Overview

Individual difference (ID) research has been a vibrant field of applied linguistics for the past several decades. Its main aim is to investigate the rate of individuals’ overall success in second/foreign language learning, based on learner characteristics (Dörnyei & Ryan, 2015). The general popularity of survey research in this field is undeniable, as large groups of learners can be investigated with the help of consistent and quantifiable data, in order to draw generalizable conclusions (Dörnyei & Ryan, 2015). However, the use of surveys also depends on what kind of individual variable is mapped. This is because research of ID variables has developed in different ways, depending on the extent researchers will accept what survey strategies can offer or whether they would rather opt for other ways of collecting data.

Surveys have been a traditional and powerful research strategy to collect data in various fields, including applied linguistics. Surveys can be defined as “a systematic method of gathering information from (a sample of) entities for the purpose of constructing quantitative descriptors of the attributes of the larger population of which the entities are members” (Groves et al., 2009, p. 2). Surveys can include different forms of data collection instruments, such as questionnaires, structured observation guides, and even in some cases, structured interviews (Saunders et al., 2009). Nevertheless, the aim of the present chapter is to summarize information on questionnaire studies used in the investigation of ID variables in foreign/second language (L2) learning processes because questionnaires are the most frequently used data collection instruments in survey research. They are not just a collection of seemingly relevant questions but measure underlying constructs, preferably with multi-item scales that are scientifically rooted in the given field.

As each research method, including questionnaire surveys, has its strengths and weaknesses, it is difficult to provide guidelines on how to best go about investigating the role of ID variables in L2 language learning processes without providing some background of the general use of questionnaires. Therefore, first, we provide an overview of the scientific use of questionnaires, in general. When discussing issues related to quality control, we focus on reliability and validity and their relations to sampling and analytical procedures in order to reduce bias. As for the data collection procedure, we describe the full process of using questionnaires, from conceptualization to data analysis, and explain the advantages of questionnaire studies. Next, we detail the characteristics of questionnaire surveys that pertain to linking the research questions to operationalization and to the type of data collected. We then provide a short summary of how the survey strategy has been used to investigate various ID variables over the years. Finally, future research directions are outlined.
Survey research found its way into ID research via social sciences and market research. These disciplines fall back on public opinion research (election polls) and social condition research (social condition of workers). The history of surveys can be traced back to the beginning of the 20th century in countries such as the US and England. However, the first documented mail survey was conducted in Scotland as early as 1788 (De Heer et al., 1999). Surveys permeated various research fields, including applied linguistics. They are especially prevalent in IDs research, despite the fact that it is fairly difficult to develop questionnaires that yield good-quality data (Plonsky & Derrick, 2016).

The overall purpose of survey research using standardized questionnaires in any scientific field is to explore respondents' opinions and dispositions as well as to understand a concept or to reveal new phenomena by questioning a limited number of individuals in a generalizable way. Surveys are also useful when the underlying construct or phenomenon of interest cannot be tapped into directly. In ID research, this entails the investigation of various ID variables in diverse contexts to characterize learners and the role of IDs in L2 learning processes as well as to map possible relationships among these ID variables (Dörnyei & Csizér, 2012; Dörnyei & Ryan, 2015). In large-scale survey studies (N > 100), individual differences in learners are explored from a trait or state perspective (Csizér & Albert, 2021). More precisely, language learners’ intended or self-reported behavior, such as their effort, is explored (Dörnyei, 2010; Gardner, 2020), or their learning strategy use is investigated (Griffith, 2018; Pawlak & Oxford, 2018). Participants' feelings and beliefs about certain issues connected to second language acquisition are mapped, such as their positive or negative emotions (Dewaele & MacIntyre, 2016; Teimouri, 2017). Their attitudes related to languages, their speakers, or aspects of the learning processes are also investigated in this way (Gardner, 2020).

The main aim of any survey study should be to obtain replicable and generalizable results. Generalizability in survey research means that even though the research is based on a limited number of participants (sample) selected from a predefined population (e.g., secondary school students in a given city or country), the results can be generalized to fit the whole target of the research (population). Therefore, a clear definition of the members of the population and the sampling procedures (i.e., how participants were selected for the purpose of the study) must be provided in order to know the extent of generalizability in a given study.

Several principles should inform questionnaire studies when designing generalizable and replicable survey research. One of the most important principles of survey research is to ensure the validity and reliability of results. Validity, in this context, means that the measurement tools used in the survey measure exactly what they are intended to measure (e.g., Likert scales can measure self-reported intended behavior, not actual behavior), while reliability refers to the consistency of the measurement (Babbie, 2016). To ensure validity and reliability in a questionnaire survey, one should carefully link the investigated topics in the questionnaire to previous theoretical and empirical studies, systematic sampling procedures, and validating the questionnaire (see Dörnyei & Csizér, 2012 for more details). A theoretically relevant background is an important principle for measuring a concept or revealing a phenomenon that demands the formulation of clear and arguable research questions. Concerning systematic sampling procedures, in order to achieve generalizability to a given target population, the target population should be defined in close relation to the research question, steps of the sampling should be systematic, and the sample should be of an appropriate size. The steps taken to validate a questionnaire include the expert judgment of items, think-aloud protocols with prospective participants, as well as a full pilot study.

The survey process starts with conceptualization: in order to answer the research question, we need to develop the definitions of the concepts that we want to measure by reviewing the relevant literature. For example, to measure language learners’ motivation, one needs to consider how motivation was conceptualized in earlier studies, what researchers think the complex construct of motivation consists of, and what some of the current challenges are (Dörnyei, 2020). Creating the items in a questionnaire is a step towards making the concept measurable during operationalization and
Surveys

should be accompanied by a validation process. Preparing a questionnaire is especially demanding in contexts where researchers’ or participants’ mental schematas are different (Illés, 2020), such as comparative international research involving researchers from various contexts, or when researching special target groups whose cultural background differs from that of the researcher’s (Illés, 2020). As a direct continuation of the operationalization processes, a sample is drawn from the defined target population with a method chosen according to both the specialties of the target population and the planned method of data collection. The preferred mode is random sampling (i.e., each member of the population has an equal and defined chance of being chosen as a sample participant; Dörnyei, 2009; Groves et al., 2009), whereas non-random but systematic sampling is also an alternative (Babbie, 2016; Saunders et al., 2009). Sampling can be a difficult task when the population is rare or elusive, as is the case with special needs students (see, for example, Csizér and Kontra [2020]). The process continues with data collection using a selected method (e.g., self-administered online or paper-and-pencil questionnaires) defined by the target population as well as by the research budget. The following step of the survey process is data cleaning and missing data imputation (i.e., replacing missing data with estimated responses; Groves et al., 2009). As for some complex statistical procedures, datasets without any missing data are required. The survey process in any scientific field is completed by data analysis and interpretation of the results, which are presented in a publication or at a conference.

By doing survey research, the facts, attitudes, intentions, and personality characteristics of individual members of a target population can be measured concurrently or retrospectively, based on self-report. While measuring the facts about individuals seems to be simple (it is not always the case), measuring psychological factors as well as competencies is much less obvious because these factors can be measured mainly by indirect means. In such cases, predefined scales and previously validated tools should be used rather than the researcher’s own constructions (see IRIS digital repository for survey instruments: https://www.iris-database.org). The measurability is always limited due to the fact that surveys are based on self-report. It means that the answers can be biased by many factors, including social expectations, the interviewer, or simply memory. Moreover, the goal of the researcher might not only be to measure the above-mentioned factors but to reveal causal relationships among them. However, it might be challenging to design cross-sectional surveys to prove causal relationships as causality has to be established through observations and experiments, in which the cause can actually precede the effect in time. Alternatively, causality can be researched by panel surveys, which are designed to deal with time-series data (i.e., collecting data from the same group of participants multiple times; Dörnyei, 2009). These types of studies come with their own methodological problems, such as attrition and panel conditioning, not to mention the expense of higher costs (Dörnyei, 2009).

**Technical Features**

The following section covers the most important steps of survey research and provides a practical introduction to formulating research questions, publishing the results, and measuring the quality of data.

**Theoretical Background of Survey Research**

Establishing the theoretical background for any survey research should provide the foundations for both the research question and the definition of target population. This first phase of the process is highly variable; therefore, there is no unified methodology established for it. Moreover, theory pertaining to survey strategies are paid much less attention than theories related to other research strategies (De Leeuw et al., 2008). Most survey handbooks do not go beyond discussing the important requirements of clear and well-defined research questions. Dörnyei (2007) suggests choosing
a research topic at first, then narrowing it down to the research question, based on the researcher’s background and the selected literature in order to establish a research gap. Babbie (2016) formulates a similar logic when defining deductive theory construction, which is based on selecting a research topic, followed by constructing a theory and a testable hypothesis. A proper definition of the target population, closely connected to the research question, is also a strong prerequisite for the subsequent parts of the process. While defining the target population (the group of people from which the sample is drawn and for which the results are generalized) seems to be simple on paper, theoretical and ethical issues can make it more complicated in practice.

Theoretical advancement supporting survey processes can be well illustrated with the development of the L2 motivational self-system theory, which is a recent example of how survey studies might contribute to theory building and theory testing. Based on the seemingly counterintuitive results of a large-scale Hungarian study, based on the work of Robert Gardner and Richard Clément (Dörnyei et al., 2006), Dörnyei proposed the reconceptualization of the notion of integrative motivation and developed the L2 motivational self-system (Dörnyei, 2005, 2009). The proposed theory was tested in various international contexts, and its components were partially verified (Kormos & Csizér, 2008; Ryan, 2008). A second wave of survey studies set out to fine-tune the proposed model (e.g., Papi et al., 2019) and once the number of studies had reached a sufficient level, a meta-analysis was carried out (Al-Hoorie, 2018). As this meta-analysis revealed some inconsistent results, it must be mentioned that theory building and theory testing are an ongoing and ever-evolving endeavor.

**The Preparation and Validation of a Questionnaire**

The formulation of a questionnaire starts with the definition of measurable dimensions (concepts) related to the research question (conceptualization). The available literature about the given research topic should inform this process. Based on the defined dimensions, the survey questions/statements can be formulated for the purpose of the study or adapted from previous investigations. The main principle of question formulation closely relates to validity and reliability: the primary goal is to measure the dimensions that were defined for all possible respondents in the target population. The main reason for keeping the questionnaire simple and straightforward is to avoid misunderstandings or biased questions in any kind of cultural context. Here is some practical advice on how to formulate questions, based on “conventional wisdom” (e.g., Marsden et al., 2010):

- Use simple, familiar words and simple syntax (according to the target population).
- Avoid ambiguity, be specific, and use concrete wording.
- Create exhaustive and mutually exclusive responses (except for multiple choice).
- Avoid leading or loaded questions that can influence the answer and ask one thing in one question.
- Avoid any kind of negation.
- The questionnaire should start with an easy and on-topic question.
- Questions on the same topic should be grouped together and proceed from general to specific.
- Filter questions (i.e., questions that help to decide who needs to answer some of the subsequent items) should be used in order to allow participants to skip non-relevant questions.
- Questions that are difficult to answer or are about a sensitive topic should be placed towards the end of the questionnaire.

Instead of developing new questions/statements, questions from previous studies should always be considered, as should previously validated scales if they are available and suit the purpose of the research. By using previously employed questions/statements, the questionnaire gains further validity (as errors revealed in previous research can be amended) and comparability.
The basic unit of the questionnaire is called an item. An item is a question/statement that requires an answer from the respondent. There are several different item types that can be used in different situations. For example, there are closed-ended question types, where the respondent can choose from predefined answers, and open-ended question types, where the respondent may answer freely or in a certain number of words or characters. The length of the answer should be limited because the categorization of longer, open-ended questions may cause difficulties for analysis. Moreover, giving longer answers takes much longer which is counterproductive for the limited optimal length of the questionnaire. In closed-ended question types, there are simple questions such as single-choice items, multiple-choice items, ranking, and numeric questions. Such types of questions can be used for directly measurable features, mainly facts.

The features that cannot be measured directly (such as attitudes, competencies, psychological factors, and other concepts) can be assessed with multi-item scales, which use several items to measure one dimension. Items in multi-item scales can be of a different type: Likert scales, semantic differential scales, and numerical rating scales are the most widely used types. The Likert scales usually represent agreement or disagreement with a statement on a defined-number points scale, such as an item from the listening class anxiety scale (Kutuk et al., 2019), “I feel nervous during listening activities in class”, and the possible answers: strongly disagree/disagree, neither agree nor disagree, and agree/strongly agree. The semantic differential scales measure the opinion of the respondent about a pair of opposites on a scale (e.g., learning Hungarian as a second language is easy or hard on a seven-point scale, for example, where 1 represents very easy and 7 represents very hard). Finally, numerical rating scales allow the respondent to choose one number, corresponding to a pre-ordered category (Dörnyei, 2007). Besides choosing the type of response scales, the number of response categories can also be problematic when developing a questionnaire. If the number of categories is even, the respondent is obliged to take a side even if their opinion is ambiguous, which may cause larger item non-response. If the number of categories is odd, those who do not want to answer tend to choose the central category repeatedly, regardless of their opinion. If there is a large number of categories, respondents cannot really differentiate between exact values, while if there is a small number of categories, respondents may struggle to differentiate between their answers. According to an empirical review, the ideal number of categories is seven, nine, or ten (Preston & Colman, 2000).

The optimal length of the questionnaire largely depends on the data collection method. When data collection happens face to face, the instrument can be as long as 40–45 minutes; however, when employing online instruments, the optimal length is much shorter and it is advised to take approximately 15 minutes (about four items per minute, when using closed-ended questions).

In order to examine the comprehensibility and interpretability of the questionnaire, it is advisable to conduct cognitive interviews pertaining to the survey questionnaire. The goal of cognitive interviews is to “reveal the thought processes involved in interpreting a question and arriving at an answer” (Presser et al., 2004). During cognitive interviews, a few members of the target group are invited to answer the survey questionnaire by using a special method. There are two main paradigms for cognitive interviews: think-aloud protocol and probing (Beatty & Willis, 2007). During probing, respondents are invited to answer further questions about their understanding of the questions and how difficult it was for them to respond to each item. During think-aloud protocol, respondents are asked to comment on what they think while answering each item. Based on cognitive interviews, items can be rephrased or even omitted from the questionnaire.

Besides cognitive interviews, pilot surveys or pretests should be considered to see how the questionnaire will work in practice, i.e., data collection duration of survey, non-response, etc.

**Sampling Procedure**

The purpose of sampling is to provide a selection drawn from the population that makes an unbiased estimation possible for the target population: the sample should resemble the features of the
target population as much as possible. A further requirement for sampling is that the reliability of the estimation should be measurable (in the form of sampling error). The sampling frame from the sample drawn should ideally match the total target population exactly (coverage).

The only sampling type that fulfills all the requirements mentioned above is random sampling, based on a population census (see Vannette et al., 2017 on the importance of probability sampling). The basic idea of random sampling is to randomly select an appropriate number of members from the sampling frame. The easiest form of random sampling is simple random sampling (SRS), when the sample is drawn from a sampling frame, and each member has an equal chance to be included in the sample. The sampling error can be decreased by stratification (stratified random sample). During stratification, some of the characteristics of the target population with known distribution (e.g., distribution of eighth graders in a primary school according to the type of settlement where the schools are located) are used to select a sample that yields exactly the same distribution for the given characteristics in the sample as in the population (some strata can be over- or under-represented if they fit the research question). In order to decrease the cost of research (when collecting data face to face), multistage clustered samples can be used (e.g., for the same target population as above, choosing schools randomly in the first stage, and choosing students in the given settlements in the second stage). In multistage cluster sampling, first, a random sample is drawn from primary sampling units (e.g., municipalities, schools, and language schools) followed by another random sample being drawn from the target population (secondary sampling units). While multistage cluster samples decrease the cost of surveys (by decreasing the number of places interviewers should visit or the number of schools they should contact), it also increases the probability of sampling error. Stratification and clustering can be combined; however, sampling error decreased by stratification might not compensate for the increased effect of clustering, in most cases.

Having a sampling frame (a list of the target population or a list of larger units with a known size of the target population, such as a list of schools in a given district) is a strict condition for drawing a random sample. If, and only if, a sampling frame does not exist, alternative methods of sampling should be used. A common possible alternative to random sampling is quota sampling. When using quota sampling, the sample is drawn from the target population according to a quota table which is based on the known distribution of the target population (e.g., if a joint distribution of gender and age groups is used, then the number of 15–21-year-old women, 15–21-year-old men, etc., should be calculated according to external data such as a census). While the distribution of a quota sample will be the same as the target population, according to the known characteristics, the distribution of other characteristics can be different in unknown ways. Due to these circumstances, estimations based on quota sampling might be biased and no valid standard error can be calculated for the estimations.

One of the most widely used alternatives to random sampling is convenience sampling. Convenience sampling uses a selection of the target population that the researcher can access easily (e.g., the questionnaire is uploaded for a group on social media and members are asked to complete it). This selection, however, may cause bias in the estimations and make it impossible to estimate sampling error. In the example above, if any group members are not on social media, they have zero probability to be included in the research.

**Data Collection Methods**

In most cases, data collection is budget dependent. However, one must realize that the method of data collection influences the quality of data, the efficiency of sampling (by introducing bias), and the efficient length of the questionnaire.

In terms of data collection methods, one must differentiate between face-to-face and self-administered questionnaires. Although face-to-face questionnaires (when the researcher or interviewer asks the questions and records the answers) remain the most reliable data collection method,
they are also the most expensive (West & Blom, 2017). Hence, self-administered questionnaires are frequently used as a survey tool in the field of SLA, as they are associated with lower costs and less social pressure, resulting from the absence of the researcher. Self-administered questionnaires are employed either offline, online, or as a combination of the two. Online data collection, although increasingly popular, has its own challenges. While it can be a very cost-effective and quick way of collecting data with no social pressure on the researcher, controlling the actual process of data collection is out of the researcher’s hands, and sampling possibilities are very limited (except for schools). What is more, methodologically, online data collection results in differing levels of reliability and validity (Couper & Miller, 2008). For example, if an online survey is paired with convenience sampling, the results are very unreliable. On the other hand, if an online survey that uses random sampling is used in a school setting, where research staff can supervise students completing the questionnaire, it can almost be as reliable as a face-to-face survey.

**Data Preparation**

If data collection is aided by a computer (e.g., online questionnaire), data cleaning automatically becomes a part of data collection because acceptable codes are set prior to the data collection. However, if paper-based questionnaires are used, data cleaning follows data collection. The goal of data cleaning is to repair the data as much as possible in order to avoid inconsistencies and impossible or improbable (outlier) values.

Item non-response causes missing data which can yield a reduced sample size, especially in multivariate analysis. Therefore, it may cause reduced reliability and large confidence intervals, while dependency of non-response on the given and/or other variables may cause bias. If the amount of missing data is small, available data can be used without any treatment (complete case analysis). If the amount of missing data is considerable, there are several methods of dealing with this problem, depending on the type of missing data. If item non-response causing missing data is independent of any variable in the database (missing completely at random, MCAR), using available data gives unbiased estimation (available case analysis or pairwise deletion). If non-response is dependent on some other variables, but not on the variable where non-response occurred (missing at random, MAR), some types of imputation can help to avoid bias. Data imputation is a method for estimating missing data, according to non-response using the available information about the given respondent and all other respondents. Most statistics programs and packages can be used for imputation, but while single data imputation provides an estimation to missing data, it can arbitrarily reduce variance, which may in turn influence the analysis. Therefore, it is advisable to use multiple imputations instead of single data imputation (Piggot, 2001).

As part of data preparation for analysis, different types of weight can be constructed. In the case of a random sample and different (but known) sampling, probabilities are employed. For example, in a study investigating the variability of foreign language courses in American high schools, the schools with a high enrolment rate of students of African-American and Hispanic origin were oversampled to increase the reliability of statistics for these groups. During analysis, data was weighted thus the estimates for the total high school population became unbiased. A weighting procedure should be applied to counteract the effect of biases concerning different sampling probabilities. If there is a considerable amount of non-response, the non-response weight can be applied, based on the characteristics of the known non-respondents in order to decrease the biasing effect of non-response. For example, when doing survey research involving schoolchildren, the researcher might have some information about the children who are absent during data collection.

In any case, post-stratification weight can be applied in order to achieve a representative sample for those characteristics of the target population whose distributions are available. Large weights should be avoided because the weighting procedure itself always increases the error of the estimations.
Data Analysis

The primary aim of data analysis is to answer the research question. The main concerns are as follows: 1) one needs to take into account the sampling design in order to find valid statistical tests and confidence intervals; 2) outlying data points should be investigated to see whether they are artifacts of the survey procedure or valid data; and 3) external and internal validity of data should also be checked (see below).

A full statistical toolkit used for survey analysis in SLA research cannot be systematically reviewed here, so we will highlight some of the more widely used tools. In most SLA research, multiscale measurement tools are employed (either the researcher’s own constructions or adapted versions from previously evaluated scales). These measurement tools assess a latent phenomenon that cannot be measured directly. In order to identify the latent dimensions of measured variables, data reduction tools can be used such as principal component analysis, exploratory factor analysis (EFA), or confirmatory factor analysis (CFA). EFA can be used as an exploratory tool when there is no solid theoretical background (assumptions) about the possible latent dimension. However, when using EFA, the resulting factors should be validated in further research or further validation steps should be taken. On the other hand, with CFA, the measurement element of structural equation modeling (SEM), solid theoretical assumptions can be tested with a single dataset, and the construction of latent variables is more flexible in a statistical sense. In addition, the reliability and validity of any dimension of the research can be checked by various reliability measures, item total correlation (internal validity), and the correlation between theoretically related constructions. While Cronbach’s alpha is the most commonly used method for reliability analysis (Derrick, 2016), the assumption of this tool is not met in many cases; therefore, other tools are recommended, such as the different versions of omega (Hoekstra et al., 2019; McNeish, 2018). A widely cited example for using CFA is the validation of the second language speaking anxiety scale (SLSAS), where the author tested a dual conceptualization of anxiety (in-class and out-of-class) with CFA (Woodrow, 2006).

In complex research projects, the research questions usually relate to causal inferences. However, statistical tools based on survey research can only reveal associations that are obviously not equivalent to causal relations (Arjas, 2001). In order to draw cause-and-effect-related conclusions, at least three conditions should be met: first, there should be statistical associations between the two measured phenomena; second, there should not be any other confounding variables to which the given statistical association could be attributed to; and third, the cause should precede the effect. While for the latter condition logical tests can be used, the first two conditions can be checked by carefully built statistical models, measuring statistical association and controlling possible confounding variables. There are several statistical tools and concepts that can achieve causal conclusions: propensity score matching, (multilevel) generalized linear models, and structural equation models.

Survey Administration and Special Cases

Data quality has gained increased importance in the field of survey research. One of the most widely used frames for quality management is the total survey error paradigm (Gideon, 2012). The total survey error not only refers to a way of calculating and adding up error terms but also to a frame for understanding the possible sources of error, through the survey process. The total survey error can be calculated in the following way:

\[
\text{Total survey error} = \text{coverage error (how much the sample frame covers the whole target population)} + \text{sampling error (according to randomness of sampling)} + \text{non-response error (according to differences between responding and non-responding members of the target population)} + \text{measurement error (according to bias or misunderstanding of the questionnaire)}.
\]
The main ethical issues in survey research include data and privacy protection as well as publication bias (Babbie, 2016). If surveys use personal data, prior permission should be sought in line with the requirements of the given context. For example, according to EU legislation, personal data can be processed if “processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data” (GDPR Art. 6 paragraph 1 f.) Ensuring maximum protection of personal data is a must and so is obtaining informed consent as soon as possible. However, privacy protection legislation can differ from country to country, which makes it harder to give advice on this matter (Lessof & Sturgis, 2018).

There are some special cases of survey research that can influence the above-mentioned procedures. We will briefly mention two that can occur in SLA research: international comparative research and research in schools. In international comparative research, the role of question formation, similarities and dissimilarities in sampling, and data collection increase together with the importance of validity checks of measurement tools (Harkness, 2003). The conceptualization should reveal dimensions that equally exist in all countries (cultures) where the research is conducted, and questionnaires should be translated so that they mean the same thing in each country (double translations, back translations, and adjudication are widely used methods). The most reliable but similar methods available in all participating countries should be chosen for sampling and data collection. When analyzing data, the cross-country validity of the scales should be checked (CFA and SEM are appropriate tools). In research in schools, there are many circumstances that can influence the survey process (De Leeuw et al., 2008). Wording the questionnaire should be adjusted to the age of the target group, while the length of the questionnaire should be adjusted to the normal timeframe of the class. Multistage sampling is applicable in most cases, due to the availability of an official school register. Cooperating with the schools can make the process smoother, for example, in timing the research as well as collecting letters of consent from the parents. The latter is a legal obligation, or at least it is strongly advisable. The institutions can determine the data collection method employed: be it a self-administered paper-based questionnaire or an online questionnaire completed in a computer lab. For analysis, cluster sampling (refer to the Sampling Procedure section) should be considered, and the use of multilevel analysis is advisable (see more about multilevel analysis in the field; Hiver & Al-Hoorie, 2019, Chapter 13), as most multivariate analytical tools assume the independence of the sampled members.

**Contributions to ID Research**

In the field of ID research, one can safely say that most of the variables are measured by using the survey strategy if information on groups of learners is needed. It is outside the scope of this chapter to provide a succinct summary of all the questionnaire-related results in ID research. Therefore, we intend to illustrate the ways surveys are used to investigate such variables without providing a full picture or dwelling on conceptualization issues or disagreements. For more details on the ID variables, see earlier chapters in this volume. In addition, some influential data collection instruments are included in Appendix 1.

To investigate second/foreign language learning motivation, surveys were used in the first major studies that dealt with the socio-educational aspects of motivation (Gardner, 1985; Masgoret & Gardner, 2003). However, at the turn of the century, this field began to witness a change in direction and saw more qualitative and longitudinal investigations (Dörnyei & Öttö, 1998; Ushioda, 2001) that paved the way for employing the principles of complex dynamic theories (Dörnyei et al., 2015) and using more situated and contextualized studies (Ushioda, 2020). Nonetheless, second/foreign language motivation surveys provided the most important theoretical advancements in this field. Such was the emergence of the notion of integrativeness, i.e., to learn a language in order to integrate into the second language speaking community (Gardner, 1985, 2020). Similarly,
a large-scale survey study by Dörnyei et al. (2006) inspired the formation of one of the most popular second language motivation theories in recent years (Boo, Dörnyei, & Ryan, 2015): the second language motivational self-system (Dörnyei, 2005, 2009, 2010). Research into the notion of willingness to communicate followed a somewhat similar path. Once the theoretical foundations were established and investigated (MacIntyre et al., 1998), the survey strategy complemented idiodynamic investigations initiated by MacIntyre (e.g., MacIntyre & Legatto, 2011) and additional qualitative approaches (Dörnyei & Ryan, 2015; Yashima, 2020).

Survey research into topics pertaining to language learning anxiety also abounds. Anxiety defined as “worry and negative emotional reaction aroused when learning or using a second language” (MacIntyre, 1999, p. 27) has undergone at least three distinct ways of conceptualization. In one approach, language use anxiety was mapped both within and outside the classroom (Horwitz et al., 1986). Another research approach investigated how anxiety influenced input, process, and output issues when using the language (MacIntyre & Gardner, 1994; Piniel, 2012). In addition, as second/foreign language learning subsumes different skills, it was also important to explore how the various uses of skills related to anxiety. For example, Cheng (2004) developed an instrument and investigated the multidimensionality of writing anxiety and thus created an instrument that could provide fine-tuned relational information on anxiety and various measures of writing performance.

Positive emotions also impact the learning process, something which has been acknowledged by the recent trend of studies conducted under the aegis of positive psychology (MacIntyre et al., 2016). Both the positive turn in SLA research as well as the need to understand the workings of emotions in learning processes necessitate survey research. Currently, a questionnaire is being developed that measures the full array of both positive and negative emotions (Albert, Lajtai & Piniel, 2020) as part of several pilot studies as well as a large-scale nationwide investigation. The instrument is to measure 11 discrete emotions: enjoyment, hope, pride, curiosity, excitement, anxiety, boredom, apathy, confusion, anger, and shame, and their relations to other individual difference variables.

**Future Directions**

In terms of future research, we think that questionnaire studies will remain essential for several reasons. However, the meaningful future of questionnaire studies hinges on the quality of the scales used in various studies. In order to conduct good-quality questionnaire studies, more care needs to be taken in establishing the overall quality of scales measuring the various aspects of ID variables. One possible way to do this is to carry out meta-analytical studies that can provide summative information on previous questionnaire studies and uncover issues related to validity evidence or the lack thereof (Plonsky & Derrick, 2016). Meta-analytical studies would also be welcome as very little information exists on how the same constructs or scales work in various contexts, or whether results are locally or globally relevant. Comparative cross-contextual studies could also shed light on what the core content of constructs might be relevant in various contexts and what the locally relevant aspects are.

Questionnaire surveys can provide a detailed understanding of the internal structure of ID variables and the relationships between them by using multilevel statistical analyses. When using statistical tools for modeling, they can provide information on the internal relationships of concepts and, at the same time, group-related differences can also be examined. The importance of modeling relationships is a methodological issue as well as a research content-related problem. Individual difference research has often been criticized for investigating ID variables in isolation (Dörnyei & Ryan, 2015; Ryan, 2020) even though learners’ profiles are much more complex than a single measure of an ID variable. Various modeling approaches also make it possible to investigate these variables in constellations (Piniel & Csizér, 2013; Ryan, 2020). Furthermore, cluster analysis seems to be a useful tool for profiling learners for various individual differences (see Csizér & Jamieson, 2013 for a general description of the statistical technique; and Papi & Teimouri, 2012 for an example on L2 motivation).
In addition, questionnaire surveys provide a useful platform to test various theories, thus contributing to the theoretical advancement of our field. As for the complexity of ID variables, questionnaire surveys are also useful to investigate how much these characteristics represent enduring traits or fleeting states (Csizér & Albert, 2021); longitudinal designs could help to establish learners’ traits while situation-specific measures can be added periodically to the dataset, thus combining the two perspectives. In order to understand any phenomenon fully, it is advisable to use mixed-methods procedures where the strengths and weaknesses of the various strategies enhance and complement each other.

Our understanding of current research protocols is that the importance attached to survey questionnaires is changing, and alternative data collection techniques are increasingly used to investigate smaller groups of learners in various settings. This, however, does not mean that questionnaire studies should not be used for what they do best. In a recent publication, Csizér and Albert (2021) argued that many ID variables have both trait and state components, and that questionnaire surveys seem to be one of the most optimal techniques to map learners’ trait characteristics. Future research should consider surveying individual difference variables in various constellations (Ryan, 2020) to obtain a more realistic understanding of the learner within the learning processes.

**Note**

The first author was supported by the NKFIH – 129149 research project in writing this chapter.

**References**


Surveys


### Appendix A1 Summary of Main Individual Difference Variables and Corresponding Instruments in Applied Linguistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Sample Instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>AMTB (Gardner, 2020)</td>
</tr>
<tr>
<td></td>
<td>Various motivational scales including scales of L2 Motivational self-system (Dörnyei, 2010)</td>
</tr>
<tr>
<td></td>
<td>Self-guides (Papi et al., 2019)</td>
</tr>
<tr>
<td>Anxiety</td>
<td>FLCAS (Horwitz et al., 1986)</td>
</tr>
<tr>
<td></td>
<td>Input/output/process anxiety (MacIntyre &amp; Gardner, 1994)</td>
</tr>
<tr>
<td></td>
<td>Skills-related anxiety (Piniel, unpublished)</td>
</tr>
<tr>
<td></td>
<td>Writing anxiety (Cheng, 2004)</td>
</tr>
<tr>
<td>Perceptual styles</td>
<td>E&amp;L inventory (Ehrman &amp; Leaver, 2003)</td>
</tr>
<tr>
<td></td>
<td>ILLS (Griffiths, 2012)</td>
</tr>
<tr>
<td></td>
<td>SR.Cvoc (Tseng et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>Emotion regulation (Gkonou &amp; Oxford, 2016)</td>
</tr>
<tr>
<td></td>
<td>Listening strategies (Zeng &amp; Goh, 2018)</td>
</tr>
<tr>
<td></td>
<td>MARSI (Mokhtari et al., 2018)</td>
</tr>
<tr>
<td></td>
<td>Vocabulary learning strategies (Gu, 2018)</td>
</tr>
<tr>
<td></td>
<td>Grammar learning strategies (Pawlak, 2018)</td>
</tr>
<tr>
<td>Emotions</td>
<td>Enjoyment (Dewaele &amp; MacIntyre, 2016)</td>
</tr>
<tr>
<td></td>
<td>Joy and shame (Teimouri, 2017)</td>
</tr>
<tr>
<td></td>
<td>Grit (Teimouri et al., 2020)</td>
</tr>
<tr>
<td>Willingness to communicate</td>
<td>WTC (MacIntyre et al., 2001; Yashima, 2002, 2009)</td>
</tr>
<tr>
<td>Language learning autonomy</td>
<td>LLOS (Noels et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>ELAQ (Hyland, 2004)</td>
</tr>
<tr>
<td></td>
<td>Learner autonomy (Murase, 2015)</td>
</tr>
<tr>
<td>Flow</td>
<td>Flow in FL classroom (Egbert, 2003)</td>
</tr>
<tr>
<td></td>
<td>Flow experience (Czimmermann &amp; Piniel, 2016)</td>
</tr>
<tr>
<td>Learner beliefs</td>
<td>BALLI (Horwitz, 1988; Jafari &amp; Shokrpour, 2012)</td>
</tr>
<tr>
<td></td>
<td>Epistemological beliefs (Mori, 1999)</td>
</tr>
</tbody>
</table>