards in order to ensure the quality of statistical surveys and their analyses.

References

Brown, JD 2001, *Using Surveys in Language Programs*, Cambridge University Press, Cambridge.

Chamot, A & Küpper, L 1989, 'Learning strategies in foreign language instruction', *Foreign Language Annals*, vol. 22, no. 1, pp. 13–24.

Dörnyei, Z 2001, *Teaching and Researching Motivation*, Longman, Harlow.

Dörnyei, Z withTaguchi, T 2010, *Questionnaires in Second Language Research: Construction, Administration, and Processing*, 2nd edn, Routledge, New York, NY.

Fink, A 1995, *How to Sample in Surveys*, Sage, Thousand Oaks, CA.

Fowler, F, Jr. 1993, *Survey Research Methods*, 2nd edn, Sage, Newbury Park, CA.

Gardner, R 1985, *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation*, Edward Arnold, London.

Gillham, B 2007, *Developing a Questionnaire*, 2nd edn, Continuum, London.

Horwitz, E 2010, 'Foreign and second language anxiety', *Language Teaching*, vol. 43, no. 2, pp. 154–167.

Kormos, J, Kiddle, T & Csizér, K 2011, 'Systems of goals, attitudes, and self-related beliefs in second-language-learning motivation', *Applied Linguistics*, vol. 32, no. 5, pp. 495–516.

Mackey, A & Gass, S 2005, *Second Language Research: Methodology and Design*, Lawrence Erlbaum, Mahwah, NJ.

Nunan, D 1992, *Research Methods in Language Learning*, Cambridge University Press, Cambridge.

Perry, F 2011, *Researching Applied Linguistics: Becoming a Discerning Consumer*, 2nd edn, Lawrence Erlbaum, Mahwah, NJ.

Phakiti, A 2003, 'A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading comprehension test performance', *Language Testing*, vol. 20, no. 1, pp. 26–56.

——— 2014, 'Questionnaire development and analysis', in A Kunnan (ed.), *Companion to Language Assessment*, Wiley-Blackwell, Oxford.

Purpura, J 1999, *Learner Strategy Use and Performance on Language Tests: A Structural Equation Modeling Approach*, Cambridge University Press, Cambridge.

Vogt, W 2007, *Quantitative Research Methods for Professionals*, Pearson, Boston, MA.